NUMERICAL APPLICATIONS
OF THE
GENERALIZED METHOD OF STEEPEST DESCENTS

by
Jean-Marie Clarisse
Ingénieur des Arts et Manufactures,
Ecole Centrale de Paris, 1987
M.S. in Naval Architecture and Marine Engineering
M.I.T., 1989

Submitted to the Department of Ocean Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Hydrodynamics

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1992

© Massachusetts Institute of Technology, 1992
NUMERICAL APPLICATIONS
OF THE
GENERALIZED METHOD OF STEEPEST DESCENTS

by

Jean-Marie Clarisse

Submitted to the Department of Ocean Engineering
on May 8, 1992, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Hydrodynamics

Abstract

Formulated by the beginning of the XX-th century, the method of steepest descents is the most classical among asymptotic methods for approximations of integrals. Used extensively in various areas of physics dealing with wave phenomena, this method was extended in the 1960’s to the uniform treatment of several coalescing saddle points. Soon it benefitted from recent results of pure mathematics in what is known as catastrophe theory. Since then efforts have been devoted to the computation of the canonical integrals involved in these generalized steepest descent expansions. With the emergence of symbolic computation, it is then quite natural to launch the idea of systematic implementations and uses of the method of steepest descents for numerical computations.

Steepest descent expansions, being asymptotic, possess the distinctive advantage over numerical integration techniques of providing robustness, a property needed in certain applications. We study here two practical problems on which the robustness of these implementations depends.

Hence we address the problem of finding the relevant saddle points of an integral for which we propose a rigorous formulation. Although this formalism seems promising, much efforts remain to be done both at practical and fundamental levels.

Regarding the task of computing numerically the coefficients of the expansions, we study the algorithms of the ordinary method of steepest descents. We show that the robustness of the expansions can be conveyed in numerical applications, and that implementations in the frame of symbolic computation preserve the generality of the method. The simplest case of a generalized steepest descent expansion, i.e. a fold catastrophe expansion, is then considered. For this case, we provide algorithms and guidelines for the building of robust numerical procedures. The composite nature of these algorithms forbids us of presenting general purpose forms for these procedures.
The effectiveness of these results is demonstrated in treating the Kelvin wave source potential in hydrodynamics. Finally we consider the application of the method of steepest descents to the three-dimensional Cauchy-Poisson problem in finite depth. We show there that more powerful results from asymptotics of multi-dimensional integrals and classification of catastrophes are necessary.

Thesis Supervisor: J. Nicholas Newman
Title: Professor of Naval Architecture
Acknowledgments

I wish to thank my advisor, Professor J. Nicholas Newman, for his support and confidence all along this work, and for his patience when this research seemed to depart so much from hydrodynamics. I also would like to associate the members of the committee: Professors Henrik Schmidt, Paul Sclavounos and Dick Yue for their valuable comments and advice on this thesis.

This research has been strongly influenced by Professor J.N. Newman’s philosophy of numerical computations, and by Professor Fritz Ursell (University of Manchester) through his work and the exchanges I had with him. Other persons have also stimulated this work through discussions or lectures: Professors N.L. Trefethen (Cornell University), J.N.L. Connor (University of Manchester) and M.V. Berry (University of Bristol).

I would add a special mention for the people composing the “free-surface hydrodynamics group” of the Ocean Engineering Department at M.I.T., for providing a friendly atmosphere.

I am deeply grateful to the M.I.T. Artificial Intelligence Laboratory for granting me unrestricted access to its resources.

Support for this research was provided by The Office of Naval Research (ONR contract N00014-88-K-0057) and by The National Science Foundation (NSF grant CTS-8921011). The French Ministry of Defense should also be mentioned for letting me accomplish my national service at M.I.T. for the year 1989-1990.
Contents

Introduction

1 The method of steepest descents
   1.1 The ordinary method of steepest descents
   1.1.1 Assumptions
   1.1.2 Some definitions
   1.1.3 The procedure
   1.1.4 The asymptotic nature of the expansions
   1.1.5 The failures of the method
   1.2 The generalized method of steepest descents
   1.2.1 The changes of variables
   1.2.2 The local expansions
   1.2.3 The canonical integrals

2 Practical aspects
   2.1 Some inherent difficulties
   2.2 Why numerical applications?
   2.2.1 Numerics and physical understanding
   2.2.2 Robustness and efficiency guaranteed by the theory
   2.2.3 The characteristics of the method
   2.3 The treated and untreated practical aspects of the method
   2.4 The frame of the proposed implementations
5.3.2 The computation of the sequences \((p_r)\) and \((q_r)\) .............. 155
5.3.3 The computation of \((G_1^{(k)})\) and \((G_2^{(k)})\) ......................... 162
5.3.4 The unfolding of the fold catastrophe ............................................. 162
5.3.5 The schematic procedure ................................................................. 169
5.4 Some inevitable choices ................................................................. 170
5.4.1 The definition of \(\zeta^{1/2}\) ................................................................. 171
5.4.2 The choice of the correct branch ....................................................... 172
5.5 Numerical characteristics ................................................................. 174
5.6 The building of an implementation ..................................................... 176
5.6.1 The numerical inversion of the Levinson’s transformation ................. 177
5.6.2 The computation of \((G_i^{(k)})\) ......................................................... 179
5.6.3 The uniform computation of the sequences \((p_r), (q_r)\) .................. 179
5.6.4 The truncations of the asymptotic series .......................................... 180
5.6.5 Assembling the final expansion ......................................................... 181
5.7 Conclusion ............................................................................................. 184

6 The Kelvin wave source potential ......................................................... 187
6.1 The “single integral” part of the Kelvin wave source potential ............ 188
6.2 The domains of definition ..................................................................... 191
6.3 The relevant col problem ..................................................................... 192
6.4 The isolated col expansions .................................................................. 199
6.5 The fold catastrophe .............................................................................. 205
6.5.1 The building of the algorithm ............................................................ 205
6.5.2 Numerical characteristics .................................................................. 215
6.5.3 Domain of validity ............................................................................ 219
6.5.4 Fundamental properties of catastrophe expansions ........................ 221
6.6 Some results on Kelvin’s ship wave pattern ........................................ 227

7 The impulse response Green function in finite depth ......................... 234
7.1 A mathematical problem ..................................................................... 235
7.2 The result of Ursell .............................................................................. 239
7.3 An asymptotic expansion in terms of the butterfly integral . . . . . . 242
  7.3.1 The relevant cols . . . . . . . . . . . . . . . . . . . . . . . . . . . 243
  7.3.2 The butterfly catastrophe expansion . . . . . . . . . . . . . . . . 246
  7.3.3 The algorithms . . . . . . . . . . . . . . . . . . . . . . . . . . . . 249
  7.3.4 Numerical results . . . . . . . . . . . . . . . . . . . . . . . . . . . 255
7.4 A uniform asymptotic expansion in terms of the hyperbolic umbilic
  integral . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 258
7.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 263

Conclusion 265

Bibliography 268

A On some properties of the sets $\mathcal{O}_i$ 273

B On certain multiple entire series 275

C Root extractions in the complex plane 280

D The computation of the coefficients in the multiple entire series 282

E Higher order terms of the series $A, B, C, D$ 286
List of Figures

3-1 Cols, crest lines and basins on a surface ........................................... 80
3-2 Cols, crest lines and basins in the z-plane ........................................... 80
3-3 An example of valleys ................................................................. 81
3-4 Example of ramification (1) .......................................................... 88
3-5 Example of ramification (2) .......................................................... 89
3-6 Example of ramification (3) .......................................................... 90
3-7 Example of ramification and catastrophe .......................................... 91
3-8 Function with a pole ................................................................. 96
3-9 Graph with generating finite subgraphs: example 1 ............................ 103
3-10 Graph with generating finite subgraphs: example 2 ............................. 104
4-1 Example 1: analytical formula of $f$ ................................................. 139
4-2 Example 1: analytical formula of $g$ ................................................. 140
4-3 Example 2: result ............................................................................. 141
5-1 Information flow .............................................................................. 182
6-1 The sets described by the roots in the first quadrant, for $4^\circ \leq \theta \leq 50^\circ$,
       $0^\circ \leq \varphi \leq 90^\circ$ ................................................................ 194
6-2 The images by $f$ of the sets 1 and 2 of Figure 6-1 ............................ 195
6-3 The vicinity of $\Re w = 0$ .................................................................. 197
6-4 The basins of $f$ for $(\theta, \varphi) = (10^\circ, 80^\circ)$ .............................. 198
6-5 The uniform choice of the branch for $z_{21}^{(1)}$ and $z_{22}^{(1)}$ ............... 200
6-6 Values of $|G_z^{(2k)}/k! (2N)^k$ for $k = 0$ and $k = 3$, at $R = 16$ ........ 201
6-7 Values of $|G_{2z}^{(2k)}/k!(2N)^k$ for $k = 0$ and $k = 5$, at $R = 16$ ........... 201
6-8 $\zeta^{1/2}$ for $(\theta, \varphi)$ in $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$ ......................... 207
6-9 $f_1^{(3)} z_1^{(1)}$ and $f_2^{(3)} z_2^{(1)}$, in $8^\circ \leq \theta \leq 40^\circ$, $40^\circ \leq \varphi \leq 90^\circ$ ............ 208
6-10 $|A_s|/N^2$, $s \leq 3$, at $R = 16$ for $8^\circ \leq \theta \leq 40^\circ$, $40^\circ \leq \varphi \leq 90^\circ$ ............ 213
6-11 $|B_s|/N^2$, $s \leq 2$, at $R = 16$ for $8^\circ \leq \theta \leq 40^\circ$, $40^\circ \leq \varphi \leq 90^\circ$ ............ 214
6-12 $|C_s|/N^2$, $s \leq 2$, at $R = 16$ for $8^\circ \leq \theta \leq 40^\circ$, $40^\circ \leq \varphi \leq 90^\circ$ ............ 215
6-13 $|D_s|/N^2$, $s \leq 3$, at $R = 16$ for $8^\circ \leq \theta \leq 40^\circ$, $40^\circ \leq \varphi \leq 90^\circ$ ............ 216
6-14 $|A_s|/N^6$ and $|p_s|$, at $R = 16$ for $18.7^\circ \leq \theta \leq 20.3^\circ$, $87^\circ \leq \varphi \leq 90^\circ$ .......... 217
6-15 $|D_s|/N^6$ and $|q_s|$, at $R = 16$ for $18.7^\circ \leq \theta \leq 20.3^\circ$, $87^\circ \leq \varphi \leq 90^\circ$ .......... 217
6-16 Arrangement of the regions of validity of (AS1), (SD1), (SD2) and (CS1) in terms of $\theta$ and $\varphi$ ........... 220
6-17 The potential $w$ for $Z = 0$, in the sector $8^\circ \leq \theta \leq 30^\circ$ and 

$16 \leq R \leq 40$ ........................................... 221
6-18 The component in $A_i$ of $w$ at $Z = 0$ for $8^\circ \leq \theta \leq 30^\circ$ and 

$16 \leq R \leq 40$ ........................................... 222
6-19 The component in $A_i'$ of $w$ at $Z = 0$ for $8^\circ \leq \theta \leq 30^\circ$ and 

$16 \leq R \leq 40$ ........................................... 222
6-20 The real part of $\sum_{s=7} a_s(\theta, \varphi)/N^s$ at $\varphi = 90^\circ$ for $8^\circ \leq \theta \leq 30^\circ$ and 

$16 \leq R \leq 40$ ........................................... 223
6-21 The real part of $\exp[N f(z_1, \theta, \varphi)] \bar{A}(N^2/3 \, e^{2iv/3} \, \zeta)$ at $\varphi = 90^\circ$ for 

$8^\circ \leq \theta \leq 30^\circ$ and $16 \leq R \leq 40$ ........................................... 223
6-22 The real part of $\bar{A}(N^2/3 \, e^{2iv/3} \, \zeta)$ for $\varphi = 90^\circ$, $8^\circ \leq \theta \leq 30^\circ$ and 

$16 \leq R \leq 40$ ........................................... 224
6-23 The real part of $\exp[N f(z_1, \theta, \varphi)]$ for $\varphi = 90^\circ$, $8^\circ \leq \theta \leq 30^\circ$ and 

$16 \leq R \leq 40$ ........................................... 224
6-24 The potential $w$ for $\varphi = 90^\circ$, $8^\circ \leq \theta \leq 23^\circ$, $95 \leq R \leq 135$ .......... 229
6-25 The potential $w$ at $d = 0$ over the domain $4 \leq Y \leq 15$, 

$-16 \leq X \leq -40$ ........................................... 230
6-26 The potential $w$ scaled by $\exp[d/2]$ at $d = 2\pi/100$ over the domain

$4 \leq Y \leq 15$, $-16 \leq X \leq -40$ ........................................... 231
6-27 The potential $w$ scaled by $\exp[d/2]$ at $d = 2\pi/10$ over the domain
$$4 \leq Y \leq 15, -16 \leq X \leq -40$$

6-28 The potential $w$ scaled by $\exp[d/2]$ at $d = 4\pi/10$ over the domain
$$4 \leq Y \leq 15, -16 \leq X \leq -40$$

7-1 Real and imaginary parts of $\tilde{\omega}'$ in the strip $0 \leq \Re w \leq 10$,
$$-0.5 \leq \Im w \leq 0.5$$

7-2 Global arrangement of the algorithm

7-3 $F(X, T, 0)$ for $T = 20, |\zeta| < 0.4$: numerical integration versus asymptotic series

7-4 $F(X, T, 0)$ for $T = 40, |\zeta| < 0.4$: numerical integration versus asymptotic series

7-5 $F(X, T, 0)$ for $T = 80, |\zeta| < 0.4$: numerical integration versus asymptotic series

E-1 $|A_s|/N^{2s}$ at $R = 16$, for $4 \leq s \leq 9$ and $(\theta, \varphi)$ in $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$...

E-2 $|B_s|/N^{2s}$ at $R = 16$, for $3 \leq s \leq 7$ and $(\theta, \varphi)$ in $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$...

E-3 $|C_s|/N^{2s}$ at $R = 16$, for $3 \leq s \leq 9$ and $(\theta, \varphi)$ in $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$...

E-4 $|D_s|/N^{2s}$ at $R = 16$, for $4 \leq s \leq 10$ and $(\theta, \varphi)$ in $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$...
List of Tables

1.1 The first elementary catastrophes. ................................................. 44

5.1 Feedback information for each series ............................................. 182

6.1 Explorations of the intersection between the domains of validity of
       (AS1) and (SD1) ................................................................. 203

6.2 Explorations of the intersection between the domains of validity of
       (CS1) and (SD1) ................................................................. 204

6.3 Determination of \((\tau_n)\) ......................................................... 209

6.4 Truncations of the power series for \((p_r)\) and \((q_r)\) ..................... 210

6.5 Determination of the radii \((\sigma_{p_r})\) and \((\sigma_{q_r})\) ................. 211

6.6 Explorations of the intersection between the domains of validity of
       (AS1) - (SD1) - (CS1) and (SD2) ......................................... 220
Notations

We define here the main notations used throughout this work and refer at the same time to definitions given therein. However basic notions of set theory, topology, theory of analytic functions and graph theory are not detailed.

We use the notations:

\( \mathbb{N} \), for the set of positive integers

\( \mathbb{Z} \), for the set of integers

\( \mathbb{Q} \), for the set of rational numbers

\( \mathbb{R} \), for the set of real numbers

\( \mathbb{C} \), for the set of complex numbers

If \( A \) and \( B \) are two sets, \( B \) being a subset of \( A \), we denote in general:

\( A \setminus B \), the complementary of \( B \) in \( A \)

for any integer \( n \), \( A^n \) the product set \( A \times A \ldots \times A \) (\( n \) times)

\( \overline{A} \), the closure of \( A \) (in some superset to be specified)

\( \overset{\circ}{A} \), the interior of \( A \)

\( \partial A \), the boundary of \( A \), defined as \( \overline{A} \setminus \overset{\circ}{A} \)

\( A \) is a domain, if it is open and connected.
We use the following conventions and their variations for intervals of \( \mathbb{R} \):

\[ [a, b] \] the set of elements \( x \) of \( \mathbb{R} \) such that \( a \leq x \leq b \) (\( a \) and \( b \) in \( \mathbb{R} \))

\[ ]a, b[ \] the set of elements \( x \) of \( \mathbb{R} \) such that \( a < x \leq b \) (\( a \) might be in \( \mathbb{R} \) now)

and the notation \( D_o(a, \rho) \) for the open disc of \( \mathbb{C} \) centered at \( a \) and of radius \( \rho \).

Any element \( (z_1, \ldots, z_n) \) of \( \mathbb{C}^n \) is denoted \( \vec{z} \), \( \mathbb{C}^n \) being considered as a \( \mathbb{C} \)-vectorial space of dimension \( n \). At times, \( \mathbb{C}^n \) will also be identified to its structure of \( \mathbb{R} \)-vectorial space, or \( \mathbb{R}^{2n} \). If \( z \) is a complex number, \( \overline{z} \) stands for its conjugate (with of course the same convention for a vector \( \vec{z} \)).

We denote indifferently the mapping \( f \) from a set \( E \) to a set \( F \), \( f : E \rightarrow F \) or \( x \mapsto f(x) \). We say that \( f \) is \textit{injective} if \( f \) is one-to-one, \textit{surjective} if one-onto-one, \textit{bijective} if injective and surjective. We recall that \( f \) is a \textit{homeomorphism} if it is continuous and has a continuous inverse, a \textit{diffeomorphism} if it is differentiable and has an inverse which is differentiable. When dealing with functions of several complex variables, we will always consider functions that are \textit{holomorphic} in some domain. (A function \( f : U \rightarrow \mathbb{C} \), \( U \) domain of \( \mathbb{C}^n \), is holomorphic, if it is continuously differentiable and if \( df = \sum_{k=1}^{n} \partial f / \partial z_k \; dz_k \). \( f \) holomorphic in \( U \) is equivalent to \( f \) analytic for all \( \vec{z} \) in \( U \).)

Let \( f \) be analytic at \( z = z_0 \), then we denote:

\[ \sum_{n=0}^{\infty} a_{f,n} (z - z_0)^n, \] its entire series at \( z = z_0 \)

\[ f_{z_0}^{(d)}, \] the \textit{value} of its \( d \)-th derivative at \( z = z_0 \)

Several sequences or series are defined and used extensively:

the sequence \( (P_k^{(p)}) \), or \textit{Bleistein sequence} (§ 1.2.2 a) and § 1.2.2 b))

the sequences \( (p_r), (q_r) \), or the sequences of Chester, Friedman and Ursell (Section 5.1)
the series \((A_s/N^{2s}), (B_s/N^{2s}), (C_s/N^{2s}), (D_s/N^{2s})\) which define the expansion of Chester, Friedman and Ursell (Section 5.1)

Certain notations and definitions relative to a function \(f\) of the complex variables \((z, \bar{\alpha})\) are specific to this work:

- \(Z\), the set of cols or saddle points of \(f\) (Section 1.1.2)
- \(B\), a basin of \(f\) (Section 3.1.1)
- \(G_f\), the graph associated to \(f\) (Section 3.1.1)
- \(K\), the set of catastrophes of \(f\) and \(K_q\), the set of catastrophes of \(f\) of co-dimension at least \(q\) (Section 3.1)
- \(R\), the set of ramifications of \(f\) (Section 3.1.2)

and referring to a contour integral \(I = \int_C g(z, \bar{\alpha}) \exp[N f(z, \bar{\alpha})] \, dz\):

- \(Z_R\), the set of relevant cols of \(f\) (§ 1.1.5 1 and Section 3.3.1)
- \(B\), the set of singular points of \(f\) and \(g\) (Section 3.2)
- \(C\), the set of critical values of \(\bar{\alpha}\) for \(I\) (Section 3.2)
Introduction

Integrals with a large parameter are encountered in many areas of physics. They appear as solutions of various problems such as: propagation equations in optics, Schrödinger equations in quantum mechanics, boundary value problems of Laplace's equation in hydrodynamics or acoustics, etc... Their common ground is to model the propagation of certain forms of energy. Interpreted as waves, their behavior can be as opposite as one of pure exponential decay, or one of ever increasing frequency oscillations. In order to infer such behaviors from integral expressions in their large parameter limits, methods were developed by the end of the XIX-th century. Taking into account the computational means of that period, these methods made use of accuracy properties provided by certain diverging series. Historically, the first of these methods – the principle of stationary phase – was formulated in 1887 by Lord Kelvin, probably influenced by Stokes, for the study of waves in dispersive media, and free-surface waves in hydrodynamics in particular (Kelvin, 1887a). Some twenty years later, for the purpose of treating Bessel functions of large orders, Debye introduced the method of steepest descents (Debye, 1909). Encompassing the principle of stationary phase and making use of the power of complex analysis, this method became the most classical and probably most powerful tool for obtaining asymptotic approximations of integrals.

However, the method of steepest descents demonstrated its limitations when applied to the problem of the wave pattern generated by a ship. Hence the method failed, in the same manner as Kelvin’s principle of stationary phase (Kelvin, 1887b), to approximate the transition between the pure exponentially decaying and the pure
oscillatory behaviors present in a ship wave pattern. This transition, a well known phenomenon in optics, had already been modeled in 1838 by Airy. In 1957, motivated by the ship wave problem, Chester, Friedman and Ursell extended the method of steepest descents to the uniform treatment of this transition, or the coalescence of two saddle points. Further developments followed in the 1960's which soon resulted in the generalization of the method to the uniform treatments of several coalescing saddle points and neighboring singularities. Around the same period (1955) but in differential topology, was formulated the idea, in the spirit of Poincaré's work, of a classification of coalescences of critical points: singularity theory was laid. A product of this new theory and of bifurcation theory, catastrophe theory was proposed by the beginning of the 1970's as a universal theory of evolution of systems (Thom, 1977). Surrounded by much controversy, this theory received however a special attention from physicists in quantum mechanics who soon found quantitative applications to it through the generalized method of steepest descents. Since then, efforts have been devoted to the numerical computation of the special functions required by these generalized steepest descent expansions. Currently and following the revision by Dingle (Dingle, 1973) of the concept of asymptotics enunciated by Poincaré, the method of steepest descents is being extended to the field of hyperasymptotics.

Hence despite the use of computers allowing the numerical evaluations of integrals with unprecedented facility, there is a persisting interest for the method of steepest descents. It is then natural to consider its applications to numerical evaluations of integrals with a large parameter. If in most cases numerical integration techniques are sufficient, certain applications necessitate more powerful methods. This corresponds by example to integrals whose integrand presents highly oscillatory behaviors which worsens as its large parameter increases in magnitude. These integrals may not be absolutely convergent in their original definition, such an example is the Kelvin wave source potential in hydrodynamics (see Chapter 2 and 6), or the potential of a source moving with steady forward speed in otherwise calm water.
In dealing with numerical evaluations two aspects intervene: the accuracy of the results and the time required by the implemented algorithm to obtain them. Efficient techniques are of course those for which the former is high and the latter small. With “pathological” integrals such as the ones we are concerned with, there exists an obvious trade-off between the simplicity of the method and its efficiency measured in terms of computational time versus accuracy. Moreover when evaluations are needed for a wide or unbounded range and for any value of some parameters, bounds on the accuracy and time are needed for whole regions of the parametric space. Tentatively, we would define the robustness of a method as the ability to ensure a fixed error, uniformly over a given region.

The association of efficiency and robustness is of particular importance for certain applications of acoustics and hydrodynamics where evaluations of such integrals are done on a regular basis. In solving a realistic problem of hydrodynamics, a Green’s function in integral form is evaluated of the order of a 1 000 000 times prior to obtaining any solution (Newman, 1990). Other areas of physics have different requirements: a single evaluation of such an integral might yield a complete solution (e.g. in quantum mechanics). In general, computations of an integral with a large parameter appear mainly under three forms and their combinations:

a) the evaluation of a time dependent solution for large times

b) the evaluation of the high frequency components of a solution

c) the evaluation of a solution at large distances

In hydrodynamics, a) is becoming more common as time-domain studies are of increasing interests. c) has gotten recent attention with the problems of ship wake detection, although it was already involved in wave resistance prediction. It must also be noted that since distances are intrinsically non-dimensionalized in formulations of Green’s functions, a “large” distance may well correspond to the description of a part of a body. In acoustics the three cases a), b) and c) may be encountered. The case of b) is of crucial interest for optics and quantum mechanics. We
should not omit the case of a large non-dimensional parameter which can arise in local solutions such as boundary layer solutions, solutions of differential equations with turning points (e.g. Schrödinger's equation). This covers investigations of unstable wave solutions, boundary layers in fluid flows, etc...

In the numerical treatment of a given problem, it is necessary to distinguish the concept of uniform global validity as opposed to ad hoc considerations. In general, it is desirable not to sacrifice the physical hypotheses of a model for the sake of numerical convenience. Discarding regions of a physical domain, because of the failure there of a numerical technique, is highly suspicious. The linear ship wave pattern problem in hydrodynamics is a famous example of such expedients. The advantage of asymptotic methods in general, and of the method of steepest descents in particular, over any numerical technique, is to guarantee robustness in open (in the topological sense) regions of a parametric space. Naturally this particular feature must be preserved in any implementation of the method in order to be of interest in numerical applications.

Hence we launch the idea of systematic implementations of the method of steepest descents for robust numerical evaluations of integrals. Our purpose is then to demonstrate that such implementations which preserve the robustness as well as the generality of the method of steepest descents, are possible.

This requires us, at first, to review the procedures involved in the ordinary method of steepest descents as well as in its generalization. This review, well suited for our purpose but far from being complete, leads us to define a theoretical frame for robust and general implementations of the method (Chapter 1).

The extent of the practical difficulties inherent to the method of steepest descents necessitates an evaluation of its characteristics as well as its treated and untreated aspects (Chapter 2). From there we focus on two problems encountered in any application of the method and which are not yet fully treated.
Hence we address the crucial but still untreated problem of determining the relevant saddle points in a steepest descent expansion of an integral (Chapter 3). For the first time a formalism for the rigorous formulation of this problem is proposed. Although this approach seems promising and well suited to implementations on computers, much work remains to be done.

The second aspect we are concerned with, consists in obtaining numerically the coefficients of the expansions. We start first by considering the cases of ordinary steepest descent expansions. In the same spirit as Dingle (1973) but benefitting from symbolic computation, we present the algorithms for the ordinary method of steepest descents and study their numerical properties (Chapter 4). These algorithms, characterized by possible cancellations, preserve however the robustness and generality of the method. Moreover, if implemented in the frame of symbolic computation, they allow us to obtain analytical formulas as well as numerical results of arbitrary accuracy.

More arduous is the corresponding task for generalized steepest descent expansions. We treat here the simplest case of the uniform treatment of the coalescence of two saddle points, or fold catastrophe (Chapter 5). Hence, for the first time, algorithms and guidelines are proposed for robust numerical evaluations of fold catastrophe expansions. However contrary to the ordinary steepest descent expansions, implementations of these algorithms are necessarily case specific. Moreover they are limited to isolated fold catastrophes.

As a direct demonstration of the robustness provided by both types of algorithms, we apply these to the Kelvin wave source potential in the far-field regime (Chapter 6). In particular we conclude that the corresponding fold catastrophe expansion, originally given by Ursell in 1960 (Ursell, 1960), is valid in more than half of the Kelvin wedge. More importantly we outline a fundamental property of catastrophe expansions which could lead to highly efficient numerical implementations as well as powerful numerical tools.

Finally, we discuss the application of the method of steepest descents to another example taken from hydrodynamics: the three-dimensional Cauchy-Poisson problem.
in finite depth in the vicinity of the wave front (Chapter 7). Following previous works on the subject, we present an asymptotic expansion in terms of the butterfly integrals as well as specific algorithms for its evaluation. However this expansion fails to be uniform in the vicinity of the wave front, evidencing the inappropriateness of the method of steepest descents for this case. Consequently we propose a new approach based on asymptotics of multi-dimensional integrals and on the classification of catastrophes, yielding an expansion uniformly valid in the neighborhood of the wave front. A complete justification of this result as well as practical methods for obtaining this expansion are yet to be produced.
Chapter 1

The method of steepest descents

Before proceeding to the description of the practical aspects of the method of steepest descents, it is desirable to give a review of the ordinary method (Section 1.1) and of what we call the generalized method of steepest descents (Section 1.2). This review, based on results from numerous works on the subject by different authors, is sufficiently thorough for allowing us to define the theoretical frame of our implementations. However many details are omitted since a complete description of the method would certainly require a whole book on the subject.

The method of steepest descents belongs to this part of applied mathematics called asymptotic methods. Under this name is in fact gathered many different “methods” often regarded just as simple tools, if not “recipes” ready to be applied to obtain asymptotic expressions. They share the particularities of providing very accurate results with very few terms extracted from divergent series (quite the opposite of what can be obtained from convergent series) for usally small regions of validity, and of being able to fail in very perverse ways, that is in fact being asymptotic. The method of steepest descents – “méthode du col” in French or “Methode der Sattelpunkte” in German – follows this pattern although it is a development of the theory of complex contour integration. Its first appearance is due to Debye who formalized and applied this theory to Bessel functions of large order to obtain asymptotic expansions (Debye, 1909). But a first sketch of the method can be traced back to Riemann in a posthu-
mous work (Riemann, 1892). A noticeable contribution to the theory was made by Watson in 1918, through his proof of the asymptotic nature, in the Poincaré sense, of steepest descent expansions, using what is now called Watson’s lemma. An account of the method as well as many remarkable examples of its applications are found in his Treatise on the Theory of Bessel Functions (Watson, 1944). Other contributions were subsequently made, mainly by applied mathematicians, on specific aspects of the method as well as on the problems of its extensions to more general cases. They will be referred to, all along this presentation. Other general references on the method can be found in various books on asymptotic methods: Copson (1965), Erdélyi (1966), Dingle (1973), Olver (1974), Wong (1989). Among asymptotic methods for evaluations of integrals, the method of steepest descents is undoubtedly the most frequently treated, so numerous are the articles on the subject. However, as noted in Wong (1989), the descriptions of the method are always too sketchy to be appropriate. The formulation of a complete mathematical theory is difficult as the method intersects with fundamental problems of mathematics far from being completely treated. Such an attempt is made in Wyman (1964), in the form of a tentative unification of the methods of stationary phase and steepest descents as subsets of Laplace’s method. However opinions vary with different authors and, by example, Olver (1970) gives another point of view on the subject.

Both articles by Wyman and Olver prove more general theorems than Watson’s lemma. The article of Olver, “Why steepest descents?”, is of particular interest since it outlines the importance of the paths of steepest descent, distinguishing the method of steepest descents, or the procedure of Riemann, Debye and Watson, from what is sometimes called the saddle point method or Perron’s method (Copson, 1965, § 8.36, or Wong, 1989, § II.5), where paths of descent, not necessarily steepest, are considered as approximations to steepest descent paths. We make the same distinction here when speaking of the method of steepest descents. On the more general topic of asymptotic expansions and their peculiarities, the work of Dingle (1973) goes particularly in depth, introducing the concepts of what is now called surperasymptotics.
and hyperasymptotics (Berry, 1992). Those leave interesting prospects for improvements in accuracy of numerical implementations of the method of steepest descents. Modestly we will stay in this work at the stage of simple asymptotics.

1.1 The ordinary method of steepest descents

1.1.1 Assumptions

For the sake of clarity, and applicability, it is necessary to impose some restrictions to the class of integrals considered. The mathematical assumptions stated here will pertain to the rest of the discussion including the numerical implementations of the method although we will mention at times practical restrictions.

We shall consider complex contour integrals of the form:

\[ I(N, \alpha) = \int_C g(z, \alpha) \exp[Nf(z, \alpha)] \, dz \]  

(1.1)

where:

(i) \( N \in \mathbb{R}^+ \)

(ii) \( z \in \mathbb{C} \) and \( \alpha \in \mathbb{C}^l, l \in \mathbb{N} \)

(iii) \( f \) and \( g \) are holomorphic in \( W \times D \), \( W \) and \( D \) being domains of \( \mathbb{C} \) and \( \mathbb{C}^l \) respectively.

(iv) \( C \) is a contour included in \( W \) such that \( I(N, \alpha) \) exists in the Riemann sense. If the end-points of \( C \) are located at infinity, we require uniform convergence of \( I \) with respect to \( (N, \alpha) \) in \([\delta, +\infty[ \times D, \delta > 0 \).

* We have chosen to restrict \( N \) to be real since the more general case of \( N \) complex can be treated by including the argument of \( N \) in the definition of \( f \). Reminiscent of
the method of stationary phase some definitions use $iN$, or $i/h$ as $h \to 0^+$, instead of $N$ but we prefer to use a real positive parameter.

- In addition to (iii) we require that the singularities of $f$ and $g$ be discrete in $\mathbb{C}^{t+1}$, and that any singularity $z_0$, of $f$ or $g$ as functions of $z$, be such that:

$$\exists r_0 \in \mathbb{R} \quad / \quad f(z) = (z - z_0)^{r_0} h(z)$$

with $h$ holomorphic in the disc $D_0(z_0, \rho)$, $\rho > 0$. This covers the cases of poles and branch points, but avoids the consideration of pathological cases such as $\exp[1/z]$.

- Finally the specification on $C$ is necessary to have an integral which has a sense. The requirement for the integrand to be Riemann integrable ensures us of a numerical meaning.

The principle

We are interested in obtaining evaluations of $I$ when $N$ becomes large. Momentarily we make the assumption that indeed $W$ is $\mathbb{C}$. The method of steepest descents then states that for a fixed $\alpha$, an expansion asymptotic in the Poincaré sense, is obtained as the sum of local contributions from saddle points of $f$, plus eventually the contributions from the end-points of the contour $C$. Each of these contributions is in fact asymptotic in $N$ and is obtained by integrating the integrand along paths of steepest descents originated from the saddle points or end-points. (For a review of the method see Watson (1944, chap. 8) and Wong (1989, II.4).)

Before going into further details, we need to give some definitions and notations used throughout this work.

1.1.2 Some definitions

We have already mentioned the terms saddle point and steepest descent, and we add here other notions as well as notations to be used in different parts of the discussion.
Extended definitions will be given in Chapter 3.

**saddle point or col:** A zero of \( \partial f(\cdot, \bar{\alpha})/\partial z \) as function of \( z \) only, denoted \( z(\bar{\alpha}) \).

Since \( f \) is holomorphic saddle points are isolated points and the set of saddle points is a discrete set. Hence we can index this set with the set of positive integers \( \mathbb{N} \): \( Z = \{ z_i(\bar{\alpha}) : i \in \mathbb{N} \} \).

**order of a col:** The order of \( z(\bar{\alpha}) \) as a solution of the equation \( \partial f(z, \bar{\alpha})/\partial z = 0 \).

The original method of steepest descents is concerned with simple cols, or first order cols, or cols of order 1.

**path:** The definition of a mapping \( \gamma : J \rightarrow \mathbb{C} \), continuous on \( J \), an interval of \( \mathbb{R} \).

\( \gamma(J) \) can be a segment, a curve, etc... but is generally referred as a **contour** or **arc**.

A contour is the image of an infinite number of paths.

**path of descent (ascent):** A path \( \gamma \) for which the function \( t \mapsto \Re\{ f(\gamma(t), \bar{\alpha}) \} \), for \( \bar{\alpha} \) fixed, is strictly decreasing (increasing) in \( J \).

**path of steepest descent (ascent):** A path of descent (ascent) \( \gamma \) for which the function \( t \mapsto \Im\{ f(\gamma(t), \bar{\alpha}) \} \) is constant.

**Notation:** We will use the letter \( S \) to name such paths in general, and \( D \) (\( A \) respectively) in particular.

- This definition comes from the fact that \( f \) is holomorphic. Thus paths for which \( t \mapsto \Re\{ f(\gamma(t), \bar{\alpha}) \} \) is the most decreasing (increasing) – i.e. for which its
derivative with respect to $t$ is maximum or minimum – are paths along which $\Im\{f(\gamma(t), \alpha)\}$ is constant.

**level curve**: A path $\gamma$ for which the function $t \mapsto \Re\{f(\gamma(t), \alpha)\}$ is constant.

Notation: $\mathcal{L}$ will stand for level curves.

We have the duality between $f$ and $if$: level curves of one are the steepest paths of the other.

Among level curves we will consider in particular:

**curves of $-\infty$-level**: These are level curves $\gamma$ for which $\Re\{f(\gamma(t), \alpha)\} = -\infty$, noted $\mathcal{L}_{-\infty}$.

The ideas behind the consideration of paths of descent and steepest descent can be visualized more easily if one considers for a fixed $\alpha$ the function of the two real variables $(x, y)$:

$$W \rightarrow \mathbb{IR}$$

$$(x, y) \mapsto \Re\{f(x + iy, \alpha)\}$$

The graph of this function defines a surface $\Sigma$ in $\mathbb{IR}^3$ where the saddle points are passes, hence the terminology of cols (see Figure 3-1). This surface does not possess any local extrema in $W$ identified as a subset of $\mathbb{IR}^2$, since $f$ is holomorphic. As $N$ increases the relief steepens. The analogy with a topographical relief can be further pursued (see Chapter 3).

Meanwhile some remarks follow:

- If we arbitrarily name valleys the regions of $\Sigma$ for which $\Re\{f(x + iy, \alpha)\}$ is negative, it is obvious that the end-points of $C$, say $a$ and $b$, must be in such valleys for the integral to exist as $N \rightarrow +\infty$.  

27
— Intuitively, the method states that when going from $a$ to $b$ the main contributions to the integral as the relief steepens, come from the vicinity of the end-points of the path and of the cols the path has to go through in order to join $a$ to $b$. The total and exact contribution stays of course the same whatever the path since the integrand is holomorphic.

— Obviously the path joining $a$ to $b$ need not pass through any cols of $f$ if the two points are in the same valley or basin (used here in the same meaning as an hydrologic basin in geography). One then could think of the method as finding the “minimal” path joining $a$ to $b$, minimal in some sense of a “least effort” path similarly to what a mountain climber would seek in a massif.

Another way of looking at paths of descent is to consider their image by $f$ in $\mathbb{C}$. There, a steepest path is a straight line parallel to the real axis whereas a level curve is a straight line parallel to the imaginary axis (the duality of $f$ and $if$ is there obvious). This image space was first used by Van der Waerden (Van der Waerden, 1950) in order to formalize the problem of the deformation of $C$.

1.1.3 The procedure

We now detail the procedure of Riemann, Debye and Watson. In this part we assume that $\bar{\alpha}$ is fixed and that in fact $W$ is the whole complex plane $\mathbb{C}$ so that $f$ and $g$ are holomorphic everywhere. Furthermore $f$ is assumed to only have cols of order 1. The set of cols is indexed by a subset of $\mathbb{N}$, $[1..J]$ ($J$ might be $\infty$) and we denote it: $Z = \{z_j : 1 \leq j \leq J\}$ or $(z_j)_J$.

To keep notations simple we drop the dependence on $\bar{\alpha}$ throughout this section.
A) The deformation of $C$

Using Cauchy theorem we deform the initial contour $C$ into $C'$ such that $C'$ is composed only of paths of steepest descent issued from the end-points $a$ and $b$, of paths of steepest descent issued from saddle points and of curves of $-\infty$-level. Using the notations: $S_a$ and $(-S_b)$ for the paths issued from the end-points, $S_{z_j}$ for the paths issued from the col $z_j$, and $\mathcal{L}_{-\infty,i}$ for curves of $-\infty$-level, the new path $C'$ is written:

$$C' = S_a \cup \bigcup_{j=1}^{J_r} S_{z_j} \cup \bigcup_{i=1}^{J_r+1} \mathcal{L}_{-\infty,i} \cup (-S_b) \quad (1.2)$$

It is easy to see that the contributions from any of the level curves $\mathcal{L}_{-\infty,i}$ are exactly zero. Hence $I$ is written:

$$I(N) = \int_{S_a} g(z) \exp[Nf(z)] \, dz + \int_{-S_b} g(z) \exp[Nf(z)] \, dz$$

$$+ \sum_{j=1}^{J_r} \int_{S_{z_j}} g(z) \exp[Nf(z)] \, dz \quad (1.3)$$

assuming that the contributions from the end-points are not zero.

• Note that the union and the sum over $j$ range from 1 to $J_r$, an integer smaller than $J$ which we do not specify. (The cols are supposedly reindexed from 1 to $J_r$ if necessary.)

So far we have only used the fact that $C'$ is homotopic to $C$ in $W$ and that the integrand is holomorphic in $W$. The result obtained is therefore exact.
B) The restrictions of the paths to neighborhoods of their highest points

Since each of the paths now considered is a steepest path of descent issued from its highest point – the highest in the sense of \( R\{f(z)\} \) – we can restrict the path of integration to a neighborhood of this point. The argument for performing this restriction is entirely similar to what is used in Laplace’s method for integrals of the real variable: the contributions from the remaining portions of the paths are exponentially negligible in \( N \) (see Wong, 1989, § II.1 by example). Remember that on steepest paths the imaginary part of \( f \) stays constant and the integral can be considered as of the real variable \( R\{f(z)\} \). In other words as the oscillating part of the exponent is constant we have avoided the complex interferences which occur as \( N \) gets large (Watson, 1944).

These new paths being noted \( \mathcal{S} \) we have:

\[
I(N) \sim \int_{\mathcal{S}_a} g(z) \exp[Nf(z)] \, dz + \int_{-\mathcal{S}_b} g(z) \exp[Nf(z)] \, dz \\
+ \sum_{j=1}^{J} \int_{\mathcal{S}_{z_j}} g(z) \exp[Nf(z)] \, dz
\]

(1.4)

where \( \sim \) implies that terms of order \( \exp[-Ne], e > 0 \) have been neglected.

C) The changes of variables

Hence \( I \) is now reduced to a sum of local integrals in neighborhoods of end-points and/or cols. For each integral we perform a change of variable which will reduce it to a Laplace kind integral. The cases of end-points and saddle points must be treated differently.

a) Change of variable at an end-point \( e \)

We assume that the end-point \( e \) of \( C' \) is “sufficiently” distant from any cols of \( f \).
The notion of “sufficiently” is totally arbitrary but let us say that they are separated by a certain non-zero distance $\delta$. Then the local change of variable – we perform this change only on the integral over $\mathcal{S}_e$ – is defined by the mapping: $u : z \mapsto u$, bijective in a neighborhood of $e$ and defined by the equation:

$$f(z) - f(e) = -u$$  \hfill (1.5)

The integral along $\mathcal{S}_e$ after applying the change of variable is written:

$$\int_{\mathcal{S}_e} g(z) \exp[Nf(z)] \, dz = \pm \exp[Nf(e)] \int_0^{\delta(e)} G_e(u) \exp[-Nu] \, du$$  \hfill (1.6)

with $G_e(u) = g(z) \, dz/du$.

- The sign $\pm$ depends whether the path $\mathcal{S}$ (hence $\mathcal{S}_e$) is a path of descent (+) or ascent (−). ($\mathcal{S}_a$ is a path of descent, but $-\mathcal{S}_b$ is a path of ascent ending at $b$.)

- $\delta(e)$ is the image by $u$ of the lowest end-point of $\mathcal{S}_e$, and is positive. The condition on $e$ and any other critical points of being “sufficiently” distant would be that $\delta > 2\delta(e)$ for example.

b) Change of variable at a col $z_j$

Here we require the cols to be “sufficiently” distant from each other and from the end-points of $C'$. The change of variable is now defined after the mapping $u : z \mapsto u$, again bijective in a neighborhood of $z_j$, defined by the equation:

$$f(z) - f(z_j) = -\frac{u^2}{2}$$  \hfill (1.7)

Performing the change of variable $z \mapsto u$, the integral along $\mathcal{S}_{z_j}$ becomes:

$$\int_{\mathcal{S}_{z_j}} g(z) \exp[Nf(z)] \, dz = \exp[Nf(z_j)] \int_{-\delta(z_j)}^{\delta(z_j)} G_{z_j}(u) \exp[-Nu^2/2] \, du$$  \hfill (1.8)
where $-\delta(z_j)$ and $\delta(z_j)$ are the images by $u$ of the starting point and the ending point of $\mathcal{S}_z$, respectively ($\mathcal{S}_z$ being chosen accordingly).

- The use of the quadratic polynomial in $u$ is necessary to have a bijective mapping.

D) The expansions of the local integrals

We now use the fact that each of the newly defined functions $G$ can be expanded in entire series of $u$. Again we need to distinguish the cases of end-points from the cases of cols.

a) Expansion of the integral at an end-point $e$

By the choice of $u$ there exists a radius $\rho(e)$ say greater than $\delta(e)$, such that the series:

$$\sum_{k=0}^{\infty} a_{G,k} u^k = G_e(u)$$

is convergent in $D_0(0, \rho(e))$.

Then we can write:

$$\int_0^{\delta(e)} G_e(u) \exp[-Nu] du = \sum_{k=0}^{K_e} a_{G,k} \int_0^{\delta(e)} u^k \exp[-Nu] du + E_{1,K_e}(e)$$

(1.9)

where $K_e$ is arbitrary but finite and $E_{1,K_e}(e)$ denotes the error due to the truncation of the series, that is:

$$E_{1,K_e}(e) = \int_0^{\delta(e)} \left( \sum_{k=K_e+1}^{\infty} a_{G,k} u^k \right) \exp[-Nu] du$$

(1.10)

b) Expansion of the integral at a col $z_j$

Similarly, the definition of the mapping $u$ (in fact its inverse $z$) allows the expan-
sion of $G_{z_j}$ in an entire series in $u$ in a disk $D_0(0, \rho(z_j))$:

$$G_{z_j}(u) = \sum_{k=0}^{\infty} a_{G,k} u^k$$

On substituting this expansion in the integral, we obtain:

$$\int_{-\delta(z_j)}^{\delta(z_j)} G_{z_j}(u) \exp[-N \frac{u^2}{2}] du = \sum_{k=0}^{K_{z_j}} a_{G,k} \int_{-\delta(z_j)}^{\delta(z_j)} u^k \exp[-N \frac{u^2}{2}] du$$

$$+ \mathcal{E}_{1,K_{z_j}}(z_j)$$

(1.11)

where $K_{z_j}$ is again arbitrary and $\mathcal{E}_{1,K_{z_j}}(z_j)$ stands for the error due to the truncation of the series.

E) The extension of the paths of integration

Instead of considering the finite integrals in the new expressions, their paths are extended to infinity with additional errors. This allows the expression of each series in terms of powers of $N$ without referring to the arbitrary values of the finite bounds of the integrations.

Hence, for an end-point we have:

$$\int_{0}^{\delta(e)} G_\epsilon(u) \exp[-Nu] du = \sum_{k=0}^{K_\epsilon} a_{G,k} \int_{0}^{+\infty} u^k \exp[-Nu] du$$

$$+ \mathcal{E}_{1,K_\epsilon}(e) + \mathcal{E}_{2,K_\epsilon}(e)$$

(1.12)
and for a col:

\[
\int_{-\varepsilon(z_j)}^{\varepsilon(z_j)} G_{z_j}(u) \exp[-N \frac{u^2}{2}] du = \sum_{k=0}^{K_{z_j}} a_k(z_j) \int_{-\infty}^{\infty} u^k \exp[-N \frac{u^2}{2}] du
\]

\[+ \mathcal{E}_{1,K_{z_j}}(z_j) + \mathcal{E}_{2,K_{z_j}}(z_j) \] (1.13)

where the terms \( \mathcal{E}_2 \) account for the errors made by extending the ranges of integration to infinity.

The integrals involved are now depending only on \( N \) and can be expressed in terms of the Gamma function \( \Gamma \). Thus we have:

\[
\int_0^{+\infty} u^k \exp[-Nu] du = \frac{k!}{N^{k+1}} \] (1.14)

and:

\[
\int_{-\infty}^{\infty} u^k \exp[-N \frac{u^2}{2}] du = \begin{cases} 
0 & \text{if } k \text{ odd} \\
\left( \frac{2}{N} \right)^{(k+1)/2} \frac{k!}{\Gamma(\frac{k+1}{2})} & \text{if } k \text{ even}
\end{cases} \] (1.15)

Finally, the two kinds of integrals are written:

\[
\int_{\mathbb{S}_e} g(z) \exp[N f(z)] dz = \exp[N f(e)] \left[ \sum_{k=0}^{K_z} a_{G,k} \frac{k!}{N^{k+1}} + \mathcal{E}_{1,K_z}(e) + \mathcal{E}_{2,K_z}(e) \right]
\]

and:

\[
\int_{\mathbb{S}_{z_j}} g(z) \exp[N f(z)] dz = \exp[N f(z_j)] \left[ \frac{\sqrt{2\pi}}{N^{1/2}} \sum_{k=0}^{K_{z_j}} a_{G,2k} \frac{(2k-1)(2k-3)\ldots1}{N^k} \\
+ \mathcal{E}_{1,K_{z_j}}(z_j) + \mathcal{E}_{2,K_{z_j}}(z_j) \right]
\]
It is shown by Watson's lemma that indeed both types of series are asymptotic in \( N \), that is we have:

\[
\int_{S_1} g(z) \exp[Nf(z)] \, dz \sim \exp[Nf(e)] \frac{1}{N} \sum_{k=0}^{K_z} a_{G,k} \frac{k!}{N^k}
\]  

(1.16)

and:

\[
\int_{S_{z_2}} g(z) \exp[Nf(z)] \, dz \sim \exp[Nf(z_j)] \frac{\sqrt{2\pi}}{N^{1/2}} \sum_{k=0}^{K_{z_2}} a_{G,2k} \left( \frac{(2k-1)(2k-3)\ldots}{N^k} \right)
\]  

(1.17)

in the Poincaré sense, when \( N \) is sufficiently large.

- The form of (1.16) is reminiscent of the results of Laplace’s method, and in fact is not different, however the change of variables as well as the other steps differ from Laplace’s method.
- In fact Watson’s lemma as well as the improved theorems proved in Wyman (1964) and in Olver (1970), covers the last two steps D) and E) of the procedure.

The total expression for \( I \) is then obtained by replacing the different integrals in (1.3) by their expansions as given by (1.16) or (1.17).

### 1.1.4 The asymptotic nature of the expansions

The asymptotic nature of the results is proved by applying Watson's lemma. The paths we have used are steepest paths, but similar results can be obtained by considering a more general class of paths as shown in Wyman (1964) and Olver (1970). Results are then obtained by using paths that are much easier to determine than steepest paths. This method, or saddle point method, is in most cases more practical than the method of steepest descents and can be applied to numerical evaluations by using numerical integration along these approximated paths. However Olver showed that considering steepest paths gives the minimal errors for \( E_1 \) and \( E_2 \). The detailed bounds for these errors are found in Olver (1970, eq. 3.18 & 3.21).
1.1.5 The failures of the method

So far we have made some restrictive assumptions regarding the integral $I$: $\bar{a}$ was held fixed, the domain $W$ was assumed to be the whole complex plane $\mathbb{C}$, cols and end-points were supposed to be "sufficiently" distant and we were considering only cols of order 1. What happens if these assumptions are not made? Can we expect to obtain an asymptotic expansion in similar ways?

More precisely, in many problems $\bar{a}$ is allowed to describe the whole domain $D$, and asymptotic expansions are then needed for any values of $\bar{a}$. In what subdomain of $D$ is a given expression an asymptotic expansion of $I$? Can a covering of $D$ by such domains, each of them associated with a certain expression, be found, thus furnishing an asymptotic expansion of $I$ for every $\bar{a}$ in $D$?

This raises the problem of the uniformity, with respect to $\bar{a}$, of these expansions, and the search for those subdomains of $D$ on which corresponding expressions are uniformly valid, i.e. uniformly asymptotic as $N$ gets large. Indeed, the method of steepest descents fails to provide such a uniformity in several ways. In asymptotic methods, such failures are known under the general term of Stokes phenomenon. They occur when $\bar{a}$ approaches the boundary of one of these subdomains. There the corresponding expression ceases to be an asymptotic expansion of $I$ – the error made when discarding the terms $E$ is no longer negligible – and another expression is to be sought. In fact for practical applications one looks rather for non-empty intersections of these subdomains in order to ensure matching between several expressions, hence the uniformity of an asymptotic treatment of $I$.

Problems that arise in different areas of physics motivated further efforts in order to circumvent these failures and resulted in extensions of the method which we group under the term of generalized method of steepest descents. Prior to the presentation of this generalization we describe each one of these failures.
1. The relevance of the cols

Probably the most subtle way in which an asymptotic expansion obtained by the method of steepest descents fails, is through modifying the set of relevant cols. Recall that we have been totally unclear about the value of $J_r$ in (1.2) and the way we could determine which of the cols the deformed path $C'$ has to go through. From (1.3) it is also obvious that, by adding a $(J_r + 1)$-th saddle point to the sum over the cols, the result will be changed. In fact at this stage of the method we have not used yet any kind of expansion and so this problem of adding a col has indeed very little to do with the Stokes phenomenon. However it does change the result since the equality in (1.3) might not hold anymore! By chance the contribution from this new col could be subdominant and therefore would not modify significantly the result, but this might not be the case as well!

Since the cols $z_j(\tilde{\alpha})$ are functions of $\tilde{\alpha}$, the set of relevant cols, say $Z_R$, might change and with it the asymptotic expansion. But how is this set of relevant saddle points defined? Is this set well determined – unique – for a given value of $\tilde{\alpha}$? How can it be modified as $\tilde{\alpha}$ varies?

This problem of the relevance of saddle points is never treated and not even mentioned in the corresponding literature, except for an article by Ursell (1970). None of the books on asymptotic methods or discussing the method of steepest descents, mentions the problem of determining this set despite its cruciality for the correctness of the result. The principles for deforming the original path are given but nothing more. Even Watson in his treatise if he stipulates that the paths should be deformed in a certain way by considering certain cols, never mentions any rule for determining which cols should be considered. In the many examples he gives this step always stays in the dark.

- It might be argued that an obvious way of selecting a col $z_{j_0}$ is to compare the quantity $\Re\{\exp[f(z_{j_0}, \tilde{\alpha})]\}$ – the "altitude" of $z_{j_0}$ – to the same quantities from other cols. But this would only determine if $z_{j_0}$ is subdominant or not with respect to other saddle points – i.e. if $z_{j_0}$ is of lower altitude than all of the other cols – and $C$ might
well be such that \( z_{j_0} \) is the only col to consider in the expansion!

A classical example of the modification of the set of relevant cols and its influence on the results, is the case of the Airy function:

\[
\text{Ai}(N^{2/3} \zeta) = \int_{0e^{-i\pi/3}}^{\infty e^{i\pi/3}} \exp[N(\frac{z^3}{3} - \zeta z)] \, dz
\]

and was treated in Ursell (1965). The asymptotic expansions for large \( N \) present discontinuities along the lines (Stokes lines): \( \arg \zeta = 0, 2\pi/3, 4\pi/3 \), delimiting regions where successively the first col, then both and then the second one are relevant. Meanwhile the results stay continuous up to exponentially small errors. This is a consequence of the set of relevant cols being modified. Note that this happens for certain values of \( \zeta \) which define the boundaries of three regions in the \( \zeta \)-space. For practical application the determination of this set is crucial and needs to be studied in more details.

2. The coalescence of several cols

The second kind of failures showed by steepest descent expansions occurs when for a certain value of \( \bar{\alpha} \), say \( \bar{\alpha}_0 \), two or more cols are coalescing, that is: \( z_1(\bar{\alpha}_0) = z_2(\bar{\alpha}_0) = \ldots = z_m(\bar{\alpha}_0) = z_0 \). Then the saddle point \( z_0 \) is a col of order \( m \). Although the usual method of steepest descents can be modified to treat cols of higher order by defining the mapping \( u \) using the equation:

\[
f(z) - f(z_0) = -\frac{u^{m+1}}{m+1}
\]

the expansions thus obtained fail as soon as \( \bar{\alpha} \neq \bar{\alpha}_0 \). The need for an expansion uniformly valid in a neighborhood of \( \bar{\alpha}_0 \) motivated the pioneering work of Chester, Friedman and Ursell, “An extension of the method of steepest descents” (Chester et al., 1957) for the case of two nearly coalescent cols. The principle was to use the Airy function in place of the Gamma function, initiating the idea of using generalized Airy
functions or canonical integrals for two or more coalescing cols. This generalization to the case of many cols was made possible by the work of Levinson (1960a, b, 1961) on the reduction of functions with \( m \) coalescing saddle points to canonical polynomials of degree \( m + 1 \), stimulating several works on the topic: Bleistein (1967), Ursell (1972) and Martin (1973).

The values of \( \alpha \) for which coalescence of several cols occurs, form a subset of the values of \( \alpha \) for which the set \( Z_R \) might be modified. In a recent theory of differential topology – catastrophe theory, Thom (1977) – this subset is named the set \( K \) of catastrophes of \( f \). The first applications of catastrophe theory to the evaluation of integrals with a large parameter are due to Connor (1976). Connor uses the theory to broaden the applications of the generalized method of steepest descents to integrals in several dimensions through the classification of canonical integrals into Thom’s seven elementary catastrophes. The results of Connor (1976) are noticeable in the fact that they are applied to practical examples in quantum mechanics. In the same context the works of Berry in optics should be mentioned (see Berry, 1976).

3. The coalescence of cols with algebraic singularities and end-points

We allow \( I \) to present more complex features as we do not assume anymore that \( W \) is the entire complex plane \( \mathbb{C} \) and that the end-points of \( C \) are “sufficiently” distant from any of the cols. We also include the cases where zeros of the function \( g \) coalesce with saddle points of \( f \) in order to treat the most general cases. As \( W \) can now be any kind of domain of \( \mathbb{C} \) satisfying (iii) in 1.1.1, algebraic singularities of \( g \), that is poles and branch points, can approach cols of \( f \). (The treatment of algebraic singularities apart from this problem of coalescence is part of contour integration and is understood to be taken care of in such a way: necessary residus appearing when \( C \) is deformed into \( C' \) are supposed to be added to (1.3).)

Modifications of the usual method of steepest descents are necessary for the simplest cases of the coalescence of a col and an end-point (Wong, 1989, § VIII.3), and
of a col and a pole (Van der Waerden, 1950, or Wong, 1989, § VIII.2). As the number of coalescing critical points (i.e. cols, zeros, poles, branch-points and end-points) involved increases, the practicality of these uniform expansions becomes doubtful (Bleistein, 1967). But they provide the greatest simplicity that can be hoped for if the full generality and uniformity of $I$ need to be retained in its asymptotic expansion.

In addition to the modifications due to the coalescence of, say $m$ saddle points of $f$, additional changes are necessary, based upon the following ideas:

— When a pole, zero or branch point of $g$ is involved, the canonical integrals are modified to include a pole, zero or branch point being the image of the corresponding point of $g$. These functions are then called *generalized Weber functions* (Bleistein, 1967).

— When an end-point is involved (thus not at infinity), the canonical integrals are defined along a contour with a finite end-point and called *incomplete* canonical integrals. The contribution from the end-point is to be added in the expansion. (The case $m = 1$ is special in the fact that a linear term is to be included in the definition of the mapping as given by (1.7).)

Here again, as for the case of the coalescence of many cols, we must emphasize that the difficulties do not arise when all the critical points are coalescent, but when seeking a uniform expansion as they coalesce.

### 1.2 The generalized method of steepest descents

We now make the assumptions of 1.1.1 without any other restrictions. As we have seen, certain steps of the procedure require a different treatment and we detail them hereafter, while the others stay unchanged, namely the steps A) and B).
1.2.1 The changes of variables

For the mapping $u$ to be holomorphic and bijective we need to modify its definition equation by taking into account the coalescence of many cols. Let $(z_j(\bar{\alpha}))_{1\leq j \leq m}$ be $m$ cols of order 1 of $f$ which are coalescing only, and only if $\bar{\alpha} = \bar{\alpha}_0$. That is we have:

\[
\begin{align*}
    z_1(\bar{\alpha}) &= z_2(\bar{\alpha}) = \ldots = z_m(\bar{\alpha}) = z_0 & \Leftrightarrow & & \bar{\alpha} = \bar{\alpha}_0 \\
    \frac{\partial f}{\partial z}(z_j(\bar{\alpha}), \bar{\alpha}) &= 0 & \forall \bar{\alpha} \text{ in a neighborhood of } \bar{\alpha}_0 \\
    \frac{\partial^k f}{\partial z^k}(z_j(\bar{\alpha}), \bar{\alpha}) &= 0 & 2 \leq k \leq m & \Leftrightarrow & & \bar{\alpha} = \bar{\alpha}_0 \\
    \frac{\partial^{m+1} f}{\partial z^{m+1}}(z_0, \bar{\alpha}_0) &\neq 0
\end{align*}
\]

In fact we could have a fewer number of cols of higher orders and the results would not be changed. (What matters is the total number of cols of $f$ in a neighborhood of $z_0$ counted with their order of multiplicity.) The simplest form for such a function $f$ would be defined by:

\[
\frac{\partial f}{\partial z}(z, \bar{\alpha}) = \prod_{j=1}^{m} (z - z_j(\bar{\alpha}))
\]

It is a known result - Cartan (1961) - that for $\bar{\alpha} = \bar{\alpha}_0$, the mapping defined by $(1.18)$ is bijective and holomorphic in an open disk $D_0(z_0, \rho)$. Therefore we can expect a similar result with a full polynomial of degree $m + 1$ for $f$ for $\bar{\alpha} \neq \bar{\alpha}_0$. This transformation of $f$ into a canonical form - or Levinson's transformation - is given by applying Levinson's theorem; see Levinson (1961). We then have the following results (Martin, 1973):

*There exists a neighborhood of $z_0$ in $W$, $W_1$, a neighborhood of 0 in $C$, $U_1$, and a neighborhood of $\bar{\alpha}_0$ in $C^1$, $D_1$, such that there exists:*

\[
\begin{align*}
    u : & \quad W_1 \rightarrow U_1 \\
    z & \mapsto u(z, \bar{\alpha}) = u
\end{align*}
\]
uniformly regular and bijective, and its inverse:

\[ z : U_1 \rightarrow W_1 \]

\[ u \mapsto z(u, \alpha) = z \]

such that:

\[
f(z, \alpha) = \frac{u^{m+1}}{m+1} + \sum_{i=1}^{m-1} \xi_i(\alpha) u^i + A(\alpha) \equiv p_{m+1}(u, \xi(\alpha)) + A(\alpha)
\]

(1.19)

The functions \( \tilde{\xi} = (\xi_i) \) and \( A \) are regular functions of \( \alpha \) in \( D_1 \), and \( \tilde{\xi}(\alpha) = 0 \Leftrightarrow \alpha = \alpha_0 \). \( D_1 \) may be chosen small enough so that \( W_1 \) is a neighborhood of all the cols \( z_j(\alpha) \), \( 1 \leq j \leq m \).

On applying this result, we can perform the change of variable \( z \rightarrow u \) in the integral:

\[
\int_{\Sigma(z_j)_m} g(z, \alpha) \exp[Nf(z, \alpha)] \, dz
\]

and obtain:

\[
\exp[NA(\alpha)] \int_{\gamma_m} G(u, \alpha) \exp[Np_{m+1}(u, \tilde{\xi})] \, du
\]

(1.20)

with:

\[ G(u, \alpha) = g(z, \alpha) \frac{dz}{du} \]

(1.21)

\( \gamma_m \) being the image by \( u(., \alpha) \) of \( \Sigma(z_j)_m \), a path included in \( U_1 \) and passing through the relevant cols among \( (z_j)_m \).

- The regularity of \( z \) is required to ensure that \( \partial z / \partial u(u, \alpha) \) stays finite as \( \alpha \rightarrow \alpha_0 \), and necessitates the use of a polynomial of degree \( m+1 \), the simplest function to have \( m \) coalescing saddle points. At this stage we do not need to consider the coalescence of any other critical points.

42
The unfolding of a singularity

This process of transforming $f$ into a canonical form, is an important part of catastrophe theory (Thom, 1977) and the central topic of singularity theory (due to H. Whitney in 1955). It is called the unfolding of a singularity. (The term singularity here is not to be mistaken with the singularities of $f$ or $g$ as functions of complex variables. Subsequently we will use the term catastrophe to avoid any confusion.) $\alpha_0$ is then an element of the set of catastrophes $\mathcal{K}$ of $f$, that is the set of values of $\alpha$ for which there is coalescence of saddle points of $f$. Although catastrophe theory is entirely qualitative, it places the problem of the coalescence of saddle points in the method of steepest descents in a much broader perspective. Hence the results valid for one dimensional integrals can be extended to higher dimensional integrals using Morse's lemma and the Splitting Lemma.

The dimension of the space of variables in the mapping (here 1) is called the co-rank of the catastrophe whereas the number of parameters involved in the definition of the unfolding (here the dimension of the $\xi$-space, $m - 1$) is the co-dimension of the catastrophe. Catastrophe theory classifies the singularities into elementary catastrophes and this is a strong result of the theory that this classification only depends on the co-rank, $k$ and the co-dimension of the catastrophe. Hence $n$-dimensional integrals can be reduced to expansions involving $k$-dimensional canonical integrals in the neighborhood of a catastrophe.

Thom has classified seven elementary catastrophes for real mappings (six for complex mappings since sign differences disappear in complex variables) of which we reproduce the most classical ones found in asymptotic evaluations of integrals in Table 1.1.

These are the first three out of four cuspoid catastrophes (singularities of co-rank 1), the others considered by Thom, being the first two umbilic catastrophes (singularities of co-rank 2).

All of these catastrophes correspond to physical phenomena that can be observed in nature. By example the hyperbolic umbilic (co-rank 1 and co-dimension 3) can
catastrophe | unfolding | canonical integrals
--- | --- | ---
fold | $p_3(u, \xi_1) = \frac{u^3}{3} + \xi_1 u$ | Airy function

cusp | $p_4(u, \xi_1, \xi_2) = \frac{u^4}{4} + \xi_2 u^2 + \xi_1 u$ | Pearcey's integral

swallowtail | $p_5(u, \xi_1, \xi_2, \xi_3) = \frac{u^5}{5} + \xi_3 u^3 + \xi_2 u^2 + \xi_1 u$ | swallowtail integral

Table 1.1: The first elementary catastrophes.

be seen in a breaking wave crest (Thom, 1977, p. 78). The Airy function is present in numerous solutions to physical problems like in semi-classical collision theory in quantum mechanics, in optics (Berry, 1976), in acoustics and in hydrodynamics as we will see later. The cusp integral appears in the work of Pearcey on electromagnetic fields near a "cusp". The swallowtail integral plays an important role in optics: see Gilmore (1981).

1.2.2 The local expansions

Once the mapping $u$ is defined, it remains to expand the new function $G(u, \alpha)$ in terms of $u$. If for $m = 1$ the entire series is used, as soon as $m \geq 2$ different expansions have to be used in order for the final expansion to be asymptotic. The first use of such expansion is found in Chester et al. (1957) for the case $m = 2$. The procedure for an arbitrary $m$ was later formalized and studied in Bleistein (1967) and is known as the Bleistein sequence. Here again we need to distinguish the cases of neighboring algebraic singularities since then the canonical integrals differ.
a) The coalescence of several cols

When \( m \) cols coalesce with no neighboring critical points, the expansion of \( G \) is obtained by defining the following function sequences \((G_k(u, \bar{\alpha})), (H_k(u, \bar{\alpha})), \) and \((P_\nu^{(p)}(\bar{\alpha}))\) for \( u \) in \( U_1, \bar{\alpha} \) in \( W_1 \):

\[
\begin{align*}
G_0(u, \bar{\alpha}) &= G(u, \bar{\alpha}) \\
G_k(u, \bar{\alpha}) &= \sum_{p=0}^{m-1} P_k^{(p)}(\bar{\alpha}) u^p + p_{m+1}'(u, \bar{\alpha}) H_k(u, \bar{\alpha}) \\
G_{k+1}(u, \bar{\alpha}) &= -\frac{\partial H_k}{\partial u}(u, \bar{\alpha}) \quad \text{for } k \geq 0
\end{align*}
\] (1.22)

The existence of such sequences follows from the Weierstrass Division Theorem. The choice of these sequences is motivated by the fact that the second term in the definition of \( G_k(u, \bar{\alpha}) \) vanishes at each of the saddle points of \( f \). By repeated integration by parts, we obtain from (1.20):

\[
\int_{\gamma_m} G(u, \bar{\alpha}) \exp[N p_{m+1}(u, \bar{\xi})] \, du \\
= \sum_{p=0}^{m-1} \sum_{k=0}^K P_k^{(p)}(\bar{\alpha}) N^{-k} \int_{\gamma_m} u^p \exp[N p_{m+1}(u, \bar{\xi})] \, du \\
+ N^{-(K+1)} \int_{\gamma_m} G_{K+1}(u, \bar{\alpha}) \exp[N p_{m+1}(u, \bar{\xi})] \, du
\] (1.23)

• The term in \( G_{K+1} \) is of the same nature as the integral in \( G \) except for the factor \( N^{-(K+1)} \), hence enabling the process to be repeated up to any order \( n \). The expansion makes use of the integrals:

\[
\int_{\gamma_m} u^p \exp[N p_{m+1}(u, \bar{\xi})] \, du
\]
which, if we extend $\gamma_m$ to infinity into $\Gamma_m$ and neglect the corresponding error, can be expressed as:

$$N^{-(p+1)/(m+1)} \int_{\Gamma_m} u^p \exp[p_{m+1}(u, X)] \, du$$

(1.24)

with $X = (X_i), X_i = N^{1-i/(m+1)} \xi_i$. If we denote by $U^{(p)}_m(X, \Gamma_m)$ the integral present above, we then have as an asymptotic expansion:

$$\int_{S(z, \bar{\alpha})} g(z, \bar{\alpha}) \exp[Nf(z, \bar{\alpha})] \, dz \sim \exp[NA(\bar{\alpha})]$$

$$\times \sum_{p=0}^{m-1} N^{-(p+1)/(m+1)} U^{(p)}_m(X, \Gamma_m) \sum_{k=0}^K P^{(p)}_k(\bar{\alpha}) N^{-k}$$

(1.25)

Contour integral expressions of the sequences are also given by Bleistein:

$$H_k(u, \bar{\alpha}) = \frac{1}{2i\pi} \oint \frac{G_k(t, \bar{\alpha})}{(t-u) P_{m+1}'(t, \xi)} \, dt, \quad k \geq 0$$

(1.26)

and:

$$R_k(u, \bar{\alpha}) = \sum_{p=0}^{m-1} P^{(p)}_k(\bar{\alpha}) u^p = \frac{1}{2i\pi} \oint \frac{P_{m+1}'(t, \bar{\xi}) - P_{m+1}'(u, \bar{\xi})}{(t-u) P_{m+1}'(t, \bar{\xi})} G_k(t, \bar{\alpha}) \, dt, \quad k \geq 0$$

(1.27)

where the contours should include all of the points $u$ and $u(z_j(\bar{\alpha})), 1 \leq j \leq m$.

In the same manner one could obtain the coefficients $P^{(p)}_k(\bar{\alpha})$ of the expansion by contour integration around $0$ of $R_k(u, \bar{\alpha})/u^{p+1}$. These sequences are useful for they define the expansion for any $m$ in a systematic manner.

* Proofs of the asymptotic nature of the series are given in Bleistein (1967) but for regions of validity that are shrinking as $N \to +\infty$. A different proof was given in Ursell (1972) where it was showed that indeed these expansions are valid in neighborhoods of $\bar{\alpha}_0$ of constant size. This result was extended by Martin (1973) who found
that these regions have definite geometrical shape determined by the properties of the functions $f$ and $g$ and the contour $C$.

b) The coalescence of several cols and algebraic singularities

Apart from the coalescence of $m$ saddle points of order 1 for $\bar{\alpha} = \bar{\alpha}_0$, we assume here that $g$ can be written, see Bleistein (1967, eq. 3.1):

$$g(z, \bar{\alpha}) = \prod_{\mu=1}^{q} (z - b_\mu(\bar{\alpha}))^\nu h(z, \bar{\alpha})$$

(1.28)

where $r_\mu \in \mathbb{R}$, $h$ and $1/h$ are holomorphic in $W_1 \times D_1$.

Consequently the integral to be considered in $U_1$ is:

$$\int_{\gamma_m} \prod_{\mu=1}^{q} (u - \beta_\mu(\bar{\alpha}))^\nu G(u, \bar{\alpha}) \exp[N \mathbf{p}_{m+1}(u, \xi)] \, du$$

(1.29)

where $G$ is defined by:

$$G(u, \bar{\alpha}) = \prod_{\mu=1}^{q} \left( \frac{z - b_\mu(\bar{\alpha})}{u - \beta_\mu(\bar{\alpha})} \right)^\nu h(z, \bar{\alpha}) \frac{dz}{du}$$

(1.30)

The definition of $G$ is such that all of its poles and branch points have been removed by introducing the points $\beta_\mu(\bar{\alpha})$ defined by the (mapping) conditions:

$$f(b_\mu(\bar{\alpha}), \bar{\alpha}) = \mathbf{p}_{m+1}(\beta_\mu(\bar{\alpha}), \xi(\bar{\alpha})) + A(\bar{\alpha}) \quad , \quad 1 \leq \mu \leq q$$

(1.31)
From there the expansion is defined using the following function sequences:

\[
\begin{align*}
G_0(u, \tilde{\alpha}) &= G(u, \tilde{\alpha}) \\
H_k(u, \tilde{\alpha}) &= \frac{1}{2i\pi} \int \frac{G_k(t, \tilde{\alpha})}{(t - u) \Psi(t, \tilde{\alpha}, \tilde{\beta})} dt , \quad k \geq 0 \\
R_k(u, \tilde{\alpha}) &= \frac{1}{2i\pi} \int \frac{\Psi(t, \tilde{\alpha}, \tilde{\beta}) - \Psi(u, \tilde{\alpha}, \tilde{\beta})}{(t - u) \Psi(t, \tilde{\alpha}, \tilde{\beta})} G_k(t, \tilde{\alpha}) dt \\
&= \sum_{p=0}^{m+q-1} P_k^{(p)}(\tilde{\alpha}) u^p , \quad k \geq 0 \\
G_{k+1}(u, \tilde{\alpha}) &= -\prod_{\mu=1}^{q} \frac{1}{(u - \beta_\mu)^{r_\mu}} \frac{\partial}{\partial u} \prod_{\nu=1}^{q} (u - \beta_\nu)^{r_{\nu+1}} H_k(u, \tilde{\alpha}) , \quad k \geq 0
\end{align*}
\] (1.32)

with:
\[
\Psi(u, \tilde{\alpha}, \tilde{\beta}) = \prod_{\mu=1}^{q} (u - \beta_\mu(\tilde{\alpha})) P'_{m+1}(u, \tilde{\xi})
\] (1.33)

Here the contours of integration should include all of the points \(u, u(z_j), 1 \leq j \leq m,\) and \(\beta_\mu(\tilde{\alpha}), 1 \leq \mu \leq q.\)

After substituting \(G\) by its expansion up to the order \(K\) and then extending the path \(\gamma_m\) to infinity into \(\Gamma_m,\) we get:

\[
\int_{\gamma_m} g(z, \tilde{\alpha}) \exp[N f(z, \tilde{\alpha})] dz \\
\sim \exp[NA(\tilde{\alpha})] \sum_{p=0}^{m+q-1} N^{-(|r|+p+1)/(m+1)} U_{m,q}^{(p)}(\tilde{X}, \tilde{\beta}', \Gamma_m) \sum_{k=0}^{K} P_k^{(p)}(\tilde{\alpha}) N^{-k}
\] (1.34)

with \(|r| = \sum_{\mu=1}^{q} r_\mu,\) and the integrals \(U_{m,q}^{(p)}(\tilde{X}, \tilde{\beta}', \Gamma_m)\) being defined by:

\[
U_{m,q}^{(p)}(\tilde{X}, \tilde{\beta}', \Gamma_m) = \int_{\Gamma_m} u^p \prod_{\mu=1}^{q} (u - \beta_\mu')^{r_\mu} \exp[p_{m+1}(u, \tilde{X})] du
\] (1.35)
where \( \bar{\beta}' = N^{1/(m+1)}(\bar{\beta}_\mu) \) and \( \bar{X} = (X_i) \) with \( X_i = N^{1-i/(m+1)}\xi_i \).

- Proof of the asymptotic nature of the expansion as well as the expressions of the function sequences are given in Bleistein (1967) (apart from an error of sign in the expression of \( G_{k+1} \)).

c) The coalescence of critical points with end-points of \( C \)

Apart from the case \( m = 1 \), where the mapping \( u \) has to be redefined (see Section 1.1.5) by:

\[
f(z, \bar{\alpha}) - f(z_1(\bar{\alpha}), \bar{\alpha}) = -\frac{u^2}{2} + \xi_1(\bar{\alpha})u
\]

to account for the coalescence of \( z_1 \) and \( e \), the definition of \( u \) is unchanged for \( m \geq 2 \) and so are the definitions of the function sequences \( (G_k), (H_k), (P_k^{(p)}) \) for both cases a) and b). However another sequence \( (B_k(\bar{\alpha})) \) must be introduced to account for the proximity of \( e \), and is defined by the relations:

\[
B_k(\bar{\alpha}) = -(\pm)H_k(u(e), \bar{\alpha}) \exp[Np_{m+1}(u(e), \tilde{\xi})] \tag{1.36}
\]

for the case a), and:

\[
B_k(\bar{\alpha}) = -(\pm) \prod_{\mu=1}^q (u(e) - \beta_\mu)^{r+1} H_k(u(e), \bar{\alpha}) \exp[Np_{m+1}(u(e), \tilde{\xi})] \tag{1.37}
\]

for the case b). The sign of \( \pm \) is determined similarly as in (1.6) depending on wether \( e \) is a or b; and \( u(e, \bar{\alpha}) \) is the image of \( e \) by \( u \), since \( e \) is assumed to be in \( W_1 \) for \( \bar{\alpha} \) in \( D_1 \). These sequences are obtained as the contribution from \( e \) after the integration by parts of:

\[
\int_{\gamma_m} p'_{m+1}(u, \tilde{\xi}) H_k(u, \bar{\alpha}) \exp[Np_{m+1}(u, \tilde{\xi})] du \tag{1.38}
\]

in a), and:

\[
\int_{\gamma_m} \prod_{\mu=1}^q (u - \beta_\mu)^{r} \Psi(u, \bar{\alpha}, \bar{\beta}) H_k(u, \bar{\alpha}) \exp[Np_{m+1}(u, \tilde{\xi})] du \tag{1.39}
\]
in b), at each step $k$.

The process of extending $\gamma_m$ to infinity is not possible since $u(e)$ is now one of its end-points. Therefore the extended path $\Gamma_m$ has a finite bound and the corresponding canonical integrals are then called incomplete. This changes the final expansions in (1.25) and (1.34) not only through the definitions of the functions $U_{m}^{(p)}(\cdot, \Gamma_m)$ or $U_{m,q}^{(p)}(\cdot, \cdot, \Gamma_m)$ but also by adding the contribution from the sequence $(B_k(\bar{\alpha}))$ in the form of the sum:

$$\sum_{k=0}^{K} B_k(\bar{\alpha}) N^{-(k+1)}$$

(1.40)

to (1.25) and (1.34).

* These results are again found in Bleistein (1967).

### 1.2.3 The canonical integrals

As the last step, we need to evaluate the different canonical integrals. In the ordinary method of steepest descents these are just Gamma functions. When $m \geq 2$ the functions $U_{m}^{(p)}$ are called generalized Airy functions since for $m = 2$, $p \leq 1$, they are expressed in terms of the Airy function $Ai$ and its derivative $Ai'$. Each of the family $(U_{m}^{(p)})$ and $(U_{m,q}^{(p)})$ as well as their incomplete forms can be defined through integral representations and ordinary differential equations: see Bleistien (1967).

a) The generalized Airy functions $U_{m}$

The integral representation for $U_{m}$ is given after (1.24) with $p = 0$:

$$U_{m}(\xi, \Gamma) = \int_{\Gamma} \exp[p_{m+1}(u, \xi)] \, du$$

with $p_{m+1}$ defined as in (1.19). By differentiating $U_{m}(\xi, \Gamma)$ with respect to $\xi_1$ and under the condition that $\Gamma$ is such that the integral is uniformly convergent with
respect to \( \xi_1 \), we obtain the relation:

\[
\frac{d^p}{d\xi_1^p} \mathcal{U}_m(\bar{\xi}, \Gamma) = \int_{\Gamma} u^p \exp[p_{m+1}(u, \bar{\xi})] \, du = \mathcal{U}_m^{(p)}(\bar{\xi}, \Gamma) \quad (1.41)
\]

This suggests that \( \mathcal{U}_m \) satisfies the ordinary differential equation with respect to \( \xi_1 \) (Bleistein, 1967, eq. (I4)):

\[
p_{m+1}'(\frac{d}{d\xi_1}, \bar{\xi}) \mathcal{U}_m(\bar{\xi}, \Gamma) = 0 \quad (1.42)
\]

b) The generalized Weber functions \( \mathcal{U}_{m,q} \)

When neighboring algebraic singularities are involved in the coalescence of \( m \) saddle points, the definition of the canonical integrals (1.35) gives for the particular case \( p = 0 \):

\[
\mathcal{U}_{m,q}(\bar{\xi}, \bar{\beta}, \Gamma) = \int_{\Gamma} \prod_{\mu=1}^{q} (u - \beta_\mu)^{r_\mu} \exp[p_{m+1}(u, \bar{\xi})] \, du
\]

Similarly as for the case of the generalized Airy functions, by differentiation with respect to \( \xi_1 \) and under the same conditions on \( \Gamma \), we have the relation:

\[
\frac{d^p}{d\xi_1^p} \mathcal{U}_{m,q}(\bar{\xi}, \bar{\beta}, \Gamma) = \mathcal{U}_{m,q}^{(p)}(\bar{\xi}, \bar{\beta}, \Gamma) \quad (1.43)
\]

Using this result it can be checked that \( \mathcal{U}_{m,q} \) satisfies the differential equation in \( \xi_1 \):

\[
\left\{p_{m+1}'(\frac{d}{d\xi_1}, \bar{\xi}) \prod_{\mu=1}^{q} \left( \frac{d}{d\xi_1} - \beta_\mu \right) - \sum_{\nu=1}^{q} (r_\nu + 1) \prod_{\nu \neq \mu=1}^{q} \left( \frac{d}{d\xi_1} - \beta_\mu \right) \right\} \mathcal{U}_{m,q}(\bar{\xi}, \bar{\beta}, \Gamma) = 0 \quad (1.44)
\]
The special case \( m = q = 1 \) involves a solution of Weber's equation of order \(-(r_1 + 1)\). Hence, \( U_{m,q} \) is thought of as a *generalized Weber function*.

Both families are called *incomplete* generalized Airy, or Weber functions when \( \Gamma \) is not of infinite extent although they do not satisfy anymore the differential equations (1.42) or (1.44) since the finite end-point of \( \Gamma_m, u(e, \tilde{\alpha}) \), is now a function of \( \tilde{\alpha} \).

These different relations and expressions are not only useful from the theoretical point of view but might be used for numerical computation of these canonical integrals as will be seen later on.

This concludes our presentation of the method of steepest descents and its generalization. Hereafter the latter will be understood when mentioning the *method of steepest descents*. Thus we include in our applications the search for uniform asymptotic expansions of integrals of the sort \( I \) when there is coalescence of several critical points. Although far from being complete, this description mentions all the points that might lead to difficulties in applications of the method. There remains now to effectively detail the practical problems encountered when using the method. This will be done in two steps, one with respect to works previously accomplished by different authors (Chapter 2), the other by fully treating the cases of an end-point then of an isolated col (Chapter 4), and the case of the coalescence of two cols (Chapter 5).
Chapter 2

Practical aspects

From the preceding description of the method, its complexity can be perceived, especially in the uniform treatment of the coalescence of critical points. This complexity found in the form of the expansions will undoubtedly be present in the procedures used to obtain them. By reviewing the difficulties embedded in the method of steepest descents (Section 2.1), the practical problems encountered in its applications may be perceived. In view of this complexity, it is then natural to question the relevance of applying this method to numerical computations. However, the method of steepest descents, through its mathematical nature and the asymptotic properties of its results, presents some irreplaceable features which are needed for certain applications in various areas of physics (Section 2.2). Prior to any implementation we review the treated and untreated practical aspects of the method (Section 2.3) and define our frame of assumptions for practical implementations (Section 2.4).

2.1 Some inherent difficulties

Difficulties exist at each step of the procedure, except for the one concerned with the restrictions of the paths of integration (§ 1.1.3 B)) which is purely theoretical.
They can be detailed accordingly:

(i) **Determining the relevant cols** (§ 1.1.3 A)):

As noted earlier this problem, being omitted in reference books, is rarely mentioned in practical applications. However there is a need to clarify the “mystery” that surrounds the determination of such cols, as pointed out in Choudhari and Kerschen (1990). Ad hoc procedures based on subdominance arguments are not only unsatisfactory but can be erroneous and should be ruled out. The fact that this problem is a *global* question, asserts its difficulty: see Ursell (1970).

(ii) **The definitions of the mappings** (§ 1.1.3 C) and Section 1.2.1):

Here the difficulty is two fold: the appropriate mapping, or unfolding, must be determined, and this mapping must be defined through the computation of $\xi$.

The identification problem implies that the relevant cols, or in other words certain solutions of the equation, $\partial f/\partial z = 0$, have to be found. This task in itself is uneasy and necessitates some suitable analysis for each case in order to make it as simple as possible.

The definition of the mapping through the computation of $\xi$ is directly depending on the dimension of the $\xi$-space, or the co-dimension of the catastrophe. Thus for the cases of an end-point, an isolated col or even the fold catastrophe (at least in theory) there are no major difficulties. However for the cusp (co-dimension 2), the swallowtail (co-dimension 3) and higher order unfoldings, the practical problems raised by the determination of $\xi$ become rapidly complex.

(iii) **Obtaining the local expansions** (§ 1.1.3 D) and Section 1.2.2):

Closely related to the choice of the mapping and to the coalescence of any critical points other than cols, the process of obtaining the coefficients $P^p_k(\vec{a})$ of the Bleistein sequence inherits the same progression of complexity as the number of critical points increases. In principle these coefficients can be computed by solving a triangular system of linear equations and present no difficulty for the cases of end-points or isolated cols. But as soon as a uniform treatment near a
catastrophe is required, this case disappears. Moreover this system of equations being defined from (1.21) or (1.30), the mapping $u$ has to be inverted since $dz/du$ is needed. Except for particular cases, this inversion can be problematic as noted in Copson (1965, p. 91-92).

(iv) The evaluation of the canonical integrals (§ 1.1.3 E) and Section 1.2.3):

Last but not least, the canonical integrals involved in the expansions (see equations (1.16), (1.17), (1.25), (1.34)) whether they are incomplete or not, have to be computed for various values of their parameters. Again the task is fairly limited for end-points (1.14), isolated cols (1.15), and even for the coalescence of two cols since the Airy function and its derivative are well known (see Abramovitz and Stegun, 1964). This is a totally different matter for Pearcey's integral, the swallowtail integral and so on, as these functions are not known. Special procedures must be found to be able to compute such functions and their derivatives with a relative facility. Nevertheless, a degree of ease comparable to the computation of the Airy function cannot be expected, for these functions present higher degrees of complexity.

2.2 Why numerical applications?

Needless to say, that after this short review the prospect of applying the method to practical cases is not exactly the most appealing. This method was formulated long before (nearly a century) the computer era. At this time the only access to new functions was either to recognize and study them in the class of special functions (Bessel functions, error functions, etc...), or to look for their decompositions onto sets of well known functions. After all, this last idea is at the foundation of Fourier analysis and all integral transforms (whether they be Fourier, Laplace, Mellin or Hankel transforms) and seems, at least now, rather natural. In fact at that time – one has to go back to the XVIII-th century – the interest was much more to capture the
behavior(s) of a solution to a particular problem, than to investigate the legitimacy of such decompositions. This is precisely what was at the root of the use of asymptotic approximations. After nearly two centuries of wonder by mathematicians such as Euler, Laplace or Legendre, about the accuracy resulting from considering only the first few terms of certain diverging series, the formalism of asymptotic expansions finally emerged in the work of Poincaré in 1886. Hence asymptotic approximations were widely used before their concept was defined (although the term asymptotic is still subject to discussion), and this primarily to obtain numerical results (the computation of Euler’s constant γ up to 15 decimal places is an example). Even today one can admire the wonderful properties of such approximations.

But since the introduction and the generalized use of digital computers, “unknown” functions and especially integrals can be accessed through numerical computations. Using the fact that elementary operations can be performed without error (apart from the finite representation of numbers) in a short time, various algorithms have been devised to perform numerical integration. Ranging from the simplest ones to more elaborate ones, they are often applications of well known and fairly old (from the XIX-th century and earlier) methods of numerical analysis. Problems arise of course when singularities are present in the integrand over the range of integration, or when this range is of infinite extent. There, more analysis is often required to deal with these singularities (to remove them for example) and more sophisticated algorithms have to be devised which accelerate the speed of convergence of partial sums, such as: Shanks transformation (1955), Richardson extrapolation, Padé approximants, use of continued fraction representations. Some of these methods are much more recent and really belong to computational methods. The situation becomes more complex when the integral is slowly converging and the integrand is highly oscillatory over part of integration range. Such an example is famous in hydrodynamics and is encountered in the problem earlier mentioned (see Introduction) of the Kelvin wave source potential.
Example 1:

The wave like disturbance of the Kelvin wave source can be expressed, see Newman (1987,b , eq. (4) with $y \neq 0$), as:

$$f(X,Y,Z) = -8 \int_0^\infty \sin X\sqrt{t^2 + 1} \cos Yt\sqrt{t^2 + 1} \exp[-Z(t^2 + 1)] \, dt \quad (2.1)$$

which must be evaluated for $X < 0$, $\rho = \sqrt{Y^2 + Z^2} > 0$, and $Z \geq 0$. The case of $Z = 0$ (when both the source and the field point are on the free-surface) and $X$ and $y$ "large", is of course the worst, as the integral is not even absolutely convergent and the integrand highly oscillatory. Numerical integrations are limited by cancellation errors between contributions from successive humps and hollows, the picture getting worse as the ‘pulsations’ $x$ and $y$ increase (see Baar and Price (1988) for a graph of the integrand). The fact that the range of integration is infinite is spurious since by letting $t = \tan \beta$ the range becomes then finite, but bringing no improvements. Some numerical techniques, specifically developed for very oscillatory functions and based on non-linear extrapolations could be used for $Z > 0$ (Sidi, 1982) but not for the case $Z = 0$ as they require an absolute convergence of the integrand.

- Kelvin treated his expression of (2.1) by using his principle of stationary phase (Kelvin, 1887b) for large distances away from the source. Since then many efforts have been made to compute this integral or related forms of it.

- A procedure used by Tuck, Collins and Wells to assess ship wakes using thin-ship theory, “mimicked” numerically the method of stationary phase: Tuck et al. (1971, App. 1). The integrand was integrated along the real axis in the neighborhoods of the stationary phase points, and a fading factor was included at the edges of the intervals of integration to avoid dependence of the results on these arbitrary end-points. The problem posed by the coalescence of the two stationary phase points where the intervals would overlap, was treated by a simple test. When there was no stationary phase point, the integration was performed at the inflection point of the phase function. The fading factor was determined experimentally. Another characteristic is that the integral was evaluated at points of a fixed discrete grid in $X$ and $Y$.
• Baar and Price showed that applying to numerical computations expressions derived by Bessho (namely series expansions in terms of Bessel functions), was more efficient than numerical integration techniques (Baar and Price, 1988). Thorough analysis of the derivations of these series are found in Ursell (1988) and Clarisse (1989), and numerical procedures are given in Newman (1988) and Clarisse (1989). Apart from leaving a domain of the parameter space untreated, these expressions are very specific and cannot be applied to other cases. But they demonstrate the relevance of asymptotic methods in the treatment of a singular behavior (here for \( Z = 0 \) and \( \rho \to 0 \)) that no numerical procedure could capture.

• More recently Iwashita and Ohkusu have proposed a numerical steepest descent integration – similar to an implementation of Perron’s method – of the same function (although in a different form); see Iwashita and Ohkusu (1989). Their idea was to integrate numerically the function along its paths of steepest descent, in order to take advantage there of its minimal oscillations and maximal convergence in modulus. This approach however raises the problem of determining such paths – a task that must be done at each evaluation – far enough from the saddle point to ensure a minimal truncation error. The accuracy achieved is as high as 8 significant digits, but the lowest value of the submergence \( Z \) considered by Iwashita et al. is \( 10^{-3} \). It must be noted that the singular behavior of the integral as well as the problem of the coalescence of the two saddle points which occurs for \( Z = 0 \) and \( X/Y = 2\sqrt{2} \) and which would certainly result in numerical problems in determining the paths, are thus avoided.

Another example taken from hydrodynamics is concerned with the solution to the time-domain problem, or Cauchy-Poisson problem, in finite depth.
Example 2:

In linear free-surface wave theory, the response to a Dirac perturbation of the free-surface elevation at the origin and at \( t = 0 \), is given by (Whitham, 1974, eq. 13.33):

\[
\eta(R,t) = \frac{1}{2\pi} \int_0^\infty J_0(kR) \cos \omega(k) t \, k \, dk \quad t > 0
\]  

(2.2)

where \( \omega(k) = \sqrt{gk} \tanh kh \), \( h \) being the depth and \( R \) the radial distance from the origin.

Evaluations of \( \eta \) (or of the corresponding potential) are needed for any value of \( t \) and \( R \). The use of panel methods for solving realistic problems impose the same constraints of robustness and efficiency as for the Kelvin wave source potential. Efficiency is even more important since solutions are obtained by convolution in time. Asymptotic methods are then appropriate for large times \( t \) and distances \( R \).

Some features need to be highlighted. The two-dimensional solution shows the coalescence of two cols at the wave front characterized by the presence of an Airy function (see Rayleigh, 1909, Whitham, 1974). A similar result could be expected for (2.2) but in fact it appears to be more complex: see Kajiura (1963), Newman (1990), Ursell (1991) and Chapter 7.

The above examples are specific to hydrodynamics, but similar problems appear in other fields (see Introduction).

2.2.1 Numerics and physical understanding

Each application of these various fields has its own requirements and peculiarities. For instance semi-classical theory in quantum mechanics is specifically concerned with uniform treatments of coalescences of several cols – catastrophes – for construction of interference figures and assessments of semi-classical trajectories in particle collisions: see Mount (1973), Connor (1976). In fact the method of steepest descents and
its numerical implementation have found numerous applications in this field, first by applying the results of Chester et al. (1957), then the more general results of Bleistein (1967) and Ursell (1972). There the interest was double in obtaining accurate numerical results with efficiencies that proved to be higher than numerical integration techniques in the limits of large $N$'s (in fact $N = 1/h$ with $h \to 0$), and in providing theoretical indications for the understanding of the physical phenomena studied: short wave-length scattering or caustics. Robustness is necessary but in a restricted domain (the vicinity of the catastrophe), efficiency is not crucial as these evaluations are final answers, but appear to be better than other methods (Mount, 1973). It must be emphasized that only the computations of the terms $P_0^{(p)}(\tilde{\alpha})$ are needed.

The role of this theoretical knowledge is probably as important if not more than the numerical capabilities of the method for these specific applications. This became especially true after the emergence of catastrophe theory: see Berry (1976), Connor (1976). However as these functions present highly complex behaviors and are not known, the only access to their features is through numerical values (see Connor (1980, Fig. 2, 3 & 4) for an example of such a function). This duality of numerical efficiency and theoretical power is uniquely combined in the method of steepest descents (among methods for computing integrals with a large parameter). Current investigations are aiming at extremely high accuracy through "exact" asymptotics or hyperasymptotics (Berry, 1992).

A steepest descent analysis is also particularly suitable to the study of instabilities as shown in Choudhari and Kerschen (1990). In fact such analysis are applicable to more general problems of acoustics. Their relevance can be perceived from simple remarks.

If a wave solution is obtained as an integral of the form $I$, its behavior for large values of $N$ – whether distance or time – can be assessed by the method of steepest descents. Expressed now as the sum of contributions from certain saddle points, this solution can be decomposed as the superposition of "distinct" waves. Indeed, to
each col $z_j$ is associated, along the direction of increasing $N$'s, a growth rate, namely $\Re\{f(z_j, \alpha)\}$, and a frequency, $\Im\{f(z_j, \alpha)\}$. Each of these waves is then given, in first approximation, by the leading term of the contribution of the corresponding col. Strong interferences occur when two (or more) waves present similar growth rates and frequencies in the same region of the $\alpha$-space. This simply corresponds to the vicinity of a catastrophe, i.e. the coalescence of saddle points. Instabilities are identified with components of positive growth rate.

The determination of the relevant cols is obviously crucial and indicates the regions of the parametric space ($\alpha$-space) in which unstable solutions exist. The robustness of the computations depends of course on the ability to solve this problem. It must be noted that this is not the case in the uniform treatment of catastrophes, and hence in quantum mechanic applications. The same remark as above concerning the efficiency of the method can be made as integrations are performed for the computations of final quantities. Only the leading terms in the expansions are needed although higher order terms may be required for improved accuracy and validity.

If the physical interpretations here rely on much simpler concepts than catastrophe theory (note that the study of catastrophes pertains as well to this approach), they are of equal importance for acoustics (and even hydrodynamics). Future developments in these fields might show an increasing interest in the uniform treatment of catastrophes.

A different aspect of the method has been emphasized, namely the importance of the question of the relevant cols. But the combination of theoretical and numerical results remain a characteristic of the application of the method.

2.2.2 Robustness and efficiency guaranteed by the theory

The requirements in hydrodynamics are much more technical. Thus for the Kelvin wave source potential, numerical evaluations of the potential and its derivatives with
respect to $X$, $Y$ and $Z$ are needed with a certain accuracy throughout the physical domain. Procedures must be developed to satisfy these criteria with the additional constraints of efficiency since they will be used in panel codes for forward speed problems. Typically such codes involve the construction of a $P \times P$ matrix, where $P$ is the number of panels discretizing the body. Each coefficient of the matrix is obtained by computing an integral containing (2.1). Since the number of panels can be of the order of 1 000, speed is necessary. As said earlier, special expressions have been developed to compute the wave like part of the Kelvin wave source potential in various regions of the physical domain allaying robustness and high efficiency. However a part of the domain for large distances downstream of the source remains untreated and appears to be suitable for the application of the method of steepest descents. This has been done analytically by Hogner (1923), Peters (1949) and Ursell (1960), for $Z = 0$.

Robustness requires to be able to compute the coefficients $P_k^{(p)}(\tilde{a})$ up to any order $k$ for the desired accuracy to be satisfied. Once this accuracy is ensured for the smallest values of $N$, robustness is guaranteed throughout the whole domain by the asymptotic nature of the expansions. These very same asymptotic properties ensure a bound on the computational time over this domain, thus providing a minimum efficiency. Moreover, if the method is applied by following the theory, explicit bounds on the errors are provided. This association of analytical formulas with their corresponding numerical results and error bounds are then useful for deriving procedures to compute derivatives of the potential.

While totally absent from hydrodynamics, catastrophe theory through some of its results could lead to a better physical understanding of the regions of the critical lines $|\theta| = 19^\circ 28'$ in ship wave patterns, or of the wave front in the Cauchy-Poisson problem. Interpretations similar to the ones mentioned in acoustics can also be applied to ship wave patterns and the problem of their detection (see Tuck et al., 1971).
2.2.3 The characteristics of the method

We can now characterize numerical applications of the method of steepest descents from the different examples mentioned above.

1. Efficiency and robustness

The first group of features is closely linked to the mathematical properties of steepest descent expansions. These are:

— the robustness of the implementation on a domain of the $\bar{\alpha}$-space on which a given expansion is valid. The "global" robustness is then ensured by the uniform treatment of the coalescence of critical points, by solving the problem of the relevance of the cols and by the matching of the different expansions on intersections of their regions of validity. This pertains to the uniformity of the expansions with respect to $\bar{\alpha}$.

— the efficiency guaranteed by a minimum value over each domain of validity. This minimum efficiency is given by the accuracy/computational time ratio obtained for the portion of the border of this domain for which $N$ is minimal. For larger values of $N$, higher accuracy will be achieved and/or fewer terms will be needed at constant accuracy. This is a natural consequence of the use of asymptotic expansions. The control of the error over the $(N, \bar{\alpha})$-space is both numerical and theoretical through the analytical formulas provided for its bounds.

2. "Classical" wave analysis

This aspect is imparted by the class of physical problems for which the method can be used. However the validity of such analysis is entirely deriving from the power of the method of steepest descents to reduce a global integration into local expansions.
3. The understanding of special phenomena

This is probably the most striking quantitative application of catastrophe theory, which is initially a qualitative theory. The emergence of this theory despite its recent age, induced many developments in different fields whether in mathematics, physics (quantum mechanics and optics) or biology. The method of steepest descents has and will benefit from this interaction as well as other fields for which catastrophe theory can intervene through the use of the method. This leaves open a whole spectrum of future applications.

4. The generality of the method

Finally the method is applicable to a wide variety of functions. This includes the treatment of partial derivatives of a given function once it has already been treated by steepest descent. Of course this generality is limited by the applicability of the method.

Requirements of certain applications

The risk in choosing to implement a method of applied mathematics to numerics, is often to devote much effort in taking care of the theoretical "details". The irony is then that a certain numerical integration method might give comparable results of robustness and efficiency in a much simpler way. But even in terms of efficiency, this effort could turn out to be very rewarding.

When applicable, if a theoretical frame in the form of analytical formulas associated with numerical results is preferable, then the application of the method of steepest descents is highly recommended. This of course corresponds mainly to numerical considerations. Moreover, as soon as the use of the method appears to contribute to the understanding of physical phenomena, we believe, its use is invaluable.

Quite the opposite to many numerical integration techniques (although the ones
useful for pathological integrals are necessarily specific), the method of steepest descents requires knowledge of its theory in order to be applied. Outside this rigid frame — such as in the case of Perron's method — the benefits of having theoretical results (formulas and bounds on the errors) along with the validity of their interpretations — namely the advantages of the method — are lost (at least partially).

Therefore rigor is necessary in the implementation, if we wish to take full advantage of the method and not to reduce it to a simple numerical technique prone to failures in delicate situations. The difficulty is then to formulate the necessary knowledge in a form suitable for computer implementations. We certainly cannot eliminate yet the necessary analytical work that a scientist or an engineer has to furnish in order to render a given problem suitable for computations (whether by numerical techniques or steepest descent)! Of course additional work is required when using the method of steepest descents (e.g. Chapter 7), and we propose to produce tools to facilitate this analysis as well as the rest of the procedures.

Many contributions have been made toward easier applications of the method and we certainly wish to include their results in our implementations.

2.3 The treated and untreated practical aspects of the method

If we consider our dissection of the method as of Section 2.1, most of the efforts toward its practical applications have been focusing on (ii) and (iv). The question of the relevance of the cols — (i) — is still untreated and we will devote a whole chapter (Chapter 3) to the development of a formalism and of tools (both mathematical and computational) in order to facilitate its treatment.

- In (ii) the problem of locating the roots of the equation, \( \partial f/\partial z = 0 \), cannot be treated in a general context. In fact appropriate analysis is required for each case,
although iterative root finders or fixed point algorithms could be used. In any case it is implied when applying the method, that the locations of these points are accessible.

The second part of (ii), or the definition of the mapping, does not present difficulties for the simplest cases. The case of the fold catastrophe is given in Chester et al. (1957), at least in theory. Much more effort is needed in order to define the cusp map and higher order maps. The general procedure can be described as follows.

From the equation of the mapping (1.19), the necessary condition for the mapping to be bijective, is written for the \( m \) cols:

\[
f(z_j, \bar{\alpha}) = p_{m+1}(u_j, \bar{\xi}) + A(\bar{\alpha}) \quad \text{for} \quad 1 \leq j \leq m
\]

where \( u_j = u(z_j, \bar{\alpha}) \). (Note that prior to this a problem of indexing of the cols needs to be addressed.)

Then, the values of the \( \xi_i \)'s and \( A \) are obtained by solving this system of \( m \) nonlinear equations in the \( m \) unknowns \((\bar{\xi}, A)\), the \((u_j)\) being obtained as solution of the equation, \( p'_{m+1}(u_j, \bar{\xi}) = 0 \).

No explicit solutions are known for the case of arbitrary \( m \). Using the theory of equations and the theory of symmetric polynomials, results for the cusp map \((m = 3)\) have been obtained in Connor and Farelly (1981a). The case of \( m = 4 \) (swallowtail map) is found in Connor, Curtis and Farelly (1984a), by using this algebraic method and an iterative method. The formal generalization of these procedures for arbitrary \( m \) is described in Connor and Curtis (1984b). So far the cases \( m \geq 5 \) have not been treated.

- The determination of the coefficients of the expansions \(- (iii) -\), or the terms of the Bleistein sequence, is in principle straightforward as they are obtained by solving a triangular set of linear equations. Full details are given in Dingle (1973) for the ordinary steepest descent expansions and certain special expansions. The principles are given in Chester et al. (1957), for \( m = 2 \), and can easily be generalized for an
arbitrary $m$. For an isolated col an alternative procedure consists in using contour integration about the image of the col (Copson, 1965, p. 69). One could also consider applying the contour integral formulas given by Bleistein. A discussion of the corresponding problems will be given for $m = 1$ and $m = 2$ in Chapters 4 and 5, respectively.

- The method of steepest descents can be seen as a decomposition of a function (the integral $I$) onto a set of simpler functions, the canonical integrals and their derivatives. This decomposition is in fact a sequence of local projections of the function onto some suitable sets (the germs). Each set is determined from the local "degree of complexity" of the exponent function $f$, i.e. whether there is no col (isolated end-point), one col, two cols (fold catastrophe), three cols (cusp catastrophe), etc... This degree of complexity is embedded in the co-dimension of the catastrophe and appears clearly in the graphs of the different canonical integrals (e.g. Connor, 1983, Fig. 2, 3, 4). The power of the method is then to reduce a global behavior (the integral is over a whole range) into a sequence of local behaviors. From there the local treatments pertain in fact to a completely different part of mathematics, namely singularity theory in differential topology (Duistermaat, 1974). It then appears inevitable to be able to compute efficiently and accurately these canonical integrals, if one wants to access this complexity.

The case of the Airy function is well known (since 1838) and efficient algorithms have been developed to this end; see Schulten, Anderson and Gordon (1979). The task for other values of $m$ ($m \geq 3$) has been undertaken only recently by Connor and others, once the importance of such integrals in quantum mechanics had been revealed. The procedures used by Connor to compute numerically these functions are based on two approaches:

- deformation of the path of integration into steeper paths, followed by numerical integration of the integrand along these paths where its oscillations are less detrimental
— use of the partial differential equations satisfied by each canonical integral (see Section 1.2.3)

These results however are limited to real values of the parameter \( \xi \). Apart from the earlier work of Pearcey, the evaluation of Pearcey's integral – or cusp integral – is treated along these lines in Connor et al. (1981a, b & 1983). Similar procedures for the swallowtail integral are given in Connor et al. (1983 & 1984a).

In view of completing these results, Kaminsky has studied the asymptotic expansions of such canonical integrals, using, of course, the method of steepest descents. From this, it appears that for the study of a given canonical integral (say the swallowtail integral), the results from the analysis of canonical integrals of lower complexities (the Airy function and the cusp integral) are necessary.

In all of the cases the difficulties are directly proportional to the richness contained in these functions and cannot be avoided.

### 2.4 The frame of the proposed implementations

We now propose to define the frame within which implementations of the method of steepest descents will be considered hereafter. To the assumptions of 1.1.1 we add the further restriction that \( f \) and \( g \) be holomorphic in \( C^{l+1} \) (that is, all of their algebraic singularities are at infinity). The cases of singular points at finite distance will not be treated in general, although their impact on the results and the corresponding necessary modifications will be mentioned.

**The goal**

Upon this, the purpose is to obtain, once the relevant cols are determined, the local expansions numerically in the form of the values of their coefficients and the values of the canonical integrals they contain.
The conditions

Toward this aim, we impose:

I - the space of the parameter $\alpha$ to be of "reasonable" dimension $l$ (and in any case finite!), say: $l \leq 10$

II - to be able to locate all of the cols, or at least the relevant ones, of $f$ for any value of $\alpha$ in $D$ (we can always restrict $D$ to be so)

III - that the cols present in the expansions be in a finite and "reasonable" number or at least that, by considering such a number of them, the errors on the results are negligible. (In general an infinite number of cols would require analysis and formal computations which will not be developed here.) This does not imply that the procedures described in what follows could not be applied if this condition was not satisfied.

IV - to have access to the numerical values of the derivatives of $f$ and $g$, with respect to $z$, taken at the cols up to any order (of course finite and "reasonable")

- The qualification of being "reasonable" is imposed by computational constraints. Certain programming environments can handle the concept of infinity and this certainly could be used, but the constraints on the numerical computations necessitate this restriction.

Under these conditions, the goal as stated above can be achieved and the algorithms for the cases of:

- an end-point (Chapter 4)

- an isolated col of order 1 (Chapter 4)

are given, and guidelines are proposed for the case of two coalescing cols of order 1, or fold catastrophe (Chapter 5).
We finally apply these algorithms to our main example of the Kelvin wave source potential (Chapter 6) and discuss the case of the Cauchy-Poisson problem in finite depth (Chapter 7).
Chapter 3

Of basins and cols

We propose here to study the problem of determining the set of relevant saddle points for a given integral $I$. We do not claim by any means to provide a complete treatment of the question. (Any such attempt is way beyond the scope of a single chapter.) For the first time, a rigorous formulation of the problem of deforming the initial path of integration is proposed. Hence the problem of finding the relevant saddle points is shown to be equivalent to the finding of a path in a certain connected graph. Except for values of $\alpha$ in a certain subset of $D$, this graph, denoted $G_f$, is shown to be a tree. Being an intrinsic property of $f$, $G_f$ is modified only through combinations of certain elementary events. Hence it suffices to know $G_f$ for a given $\bar{\alpha}_0$ in order to infer $G_f$ for any other value of $\bar{\alpha}$. We are thus reduced to the problem of determining $G_f$ for a certain $\bar{\alpha}$. At first we treat the case of $f$ holomorphic in $\mathbb{C} \times D$ (Section 3.1), and then study how these results could be extended to the cases of $f$ with singular points (Section 3.2). Finally we review the practical aspects and restrictions in eventual applications. This brings us back to a still unanswered and fundamental question of characterizing certain regions of the complex plane (Section 3.3).

The only previous attempt of treating this problem, at least in relation with the method of steepest descents, can be found in an article by Ursell “Integrals with a large parameter: paths of steepest descent and conformal mapping” (Ursell, 1970).
Therein the problem is formulated as follows:

*Given 2 points \( z = \alpha \) and \( z = \beta \) where \( \Re f(\alpha) > \Re f(\beta) \), can a path of
descent be drawn from \( \alpha \) to \( \beta \)? If also \( \Im f(\alpha) = \Im f(\beta) \) can a path of
steepest descent be drawn from \( \alpha \) to \( \beta \)?*

and its global character is emphasized. Ursell presents a practical method for answering this question based on properties of conformal mappings and on the method of univalent dissection. His approach uses the correspondence between \( W \) and \( f(W) \), and the simple geometry of steepest paths in \( f(W) \). However when the mapping is non-univalent on a given subdomain, an understanding of Riemann surfaces is necessary. This method is of great help when the tracing of steepest paths is unpractical (typically and in most cases when doing it *by hand*). But the use of digital computers renders the task of curve tracing much easier than Ursell's method of conformal mapping is.

In both approaches the action of *tracing* or *drawing* curves is central. Such process simply corresponds to the property of *connectedness* under certain constraints (monotonous variation of a certain function along the desired path). The connectedness of a region of simple shape in two or three-dimensional spaces is easily perceived by a human being. On a computer however, paths must be numerically generated as sequences of numbers in order to infer the connectedness of two points in a given region, unless of course its shape is trivial. Assuming that this problem of finding the relevant cols is solved for a given value of \( \tilde{\alpha} \), the parameter of \( f \), by means of curve tracing or conformal mapping, the whole process would have to be repeated a priori for any other value of this parameter. The task would then be equivalent or even more expensive than any numerical integration of the integrand along these paths of steepest descent. In short there would be little interest in considering steepest descent expansions.

It is worth noting that other problems closely related to the one we are concerned with, are encountered in different areas of physics. Naturally the difficulties faced in their solving are amazingly similar.
Several facets of the same problem

We have seen in Section 1.1 an intuitive interpretation of the process of deforming the initial contour of integration. One can in fact think of several different ways of formulating this problem.

Hence if one considers $z = x + iy \mapsto f(z) = \Re f(z) + i\Im f(z)$ as giving a potential flow in two dimensions, $\Re f$ is the potential function whereas $\Im f$ is the stream function. The saddle points of $f$ are now stagnation points of the flow and the steepest paths, the streamlines. Posing the fundamental question in terms of steepest paths is then asking whether or not two points are on the same streamline. The streamlines ending or originating at the stagnation points play an important role in the physical aspect of the flow, they are the *dividing streamlines* and actually divide the flow in an arrangement of regions usually called the “topology” of the flow. Knowing this topology is extremely close to solving our problem. However this analogy cannot be pursued since we allow points to be connected not only by steepest paths.

Almost identical is the consideration of the gradient system governed by the autonomous system of differential equations:

$$
\dot{x} = \frac{\partial \varphi}{\partial x}(x, y) \\
\dot{y} = \frac{\partial \varphi}{\partial y}(x, y)
$$

with $\varphi(x, y) = \Re f(x + iy)$.

The trajectories are then given by the level curves of $\varphi$, that is the level curves of $f$ as defined in Section 1.1.2. The critical points of the phase flow in the space $(x, y, \dot{x}, \dot{y})$ are all saddle points (if $f$ is holomorphic in $\mathbb{C}$) and are the cols of $f$. The trajectories going through these critical points (i.e. the level curves of $f$ going through its cols) are then the projections of the separatrices of the phase flow. Those define the basins of attractions of the flow. (Note that any attractor is at infinity if $f$ is holomorphic in $\mathbb{C}$.) Again we are interested in the global arrangement of these basins (or at least
Similarly, \( \varphi \) can be assumed to be the potential energy of a unit mass particle. The Hamiltonian of this conservative system is then:

\[
H(x, y, \dot{x}, \dot{y}) = \frac{1}{2} (\dot{x}^2 + \dot{y}^2) - \varphi(x, y)
\]

This defines a phase flow in the phase space \((x, y, \dot{x}, \dot{y})\), satisfying the differential equations:

\[
\begin{align*}
\dot{x} &= \frac{\partial H}{\partial \dot{x}} \\
\dot{y} &= \frac{\partial H}{\partial \dot{y}} \\
\ddot{x} &= -\frac{\partial \varphi}{\partial x} \\
\ddot{y} &= -\frac{\partial \varphi}{\partial y}
\end{align*}
\]

The level sets of the potential energy in the \((x, y)\)-plane are the level curves of \(f\), the trajectories of zero initial velocity being then the paths of steepest descent of \(f\). This model represents by example the dynamics of a point mass sliding without friction on the surface \(\Sigma = \{(x, y, \varphi(x, y))\}\). Identically to the above interpretation, the matter is to characterize the basins of attraction of this flow.

More generally, \(f\) is treated as a mapping from \(W\) to \(\mathbb{C}\). The saddle points of \(f\) are then the points of \(W\) of index of ramification \(> 1\) (see Cartan, 1961). The study of these points belongs to the study of singularities of differentiable mappings as already mentioned in Chapter 1. Singularity theory and catastrophe theory, which have obvious applications for the above problems and for the method of steepest descents (see Section 1.2.1) are local theories. This locality is certainly not useless in view of the recent developments in “non-linear” dynamics of systems, nor it is useless in the method of steepest descents. Furthermore catastrophe theory as being the association of singularity theory and bifurcation theory has important consequences for the study of systems and their evolution. The key concept of catastrophe theory,
or the idea of *forms* and the breaking up of their generation into elementary events – the catastrophes – is relentlessly present in the above examples if one studies their evolution with respect to some parameters. We consider the same situation in the method of steepest descents when we raise the question of identifying the set of relevant saddle points and its modifications with the parameter \( \tilde{\alpha} \). This idea of *form* will be more present than ever in this chapter. Hence the parallel we just proposed is more profound than it appears.

If the power of the concept of *form* is doubtless, we have to face the reality of practical applications for our global problem. There, we are reduced to the process of curve tracing which seems *unavoidable* for now. It is worth noting that the depiction of a potential flow uses streamlines, and that identifying the properties of an attractor in dynamics requires the consideration of trajectories (the geometry of this attractor being well beyond our understanding in non-trivial cases). This similitude with our case is self-explanatory.

### 3.1 Holomorphic functions

In this section, we develop the formalism corresponding to the case of \( f \) being a holomorphic function of \( z \) in \( \mathbb{C} \). This is done first, by defining certain regions of the complex plane and studying their topological properties (Section 3.1.1), and second, by deriving from there more abstract concepts (Section 3.1.2). Finally, we make the parallel between the concept of *ramification* introduced in this second step and the concepts of catastrophe and form found in catastrophe theory (Section 3.1.3).

Since \( f \) is holomorphic, its set of saddle points, \( Z \), is discrete and we suppose that there exists a value of \( \tilde{\alpha} \) in \( D \) such that all the elements of \( Z \) are of order exactly 1. This corresponds to the simplest *form*, in the sense of catastrophe theory, that \( f \) can possibly take.
The set of catastrophes is then defined as:

\[ \text{The set of catastrophes, } K, \text{ is the subset of } D \text{ such that one element, at least, of } Z \text{ is of order 2 or higher.} \]

This definition is refined with respect to the nature of the catastrophes:

\[ \text{The set of catastrophes of co-dimension } q, \ K_q, \text{ is the subset of } D \text{ such that one element at least of } Z \text{ is of order } q + 1. \]

Hence \( K = K_1 \) and we have the sequence of inclusions:

\[ K_1 \supset K_2 \supset \ldots \supset K_q \supset K_{q+1} \supset \ldots \]

The unfolding of a catastrophe of co-dimension \( q \) (note that all catastrophes in the method of steepest descents as defined in Chapter 1, are of co-rank 1) is the local immersion of \( K_q \) into a \( q \)-dimensional \( C \)-vectorial space. These results come from singularity theory.

In the method of steepest descents catastrophes correspond to the coalescence of several saddle points and it is clear that \( Z \) is “modified” when \( \alpha \) goes from \( D \setminus K \) to \( K \), or from \( K_q \) to \( K_{q+1} \). This modification is to be understood as the coalescence of two or several elements of \( Z \). (This notion will become clearer subsequently.) Simultaneously the “topology” of the potential flow as defined at the beginning of this chapter is modified. However this “topology” may change for values of \( \alpha \) not necessarily in \( K \), and may result in a modification of the set of relevant saddle points \( Z_R \). We emphasize that the modification of the latter is not systematic since \( Z_R \) is defined with respect to \( f \) and \( C \) the initial contour of integration, whereas \( Z \) and the sets \( K_q \) are defined only with respect to \( f \).

### 3.1.1 On certain regions and their properties

In this section we define certain regions of \( C \) with respect to \( f \) and its saddle points. These regions and their properties will be used in the next section for more funda-
mental notions.

Steepest paths and crest lines

Since $f$ is holomorphic, steepest paths have well defined properties. Hence a path of steepest ascent issued from a point $b$, $A_b$, is such that its image by $f$ is $\Re f(A_b) = [\Re f(b), +\infty[$, or in other words it satisfies the condition $f(z) = f(b) + t$ with $t$ in $]0, +\infty[$. (The same result bears for $D_b$ a path of steepest descent issued from $b$ with $-t$ in place of $+t$.)

Consequently any steepest path must end at $\infty$, the only singularity of $f$. This implies also that two distinct steepest paths can only intersect at a saddle point of $f$.

The subset of $\mathbb{C}$ made of the union of $Z$ and of all the paths of steepest ascent issued from its elements, or:

$$\tilde{\mathcal{C}} = Z \cup \{A_c : A_c = \{z : f(z) = f(c) + t + t \in ]0, +\infty[\} \text{ and } c \in Z\}$$

is the set of crest lines of the surface $\Sigma$. This set is composed of piecewise $C^\infty$ arcs and is closed in $\mathbb{C}$. Its complementary in $\mathbb{C}$, $\mathbb{C} \setminus \tilde{\mathcal{C}}$, is thus open.

Basins

By considering $\mathbb{C} \setminus \tilde{\mathcal{C}}$, we define the basins of $f$:

basin$^{(1)}$: The basins of $f$ are the connected components of $\mathbb{C} \setminus \tilde{\mathcal{C}}$.

Such regions satisfy the properties:

- Any basin $B$ is open and simply connected (since any path $A_z$ ends at $\infty$).
- The closure of $B$, $\overline{B}$, contains at least an element of $Z$, $c$, and at least a path $A_c$. Consequently the set of basins of $f$ is countable.
If \( c \) is a col of order \( p \), there is at least \((p + 1)\) basins such that each of their closure contains \( c \).

- A path of steepest descent issued from any element of \( B \) is unique and is contained in \( B \).

In fact this last property is characteristic of a basin and provides an equivalent definition:

\textbf{basin}\textsuperscript{(2)}: The basins of \( f \) are the connected subsets \( B \) of \( \mathbb{C} \) such that any path of steepest descent issued from any element of \( B \) is unique and contained in \( B \).

This definition excludes naturally the cols of \( f \) and any path of steepest ascent issued from a col, hence the equivalence holds.

By definition, \( f \) is locally univalent in any of its basins: all the points of a basin have their index of ramification = 1. This leads to the following conjecture, given without proof:

\textit{Conjecture:}

Given a function \( f \) holomorphic in \( \mathbb{C} \) and \( B \) one of its basins, \( f \) is univalent in \( B \).

One can verify the above conjecture on specific cases, by considering by example images by \( f \) of subdomains of \( B \). An attempt of proving this result is most difficult. The problem is easily reduced to showing that a certain mapping is univalent in the unit disc \( D_\circ(0, 1) \): it is clear that \( B \) is different from \( \mathbb{C} \) (unless \( Z \) is empty, a case of no interest here) and thus isomorphic to \( D_\circ(0, 1) \) (Riemann mapping theorem). This new mapping appears to be a slit mapping (see Duren, 1983): the adjunction of two points of \( \partial B \) to \( B \) is sufficient to obtain a non-univalent mapping. There exist sufficient criteria for the univalence of mappings defined in \( D_\circ(0, 1) \), but they require analytical knowledge of the mapping considered. Hence
we would need the explicit knowledge of the isomorphism mapping $E$ onto $D_ε(0,1)$, which cannot be obtained for general cases.

Our definition of a basin is identical to the definition of basins found in hydrology. (See Figures 3-1 and 3-2 for an illustration.)

Valleys and hills

Using a somewhat more classical point of view we now introduce definitions that are close to topographical notions. Hence come first the concepts of valleys and hills.

A valley of $f$ is defined with respect to a col of $f$ as follows:

valley: Given a col of $f$, $c$, the valleys of $f$ at $c$ are the connected components of the subset of $C \{z : Rf(z) < Rf(c)\}$, which closures contain $c$.

Naturally we can define the hills of $f$ at a col $c$ as the valleys of $-f$ at $c$. Hills inherit their properties from valleys.

These properties are:

- There are exactly $(p + 1)$ valleys at a col of order $p$.

- Valleys are obviously open subsets of $C$ and the intersection of a valley with a basin, if non-empty, is open and connected.

- The important property of valleys as defined above is the following:

  A valley contains any path of descent issued from any of its element.

We note that a path $D_z$ issued from a point $z$ of a valley is not necessarily unique: a valley may have cols for elements. (See Figure 3-3 for an example.)

The definition of valleys given here is a local definition, less restrictive than the definition of basins. Therefore this does not quite serve our purpose and leads to the definition of the property of confluence.
Figure 3-1: Cols, crest lines and basins on a surface
The symbol $\mathbf{x}$ is used for the cols here and after.

Figure 3-2: Cols, crest lines and basins in the $z$-plane
The continuous lines are the boundary of the two valleys, in white, defined with respect to $z_1$. The dashed lines are the boundary of the two valleys, in gray, defined with respect to $z_2$.

**Confluent valleys and fjords**

**Confluence:** Two valleys $\mathcal{V}$ and $\mathcal{V}'$ are confluent in a basin $B$, if $\mathcal{V} \cap B$ and $\mathcal{V}' \cap B$ are non-empty.

From this definition and the definition of basins, we conclude that two confluent valleys have non-empty intersections. The interest lies here in considering the set of all valleys which are confluent in a given basin $B$. This set is countable and we denote it $\{\mathcal{V}_j^B\}_{j \in J}$ with $J$ countable.

The next idea is to consider the intersection of all valleys confluent in $B$. Thus comes the definition of a fjord:

**Fjord:** We say that $B$ has a fjord, denoted $\mathcal{F}^B$, if the set $\mathcal{F}^B = B \cap \bigcap_{j \in J} \mathcal{V}_j^B$ is non-empty.
A basin does not always have a fjord: this might happen when the sequence of cols contained in $\overline{B}$ is unbounded (see Figure 3-9 for such an example). In such a case we say that $B$ has a fjord at infinity. This corresponds to what is usually meant by “a valley of $f$ at infinity” and which is understood as a direction along which $\Re f$ is the most decreasing. The existence of such direction is also included in the definition of a fjord when it exists.

Fjords have three important properties:

- $\mathcal{F}^B$, if non-empty, is open and simply connected.

- Any two distinct points $a$ and $b$ of $\mathcal{F}^B$ such that $\Re f(a) = \Re f(b)$, are on the same level curve of $f$.

- Two distinct points $a$ and $b$ of $\mathbb{C} \setminus \mathcal{C}$, are in the same basin $B$ of fjord $\mathcal{F}^B$ if, and only if, the intersections of the paths of steepest descent issued from $a$ and $b$ with $\mathcal{F}^B$, or $\mathcal{D}_a \cap \mathcal{F}^B$ and $\mathcal{D}_b \cap \mathcal{F}^B$, are non-empty.

Hence showing that two points $a$ and $b$ are in a basin with a fjord is equivalent to showing that there exists a path joining $a$ to $b$ made of a portion of $\mathcal{D}_a$, a level curve (in the fjord) and a portion of $\mathcal{D}_b$.

We conclude here this series of definitions and properties relative to domains of $\mathbb{C}$ which will help us to introduce more fundamental notions.

3.1.2 Graph of $f$ and ramification set

In this section we leave the study of the topological properties of regions of $\mathbb{C}$ to introduce more abstract concepts.
Passes and graph of $f$

Here we use the properties of the basins of $f$ to define a relation between cols and basins.

**pass:** Let $c$ be a col of $f$, we say that $c$ is a pass for $B$, a basin of $f$, if and only if $c \in \partial B$.

Obviously any col is a pass for at least 2 basins of $f$. More precisely, if $c$ is a col of order $p$, $c$ is a pass for at least $(p + 1)$ basins, and at most $2(p + 1)$ basins. This number of $2(p + 1)$ is obtained when the paths of steepest descent $D_c$ are also paths of steepest ascent issued from other cols of $f$.

Let $\{B_j\}_{j \in J}$ and $Z = \{z_h\}_{h \in H}$ be the sets, respectively, of basins and cols of $f$, and $V_f$ and $E_f$ the sets:

$$V_f = Z \cup \bigcup_{j \in J} B_j$$

and:

$$E_f = \{ \{B_j, z_h\} : z_h \text{ is a pass for } B_j \}$$

We define the graph of $f$ by:

**graph of $f$:** The graph of $f$ is the graph $G_f = (V_f, E_f)$ where $V_f$ is the set of vertices and $E_f$ the set of edges of $G_f$.

The term graph is to be understood in the sense of graph theory (see Wilson, 1986 or König, 1989). By construction $G_f$ is a connected, bipartite, simple graph with no loops.

**ramification set:** We define the ramification set of $f$, $R$, as the subset of $D$ such that $G_f$ has a circuit. $\bar{\alpha}$ is said to be a ramification value for $f$ if $\bar{\alpha} \in R$.

Before going further, we show the equivalence of the existence of a circuit in $G_f$ and the existence of two cols $c$, $c'$ in the same connected component of $\tilde{C}$.
Proposition 1:

\( \bar{\alpha} \) is a ramification value for \( f \) if, and only if, there exist \( c \) and \( c' \) elements of \( \mathbb{Z} \), such that \( c' \) is an element of a path \( \mathcal{A}_c \).

A proof comes as follows:

(\( \Leftarrow \)) If there exist such cols, \( G_f \) has a circuit of length 4 (see Figures 3-6, 7 for an example).

(\( \Rightarrow \)) Let us assume that \( G_f \) has a circuit \( O \):

\[ O: \quad B_1 \to c_1 \to B_2 \to \ldots \to B_1 \]

By definition of basins and passes, we can construct a closed contour in \( \mathcal{C}, \Gamma \), such that the intersection of \( \Gamma \) with \( \mathcal{C} \) is exactly the set of cols of \( O \) and such that the intersection of \( \Gamma \) with any basin of \( O \) is non-empty. \( \Gamma \) being closed, is the boundary of a domain of \( \mathcal{C} \), say \( \mathcal{C}' \). Assuming that all the cols of \( O \) are on different connected components of \( \mathcal{C} \), there exists a path, say \( \mathcal{A}_{c_1} \) contained in \( \mathcal{C}' \). Since \( \Re f(\mathcal{A}_{c_1}) = [\Re f(c_1), +\infty[ \), \( f \) has a singular point in \( \mathcal{C}' \), hence \( f \) is not holomorphic in \( \mathcal{C} \).

From there we deduce the following equivalence:

\( \bar{\alpha} \not\in \mathbb{R} \quad \Leftrightarrow \quad G_f \) is a tree

which has the strong consequence that the path joining two distinct vertices in \( G_f \) is unique unless \( \bar{\alpha} \) is a ramification value for \( f \).

The above results depend entirely on our choice for the definition of \( G_f \). Would this choice be different (e.g. considering only cols or basins as vertices) Proposition 1 and the above equivalence would not hold.

Similarly to catastrophe theory where an object is said to be structurally stable if \( K \) is nowhere dense in the space of parameters, it is natural here to wonder about
the topological properties of $\mathbb{R}$. Toward this aim, we can easily furnish a superset of $\mathbb{R}$ as:

$$R' = \{ \bar{a} : \exists c, c' \in \mathbb{Z}, c \neq c', \exists \{ f(c, \bar{a}) - f(c', \bar{a}) \} = 0 \}$$

the inclusion $\mathbb{R} \subset R'$ being possibly strict. Then the question is whether or not $\overline{\mathbb{R}}$ is nowhere dense in $D$, knowing that $R'$ (or $\overline{R}$) is an analytic set. Indeed this problem is closely related to the problem of structural stability in catastrophe theory since $\mathbb{K} \subset \overline{\mathbb{R}}$. (Regarding this question, we refer to results used in catastrophe theory; see Thom, 1977). The interest of such a question is clearer when considering the complementary of $\overline{\mathbb{R}}$, $D \setminus \overline{\mathbb{R}}$, and more precisely its connected components $\{ \mathcal{O}_i \}_{i \in S}$.

Each domain $\mathcal{O}_i$ is defined by a certain set of inequalities of the type:

$$\exists \{ f(c, \bar{a}) - f(c', \bar{a}) \} > 0 \quad c, c' \in \mathbb{Z}$$

Consequently, the set $S$ is countable. The fundamental property of the domains $\mathcal{O}_i$ with respect to $G_f$ is the following:

*Given $\mathcal{O}_i$ and $\bar{a}'$, $\bar{a}'$ two of its elements, the graphs $G_f(\bar{a})$ and $G_f(\bar{a}')$ are identical.*

This result is a consequence of Proposition 1 and of the inclusion $\mathbb{K} \subset \overline{\mathbb{R}}$.

When $\bar{a}$ and $\bar{a}'$ are in two different domains $\mathcal{O}_i$, $G_f(\bar{a})$ and $G_f(\bar{a}')$ are not identical although they might be isomorphic. This raises the question of how $G_f(\bar{a})$ is modified when going from $\bar{a}$ to $\bar{a}'$ along a path in $D$. For this purpose we investigate the properties of the vertices of $G_f$.

**Catastrophes and ramifications**

In this section we infer the properties of basins and cols considered as vertices of $G_f$, from their properties as objects of $\mathcal{C}$. Hence we have:

* The degree of a basin $B$ in $G_f$ is given by the cardinal of $\overline{\mathcal{B}} \cap \mathbb{Z}$ and is at least 1.
• The degree of a col of order \( p \) in \( G_f \) is \( (p + 1) \), unless \( \alpha \) is a catastrophe or ramification value for \( f \) at \( c \).

When \( \alpha \) is an element of \( \mathbb{R} \), we need to make the distinction between a catastrophe and a ramification, both being local in \( G_f \). The consideration of such elementary – i.e. local – events comes readily from the study of subgraphs of \( G_f \).

**Effect of a catastrophe on \( G_f \)**

Here we restrict our study to subgraphs of \( G_f \) given by two cols. More precisely if \( c \) and \( c' \) are two cols of \( f \), the graph \( G_f\{c,c'\} \) is defined by taking \( c \), \( c' \) and all the basins adjacent to them in \( G_f \). Then \( \alpha_0 \) is said to be a *catastrophe value* for \( G_f\{c,c'\} \), or for \( f \) at \( c \) (or \( c' \)), if \( c \) and \( c' \) are coalescing for \( \alpha = \alpha_0 \). Similarly, \( \alpha_0 \) is said to be a *ramification value* for \( G_f\{c,c'\} \), if \( G_f\{c,c'\}(\alpha_0) \) is not a tree. In this localization of the definition of catastrophe (respectively ramification) \( \alpha \) might still be in \( K(R) \) while not being a catastrophe (ramification) value for \( G_f\{c,c'\} \). Thus in the following we will refer to the “restriction” of \( K \) (respectively \( R \)) to catastrophe (ramification) values for \( G_f\{c,c'\} \), denoted \( K_{\{c,c'\}}(R_{\{c,c'\}}) \).

The effect of a catastrophe on \( G_f\{c,c'\} \) is characterized as follows:

Let \( \alpha \) be in \( D \setminus \mathbb{R}_{\{c,c'\}} \) and \( \alpha_0 \) in \( K_{\{c,c'\}} \) such that there exists a path joining \( \alpha \) to \( \alpha_0 \) in \( D \setminus \mathbb{R}_{\{c,c'\}} \). Then \( G_f\{c,c'\}(\alpha_0) \) is obtained from \( G_f\{c,c'\}(\alpha) \) by a collapse of the vertices \( \{c,c'\} \), \( G_f\{c,c'\}(\alpha_0) \) still satisfying the definition of \( G_f \) (i.e. one edge between adjacent vertices).

In other words \( G_f\{c,c'\}(\alpha_0) \) has one vertex, and one edge less than \( G_f\{c,c'\}(\alpha) \).

The effect of a catastrophe on \( G_f \) is then obtained by combinations of the above elementary event:

The coalescence at \( \alpha_0 \) of \( m \) cols, while \( \alpha \) stays in \( D \setminus \mathbb{R} \), corresponds to \( (m - 1) \) collapses resulting in a graph \( G_f(\alpha_0) \) with \( (m - 1) \) vertices and \( (m - 1) \) edges less than \( G_f(\alpha) \).

The edges to be deleted in this process are uniquely determined by each collapse.
Effect of a ramification on $G_f$

Using the same approach, we consider the effect of ramifications on a subgraph $G_{f\{c,c'\}}$. We have then the characterization:

Let $\vec{\alpha}_0$ be in $R_{\{c,c'\}}$ and $\vec{\alpha}$ in $D \setminus \overline{R_{\{c,c'\}}}$ such that there exists a path joining $\vec{\alpha}$ to $\vec{\alpha}_0$ in $D \setminus \overline{R_{\{c,c'\}}}$. Then $G_{f\{c,c'\}}(\vec{\alpha}_0)$ is obtained from $G_{f\{c,c'\}}(\vec{\alpha})$ by adding a specific edge.

In more details:

Let us assume that $\Re f(c) > \Re f(c')$, then as $\vec{\alpha} \rightarrow \vec{\alpha}_0$ there exists a unique path $A_{c'}$ such that the distance $d(c, A_{c'}) \rightarrow 0$. By definition, this path is the boundary between two basins adjacent to $c$, say $B'_1$ and $B'_2$. One of these basins, say $B'_1$, is already adjacent to $c$ (this basin contains one of the paths $D_c$). Then $G_{f\{c,c'\}}(\vec{\alpha}_0)$ is $G_{f\{c,c'\}}(\vec{\alpha})$ plus the edge $\{B'_1, c\}$.

The term ramification comes of course from this feature. (Two examples are given in Figures 3-4, 5, 6, and 7.)

The effect of a ramification on $G_f$ is obtained by combinations of the above elementary ramification. Since an elementary ramification gives rise to a circuit of length 4, any circuit in $G_f$ contains a subcircuit of length 4.

Our introduction of the notion of ramification appears as a complement to the concept of catastrophe in order to completely describe any change of the “topology” of $f$. (We use abusively the term “topology” in the same sense as for a potential flow here.) As we saw, this “topology” is entirely comprised by the graph of $f$.

3.1.3 Ramifications, catastrophes and forms

We have seen the distinct but determining roles of catastrophes and ramifications in $G_f$. Similarly to catastrophe theory, we may consider $G_f$ as being the form of $f$ which now is not only modified by catastrophes but also by certain other events called
Figure 3-4: Example of ramification (1)
Basins and cols [top] and corresponding graph [bottom]. Here and after o represents a basin.
Figure 3-5: Example of ramification (2)
Basins and cols [top] and corresponding ramified graph [bottom].
Figure 3-6: Example of ramification (3)
Basins and cols [top] and corresponding graph [bottom].
Figure 3-7: Example of ramification and catastrophe
Basins and cols [top] and corresponding graph [bottom]. The col on the left is of order 2 (fold catastrophe) while the one on the right is simple.
ramifications. In the study of the evolution of $G_f$ with $\bar{\alpha}$, the effect of a catastrophe or a ramification is completely defined. However the ability of “following” this evolution lies entirely on the assumption that the saddle points of $f$ are continuous functions of the parameter $\bar{\alpha}$ in $D$. The definition of $D$ excludes any singular points of $f$ as function of $\bar{\alpha}$, and $Z$ being an analytic set, its elements are locally continuous functions of $\bar{\alpha}$. Extending this result to $D$ belongs to the study of deformations of analytic sets. Here, we make the assumption that indeed any col of $f$ is a continuous function of $\bar{\alpha}$. Then the knowledge of $G_f$ for $\bar{\alpha}_0$ in some $O_i^1$ is sufficient to infer $G_f(\bar{\alpha}_1)$ for $\bar{\alpha}_1$ in $O_i^1$, $t \neq s$, through combinations of catastrophes and ramifications.

Catastrophes and ramifications of $G_f$ have corresponding effects for the “topology” of the potential flow or the global arrangement of the basins of the phase flows described at the beginning of this chapter. Their effects on the derivation of steepest descent expansions for $I$ are as remarkable. We already know (Chapter 1) that a catastrophe requires for its uniform asymptotic treatment a special form of expansion. It appears that a ramification, if affecting the path of $G_f$ corresponding to $C$, the initial contour of integration, gives rise to a Stokes phenomenon in the expansion of $I$ (see Section 3.3).

The differences between the local property of a catastrophe, which is bound to the neighborhood of a coalescence in $C$, and the global (i.e. non-local) property of a ramification seem to emphasize a fundamental difference in their nature. However their common characteristic in inducing modifications of $G_f$, suggests that a ramification might be in fact a catastrophe. Hence a ramification is characterized by the coalescence of certain points $\exists f(c, \bar{\alpha})$ with $c \in Z$. But can a mapping having these points as critical points be defined?
3.2 Functions with singular points

In the study of a function with singular points, we assume here that $f : W \times D \to \mathbb{C}$ is holomorphic, $W$ being a domain of $\mathbb{C}$. The results of the holomorphic case seem to be extendable under certain conditions.

Singular points and simple connectedness

We stipulated in Section 1.1.1 that $W$ should be a domain, i.e. an open subset of $\mathbb{C}$ simply connected. The necessity of satisfying this last constraint is the main difficulty here.

Any branch point of a function of a complex variable is associated to a branch cut of its definition set. Here we require that such cuts leave this set simply connected. Moreover we extend the necessity of having branch cuts to the cases of poles (usually this is not necessary) in order to preserve the simple connectedness of the definition set $W$. This requirement leaves a priori a lot of freedom in the choices for such cuts. But additional constraints arise if we wish to preserve the main results valid for holomorphic functions. The principal consequence of considering a function with singular points in $\mathbb{C}$ is that $\partial W$ might be at finite distances. Depending on the value of $r$ as given in Section 1.1.1, we distinguish two cases which determine the type of singularity of $f$ at $b$:

- If $r > 0$, $b$ is a branch point, is in $B_1$, and is considered as a point of $\partial W$ with of course the corresponding cut.

- If $r < 0$, $b$ might be a pole or a branch point ($r$ non entire) is in $B_2$, and is considered in any case as a branch point with its associated cut. The crucial difference with the above situation is that certain steepest paths now end at $b$.

In both cases $b$ may be a singular point of $g$, i.e. an element of $B_g$ or a singular point of $f$, an element of $B_f$.  

93
In order to preserve the main results of Section 3.1.1, we impose the following condition on the branch cuts of $f$:

*For any singular point $b$ of $f$ of $g$, its branch cut is chosen to be a steepest path ending at a singular point other than $b$.*

The existence of such a cut is granted for both cases:

- If $b \in B_1$, since $\lim f(z, \alpha) \text{ as } z \to b \text{ in } W$ exists, such a path exists.

- If $b \in B_2$, since $f$ is singular at $b$, we can find a point $z_1$ in a neighborhood of $b$, such that $z_1$ is neither a col or a branch point (by definition of the singularity at $b$), and $A_{z_1}$ (or $D_{z_1}$) ends at $b$. Then the steepest path is defined by the union $A_{z_1} \cup \{z_1\} \cup D_{z_1}$.

We emphasize that it is not necessary for every branch cut to leave $W$ simply connected. However their *union* must be such that $W$ is simply connected. The global problem of arranging the cuts in order to satisfy this condition is not obvious and would require further studies.

The changes induced by the presence of singular points necessitate an extension of the definition of basins as of Section 3.1.1.

**Splitting of basins**

Let $B = B_1 \cup B_2$ be the set of singular points of $f$ and $g$ for which we will use abusively the term *branch points*. We define the set $Z^*$ as $Z \cup B$ and consequently the set $\hat{C}^*$ obtained from $\hat{C}$ by adding the elements of $B$ and their branch cuts. In order to include the coalescence of branch points with cols of $f$, we introduce the set $C$ of values of $\alpha$ for which such coalescences occur.

The basins are then defined as in Section 3.1.1 using $\hat{C}^*$ instead of $\hat{C}$. This may result in the *splitting* of a basin as originally defined. The definitions of valleys, hills and fjords come then naturally, their properties apart from losses of connectedness due
to splittings, being identical. It is not clear at that point whether \( G_f \) should include the elements of \( B \) as vertices or not. Basins, as vertices of \( G_f \), could well contain the necessary information about their adjacent branch points and cuts without having to convey this information to \( G_f \). Three examples of graphs with singular points and using the original definition of Section 3.1.2 are given in Figures 3-8, 9, 10. For these cases this simple definition of \( G_f \) is entirely satisfactory. However one can imagine cases where a branch point lies on the path of steepest descent issued from a col. There the extension of the notion of ramification would be necessary, and the inclusion of branch points as vertices of \( G_f \) seems recommendable. Undoubtedly further studies are necessary in order to find the appropriate formalism permitting to consider the most general cases.

It remains now to relate our initial problem of determining the relevant saddle points to the results obtained so far.

### 3.3 Practical aspects: a still unsolved question

We propose here to investigate the practical aspects of applying the formalism developed in Sections 3.1 and 3.2 to the problem of finding the relevant cols of an integral \( I \). This investigation will stay at a theoretical level since no implementation of this formalism has been done. However we will try to be as scrupulous as possible in enumerating eventual difficulties.

#### 3.3.1 Relevance and Stokes phenomenon

Let us consider our initial problem of the datum of an integral \( I \). For any value of \( \alpha \) in \( D \), we are capable of determining, at least in theory, the graph \( G_f \). By definition, the end-points of \( C \), \( a \) and \( b \), must be in some basins \( B_a \) and \( B_b \) of \( f \). Since \( G_f \) reflects the simple connectedness of \( W \), the problem of finding the relevant saddle points of \( I \)
Figure 3-8: Function with a pole
Basins and cols [top], corresponding graph [bottom]. The function presented here is the exponent function of the Kelvin potential (see Chapter 6) for $\theta = 10^\circ$, $\varphi = 80^\circ$ in the domain $-2 \leq \Re z \leq 5$, $-2 \leq \Im z \leq 5$. The pole ($\bigcirc$) is located at the origin, the branch cut (- - -) splitting a basin in two.
at \( \tilde{\alpha} \), is equivalent to finding a path joining \( B_a \) to \( B_b \) in \( G_f \). The set \( Z_R \) is then given by the elements of \( Z \) contained in this very path. Unless \( G_f \) has a circuit – i.e. for \( \tilde{\alpha} \) in \( R \) – such path is unique, its existence being granted by the connectedness of \( G_f \). When \( G_f \) presents a circuit, we require that the path joining \( B_a \) to \( B_b \) is of minimal length, and if several choices exist, to select the one for which \( \sum \Re f(z) \), \( z \) cols of this path, is minimal.

As \( Z_R \) is modified through a ramification, one could wonder whether the steepest descent expansion obtained by adding the contributions from the elements of \( Z_R \) is valid. We already know that \( \tilde{\alpha} \) must be away from neighborhoods of \( C_R \cup K_R \) (here the subscript \( \alpha \) signifies that we consider the parts of these sets relevant to the subgraph \( G_f Z_R \)) for this additivity to be legitimate. The modification of the expansion of \( I \), which takes place when \( G_f Z_R \) presents a ramification, was first identified by Stokes in 1857 for Bessel functions (whence the term Stokes phenomenon). This modification appears as the disruption of the asymptotic series through the adjunction of necessary contributions from new cols.

We must emphasize that a ramification of \( G_f Z_R \) corresponds to an exact result of contour integration. A Stokes phenomenon appears only when reducing the global behavior of \( I \) (along the deformed contour) to a discrete sum of local contributions.

The justification for such a disruption was left in the dark for a long time. It is only recently (Dingle, 1973) that a rigorous treatment of the Stokes phenomenon and of asymptotic series was suggested by introducing the notion of terminants. In fact all this matter comes down to the definition of the validity of an expansion. If indeed exponentially negligible terms and error bounds are satisfactory, the disruption of an asymptotic series along a Stokes “line” (in fact \( R_R \)) is necessary and has no consequence for the continuity of the result within, of course, some error bounds. In an exact asymptotic expansion of \( I \), the additional terms naturally arise from retaining and rewriting the complete remainder of the expansion. The process of recursively looking for the asymptotics of successive remainders, or terminants, is
named hyperasymptotics. In such treatments, resulting in extremely high accuracy, the notion of "adjacent" saddle points is fundamental (Berry, 1992). This "adjacency" based on distances is different from the adjacency of graph theory. However it clearly appears that the consideration of $G_f$ would be of great use in such procedures.

After this nearly complete (we will treat later the case of neighboring end-points of $C$) inventory of the modifications that a steepest descent expansion may undergo as $\bar{\alpha}$ describes $D$, the next step considers the domains of validity implicitly defined by $\overline{R_R}$, and $C$.

### 3.3.2 Uniform asymptotic expansions and domains of validity

We already mentioned the countable collection of domains of $D$ given by the connected components of $D \setminus \overline{R}$ in the holomorphic case (see Section 3.1.1). When $f$ presents singular points, a similar collection $\{O_{s}^{1}\}$, $s$ in $S$, is defined from $D \setminus \{\overline{R} \cup C\}$. In order to be complete, we need to introduce at this stage the subset $C_C$ of $D$ for which coalescences of end-points (necessarily finite) of $C$ and elements of $Z^*$ occur. The corresponding effect of such coalescence on a steepest expansion was reviewed in § 1.2.2 c). This leads us to define in fact the collection $\{O_{s}^{1}\}$ as the connected components of $D \setminus \{\overline{R_R} \cup C_R \cup C_C\}$. We know then that within each of these domains, the asymptotic expansion of $I$ is formally identical and is given by a simple addition of the contributions from saddle points of order 1, plus eventual contributions from the end-points of $C$.

In neighborhoods of points of $K_R$, $C_R$ or $C_C$, different expansions are required which treat uniformly the various events: catastrophes, coalescences of branch points or end-points. We generically denote such neighborhoods by the collection of domains $\{O_{t}^{2}\}$, $t$ in $T$. This collection is such that for each $O_{t}^{2}$ there exists at least a domain $O_{s}^{1}$ such that $O_{t}^{2} \cap O_{s}^{1}$ is non-empty. If the collection $\{O_{s}^{1}\}$ is disjoint, the collection
\{\overline{O}_i\} furnishes a covering of \(D\): for any domain \(O^1_i\) there exists a domain \(O^2_i\) such that \(\overline{O}_i^1 \cap \overline{O}_i^2\) is non-empty and in fact included in \(\overline{R}_R\). Finally the total collection \(\{O^1_i, O^2_i\}_{i \in S, i \in T}\) gives a covering of at least \(D \setminus R_R\), with a given expansion associated to each domain. In order to obtain a covering of \(D\), the expansions corresponding to the domains \(O^1_i\) must be extended to their boundaries (i.e. to components of \(R_R\)) through Stokes phenomena.

We now have a sound basis for the various expansions of \(I\) in terms of the topological properties of their domain of validity. (An in depth study of the geometry of the domains \(O^2_i\) is found in Martin (1974).) We now focus on the practical remaining tasks: the building of \(G_f\) and the finding of paths in this graph.

3.3.3 Characterizations of basins

The apparent simplicity in obtaining the set \(Z_R\) in Section 3.3.1, relies entirely on the ability of building the graph \(G_f\), or more precisely of characterizing the basins of \(f\). Hence in fact we are brought back to the fundamental question as formulated by Ursell but in terms of steepest paths only.

The finding and tracking of cols

So far we have assumed that for any \(\bar{\alpha}\) in \(D\), we were capable of determining the positions of all the cols of \(f\) (condition III of Section 2.4). The corresponding practical task is not simple: finding the roots of analytic equations and following those roots for finite variations of \(\bar{\alpha}\) are the main difficulties. These problems are being addressed in studies of dynamical systems, but in a different form: namely the finding and tracking of fixed points of mappings. We refer to techniques used in this field since solving an analytic equation can be reduced to finding the fixed points of a mapping. Inherent also to \(G_f\) is the necessity of indexing the saddle points, and keeping track of their progress as \(\bar{\alpha}\) varies - i.e. preserving the indexing. Difficulties arise when
a catastrophe or a collision occur. However this obstacle is surmounted by taking advantage of the analyticity of the cols in a neighborhood of a catastrophe.

By using the entire series associated to each col, we are capable of following their evolution through a catastrophe. Obtaining the numerical values of the series coefficients can be done efficiently using the equivalence between Fourier series on poly-circles and entire series in poly-discs (see Appendix D for more details).

Characterizations of basins with fjords

Until now, we have not used the notions of valleys, confluence and fjords. These definitions were precisely introduced toward the practical characterization of basins.

A basin, considered as a vertex of $G_f$, is entirely determined by its passes. Hence the task of building the graph $G_f$ consists in assessing whether or not 2 cols are passes of the same basin. When this basin has a fjord, a practical criterion is given in Section 3.1.1: we know how to create a path joining two points of this basin. The actual process of showing that such points can effectively be joined, consists in generating the path. This can be done by using algorithms similar to gradient descent algorithm (all the steepest path shown in this chapter were traced using such an algorithm). In fact the marching procedure is extremely simple since we know that the imaginary (real) part of $f$ is constant along steepest paths (respectively level curves). The main difficulty lies in choosing the size of the steps since we do not know a priori whether a path ends at the border of the computed domain, at a singular point of $f$ or at a col. (The dents present in some paths of steepest ascent shown in this chapter are witnesses of this difficulty.) When branch points are involved the task is worsened by the presence of branch cuts.

It appears that we cannot escape the process of curve tracing unless another practical characterization of basins is found. This question is still open although it must be noted that basins are very special regions of $\mathbb{C}$. In fact, basins can be interpreted as maximum domains of univalence. In that respect the formalism of
Section 3.1.1 is very close to the method of univalent dissection of Ursell (1970).

The computational constraint bearing on the process of curve tracing implies a priori that only functions with finite numbers of cols may be treated. (Note that a finite set Z induces the existence of fjords for every basins.)

Similarly the practical identification of the basins (or fjords) containing the end-points of C is bound to the consideration of finite end-points. A possible overcoming of this limitation is the use of inversions of the complex plane and arguments of convexity in the vicinity of the origin. After all, this case is equivalent to the consideration of singular points of f.

Path finding

From the above restrictions, it appears that only finite graphs $G_f$ can be considered. Implementations of graphs $G_f$ corresponding to holomorphic functions do not seem to present any difficulties. On the contrary the introduction in $G_f$ of vertices representing elements of B complicates the matter. The problem of finding paths in trees or constrained paths in graphs is a classical and extensively treated topic of computer science (see Gibbons (1985) by example). Algorithms solving such problems with various efficiencies are available, and we do not go into details.

The restriction to finite graphs, may be lifted under certain conditions: namely, if $G_f$ presents a complete finite collection of generating finite subgraphs and if the problem of path finding in $G_f$ can be reduced to some given problems in these subgraphs. We call generating finite subgraph of $G_f$, any finite subgraph $G^0_f$ such that there exists an infinite subgraph of $G_f$ made entirely of an infinite union of graphs isomorphic to $G^0_f$. A collection of generating finite subgraphs is complete for $G_f$ if the total union of the generated graphs is $G_f$. This opens the possibility of treating periodic graphs (the function $f$ being not necessarily periodic) as well as using symmetry properties of $f$. This feature emphasizes the advantage of using the formalism of graphs.
Two examples of such graphs are given in Figures 3-9 and 3-10. In the first case two generating subgraphs and a finite subgraph are identified, while in the second two generating subgraphs are sufficient.

3.4 Conclusion

The formalism we defined in Section 3.1 is the first attempt to rigorously formulate the problem of relevant cols in the method of steepest descents.

The central key of this formalism is to associate to any function \( f \), a graph \( G_f \) based uniquely on topological properties intrinsic to \( f \). In the study of the evolution of this graph with the parameter \( \alpha \), the new notion of ramification arises as a complement to the notion of catastrophe which plays a crucial role in the generalized method of steepest descents. When \( f \) is holomorphic in the whole complex plane, both ramifications and catastrophes appear to be the only agents in any change undergone by \( G_f \). When looking for a steepest descent expansion of the integral \( I \), the relevant saddle points for a given \( \alpha \) come then readily as the vertices (cols) of the path of \( G_f \) joining the 2 vertices (basins) associated to \( C \). Stokes phenomena occurring in the asymptotics of \( I \) are then directly related to the crossing by \( \alpha \) of certain components of \( R \), i.e. to a ramification of \( G_f \). From this it appears that \( G_f \) contains some information which could be useful for the hyperasymptotic treatment of \( I \). Finally the domains of the \( \alpha \)-space defined with the help of \( K \) and \( R \), furnish a basis for the domains of validity of the various asymptotic series.

However the appropriate modifications to the definition of \( G_f \) required by the consideration of functions with singular points are still unclear and necessitate further studies. More importantly remains the fundamental question of a practical characterization of basins. Hence for the moment, the process of curve tracing seems inevitable. This task however can be carried out on a computer for a function with a finite number of cols. Furthermore the formalism provided by \( G_f \) is extremely well suited to computer implementations. In principle such implementations are limited
Figure 3-9: Graph with generating finite subgraphs: example 1
Cols and basins [top], and subgraph [bottom] of $\sqrt{z} \tan z - 1.2z$ (see Chapter 7) in the domain $-10 \leq \Re z \leq 10, -2 \leq \Im z \leq 2$. 
Figure 3-10: Graph with generating finite subgraphs: example 2
Cols and basins [top], and subgraph [bottom] of $-\sqrt{z \tan z} - z$ (see Chapter 7) in the domain $-10 \leq \Re z \leq 10, -2 \leq \Im z \leq 2$. 
to the consideration of finite graphs, but certain cases of infinite graphs could be treated as well.

Based on the assumptions that cols and critical points of $f$ and $g$ can be located and followed for any $\bar{\alpha}$, a complete implementation of our formalism to the relevant col problem, would then:

1) build the graph $G_f$ for a given value of $\bar{\alpha}$,

2) find in $G_f$ the path corresponding to $C$ (the initial contour of integration),

and subsequently:

3) update $G_f$, if necessary, through catastrophes or ramifications due to any variations of $\bar{\alpha}$, and find the corresponding new path in $G_f$

Toward this aim much work is still required. The motivation and justification of such efforts lay in the generality of the procedures through which steepest descent expansions are obtained (e.g. Dingle, 1973). In the following chapters, we will see how ordinary steepest descent expansions can be obtained in a systematic manner (Chapter 4), and how algorithms could be devised to obtain robust numerical results for the case of the coalescence of two saddle points (Chapter 5).
Chapter 4

The ordinary method of steepest descents

Hence we start here our detailed study of the numerical implementation of the method of steepest descents under the conditions stated in Section 2.4. This chapter presents the cases of the ordinary method of steepest descents, that is the case of an endpoint (Section 4.1), and the case of an isolated col of order 1 (Section 4.2). In both cases we show that implementations which convey the robustness of the asymptotic expansions are possible. Being at risk of giving a presentation which might seem too technical, we wish to not leave any "detail" in the dark. This is possible, only through the analysis of the procedures and their decomposition into simple tasks. A similar decomposition can be found in Dingle (1973) without of course the numerical analysis proper to computer implementations. The use of symbolic computation is emphasized both for the analysis (Sections 4.1 and 4.2) and for the implementations (Section 4.3). This allows us to present simple but illustrative examples of the properties of the corresponding algorithms (Sections 4.1.2, 4.2.2 and 4.3.4).

Preamble

We have seen that given the set \( \overline{\mathcal{R}_R} \cup \mathcal{C}_R \cup \mathcal{C}_C \) of critical values of \( \bar{\alpha} \) in \( D \), we could define a countable collection of disjoint subdomains of \( D \), \( \{ O_i \} \), such that their union
with $\overline{R_R} \cup C_R \cup C_C$ is $D$. Over a given $O_1^i$, the set $Z_R$ of relevant cols stays identical. Thus for all $\bar{\alpha}$ in $O_1^i$, the asymptotic expansion of $I$ is given, in theory, as the sum over $Z_R$ of expansions, (1.16), from isolated cols of order 1 (the elements of $Z_R$), plus possibly the expansion(s), (1.17), from one (or both) end-point(s) of $C$. This expansion fails as $\bar{\alpha}$ tends to the boundary of $O_1^i$, staying in $O_1^i$. The hypothesis of the ordinary method of steepest descents states that $D \cap \{K_R \cup C_R\}$ is empty, or in other words that $\bar{\alpha}$ is always away from any catastrophe value or value corresponding to a coalescence of cols and singular points of $f$ or $g$.

The problem associated to the numerical evaluation of $I$ for $\bar{\alpha}$ in $D$, is to find a countable (hopefully finite) collection of connected subsets of $D$, $\{\bar{O}_t\}$, such that on each $\bar{O}_t$, the corresponding asymptotic expansion is uniformly approximating $I$ with error $\epsilon$. This leads us to define a uniform approximation of $I$ in $D$ in the following sense:

*Given $\epsilon > 0$, there exists a collection $\{\bar{O}_t\}$, $t \in T$, satisfying the conditions:

- $T$ is countable (in practice finite),
- $\forall t \in T$, $\bar{O}_t$ is connected,
- if we denote by $E$ the error made by approximating $I$ by its expansion for a given $(N, \bar{\alpha})$, there exists $N_0$ positive and real such that:

$$\forall N > N_0, \forall t \in T, \forall \bar{\alpha} \in \bar{O}_t, E < \epsilon/2$$

and $\{\bar{O}_t\}$, $t \in T$ is a covering of $D$.

Then $I$ is uniformly approximated over $[N_0, +\infty[\times D$ with the error $\epsilon$.

Of course the interest lies in finding the smallest possible value of $\epsilon$. We must emphasize that $E$ contains any numerical errors made in the numerical evaluation of the expansions. Therefore the determination of the regions $\bar{O}_t$ can only be made when
for each $\bar{O}_t$ the corresponding expansion is numerically available. (See Appendix A for additional details on the $\bar{O}_t$.) This task of obtaining numerically these expansions is undertaken in this chapter and pursued in the following. The problem of assessing the shapes of the regions $\bar{O}_t$ will only be presented for the particular example of Chapter 6.

4.1 The case of an end-point

Here, we consider an end-point $e$ of the initial path of integration $C$. The problem is then to compute the coefficients in the expansion (1.16). We assume that the cols of $f$ are sufficiently distant from $e$ so that the mapping $u$ defined as in Section 1.1.3 a), from the equation:

$$f(z, \bar{\alpha}) - f(e, \bar{\alpha}) = -u$$

is valid.

For a fixed value of $\bar{\alpha}$, the bijective property of the mapping $u$ is ensured by the fact that $\partial f(e, \bar{\alpha})/\partial z \neq 0$. Moreover we know that the inverse mapping, $z$, exists and is analytic in $(u, \bar{\alpha})$ at $(0, \bar{\alpha})$. (An expansion of $f(z, \bar{\alpha})$ in its entire series in $z$ shows it.) Therefore, there exists a radius $\rho$, such that for $u$ in the open disk $D_0(0, \rho)$ and $\bar{\alpha}$ in, say, $W_1$, the series:

$$z(u, \bar{\alpha}) = \sum_{n=0}^{\infty} a_{z,n} u^n$$

is convergent.

From the fact that $g$ is analytic in $z$ at $z = e$, we conclude that $G$ defined by:

$$G(u, \bar{\alpha}) = g(z, \bar{\alpha}) \frac{\partial z}{\partial u}(u, \bar{\alpha})$$

is analytic at $u = 0$. 108
Thus there exists $\rho_1$, such that for $u$ in $D_o(0, \rho_1)$, the series:

$$\sum_{n=0}^{\infty} a_{G,n}(\bar{\alpha}) u^n$$

is convergent for all $\bar{\alpha}$ in $W_1$ and is the entire series of $G(., \bar{\alpha})$ at $u = 0$.

The coefficients $a_{G,n}(\bar{\alpha})$ correspond to the coefficients of the final expansion (see Section 1.1.3 E)). The problem is then to compute them numerically for any value of $\bar{\alpha}$. This can be done, a priori, in two ways.

### 4.1.1 The contour integral method

Using the fact that $G$ is analytic in $u$ at $u = 0$, we have the obvious result:

$$a_{G,n}(\bar{\alpha}) = \frac{1}{2i\pi} \oint G(u, \bar{\alpha}) \frac{du}{u^{n+1}}$$

where the contour should include 0 and be contained in $D_o(0, \rho_1)$. However under this form, these results are unpractical as the knowledge of $G$ is needed for $u$ describing a whole contour. (A priori $\partial z/\partial u$ is unknown as a function of $u$.) We can modify these expressions by performing the change of variable $u \rightarrow z$ in order to reduce them to known functions. Hence we have:

$$a_{G,n}(\bar{\alpha}) = \frac{1}{2i\pi} \oint \frac{g(z, \bar{\alpha})}{u^{n+1}} dz$$

$$= \frac{1}{2i\pi} \oint \frac{g(z, \bar{\alpha})}{(f(e, \bar{\alpha}) - f(z, \bar{\alpha}))^{n+1}} dz$$

where the last integral involves only known functions. (Although we have assumed that only the values of $f$ and $g$ at the points of interest – here the end-point $e$ – are known, it is reasonable to assume that values for other points are accessible.) The contour integration is then to be done on the image by $z$ of the initial contour in the $u$-plane. In fact the initial contour is of no importance as long as it contains $e$
and is included in the preimage of $D_o(0, \rho_1)$. If we could perform this integration numerically with sufficient accuracy, the results would be at hand.

If this approach is appealing by the elegance and the generality of its form, it contains some difficulties. A numerical contour integration is depending on the geometric shape of the contour – shape which is depending here on the value of $\alpha$ –, the distribution of quadrature points along this contour – which should depend on the local curvature – and the minimum of $|f(e, \alpha) - f(z, \alpha)|$, since the $(n+1)$-th power of this quantity is taken. These choices are determinant for the accuracy of the results and as they depend on $\alpha$ and $n$, they might not be robust for all functions. Last and most important, the fact that there is a choice to be made is an obstacle to a practical usefulness.

Such numerical contour integration could be done by using fast Fourier transform algorithms, but on the necessary condition that the contour be a circle in the variable of integration. (This technique will be used a propos in Chapter 5.) Here this is impossible since $(f(e, \alpha) - f(z, \alpha))$ is involved at the denominator instead of $(e - z)$. This could be overcome by approximating the image by $f$ of the circle of integration, by a circle (after all the definition of $u$ tells us that locally $f$ is not “too far” from being the identity function) in the $u$-plane. But this might require to consider values of the radius that are too small (in order to minimize the error due to this approximation) to provide the necessary accuracy as $n$ increases.

These remarks about the practicality of numerical contour integration will be valid for other cases, and we will not reproduce them subsequently. Of course there might be cases where this approach is effective. For such particular cases this method is in fact more efficient than its alternative.
4.1.2 The point inversion method

The other approach consists in using of the definition equation of the mapping, (4.1), and of course of the fact that \( g \) and \( z \) are \( C^\infty \) at \( z = e \) and \( u = 0 \) respectively. This is motivated by the fact that we have between \( a_G(n, \bar{a}) \) and \( \partial^n G(u, \bar{a})/\partial u^n \), the relation:

\[
a_G(n, \bar{a}) = \frac{1}{n!} \frac{\partial^n G(u, \bar{a})}{\partial u^n} \bigg|_{u=e} \tag{4.7}
\]

The \( n \)-th derivative of \( G \) is then obtained from the definition relation of \( G \), (4.3), by taking its successive derivatives, namely:

\[
\frac{\partial^n G}{\partial u^n}(u, \bar{a}) = \frac{d^n}{du^n} \left[ g(z, \bar{a}) \frac{\partial z}{\partial u}(u, \bar{a}) \right] \tag{4.8}
\]

The right hand side of this last expression is obtained formally for \( 0 \leq n \leq K \), with \( K \) finite. \( K \) corresponds to the index at which the asymptotic expansion is truncated. We thus have a set of \( (K + 1) \) equations, linear in derivatives of \( g \) and involving powers of derivatives of \( z \).

In using the following notations (note that we drop the dependence in \( \bar{a} \) although it remains present in all the quantities):

\[
\begin{align*}
g^{(n)}_e &= \frac{\partial^n g}{\partial z^n}(z, \bar{a}) \bigg|_{z=e} \\
G^{(n)}_e &= \frac{\partial^n G}{\partial u^n}(u, \bar{a}) \bigg|_{u=0} \\
z^{(n)}_e &= \frac{\partial^n z}{\partial u^n}(u, \bar{a}) \bigg|_{u=0}
\end{align*}
\tag{4.9}
\]

we obtain, by carrying these differentiations up to the order \( K \) and taking values at the point \( z = e \leftrightarrow u = 0 \), a set of equations, explicit in the unknowns \( G^{(n)}_e \).
Hence we have:

\[
\begin{align*}
G_{e}^{(0)} &= g_{e}^{(0)} z_{e}^{(1)} \\
G_{e}^{(1)} &= g_{e}^{(0)} z_{e}^{(2)} + g_{e}^{(1)} z_{e}^{(1)^2} \\
G_{e}^{(2)} &= g_{e}^{(0)} z_{e}^{(3)} + 3 g_{e}^{(1)} z_{e}^{(1)} z_{e}^{(2)} + g_{e}^{(2)} z_{e}^{(1)^3} \\
G_{e}^{(3)} &= g_{e}^{(0)} z_{e}^{(4)} + 4 g_{e}^{(1)} z_{e}^{(1)} z_{e}^{(3)} + 3 g_{e}^{(1)} z_{e}^{(2)^2} + 6 g_{e}^{(2)} z_{e}^{(1)^2} z_{e}^{(2)} + g_{e}^{(3)} z_{e}^{(1)^4} \\
G_{e}^{(K)} &= \ldots
\end{align*}
\]

(4.10)

Since we have assumed that \(g_{e}^{(n)}\) was known, or computable, for any \(\bar{c}\) and any \(n\), there remains to determine the values of \(z_{e}^{(n)}\). In theory this would require the knowledge of the mapping \(z\), thus implying the inversion of equation (4.1). Except for particular cases, this inversion is impossible in closed form, that is to express \(z\) as a function of \(u\) and \(\bar{c}\). However we only need point values of the derivatives of \(z\) at \(u = 0\). This corresponds to what we call a “point inversion”.

For this purpose we consider successive derivatives of (4.1), that is:

\[
\frac{d^k}{du^k} \left[ f(z, \bar{c}) - f(e, \bar{c}) = -u \right] \quad \text{for} \quad 1 \leq k \leq K + 1
\]

(4.11)
This results in a system of \((K + 1)\) function equations valid for all \(u\) in \(D_0(0, \rho), \alpha\) in \(W_1\). We then obtain a system of equations in the unknowns \(z_e^{(k)}\), by taking this system at the point \(u = 0 \leftrightarrow z = \epsilon\). Hence we get:

\[
\begin{align*}
    f_e^{(1)} z_e^{(1)} &= -1 \\
    f_e^{(1)} z_e^{(2)} + f_e^{(2)} z_e^{(1)^2} &= 0 \\
    f_e^{(1)} z_e^{(3)} + 3 f_e^{(2)} z_e^{(1)} z_e^{(2)} + f_e^{(3)} z_e^{(1)^3} &= 0 \\
    f_e^{(1)} z_e^{(4)} + 4 f_e^{(2)} z_e^{(1)} z_e^{(3)} + 3 f_e^{(2)} z_e^{(2)^2} + 6 f_e^{(3)} z_e^{(1)^2} z_e^{(2)} + f_e^{(4)} z_e^{(1)^4} &= 0
\end{align*}
\]

This system forms a set of \((K + 1)\) equations non-linear in the \((K + 1)\) unknowns \(z_e^{(k)}\), \(1 \leq k \leq K + 1\). Fortunately this set is triangular and solutions are obtained explicitly at each step \(k\) as functions of \(f_e^{(n)}\), for \(n \leq k\), and powers of \(z_e^{(n)}\), for \(n \leq k - 1\). More explicitly we have:

\[
\begin{align*}
    z_e^{(1)} &= -\frac{1}{f_e^{(1)}} \\
    z_e^{(2)} &= -\frac{f_e^{(2)} z_e^{(1)^2}}{f_e^{(1)}} \\
    z_e^{(3)} &= -\frac{3 f_e^{(2)} z_e^{(1)} z_e^{(2)} + f_e^{(3)} z_e^{(1)^3}}{f_e^{(1)}} \\
    z_e^{(4)} &= -\frac{4 f_e^{(2)} z_e^{(1)} z_e^{(3)} + 3 f_e^{(2)} z_e^{(2)^2} + 6 f_e^{(3)} z_e^{(1)^2} z_e^{(2)} + f_e^{(4)} z_e^{(1)^4}}{f_e^{(1)}} \\
    \vdots
\end{align*}
\]
We note that the forms of the solutions are generic in the sense that they all involve the first derivative of \( f \) taken at \( e, f'_e(1) \), in the denominator.

The procedure

We now have the complete procedure to obtain the coefficients \( a_{G,k}(\bar{\alpha}) \), for \( 0 \leq k \leq K \). We distinguish two main steps, to be performed for each new value of \( \bar{\alpha} \):

1. the computations of \( z^{(k)}_e \), for \( 1 \leq k \leq K + 1 \), from the computed values of the derivatives of \( f \) at \( e, f'_e(n) \) for \( 1 \leq n \leq K + 1 \), using the formulas (4.13)

2. the computations of \( G^{(k)}_e \), for \( 0 \leq k \leq K \), from the values of the derivatives of \( g \) computed at \( e, g'_e(n) \), \( 0 \leq n \leq K \), and powers, up to \( K + 1 \), of \( z^{(n)}_e \), \( 1 \leq n \leq K \). This uses the expressions in (4.10).

The coefficients are then computed from the relation:

\[
 a_{G,k}(\bar{\alpha}) = \frac{1}{k!} G^{(k)}_e
\]  

and are summed into the expansion after being multiplied by the corresponding factor \( k!/N^k \). From there, it is obvious that in fact the expansion should be rewritten in terms of \( G^{(k)}_e \) rather than \( a_{G,k}(\bar{\alpha}) \). Indeed:

\[
 \sum_{k=0}^{K} a_{G,k} \frac{k!}{N^k} = \sum_{k=0}^{K} \frac{G^{(k)}_e}{N^k}
\]  

and naturally, the right hand side is used in the computation of the expansion.

Numerical aspects

The forms of the different equations in (4.13) or (4.10) are very general in the sense that their various coefficients – understood as the quantities \( f'_e(n) \) and \( g'_e(n) \) – depend
only on the choice of the functions \( f \) and \( g \) involved in \( I \), and also on the end-point \( e \) considered. It is therefore very difficult to proceed to a thorough analysis of their numerical properties. However some remarks can be made, based on generic features.

**Floating point number representations**

Considering (4.13), we have noticed that the first derivative of \( f \), taken at the end-point, appears in the denominator. This characterizes the fact that the radius \( \rho \) of the entire series of \( z \) at \( u = 0 \), (4.2), is depending on the proximity of a saddle point of \( f \). Hence the algorithms given by these equations will fail when a col approaches \( e \). Numerically, this could translate into the handling of increasingly large values of \( z^{(n)} \) as \( n \) increases, and the risk of exceeding the range of a given floating point number representation whereas the expansion is still valid. It can be shown that, if \( |f^{(1)}_e| < 1 \):

\[
|z^{(n)}_e| > A \left| \frac{1}{f^{(1)}_e} \right|^n \quad \text{with} \quad A > 0
\]

and therefore that \( (z^{(n)}_e) \) and consequently \( (G^{(n)}_e) \) are increasing sequences of \( n \).

It is interesting to note that all of the equations are homogeneous in \( du \) (this is due to the fact that they are derivatives in \( u \) of some initial equations). Practically, this signifies that if \( z^{(1)}_e \) is multiplied by \( \lambda \), then \( z^{(n)}_e \) is multiplied by \( \lambda^n \), and so is \( G^{(n)}_e \) by \( \lambda^{n+1} \). Therefore it is possible to avoid the troubles mentioned above by choosing an appropriate value for \( \lambda \). This can be done as follows.

Consider the ratio \( |z^{(n+1)}_e/z^{(n)}_e| \). Its limit, as \( n \to \infty \), is \( 1^{-1}/\rho \) since the entire series (4.2) has \( \rho \) for its radius of convergence. Therefore, in order to ensure that \( z^{(n)}_e \) stays bounded, it suffices to take \( \lambda = \rho/2 \). (In fact a smaller value of \( \lambda \) might be needed since we are considering small indices \( n \).) However the value of \( \rho \) is not known a priori and much smaller values maybe needed. But the ratio \( |f^{(1)}_e/f^{(2)}_e| \), if \( f^{(2)}_e \neq 0 \), should give a good upper bound, as \( e \) is approaching a col of \( f \).

Once this is done, there remains to compensate for the scaling at the very end of the
computations in order to recover the right result. The value of $\lambda$ can then be included in $N$, that is $\lambda N$ would replace $N$ is the expansion. Hence this loss of magnitude of the radius of convergence $\rho$, can be entirely accounted by the consideration of smaller values of $N$, and thus by a weaker asymptotic property of the expansion.

- Cancellation errors

From the expressions of the solutions in (4.13) or (4.10), it appears that cancellation errors might occur if all, or some, of the products are large quantities of different signs. An example of this phenomenon is given by considering the simple case where:

\[
\begin{align*}
  f(z) &= -z(z+1) \\
  g(z) &= 2z + 1 \quad (\equiv -f'(z))
\end{align*}
\]

Here we have cancellations of large quantities in the computation of the $G^{(n)}$. The exact result, if $C$ is a path starting at $e$ and ending in a valley of $f$, is for any $e$ (away from the saddle point $z = -1/2$) given by $\exp[N f(e)]/N$. Numerically, this result is recovered exactly if $e$ is on the real axis, but with errors increasing with $k$ (the index of the coefficient) when $e$ is non-real. This is easily seen for $e = 0$. The path $C$ is chosen to start at 0 and to be ending in the valley $z \to +\infty$. 0 is distinct from the col of $f$, and therefore we can hope that the expansion for the end-point case will be efficient. Indeed this is the case, since the final numerical result for $K = 9$ is:

\[1.0/N\]

which is also the exact result. (Note that the expansion is here "totally" asymptotic, an ideal case, as all the other terms are zero.)

It is then interesting to study the formula giving the last term in the expansion, $G^{(9)}$. This expression would be too large to be displayed in full. But thanks to the
fact that $g_e^{(0)}$ and $g_e^{(1)}$ are the only non-zero terms, we are reduced to consider:

$$G_e^{(s)} = 1 \ z_e^{(10)} + 20 \ z_e^{(1)} \ z_e^{(9)} + 90 \ z_e^{(2)} \ z_e^{(8)} + 240 \ z_e^{(3)} \ z_e^{(7)} + 420 \ z_e^{(4)} \ z_e^{(6)} + 252 \ z_e^{(5)}^2$$

after having replaced the $g_e^{(n)}$ by their values. There remains now, to substitute the values of $z_e^{(n)}$ involved in this expression. Thus we have:

$$0 = 1 \ (-1.7643225 \times 10^{10}) + 20 \ (1.0) \ (5.189184 \times 10^8) + 90 \ (-2.0) \ (-1.729728 \times 10^7) + 240 \ (12.0) \ (665280.0) + 420 \ (-120.0) \ (-30240.0) + 252 \ (1680.0)^2$$

where it is obvious that we have cancellation of terms, although without any error!

But let us take another value of $e$, $e = 0.2 + 0.1 \ i$, for the sake of honesty. Then the final numerical result for the same value of $K$ is:

$$\left(\frac{1.0000001}{N} + 2.9802322 \times 10^{-8} \ i - 1.1920929 \times 10^{-7} \ - 2.3841858 \times 10^{-7} \ i \right) \ N^{-1}$$

$$+ \frac{2.1457672 \times 10^{-6} \ i - 4.7683716 \times 10^{-7} \ i}{N^{-2}} + \frac{7.6293945 \times 10^{-5} \ i + 3.8146973 \times 10^{-6}}{N^{-3}}$$

$$+ \frac{-9.765625 \times 10^{-4} \ i - 2.4414063 \times 10^{-4} \ i + 0.0078125 \ i + 0.0036621094}{N^{-4}}$$

$$+ \left(\frac{-0.125 \ i - 0.125 \ i + 1.4375 \ i + 1.25}{N^{-9}} + \frac{1.4375 \ i + 1.25}{N^{-10}}\right) \ exp[-N \ (0.14 \ i + 0.23)]$$

obtained with "single" precision (8 significant digits), whereas the exact result is:

$$\frac{1}{N} \ exp[-N \ (0.14 \ i + 0.23)]
The effects of cancellation errors are there quite evident in the values of the successive terms of the expansion. By example, the last term of the expansion, \(1.4375 i + 1.25\), is the result of additions and subtractions of quantities of the order of \(10^7\).

Mainly three aspects have to be considered with respect to this phenomenon.

1. Cancellation errors at each step are bound to the loss of the last significant digit in the floating point number representations of the largest quantity present at the numerator. Therefore, the use of extended number representations is of particular interest and is effective.

2. These errors are more and more acute as the order of the coefficients increases. In regard of the growing number of operations performed in order to compute these values, and of the propagation of the errors from previously computed quantities, this is not surprising. Their influence on the results is driven by the degree of cancellations which occur in the equations.

3. We must keep in mind that these expansions are asymptotic, and that consequently, only the first few terms have to be computed. Hence, if the floating point number representation is sufficient, the cancellation errors will remain marginal in the numerical result of the expansion.

Thus it appears that this problem certainly does not eliminate the numerical interest of using this method.

A general statement on the occurrence of cancellations in the formulas of (4.13) or (4.10) is most uneasy if not impossible, as it depends exclusively on the properties of the entire series of \(f\) and \(g\) at \(z = e\). In order to make this point clearer, it suffices to consider two simple examples.

With respect to the first step of the procedure (or the computations of \(z^{(n)}\)) we first consider \(f(z) = -\log(1 + z)\), and performs the inversion at \(e = z = 0\) \((\log(1 + z)\) is
analytic there). The result is of course:

\[ z(u) = e^u - 1 \]

Now by expanding \( f \) and \( z \) in entire series we obtain:

\[ f_e^{(n)} = (-1)^{n+1} (n - 1)! \quad \text{for} \quad n \geq 1 \]

and:

\[ z_e^{(n)} = 1 \quad \text{for} \quad n \geq 1 \]

Thus the value 1 for \( z_e^{(n)} \) is obtained after cancellations between terms of the order of \((n - 2)!\). On the opposite, if the inversion of \( f(z) = 1 - e^z \) was to be performed, the results indicate that no cancellation at all takes place. Hence both extremes might occur as well!

The most dramatic case is of course when \( \partial z / \partial u \) turns out to be \( 1/g(z(u, \tilde{a}), \tilde{a}) \) (which is what happens in the numerical example presented above). There the process of obtaining the \( G_e^{(n)} \) is nothing else than the product of two series, inverse one of the other! Of course this is far from being the most efficient way to get the right result, both in terms of number of operations and accuracy. But there is no other alternative, except of being able to recognize that we are dealing with such a case.

It results from such examples, that there exists no solution for treating this problem in a satisfactory manner for general cases, apart from using extended number representations. However, this leaves the freedom to adapt the above procedure to any specific case where improvements ought to be expected: by example through the use of symbolic computation on the analytical formulas of \( f \) and \( g \) (for the evaluation of the coefficients of the expansion) or by identifying trivial cases rather than computing them. This last possibility would require the elaboration of a series of tests to be performed on the chosen functions. But this undoubtedly belongs to a preliminary
analysis of the problem of evaluating the integral $I$, i.e. concerned with its formulation, the use of changes of variables, considerations of similar functions, etc... aspects which are out of the scope, so far, of any symbolic manipulation program (or expert system), and of systematic numerical computations.

- **Particular cases**

  In the case of a polynomial, $f^{(n)}_e$ is equal to zero past a certain row. The formulas giving $z^{(n)}_e$, are then simpler, and the same assumption on $g$ has similar consequences on the expressions of $G^{(n)}_e$. The consideration of simple functions of this kind constitutes good ways of testing the validity of the formal computations, and of assessing the numerical problems that might arise in various cases.

  An obvious application for this algorithm is the case of integrals for which Laplace’s method can be applied. As mentioned above, exact results for integrable cases can be recovered. Limitations in accuracy exist due to the problem of cancellation errors, however they can be handled so as to remain marginal. Of course this algorithm breaks down when the end-point approaches a saddle point of $f$. This is easily seen in the numerical results as the coefficients increase more and more drastically with $n$ as $e$ tends to a col. The uniform treatment of this case is accomplished by considering the more complex mapping of Section 1.2.2 c), but this really belongs to the next section.

  We would need now to present our implementation of this algorithm. As the considerations will be similar for the case of an isolated col, we have grouped these descriptions in Section 4.3.
4.2 The case of an isolated col of order 1

We are now concerned with the expansion for the case of an isolated col of order 1 as given in § 1.1.3 E) by equation (1.17), namely:

$$\exp[Nf(z_1)] \frac{\sqrt{2\pi}}{N^{1/2}} \sum_{k=0}^{K} a_{G,k}(\tilde{\alpha}) \frac{(2k-1)(2k-3)\ldots1}{N^k}$$  \hspace{1cm} (4.18)

for a col $z_1$.

As in Section 4.1, we assume that end-points of $C$, eventual singularities of $g$ and other cols of $f$ are sufficiently distant form $z_1$ for the mapping $u$ defined in Section 1.1.3 C) b) to be valid in some neighborhood of $z_1$. That is we have:

$$f(z, \tilde{\alpha}) = -\frac{u^2}{2} + f(z_1, \bar{\alpha})$$  \hspace{1cm} (4.19)

with $u(., \tilde{\alpha}) : z \mapsto u$ bijective and analytic at $(z_1, \bar{\alpha})$ for any $\bar{\alpha}$ in, say, $W_1$. Under these conditions, we know that (4.18) is asymptotic, and we need to be able to compute the terms of this expansion up to the power, say $(K + 1/2)$ in $N$.

We immediately conclude from (4.18), that we will probably need to compute the $(2K+1)$ terms, that is $a_{G,k}(\bar{\alpha})$ for $0 \leq k \leq 2K$, or at least to perform part of the computations for this range of the indices. In an identical manner to the case of an end-point, the coefficients $a_{G,k}(\bar{\alpha})$ correspond to the coefficients of the entire series of the function $G$ at $(0, \bar{\alpha})$. Although defined through the same formula, (4.3), $G$ differs by the choice of the mapping $z$. Here again two methods are possible to obtain these coefficients numerically.
4.2.1 The contour integral method

Quite evidently, $a_{G, n}(\vec{a})$ is given by:

$$a_{G, k}(\vec{a}) = \frac{1}{2\pi i} \oint \frac{G(u, \vec{a})}{u^{n+1}} du$$

$$= \frac{1}{2\pi i} \oint \frac{g(z, \vec{a})}{u^{n+1}} dz$$

$$= \frac{1}{2\pi i} \oint \frac{g(z, \vec{a})}{2^{(n+1)/2} (f(z_1, \vec{a}) - f(z, \vec{a}))^{(n+1)/2}} dz$$

(4.20)

where the first integral is along a contour around $u = 0$, and the two others along a contour in the $z$-plane around $z = z_1(\vec{a})$. The square roots involve the choice of an appropriate branch of the transformation $z \mapsto u$.

This approach suffers exactly the same drawbacks as its equivalent for the case of an end-point. In addition, the question of the choice of the branch has to be resolved. This may cause problems when using numerical values, since no additional information are available (like the expression of the integrand) except for the values themselves. However this difficulty cannot be avoided as it is embedded in the quadratic transformation (4.19).

Therefore we propose another procedure which follows the same principles as the method described in Section 4.1.2.

4.2.2 The point inversion method

The process for obtaining the coefficients $a_{G, n}(\vec{a})$ uses the relation:

$$a_{G, n}(\vec{a}) = \frac{1}{n!} \left. \frac{\partial^n G}{\partial u^n} (u, \vec{a}) \right|_{u=z_1}$$

(4.21)

which differs from (4.7) only through the value of the point at which the derivatives are taken (and implicitly through the definition of $G$). The expressions of the partial
derivatives of $G$ as functions of the partial derivatives of $g$ and $z$, are identical to (4.8).

Therefore, if we use the notations:

$$
\begin{align*}
&g^{(n)}_{z_1} = \frac{\partial^n g}{\partial z^n}(z, \alpha) \bigg|_{z=z_1} \\
&G^{(n)}_{z_1} = \frac{\partial^n G}{\partial u^n}(u, \alpha) \bigg|_{u=0} \\
&z^{(n)}_{z_1} = \frac{\partial^n z}{\partial u^n}(u, \alpha) \bigg|_{u=0}
\end{align*}
$$

(4.22)

we have the corresponding expressions:

$$
\begin{align*}
G^{(0)}_{z_1} &= g^{(0)}_{z_1} z^{(1)}_{z_1} \\
\left[ G^{(1)}_{z_1} = g^{(0)}_{z_1} z^{(2)}_{z_1} + g^{(1)}_{z_1} z^{(1)^2}_{z_1} \right] \\
G^{(2)}_{z_1} &= g^{(0)}_{z_1} z^{(3)}_{z_1} + 3 g^{(1)}_{z_1} z^{(2)}_{z_1} z^{(1)}_{z_1} + g^{(2)}_{z_1} z^{(1)^3}_{z_1} \\
\left[ G^{(3)}_{z_1} = g^{(0)}_{z_1} z^{(4)}_{z_1} + 4 g^{(1)}_{z_1} z^{(3)}_{z_1} z^{(1)}_{z_1} + 3 g^{(2)}_{z_1} z^{(2)}_{z_1} z^{(1)^2}_{z_1} + 6 g^{(3)}_{z_1} z^{(1)^3}_{z_1} \right] \\
&\vdots \\
G^{(2K)}_{z_1} &= \ldots
\end{align*}
$$

(4.23)

where the lines between brackets need not be evaluated (since the corresponding integrals are 0).

The evaluation of $a_{G, 2k}(\alpha)$ for $0 \leq k \leq K$, requires the computations of $g^{(n)}_{z_1}$ for $0 \leq n \leq 2K$, and of $z^{(n)}_{z_1}$ for $1 \leq n \leq 2K + 1$. Here again, we perform a point inversion of the mapping $u$, but at the col $z_1$, and by considering successive derivatives.
of the definition equation of this mapping, (4.19), namely:

\[
\frac{d^k}{du^k} \left[ f(z, \bar{\alpha}) = \frac{u^2}{2} + f(z_1, \bar{\alpha}) \right] \quad \text{for} \quad 1 \leq k \leq 2K + 2 \quad (4.24)
\]

Hence after differentiation, we have the function equations:

\[
\begin{align*}
\frac{\partial f}{\partial z} \frac{\partial z}{\partial u} &= -u \\
\frac{\partial f}{\partial z} \frac{\partial^2 z}{\partial u^2} + \frac{\partial^2 f}{\partial z^2} \frac{\partial z}{\partial u} &= -1 \\
\frac{\partial f}{\partial z} \frac{\partial^3 z}{\partial u^3} + 3 \frac{\partial^2 f}{\partial z^2} \frac{\partial z}{\partial u^2} + \frac{\partial^3 f}{\partial z^3} \frac{\partial z}{\partial u} &= 0 \\
\frac{\partial f}{\partial z} \frac{\partial^4 z}{\partial u^4} + 4 \frac{\partial^3 f}{\partial z^3} \frac{\partial z}{\partial u^3} + 3 \frac{\partial^2 f}{\partial z^2} \frac{\partial^2 z}{\partial u^2} + 6 \frac{\partial^3 f}{\partial z^3} \frac{\partial z}{\partial u^2} + \frac{\partial^4 f}{\partial z^4} \frac{\partial z}{\partial u} &= 0 \\
\frac{\partial f}{\partial z} \frac{\partial^5 z}{\partial u^5} + 5 \frac{\partial^3 f}{\partial z^3} \frac{\partial z}{\partial u^4} + 10 \frac{\partial^2 f}{\partial z^2} \frac{\partial z^2}{\partial u^3} = 0
\end{align*}
\]

\[
\vdots
\]

(4.25)

- Note that we have shown the first five differentiations of (4.19) and that the upper bound of k is \((2K + 2)\).
Now by considering these equations at $z = z_1 \leftrightarrow u = 0$, we obtain:

\[
\begin{align*}
(0 = 0) \\
&f_{z_1}^{(2)} z_{z_1}^{(1)}^2 = -1 \\
&3 f_{z_1}^{(2)} z_{z_1}^{(1)} z_{z_1}^{(2)} + f_{z_1}^{(3)} z_{z_1}^{(3)} = 0 \\
&4 f_{z_1}^{(2)} z_{z_1}^{(1)} z_{z_1}^{(3)} + 3 f_{z_1}^{(2)} z_{z_1}^{(2)} + 6 f_{z_1}^{(3)} z_{z_1}^{(1)} z_{z_1}^{(2)} + f_{z_1}^{(4)} z_{z_1}^{(4)} = 0 \\
&5 f_{z_1}^{(2)} z_{z_1}^{(1)} z_{z_1}^{(4)} + 10 f_{z_1}^{(2)} z_{z_1}^{(2)} z_{z_1}^{(3)} + 10 f_{z_1}^{(3)} z_{z_1}^{(1)} z_{z_1}^{(3)} + 15 f_{z_1}^{(3)} z_{z_1}^{(1)} z_{z_1}^{(2)} \\
&\quad + 10 f_{z_1}^{(4)} z_{z_1}^{(1)} z_{z_1}^{(2)} + f_{z_1}^{(5)} z_{z_1}^{(5)} = 0
\end{align*}
\]

(4.26)

where $f_{z_1}^{(k)}$ stands for the value of $\partial^k f(z, \alpha)/\partial z^k$ taken at $z_1$.

This constitutes a system of $(2K + 1)$ equations non-linear in the unknowns $z_{z_1}^{(k)}$, $1 \leq k \leq 2K + 1$.

- The first equation, or $0 = 0$, signifies that $u = 0$ corresponds to the col $z_1$ of $f$, by the mapping $u$. As a consequence, the $(k + 1)$-th derivative of the mapping equation (4.19) has to be taken in order to obtain $z_{z_1}^{(k)}$.

Solutions are obtained explicitly since the system is triangular. The value of $z_{z_1}^{(k)}$ is expressed in terms of the $(k + 1)$ first derivatives of $f$ at $z_1$, and of powers of $z_{z_1}^{(n)}$ for $1 \leq n \leq k - 1$. 

125
Hence we have:

\[
\begin{align*}
    z_{z_1}^{(1)} &= \frac{1}{(-f_{z_1}^{(2)})^{\frac{1}{2}}} \\
    z_{z_1}^{(2)} &= -\frac{f_{z_1}^{(3)} z_{z_1}^{(1)^2}}{3 f_{z_1}^{(2)}} \\
    z_{z_1}^{(3)} &= -\frac{3 f_{z_1}^{(2)} z_{z_1}^{(2)^2} + 6 f_{z_1}^{(3)} z_{z_1}^{(1)^2} z_{z_1}^{(2)} + f_{z_1}^{(4)} z_{z_1}^{(1)^4}}{4 f_{z_1}^{(2)} z_{z_1}^{(1)}} \\
    z_{z_1}^{(4)} &= -\left(10 f_{z_1}^{(2)} z_{z_1}^{(2)} + 10 f_{z_1}^{(3)} z_{z_1}^{(1)^2} z_{z_1}^{(2)} + 15 f_{z_1}^{(3)} z_{z_1}^{(1)} z_{z_1}^{(2)^2} + 10 f_{z_1}^{(4)} z_{z_1}^{(1)^3} z_{z_1}^{(2)} + f_{z_1}^{(5)} z_{z_1}^{(1)^5}\right) / 5 f_{z_1}^{(2)} z_{z_1}^{(1)}
\end{align*}
\]

(4.27)

- All of these solutions involve the quantity \( f_{z_1}^{(2)} \) at their denominator. Since the col \( z_1 \) is of order 1, this value is of course non-zero. Apart for the square root in the first equation, this quantity always appears in the form of the product \( f_{z_1}^{(2)} z_{z_1}^{(1)} \).

- The forms of these equations are more complex than their equivalent in (4.13).

**The procedure**

We can now decompose the procedure in two main steps, to be performed for each new value of \( \bar{\alpha} \):

1. the computations of the quantities \( z_{z_1}^{(k)} \), for \( k \) in \([1,..,2K+1]\) using the formulas of (4.27). This requires the computations of the values \( f_{z_1}^{(k)} \) for \( k \) in \([2,..,2K+2]\).

2. and then the computations of \( G_{z_1}^{(2n)} \), for \( n \) in \([0,..,K]\) by using the
corresponding formulas in (4.23). This requires the computation of \( g^{(k)}_{z_1} \), for \( k \) in \([0,..,2K]\).

The expansion is then obtained through a summation, directly from the \( G^{(2n)}_z \), having noticed that:

\[
a_{G,2k} \frac{(2k - 1)(2k - 3) \ldots 1}{N^k} = \frac{G^{(2k)}_{z_1}}{k!} \frac{1}{(2N)^k}
\]

(4.28)

Numerical aspects

The numerical problems encountered when using the equations in (4.27) and (4.23) are similar to the ones faced when treating the end-point case. Therefore we will resort to the same remedies when applicable. However, certain difficulties are specific to the use of the quadratic mapping and need to be addressed.

- The problem of the square root

Such a problem is raised by the choice of the branch of the square root in the first equation of (4.27):

\[
z^{(1)}_{z_1} = \frac{1}{(-f^{(2)}_{z_1})^{\frac{1}{2}}}
\]

(4.29)

A change in the sign of \( z^{(1)}_{z_1} \) induces a change of sign in all the quantities \( z^{(n)}_{z_1} \) for \( n \) odd, whereas for \( n \) even, the signs remain the same. This results in the change of signs of all the coefficients \( G^{(2k)}_{z_1} \) in the expansion. (This can be seen directly in the derivation of the expansion, as the choice of the branch of the mapping \( u \) signifies that the path \( S^*(z_1) \) or its opposite is chosen.) The difficulty is then not so much to choose the right branch – as this can be corrected by a change in the sign of the result – but to make this choice consistent throughout the computation. For this purpose we need to relate the local information needed – the branch to choose – to the global process of deforming the initial path of integration \( C \).
This is done as follows. For a given value of $\bar{\alpha}$ which is not a ramification value, say $\bar{\alpha}_{\text{ref}}$, we have determined the set of relevant cols (of which $z_1$ is an element), and in the process the two (and exactly two) basins, say $B_-$ and $B_+$, for which $z_1$ is a pass. When integrating along $S^*(z_1)$, $z$ goes from $B_-$ to $B_+$ through $z_1$. This must correspond to the changing of the values of $u$ from being negative to positive, in order to respect the direction of the integration along $C$. Since $\bar{\alpha}_{\text{ref}}$ is, $B_-$ and $B_+$ are the only basins to be connected by $z_1$. Then, a point $z$ of $B_-$ in a neighborhood of $z_1$ is characterized by:

$$\Re \left\{ \frac{(z - z_1)}{z_1^{(1)}} \right\} < 0$$

which determines uniquely the sign of $z_1^{(1)}$ for the value $\bar{\alpha}_{\text{ref}}$. ($z$ can be taken as the starting point of a path of steepest descent in $B_-$.) From there we choose the determination of $\arg\{-1/f_1^{(2)}(z_1^{(1)})\}$ which is in $[0, 2\pi[$. This constitutes a reference value for subsequent computations of $z_1^{(1)}$. The square root of $-1/f_1^{(2)}$ is naturally taken as of argument $\arg\{-1/f_1^{(2)}\}/2$.

For any other values of $\bar{\alpha}$ in the same region $\bar{O}_t$, the determination of $z_1^{(1)}$ is based on continuity arguments along a path joining $\bar{\alpha}_{\text{ref}}$ to $\bar{\alpha}$ (see Appendix C for details of the procedure and its implementation). Such a determination is then uniform on $\bar{O}_t$ and stays valid as long as $z_1$ is a relevant col of order 1 (thus eventually through several $\bar{O}_t$).

**Floating point number representations**

Here, we face a similar problem as the one encountered with the procedure for an end-point. Concerning (4.27), we have noticed that $f_1^{(2)}$ always appear in the denominator. In fact apart from the second formula, the square root of this quantity is involved (after replacing $z_1^{(1)}$ by its value). Again, we can relate the presence of $\sqrt{f_1^{(2)}}$ as being a measure of the proximity of another col of $f$. The procedure fails when such a col approaches $z_1$. As usual, a numerical implementation of these formulas will fail to be accurate and will exceed the capacity of a given number representation, well before being close to the actual boundary of its domain of validity.
Since these expressions are derived from the same equation, we have the same homogeneity property with respect to $du$. Hence if $z_{21}^{(1)}$ is replaced by $\lambda z_{21}^{(1)}$, $z_{21}^{(n)}$ is multiplied by $\lambda^n$ and $G_{21}^{(2k)}$ is multiplied by $\lambda^{2k+1}$. The scaling is then compensated by including $\lambda$ into $N^{1/2}$. An upper bound for $\lambda$ is here given by $|f_{21}^{(2)}/2f_{21}^{(3)}|$ (if $f_{21}^{(3)} \neq 0$). This keeps all the magnitudes of the numerical values below a given bound.

- **Cancellation errors**

  The characteristics of the formulas in (4.27) differ in no way from the ones of (4.13) with respect to the problem of cancellation errors. To convince ourselves, we just need to reconsider the example of $f$ being defined by a logarithmic function, but this time in the form of $f(z) = -\log(1 + z^2)$. Then its inversion ($f$ has indeed a saddle point at $z = 0$) using the quadratic map at $z = 0$, will give:

  $$ z(u) = (e^{u^2} - 1)^{1/2} $$

  The coefficients of the corresponding entire series show the same pathological display of cancellations, whereas the results from inverting the function $f(z) = 1 - \exp[z^2]$ will not demonstrate such phenomenon.

  The problems belonging to the computations of the $G_{21}^{(2k)}$ are unchanged, as the procedure is identical.

  Therefore, the same conclusions can be made regarding the influence of cancellation errors on the numerical computation of the coefficients. However their effects are more acute, as more terms are required to obtain the same asymptotic "quality" as a Laplace kind expansion (the expansion is in $G_{21}^{(2k)}/N^k$). It might then be necessary to increase the lower bound of $N$ for which a given accuracy is expected, in order to be able to truncate the series at a lower $k$.

  The total error in the numerical result of the expansion is accounted by the truncation error (given by the asymptotic properties of the series) and the error due to
cancellations. Both errors contribute in different manners depending on the value of $N$. For small values of $N$, the loss of the asymptotic properties of the expansion is dominant. In other words the expansion is truncated at a value of $k$ too small for cancellation errors to be significant. On the contrary, for larger values of $N$ more terms can be included in the expansion (their magnitude are now decreasing for a greater range of the index $k$) if the minimal error is sought. Cancellation errors might then be predominant since they increase with $k$, and the numerical estimate of the truncation error would be erroneous. In such cases Dingle (Dingle, 1973) advocates an asymptotic expansion of the remainder (or hyperasymptotics) rather than the search for higher order terms. This could indeed furnish a remedy to the loss of accuracy in higher order terms.

Again, we face the impossibility of formulating any general conclusion about the limitations inherited from cancellation errors, as this depends on the functions $f$ and $g$. We can only sketch some tendencies.

Let us assume that the value of the $n$-th derivative of $f$ at $z = 0$ is $n!$. The absolute error for each $n$ is at least $n! \times \text{(machine accuracy)}$. We can then form the following table based on machine floating point number representations of 8 and 16 digits.

<table>
<thead>
<tr>
<th>$n$</th>
<th>machine accuracy</th>
<th>8 digits</th>
<th>16 digits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$O(10^{-7})$</td>
<td>$O(10^{-15})$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$O(10^{-6})$</td>
<td>$O(10^{-14})$</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>$O(10^{-3})$</td>
<td>$O(10^{-11})$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$O(10^{-1})$</td>
<td>$O(10^{-9})$</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>$O(1)$</td>
<td>$O(10^{-8})$</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>$O(10^{8})$</td>
<td>$O(1)$</td>
<td></td>
</tr>
</tbody>
</table>

These values are to be multiplied by $1/N^n$ for an end-point expansion or $1/N^{n/2}$ for an isolated col expansion. They are to be compared to the actual values of the
coefficients, which could be, as we have seen, 0 or of the order of \( n! \), if not larger. In any case the truncation of the expansion is driven by the values of the terms \( z_0^{(n)}/N^n \) (or \( z_0^{(2n)}/N^n \)), and the estimates of the absolute errors are given by \( k! \times (\text{machine accuracy})/N^k \) (respectively \( 2k! \times (\text{machine accuracy})/N^k \)), for \( 0 \leq k \leq K \).

In the worst case (depending on the behavior of the \( z_0^{(n)} \)) the values of the terms might be increasing for a lower value of \( k \) than in theory, due to the overtaking of cancellation errors. It is then wise to truncate the series at the value of \( K \) for which the modulus of the corresponding term is the smallest, and not the theoretical value of \( K \) given by the analytical expression of the expansion. Even then, truncation errors can be significant (e.g. the integrable case of Section 4.1.2) with respect to the theoretical accuracy of the expansion. Nevertheless, they can be brought well below the desired numerical accuracy by choosing extended number representations.

- **Particular cases**

The treatments of simple cases for which solutions are known constitute good tests. The example of Section 4.1.2 can be used again, but this time by choosing \( C \) to be any path starting at \( \infty e^{is} \) and ending at \( \infty \) (see Section 4.3 for other examples as well).

There remains to discuss how such procedures can be implemented on a computer and what are the practical restrictions of such implementations.

### 4.3 The implementation

#### 4.3.1 The principles

The complete algorithms for obtaining the expansions (1.16) and (1.17) for a given value of \( N \) and \( \alpha \), are built along identical principles. In both cases the procedures as given in Sections 4.1.2 and 4.2.2 are used, but in a different layout. This is due
to the fact that the index $K$, at which the appropriate expansion is truncated, is not known a priori and depends on the values of the coefficients of the expansion. The value of $K$ is given by the following application of the asymptotic properties of the expansions.

1. **The truncation test**

Both expansions are asymptotic in the Poincaré sense, or in other words are such that:

*The error made in truncating the expansion is “o” of the last term included in the expansion.*

Therefore if we follow this definition and if we desire the maximum accuracy provided by the expansion, $K$ should be taken as the value of the index $k$ for which the modulus of the coefficient of index $k$ is minimum. In our cases these moduli are given by:

$$\text{mod}_k = \left| \frac{G^{(k)}_e}{N^k} \right|$$  \hspace{1cm} (4.30)

for the end-point case, and:

$$\text{mod}_k = \left| \frac{G^{(2k)}_{2,1}}{k! \frac{1}{(2N)^k}} \right|$$  \hspace{1cm} (4.31)

for the case of an isolated saddle point of order 1.

Then the truncation criterion is formulated as follows. $K$ is the largest index such that:

$$\forall k \leq K \quad \text{mod}_k > \text{mod}_K$$  \hspace{1cm} (4.32)

As we have seen in Section 4.2.2, this condition is appropriate to the requirement of minimizing the influence of cancellation errors. A better guarantee is in fact to impose an alternative condition based on a fixed error tolerance $\epsilon$. 

132
The purpose of such a constraint is in fact double:

— one is to limit the influence of cancellation and round off errors, by imposing a minimum accuracy well below the accuracy of the machine,

— the other is to bound the computational time for large values of $N$. (The theoretical accuracy of the expansion is increasing with $N$ by including more terms in the expansion and hence increasing the computational task.)

Hence $K$ is chosen such that:

\[
\forall k \leq K \quad \text{mod}_k > \text{mod}_K
\]

or

\[
\text{mod}_k < \epsilon \leftrightarrow k = K
\] (4.33)

In practice, the finite sequence $(\text{mod}_k)$, for $k \leq K$ is often a strictly decreasing sequence of $k$, whereas for $k \geq K$ it is strictly increasing. Therefore, $K$ can be chosen as the smallest index for which we have:

\[
\text{mod}_K \leq \text{mod}_{K+1}
\] (4.34)

However, it might be the case that $(\text{mod}_k)$, for $k \leq K$, is non strictly decreasing. Then criterion (4.34) might result in a truncation of the expansion for a value of $K$, too small. One can then enforce a stronger requirement, formalizing an increasing sequence of three consecutive terms:

\[
\text{mod}_K \leq \text{mod}_{K+1} \leq \text{mod}_{K+2}
\] (4.35)

Although this last criterion is not failure proof, it provides a rather reliable test. Any criterion of course can be associated with an alternative constraint such as in (4.33).
It thus appears that in fact $k + 2$ terms of the expansion are needed for applying the truncation test (4.35) at $k$. In terms of efficiency we would like to minimize the number of terms computed for a given $K$. A criterion satisfying such constraint is:

$$\mod_{K-2} \geq \mod_{K-1} \geq \mod_{K} \quad (4.36)$$

However this last example is a priori less reliable.

In any case, in order for the algorithm to be the most efficient, the value of the coefficient at each step should be obtained by reducing the computation of the quantities involved in the procedure, to the strict necessary. This is done by taking advantage of already computed quantities from previous steps.

2. Reducing the computational task

In order to reduce the additional computational task when going from the step $k$ to the step $k + 1$, the algorithms have to obey a case specific rule. The rule for each case is deduced from the forms of the coupled systems (4.13) and (4.10), or (4.27) and (4.23).

End-point case

For each $k$, are evaluated in the following order:

1. $f_e^{(k+1)}$
2. $z_e^{(d)}$ for $E(k/d) \leq p \leq E((k + 1)/d)$, with $1 \leq d \leq k$
3. $z_e^{(k+1)}$ from the corresponding formula in (4.13)
4. $g_e^{(k)}$
5. $G_e^{(k)}$ from the corresponding formula in (4.10)
Isolated col case

For each $k > 0$ – i.e. for computing $G_{z_{1}}^{(2k)}$ – are evaluated in the following order:

1. $f_{z_{1}}^{(2k+1)}$ and $f_{z_{1}}^{(2k+2)}$
2. $z_{z_{1}}^{(d)p}$ for $E(2k/d) \leq p \leq E((2k + 1)/d)$, with $1 \leq d \leq 2k - 1$
3. $z_{z_{1}}^{(2k)}$ from the corresponding formula in (4.27)
4. $z_{z_{1}}^{(d)p}$ for $E((2k + 1)/d) \leq p \leq E((2k + 2)/d)$, with $1 \leq d \leq 2k$
5. $z_{z_{1}}^{(2k+1)}$ from the corresponding formula in (4.27)
6. $g_{z_{1}}^{(2k-1)}$ and $g_{z_{1}}^{(2k)}$
7. $G_{z_{1}}^{(2k)}$ from the corresponding equation in (4.23)

Most of the computational effort, at least the one we can control, is spent in computing $z_{z}^{(k)}$ and $G_{z}^{(k)}$. From the forms of the formulas used for this purpose, it appears that a great latitude is left with respect to their encoding. But first of all, how can these formulas be obtained for an arbitrary $k$?

4.3.2 Obtaining the formulas

As the first step in elaborating the algorithms, the generation of the formulas is crucial although it is done once and for all. These formulas result from the differentiations of a composition of functions: $f \circ z$ or $g \circ z$. This task is purely algebraic and the first equations are obtained easily by hand. But getting the correct result for the order 9, by example, is a cumbersome and tedious task, by hand. However such computations were carried out by Dingle (Dingle, 1973, Chap V) for $0 \leq k \leq 8$. These formulas were obtained with a mechanical calculator! (Computations of the formulas for the first 10 terms take 5 hours in Macsyma on a Symbolics 3640 machine.) Since the process is repetitive and automatic, programs can be written in nearly any computer language to perform this specific task. Of course the more specific $f$ and $g$ are, the
easier the programming becomes. However, it is then very difficult to manipulate the results, i.e. the representations of these formulas, without “reinventing the wheel”.

Symbolic manipulation programs have been available in mature forms for about fifteen years now, and it would have been a waste not to take advantage of them. Such programs are especially useful when attempting to optimize the encoding of these formulas in view of numerical computations. Any numerical programming language is unable to handle such formulas in a systematic manner. We do not even mention the eventuality of manipulating by hand an expression that would fill more than a page!

Such programs can in fact be used for much more, if not all of the implementation. The ability of deriving formally these expressions, manipulating them systematically, for optimization and improvement of accuracy, of encoding their final forms for numerical computations; the ability of performing numerical computations with an arbitrary (although finite) accuracy, are invaluable for the investigation of the numerical properties of these algorithms. None of the usual computer languages (FORTRAN, C,...) can provide the necessary environment for performing such tasks. Symbolic manipulation is here necessary and appears as a powerful tool for the generation of the formulas, their optimization and their encoding – i.e. the design of the algorithms – and for performing numerical experiments.

The ease with which various examples can then be treated, simply by specifying the analytical expressions of the functions \( f \) and \( g \), apart for testing purposes, is particularly helpful for the understanding of the mathematical properties of such expansions. Indeed this aspect could be developed and extended to create a pedagogical, as well as practical, tool for applications of the method of steepest descents. Here again the frame of symbolic computation is unique in preserving the generality of the method in its applications (whether numerical or analytical).

The optimization for a specific case, through the analytic knowledge of the functions \( f \) and \( g \), can easily be done and tested, a task which otherwise would have to be
done by hand. Then it belongs to the user to choose whether or not to translate the algorithms into a computer language specialized for numerical computation. In fact a real opportunity for portability is offered.

As an illustration to the capabilities of such a programming environment, all of the formulas, the numerical results and their break down in Sections 4.1.2 and 4.2.2 are **TeX** versions of the outputs from various functions written and executed in Symbolics Macsyma.

### 4.3.3 Optimization of the formulas

The optimization of the formulas with respect to the number of floating point operations (or FLOPs) has to be done by considering simultaneously the systems of equations giving $z^{(k)}_2$ and $G^{(k)}_2$. From such a consideration it appears that apart from the presence of $f^{(k)}_1$ in one and $g^{(k)}_2$ in the other, the terms are identical. Each of the equations giving $z^{(k+1)}_2$ and $G^{(k)}_2$ involve the products:

$$\prod_i z^{(d_i)p_i}_2 \text{ such that } 1 \leq d_i \leq k , \quad \sum_i d_i \ p_i = k + 1$$ 

multiplied by the same integer factors. Under the cost of storing, in memory, the values of these products multiplied by their corresponding factor between the evaluation of $z^{(k+1)}_2$ and the evaluation of $G^{(k)}_2$, the number of multiplications in the computations of $G^{(k)}_2$ is significantly reduced compared to the rough number of $(2^{k+1} - 1)$.

However this gain is pertaining to the computation of $G^{(k)}_2$. The task of evaluating the products (4.37) has to be done anyway when computing $z^{(k+1)}_2$. But quantities computed at previous steps can be used in order to minimize this task. Here the problem is much more complex, as an optimization of this process might imply modifications of the algorithms for the computation of $G^{(k)}_2$. It can be shown that if all of the products (4.37) are known up to $k + 1$, the products at $k + 2$ are obtained
in exactly (number of such products\( -1 \)) multiplications. (The number of products (4.37) is rather difficult to assess.)

Tools exist in symbolic manipulation programs to perform this kind of task. For example, use of the function \texttt{optimize( )} in Macsyma can generate an optimized version of the systems of equations for a given \( K \).

Once the optimization is fixed, translation functions can be used to generate the equivalent statements in another language. (This possibility has been used in the generation of the FORTRAN subroutines for the example of Chapter 6.) The advantage is then to benefit from faster numerical computations allowed by the use of these languages.

### 4.3.4 Some examples

As illustrations, we present here some examples of an implementation of these algorithms in the frame of Macsyma. This results in a set of less than a dozen Macsyma functions which perform the necessary tasks, starting with the analytical definitions of the functions and producing the asymptotic expansion at a given point (end-point or isolated col of order 1). Interactive capabilities of Macsyma are used when prompting for the definitions of the functions \( f \) and \( g \). This is summarized by screen hardcopies (i.e. snapshots) as each example is treated, or by Macsyma \TeX{} outputs of the results: formulas (4.41) and (4.42) in Example 3.

**Example 1**

We promised at the end of Section 4.2.2 to reconsider the example of Section 4.1.2 but with a different path of integration. Namely we now consider the integral \( I \) given by:

\[
I(N) = \int_{\infty}^{\infty} (2z + 1) \exp[-N z (z + 1)] \, dz
\]

for which the result is now 0.
The first step is to give the analytical expression of \( f \) and the order of the highest derivative of \( f \) that will be necessary to consider (i.e. \( 2K + 1 \)) since we are treating the isolated col case).

The formula should be typed in Macsyma syntax. This will correspond to the definition of the Macsyma function:

\[
\text{FUNC}(f)(Z) := \{ \text{expression in } Z \text{ and parameters} \}
\]

Figure 4-1: Example 1: analytical formula of \( f \)

The second step is to define \( g \) by its analytical expression and also the order of its highest derivative (in principle \( 2K \)).

The various functions generate then the corresponding derivatives of \( f \) and \( g \), as Macsyma functions, and forms (if they are not available yet) the formulas for the computations of \( z_2^{(k)} \) and \( G_2^{(k)} \) up to the maximum order allowed by the restrictions given by the user. It remains to compute the final expansion: we know that the only relevant col to consider is \( -1/2 \) and therefore asks for the isolated col expansion at this col for \( K = 4 \) (i.e. \( 2K = 8 \)) by example. Of course we recover the exact result, i.e. 0.
g should be an ANALYTIC function of $Z$ and of some parameters.

* The same remarks as for $f$ apply.

This is the expression of $g$ that will be used hereafter:

$$g(Z) = 2Z + 1$$

Figure 4-2: Example 1: analytical formula of $g$

**Example 2**

As the last simple example, we consider $I$ defined by:

$$I(N, \alpha_1) = \int_{-\infty}^{\infty} \exp[-N z(z - 2\alpha_1)] \, dz$$

for which we know that the only col $z = \alpha_1$ is always relevant.
Following the same process, we get as a final answer:

$$\frac{2}{\sqrt{N}} \approx 1.7724638509055163$$

Figure 4-3: Example 2: result

where $\alpha_1$ stands for $\alpha_1$ and $e$ for $e$. The number 1.7724538509055163 is the value of $\sqrt{\pi}$ with an error of $5 \times 10^{-16}$. (Note that it is not necessary to use numerical values, and that the analytical form of the expansion can be obtained.) One can check that this result, apart from the round off error is the exact result.

Example 3

As a final example we treat the following Bessel function (see Watson, 1948, §8.31, 8.4):

$$J_N\left(\frac{N}{\cosh B}\right) = \int_{\infty - i\pi}^{\infty + i\pi} \exp\left[N \left(\frac{\sinh z}{\cosh B} - z\right)\right] dz$$

with $B$ real and positive.

- This example is historical. The asymptotic expansion for large $N$ was established by Debye in his article of 1909 in which he formulated, for the first time, the method of steepest descents.
The resulting asymptotic expansion given by Watson (1948, §8.4 eqs (3) & (4))
is, in a slightly different form:

\[ J_N\left( \frac{N}{\cosh B} \right) \sim e^{N(tanh B - B)} \frac{A_m}{N^{1/2}} \sum_{m=0}^{\infty} \frac{A_m}{\sqrt{2\pi} (\tanh B)^{m+1/2}} \frac{(2m - 1)(2m - 3)\ldots 1}{N^m} \]  

(4.39)

with:

\[
\begin{align*}
A_0 &= 1 \\
A_1 &= \frac{1}{8} - \frac{5}{24} \coth^2 B \\
A_2 &= \frac{3}{128} - \frac{77}{576} \coth^2 B + \frac{385}{3456} \coth^4 B \\
&\quad \vdots
\end{align*}
\]

(4.40)

Additional terms can be found in Abramovitz and Stegun (1964).

The details of the computations of the first three values of \( A_m \) are also given and
it appears that the task is rapidly becoming tedious as \( m \) increases. Our procedure
to obtain the coefficients is in no way different, apart from the use of a computer to
carry it.

This is an example where we can also produce the analytical formulas of the
coefficients up to an arbitrary order. Hence the expansion truncated at \( K = 2 \) gives:

\[ J_N\left( \frac{N}{\cosh B} \right) \sim \left( 1 + \frac{\cosh B}{8 \sinh B N} + \frac{-5 \cosh^3 B}{24 \sinh^3 B N} \right. 
\]

\[ + \frac{9 \cosh^2 B}{128 \sinh^2 B N^2} + \frac{-77 \cosh^4 B}{192 \sinh^4 B N^2} + \frac{385 \cosh^6 B}{1152 \sinh^6 B N^2} \bigg) \]

\[ \times \frac{\cosh^{1/2} B}{\sinh^{3/2} B} \frac{e^{N(tanh B - B)}}{2^{3/2} \pi^{3/2} N^{1/2}} \]

(4.41)

which is the result given by Watson.

In order to perceive the numerical properties of this asymptotic series, we can
then ask for numerical results for various values of \( B \) as long as \( B > 0 \) (in theory),
and observe its failure as $B \to 0^+$. Such a result, for $B = 1$ and $2K = 14$, is:

$$J_N\left(\frac{N}{\cosh 1}\right) \sim \left(-4.5352466 \times 10^{-10} i + 0.45713902\right)$$

$$+ \frac{4.183574 \times 10^{-10} i - 0.14056382}{N}$$

$$+ \frac{-1.4553914 \times 10^{-9} i + 0.29339802}{N^2}$$

$$+ \frac{8.851943 \times 10^{-9} i - 1.2746413}{N^3}$$

$$+ \frac{-7.2625326 \times 10^{-8} i + 8.133791}{N^4}$$

$$+ \frac{7.5040896 \times 10^{-7} i - 68.762665}{N^5}$$

$$+ \frac{-9.347177 \times 10^{-6} i + 724.7438}{N^6}$$

$$+ \frac{1.362068 \times 10^{-4} i - 9152.82}{N^7} \right) e^{-0.23840685 N}$$

(4.42)

(obtained with a 8 digit mantissa in Macsyma, on a Symbolics 3600 series machine).

We can remark that this expansion is not very good in the asymptotic sense: values of $N$ above 10 have to be considered in order to have a remainder – in keeping 6 terms – smaller than $6 \times 10^{-4}$. Round off and cancellation errors are obvious in the imaginary parts of the coefficients (which should be purely real), but are negligible in front of the magnitudes of the coefficients.

Of course one could think of many other examples. But as long as the method of steepest descents is applicable and only end-points and isolated cols of order 1 are involved, the algorithms described in this chapter provide systematic means for obtaining numerical results. This capacity, of course, is entirely depending on the
ability of locating the relevant saddle points.

4.4 Conclusion

We have presented two procedures and their implementation, for obtaining numerically the asymptotic expansions given by the method of steepest descents: one, for the contribution from an end-point of the path of integration, the other, for the contribution from an isolated saddle point of order 1. Both of these cases correspond to the application of the "classical" method of steepest descents (as opposed to the generalized method of steepest descents). Hence, we have set aside the considerations of coalescing algebraic singularities or saddle points (singularities in the sense of singularity theory). In this frame, we can state that these algorithms are robust, efficient and of general use for the purpose of numerical applications.

1. Robustness

The asymptotic nature of the expansions obtained by the method of steepest descents, guarantees the robustness of the procedures. However, this would fail if the corresponding algorithms (or implementations) did not preserve the asymptotic properties of the series. The study of these procedures showed that the main source of numerical errors is caused by the occurrence of cancellations in the computations of the coefficients in the expansions. It appeared (Section 4.1.2) that no general treatment exists to avoid such phenomena apart maybe from the possible use of hyperasymptotics. This constitutes a possible development for numerical applications. Nevertheless, influence of cancellations on the results can be rendered negligible in two ways:

— upon simplifications, in particular instances, of the formulas giving the coefficients. This includes the treatment of "integrable" cases or trivial situations like in Section 4.1.2.
— in a more systematic manner, by the use of extended number representation on a computer at the price of increasing the computational time.

In any case, these errors become significant only through the inclusion of more terms in the expansions. Since the expansions are asymptotic, this effect is limited to the search for very high accuracy.

2. Efficiency

Several aspects intervene in the efficiency of these algorithms. The cost of computing an additional term in the expansion is increasing with the index $k$ of the term. This trend is present in both algorithms, although it is much more pronounced in the algorithm for the isolated col case (an additional term requires the computation of two quantities $z_2^{(k)}$). The efficiency of these procedures depends then on the ability of optimizing the corresponding algorithms with respect to the number of floating point operations.

- For large values of $N$, as the theoretical accuracy of the expansion increases, the computational cost can be very high if more terms are included to improve the accuracy. The bound on the efficiency is then driven by the choice of the truncation test. A test which includes an alternative fixed accuracy constraint such as (4.33), will guarantee a lower bound on the efficiency for large $N$. Thus we note that the test given by (4.35) might reveal to be very expensive, as it requires the computation of 2 additional terms after the last term included in the series.

- For small values of $N$, the theoretical accuracy of the expansion is limited, and poor resulting efficiencies should not be imputed to the algorithms, but rather to the inappropriate application of the method of steepest descents.

If speed is important, then other methods for obtaining the coefficients can be considered: table look-up, approximation by basis functions. In any case, they will be based on numerical values computed by the above algorithms. Case specific improvements are possible, but not always guaranteed, through the use of symbolic
computation in order to derive analytical formulas for the coefficients.

3. Generality

In their essence, the procedures presented here, are general and applicable to a wide class of functions, namely functions for which the classical method of steepest descents can be applied in regions of their parametric space. The preservation of this generality in the corresponding algorithms is made possible by the frame specified in Section 2.4. The practical restrictions are then the ability of locating the relevant saddle points and computing the derivatives of the functions $f$ and $g$ at these points.

As already mentioned, the role played by symbolic computation is there crucial. Indeed, the generation of the algorithms and their optimization are only possible through the use of symbolic computations. Such an environment brings the additional advantages of performing numerical experiments, and the possibility of obtaining analytical results as well, preserving the generality of the method in its entirety. On a more practical level, it also provides the choice for encoding these algorithms into different "numerical" computer languages. After all, these procedures have been known for a long time and it is only with the help of symbolic computation that they are of practical use in numerical applications.

In a uniform numerical treatment of the integral $I$, it would remain to verify that the various expansions obtained by applying these algorithms at the relevant cols and end-points, match within a required tolerance (the $\epsilon$ of the preamble) along Stokes lines (the common boundaries between two domains $\tilde{O}_i$). However, this would leave the eventual cases of the coalescence of saddle points and algebraic singularities untreated (the collection $\{\tilde{O}_i\}$ would not then be a covering of $D$). These situations do occur in solutions to physical problems. Can then algorithms with similar properties of robustness, efficiency and generality be found for the uniform treatment of such cases? We propose to investigate this question by considering the "simplest" case, that is the uniform treatment of the coalescence of two cols, or fold catastrophe.
Chapter 5

The fold catastrophe

We pursue here, in the same spirit as the previous chapter, a detailed study of the algorithms for the numerical computations of the simplest case of catastrophe expansions: the fold catastrophe. This analysis, permitted again by the use of symbolic computation, is entirely new. We start by recalling the classical result from Chester, Friedman and Ursell (1957), which is at the origin of the uniform treatments of catastrophes in the evaluation of integrals with a large parameter (Section 5.1). After reviewing the contour integral method for the corresponding Bleistein sequence (Section 5.2), we detail the point inversion method for the series of Chester et al. (Sections 5.3, 5.4 and 5.5). The cornerstone of this procedure lies in the actual inversion of the unfolding of the fold catastrophe (Sections 5.3.4, 5.4 and Appendices B, C, D). This major difficulty compels us to only produce guidelines (Section 5.6) for building robust implementations in the case of isolated fold catastrophes.

Unlike Chapter 4, we now allow two cols of $f$ to coalesce for certain values of $\alpha$. Hence we consider values of $\alpha$ in a domain $O^2_{\alpha}$ (that is a domain of non-empty intersection with $K$, the set of catastrophes of $f$), but away from the sets $K_2$, $C_C$ and $C$ (i.e. only two saddle points are coalescing without neighboring end-points or algebraic singularities of $g$). We are thus in the conditions expressed in Section 1.2.1, for $m = 2$. Then, $f$ presents a fold catastrophe for say $\alpha = \alpha_0$, $\alpha_0$ in $K$. 

147
We deliberately use the terminologies of catastrophe theory and of singularity theory, although this is not entirely necessary. The aim in doing so is double. First, we wish to emphasize the difference between the preceding chapter where the parameter $\bar{a}$ was always an ordinary point of the mapping $f(., \bar{a})$, and the case we are concerned with, where $\bar{a}$ can be in the set of bifurcations of $f$. Hence, if in the regions of $D$ previously considered, no structural changes occurred in $f$ and consequently in $I$, such changes now take place for $\bar{a} = \bar{a}_0$. The notion of structural change, borrowed from catastrophe theory, is of importance in the link it makes between the mathematical event of $\bar{a}$ taking the value $\bar{a}_0$, and the visible change (i.e. the catastrophe) in $I$ as a model of a particular physical phenomenon. The second aspect is, in a sense, much more practical, as the main problem in obtaining the coefficients of the expansion, is the inversion of the appropriate Levinson’s transformation defined by the unfolding of the catastrophe at $\bar{a}_0$.

5.1 The results of Chester, Friedman and Ursell

The purpose of the pioneering paper of Chester et al. (1957) was to furnish a uniform asymptotic expansion of integrals of the form $I$, as two saddle points of order $1$, $z_1(\alpha)$ and $z_2(\alpha)$, of $f$, coalesced at $z = z_0$ for $\alpha = 0$. The concern was to find an alternative means of obtaining asymptotic expansions where the ordinary method of steepest descents failed. The main idea was to introduce the new variable of integration $u$ by using a canonical form representation for such a function $f$.

Hence the mapping $u : z \mapsto u$ was defined by the relation:

$$f(z, \alpha) = \frac{u^3}{3} - \zeta(\alpha) u + A(\alpha)$$  \hspace{1cm} (5.1)

where the parameters $\zeta(\alpha)$ and $A(\alpha)$ were chosen such that the transformation was uniformly regular and bijective for $(z, \alpha)$ in a neighborhood of $(z_0, 0)$. As mentioned in Section 1.2.1 on Levinson’s transformation, the choice of a cubic mapping is motivated
by the fact that a cubic polynomial is the simplest form of an analytic function presenting two saddle points. The necessary conditions on $\zeta$ and $A$ for the mapping to be uniformly regular and bijective, were obtained from the differentiation of (5.1) with respect to $u$ once. Thus the equation:

\[
\frac{\partial f}{\partial z}(z, \alpha) \frac{\partial z}{\partial u}(u, \alpha) = u^2 - \zeta(\alpha)
\]  

(5.2)

taken at $z_1$ and $z_2$, imposes that the points $z_1$ and $z_2$, on one hand, must correspond to the points $-\zeta^{1/2}$ and $+\zeta^{1/2}$ on the other hand. In choosing $u(z_1, \alpha) = +\zeta^{1/2}$, and so $u(z_2, \alpha) = -\zeta^{1/2}$, the definition equation (5.1) determines $\zeta$ and $A$ as:

\[
A(\alpha) = \frac{1}{2} (f(z_1, \alpha) + f(z_2, \alpha))
\]

(5.3)

\[
\zeta^{3/2}(\alpha) = \frac{3}{4} (f(z_2, \alpha) - f(z_1, \alpha))
\]

(5.4)

(where the explicit dependence of $z_1$ and $z_2$ on $\alpha$ has been dropped for convenience).

It was then showed that there exists one branch of this transformation for which $u$ is uniformly regular and bijective in a neighborhood of $(z_0, 0)$ in $W \times D$. This allows the expansion of the new function $G(u, \alpha) = g(z, \alpha) \partial z/\partial u$ in an entire series at $u = 0$. However the expansion thus obtained by formally replacing $G$ by its series in the integral, cannot be considered as an asymptotic series in $N$.

Another form of expansion was introduced, namely:

\[
G(u, \alpha) = \sum_{r=0}^{\infty} p_r(\alpha) (u^2 - \zeta)^r + \sum_{r=0}^{\infty} q_r(\alpha) u(u^2 - \zeta)^r
\]  

(5.4)
which gave, upon its formal substitution in \( I \), the expansion:

\[
I(N, \alpha) \sim \exp[N A(\alpha)] \left[ \sum_{r=0}^{\infty} p_r \int_{\Gamma} (u^2 - \zeta)^r \exp[N(u^3/3 - \zeta u)] \, du \right. \\
\left. + \sum_{r=0}^{\infty} q_r \int_{\Gamma} u(u^2 - \zeta)^r \exp[N(u^3/3 - \zeta u)] \, du \right] 
\]

The canonical integrals involved were shown to be expressible in terms of an Airy function and its derivative. Hence the asymptotic expansion was found to be of the form – Chester et al. (1957, eq. (2.5)):

\[
I(N, \alpha) \sim \exp[N A(\alpha)] \left[ \frac{\text{Ai}(N^{2/3} \zeta)}{N^{1/3}} \left( \sum_{s=0}^{\infty} \frac{A_s(\alpha)}{N^{2s}} + \frac{1}{N} \sum_{s=0}^{\infty} \frac{D_s(\alpha)}{N^{2s}} \right) \right. \\
\left. + \frac{\text{Ai}''(N^{2/3} \zeta)}{N^{2/3}} \left( \sum_{s=0}^{\infty} \frac{C_s(\alpha)}{N^{2s}} + \frac{1}{N} \sum_{s=0}^{\infty} \frac{B_s(\alpha)}{N^{2s}} \right) \right] 
\]

if \( \Gamma \) was joining \( \infty e^{-i\pi/3} \) to \( \infty e^{i\pi/3} \). The coefficients \( A_s, B_s, C_s, D_s \) were obtained from the sequences \( (p_r) \) and \( (q_r) \) through certain recursion relations.

We must mention that:

— the results concerning the mapping \( u \) were obtained before Levinson’s theorem,

— the expansion proposed by Bleistein for more general cases – or Bleistein sequence (1.22) – differs from (5.5) except for the first terms \( P_0^{(0)}, P_0^{(1)} \) on one hand and \( p_0, q_0 \) on the other.

We note that the process described here, in order to obtain the final form of the expansion, is more complex than its equivalents for the cases previously treated. Consequently, we expect the corresponding algorithm to be more complicated than those presented in Chapter 4. Once \( p_r \) and \( q_r \) are known, the coefficients in the expansion (5.6) are easily obtained using the recursion formulas. The goal is then to
obtain numerically \( p \) and \( q \), uniformly in \( \alpha \). Two approaches are again available for which both types of expansions – the expansion of Chester et al., and the expansion of Bleistein – are more or less appropriate.

### 5.2 The contour integral method

Both expansions can be computed using contour integrals. However by the form of its definition the Bleistein sequence is readily obtained from contour integral formulas. This procedure makes use of the analyticity of the mapping \( u \) as defined by (5.1) (with \( \bar{\alpha} \) in place of \( \alpha \)) in a neighborhood of \( z_0 \). This property is ensured by Levinson's theorem (see Section 1.2.2 a)). We know then, from equation (1.27), that we have, for our special case \( m = 2 \):

\[
P_r^{(0)}(\bar{\alpha}) + P_r^{(1)}(\bar{\alpha}) = \frac{1}{2i\pi} \oint \frac{t + u}{t^2 - \zeta} G_r(t, \bar{\alpha}) \, dt \quad \text{for} \quad r \geq 0
\]  

(5.7)

where:

\[
G_r(u, \bar{\alpha}) = -\frac{\partial}{\partial u} \left( \frac{1}{2i\pi} \oint \frac{G_{r-1}(t, \bar{\alpha})}{(t-u)(t^2-\zeta)} \, dt \right) 
\]

(5.8)

for \( r \geq 1 \).

Here the expression of \( p_3(u, \bar{\zeta}) \) is simple enough so that \( P_r^{(0)} \) and \( P_r^{(1)} \) are obtained explicitly as contour integrals in terms of \( G_r \):

\[
P_r^{(0)}(\bar{\alpha}) = \frac{1}{2i\pi} \oint \frac{t}{t^2 - \zeta} G_r(t, \bar{\alpha}) \, dt
\]

(5.9)

\[
P_r^{(1)}(\bar{\alpha}) = \frac{1}{2i\pi} \oint \frac{1}{t^2 - \zeta} G_r(t, \bar{\alpha}) \, dt
\]

All the contours have to enclose the points \( u , -\zeta^{1/2} \) and \( +\zeta^{1/2} \). In fact as \( u \) is just
a variable in (5.7), the only constraint comes from $-\zeta^{1/2}$ and $+\zeta^{1/2}$. This necessitates the knowledge of $G_r(t, \bar{\alpha})$ on these contours, knowledge which is only accessible through the contour integral formula of (5.8). Hence the integral in (5.8) will have to be evaluated for $u$ describing a given contour, therefore $t$ describing contours around $u$ (and $-\zeta^{1/2}, +\zeta^{1/2}$).

Starting at $r = 0$, i.e. with $G_0(u, \bar{\alpha}) = g(z, \bar{\alpha}) \partial g(u, \bar{\alpha}) / \partial u$, this method requires the knowledge of the successive functions $G_r(., \bar{\alpha})$ on whole contours, or at least on finite sets of points, of the $u$-plane and/or the $z$-plane. Similarly to Sections 4.1.1 and 4.2.1, we conclude that the problem of defining such sets of points to be uniformly valid for $\bar{\alpha}$ in a neighborhood of $\bar{\alpha}_0$, is an obstacle to the practical application of this method. We only presented formulas for the Bleistein sequence, but contour integral formulas can also be found for the sequences $(p_r)$ and $(q_r)$ which will suffer from the same drawbacks in practical applications.

5.3 The inversion method

This approach resembles the point inversion methods of Chapter 4 from which we derived the algorithms used to treat the case of an end-point or an isolated col. The idea of performing a point inversion of the mapping $u$ is still valid, except for the fact that now it must be performed at both saddle points $z_1$ and $z_2$. However, the total procedure is more complex as it involves four main steps in order to obtain the numerical values of the coefficients $A_*, B_*, C_*$ and $D_*$ of the final expansion. The last and easiest step is to relate these coefficients to the sequences $(p_r)$ and $(q_r)$. Of course this step would have been part as well of the contour integral method if we could have used it.

We only consider here the expansion of Chester et al. since the procedure corresponding to the inversion method is much simpler than for the Bleistein sequence.
5.3.1 The computation of the coefficients $A_*, B_*, C_*, D_*$

The recursion formulas to obtain these coefficients from the sequences $(p_\tau)$ and $(q_\tau)$ are given in Chester et al. (1957, eqs (4.1)-(4.4)). We here recall the various relations as they will be used in the algorithm. We first introduce the notations:

\[
\mathcal{F}_r(\zeta, N, \Gamma) = \frac{1}{2i\pi} \int_{\Gamma} (u^2 - \zeta)^r \exp[N\left(\frac{u^3}{3} - \zeta u\right)] du
\]

\[
\mathcal{G}_r(\zeta, N, \Gamma) = \frac{1}{2i\pi} \int_{\Gamma} u(u^2 - \zeta)^r \exp[N\left(\frac{u^3}{3} - \zeta u\right)] du
\]

where $\Gamma$ is one of the contours: $C_1$ from $\infty e^{-i\pi/3}$ to $\infty e^{i\pi/3}$, $C_2$ from $\infty e^{i\pi/3}$ to $\infty e^{i\pi}$, $C_3$ from $\infty e^{i\pi}$ to $\infty e^{-i\pi/3}$. These integrals are just variants of the canonical integrals $U^{(0)}_2$ and $U^{(1)}_2$ of Section 1.2.3. Hence we have the relations:

\[
\begin{cases}
\mathcal{F}_r(\zeta, N, \Gamma) = N^{-\frac{3}{2}r-\frac{1}{2}} \mathcal{F}_r(N^{2/3}\zeta, 1, \Gamma) \\
\mathcal{G}_r(\zeta, N, \Gamma) = N^{-\frac{3}{2}r-\frac{1}{2}} \mathcal{G}_r(N^{2/3}\zeta, 1, \Gamma)
\end{cases}
\]

(5.10)

The relations between the choices of $C_1$ and $C_2$ for the contour $\Gamma$ are:

\[
\begin{cases}
\mathcal{F}_r(\zeta, N, C_2) = \exp\left[\frac{4}{3}ir\pi + \frac{2}{3}i\pi\right] \mathcal{F}_r(N^{2/3}\zeta, 1, C_1) \\
\mathcal{G}_r(\zeta, N, C_2) = \exp\left[\frac{4}{3}ir\pi + \frac{4}{3}i\pi\right] \mathcal{G}_r(N^{2/3}\zeta, 1, C_1)
\end{cases}
\]

(5.11)

whereas the relations for $C_3$ are obtained by changing the sign of $i$ in the above equations. With these results all the functions $\mathcal{F}_r(\zeta, N, \Gamma)$ and $\mathcal{G}_r(\zeta, N, \Gamma)$ can be computed from the following recursion relations:

\[
\mathcal{F}_r(\zeta, N, \Gamma) = \frac{2}{N^2} (r - 1) \left[ (2r - 5) \mathcal{F}_{r-3}(\zeta, N, \Gamma) + 2(r - 3) \zeta \mathcal{F}_{r-4}(\zeta, N, \Gamma) \right]
\]

\[
\mathcal{G}_r(\zeta, N, \Gamma) = \frac{2}{N^2} \left[ (2r - 1) (r - 2) \mathcal{G}_{r-3}(\zeta, N, \Gamma) + 2(r - 1) (r - 3) \mathcal{G}_{r-4}(\zeta, N, \Gamma) \right]
\]

(5.12)
along with the first four functions of each kind for $C_1$, given by:

$$
\begin{align*}
\mathcal{F}_0(\zeta, N, C_1) &= N^{-1/3} \text{Ai}(N^{2/3} \zeta) \\
\mathcal{G}_0(\zeta, N, C_1) &= -N^{-2/3} \text{Ai}'(N^{2/3} \zeta) \\
\mathcal{F}_1(\zeta, N, C_1) &= 0 \\
\mathcal{G}_1(\zeta, N, C_1) &= -N^{-4/3} \text{Ai}(N^{2/3} \zeta) \\
\mathcal{F}_2(\zeta, N, C_1) &= 2 N^{-5/3} \text{Ai}'(N^{2/3} \zeta) \\
\mathcal{G}_2(\zeta, N, C_1) &= -2 \zeta N^{-4/3} \text{Ai}(N^{2/3} \zeta) \\
\mathcal{F}_3(\zeta, N, C_1) &= 4 N^{-7/3} \text{Ai}(N^{2/3} \zeta) \\
\mathcal{G}_3(\zeta, N, C_1) &= -10 N^{-8/3} \text{Ai}'(N^{2/3} \zeta)
\end{align*}
$$

(5.13)

As noted by Chester et al., the recursion relations are of depth four. Hence the sequences $(\mathcal{F}_r(\zeta, N, \Gamma))$ and $(\mathcal{G}_r(\zeta, N, \Gamma))$ tend to decrease only in sets of four, and it was remarked that the resulting series are not optimal. In fact it is the Bleistein sequence which is optimal. It is of interest to remark that the asymptotic nature of the final expansion (5.6) comes in fact from the asymptotic properties of each of the four series $\sum A_s/N^{2s}$, $\sum B_s/N^{2s}$, $\sum C_s/N^{2s}$, $\sum D_s/N^{2s}$. This implies that any truncation test will have to be applied independently on each of these series.

The choice of $\Gamma$ is given by the global information (with respect to the sense of integration) of the deformation of the initial path $C$, and the necessary local correspondence between the part of $\mathbb{R}$ (the set of ramification values of $f$) in the neighborhood of $\bar{\alpha}_0$, and the set formed by the three Stokes lines $\arg \zeta = 0, 2\pi/3, 4\pi/3$ in the neighborhood of $\zeta = 0$. This question involves the determination of the subset of relevant saddle points in the pair $\{z_1, z_2\}$. Once this choice is fixed, the coefficients $A_s, B_s, C_s, D_s$ are computed directly from the decompositions of the products $p_r \mathcal{F}_r(\zeta, N, \Gamma)$ and $q_r \mathcal{G}_r(\zeta, N, \Gamma)$ onto the generic functions $N^{-1/3} \text{Ai}(N^{2/3} \zeta)$, $N^{-2/3} \text{Ai}'(N^{2/3} \zeta)$, $N^{-4/3} \text{Ai}(N^{2/3} \zeta)$ and $N^{-5/3} \text{Ai}'(N^{2/3} \zeta)$. From the recursion relations it appears that for a given $s$, a finite number of $p_r$ and $q_r$ are necessary in order to compute these coefficients. The dependence of $A_s, B_s, C_s, D_s$ on $\bar{\alpha}$ comes from the presence of $\zeta(\bar{\alpha})$ in the recursion relations, and from the dependence of the terms.
$p_r$ and $q_r$ on $\bar{\alpha}$. Hence the final result will be a function of $\bar{\alpha}$ not only through the arguments of the Airy function and its derivative, but also through the coefficients of the series.

The problem of obtaining the expansion (5.6) is now reduced to the computation of a finite number of terms of the sequences $(p_r)$ and $(q_r)$.

### 5.3.2 The computation of the sequences $(p_r)$ and $(q_r)$

The sequences $(p_r)$ and $(q_r)$ introduced by Chester et al., are defined by the relation:

$$G(u, \bar{\alpha}) = \sum_{r=0}^{\infty} p_r(\bar{\alpha}) (u^2 - \zeta)^r + \sum_{r=0}^{\infty} q_r(\bar{\alpha}) u(u^2 - \zeta)^r \quad (5.14)$$

Their existence is derived from the analyticity of $G(., \bar{\alpha})$ in a neighborhood, $U_1$, of $u = 0$, $u = -\zeta^{1/2}$ and $u = +\zeta^{1/2}$, and this uniformly for $\bar{\alpha}$ in a neighborhood, $D_1$, of $\bar{\alpha}_0$ (cf Levinson's theorem Section 1.2.1). This detail is of importance since we are going to use this property of $U_1$ in two different ways in the computation of the $p_r$'s and $q_r$'s.

#### 1. The "direct" method

The method suggested by Chester, Friedman and Ursell consists in repeatedly differentiating equation (5.14) with respect to $u$ and making use of the correspondences $z_1 \leftrightarrow \zeta^{1/2}$, $z_2 \leftrightarrow -\zeta^{1/2}$. Assuming that the $p_r$'s and $q_r$'s are needed for $r$ up to a certain order $R$, and using the notations:

$$G_1^{(n)} = \frac{\partial^n G}{\partial u^n}(u, \bar{\alpha}) \bigg|_{u = \zeta^{1/2}}$$

$$G_2^{(n)} = \frac{\partial^n G}{\partial u^n}(u, \bar{\alpha}) \bigg|_{u = -\zeta^{1/2}} \quad (5.15)$$
we obtain a set of \((2R+2)\) equations, linear in the \((2R+2)\) unknowns \(p_r, q_r, 0 \leq r \leq R\), namely:

\[
\begin{align*}
G_{1}^{(0)} &= q_0 \zeta^{\frac{1}{2}} + p_0 \\
G_{2}^{(0)} &= -q_0 \zeta^{\frac{1}{2}} + p_0 \\
G_{1}^{(1)} &= 2q_1 \zeta + 2p_1 \zeta^{\frac{1}{2}} + q_0 \\
G_{2}^{(1)} &= 2q_1 \zeta - 2p_1 \zeta^{\frac{1}{2}} + q_0 \\
G_{1}^{(2)} &= 8q_2 \zeta^{\frac{3}{2}} + 8p_2 \zeta + 6q_1 \zeta^{\frac{1}{2}} + 2p_1 \\
G_{2}^{(2)} &= -8q_2 \zeta^{\frac{3}{2}} + 8p_2 \zeta - 6q_1 \zeta^{\frac{1}{2}} + 2p_1 \\
&\vdots
\end{align*}
\] (5.16)

Since this set is triangular, explicit solutions can be obtained which express \(p_r - q_r\) in terms of \(G_1^{(r)}, G_2^{(r)}\), previously computed values of \(p_n\) respectively \(q_n\) for \(n \leq r - 1\), and powers of \(\zeta^{1/2}\) up to \(r\). Thus the first few expressions of the \(p_r\) and \(q_r\)
are:

\[
\begin{align*}
p_0 &= \frac{G_2^{(0)} + G_1^{(0)}}{2} \\
q_0 &= \frac{G_1^{(0)} - G_2^{(0)}}{2 \zeta^{1/2}} \\
p_1 &= \frac{G_1^{(1)} - G_2^{(1)}}{4 \zeta^{1/2}} \\
q_1 &= \frac{G_1^{(1)} + G_2^{(1)} - 2q_0}{4 \zeta} \\
p_2 &= \frac{G_1^{(2)} + G_2^{(2)} - 4p_1}{16 \zeta} \\
q_2 &= \frac{G_1^{(2)} - G_2^{(2)} - 12q_1 \zeta^{1/2}}{16 \zeta^{3/2}} \\
&\vdots
\end{align*}
\]

(5.17)

The forms of these expressions present some generic features which can be deduced from the principle used to generate them. From the fact that equation (5.14) is repeatedly differentiated in \(u\) and taken at \(u = \frac{1}{2}\) and \(u = -\frac{1}{2}\), it can be shown that:

1. The denominator in the expression of \(p_r\) is proportional to \(\zeta^{r/2}\) while the denominator for \(q_r\) involves \((\zeta^{(r+1)/2})\).

2. The quantities \(G_1^{(n)}\) and \(G_2^{(n)}\) only appear under the form of their sum or difference. More precisely:

   - \(p_r\) with \(r\) even contains the sum \((G_1^{(r)} + G_2^{(r)})\),
   - \(p_r\) with \(r\) odd, the difference \((G_1^{(r)} - G_2^{(r)})\),
   - \(q_r\) with \(r\) even, the difference \((G_1^{(r)} - G_2^{(r)})\),

157
— and \( q_r \) with \( r \) odd, the sum \( (G_1^{(r)} + G_2^{(r)}) \).

3. The numerators involve only certain powers of \( \zeta^{1/2} \):
   
   — For \( r \) even, \( p_r \) contains only even powers of \( \zeta^{1/2} \), and for \( r \) odd, odd powers.
   
   — For \( r \) even, \( q_r \) contains only odd powers of \( \zeta^{1/2} \), and for \( r \) odd, even powers.

Such remarks are useful when trying to improve the efficiency of the corresponding algorithm.

Of course, these expressions are valid as long as \( \zeta^{1/2} \neq 0 \), i.e. as long as the cols are not coalescent. When \( \zeta^{1/2} = 0 \) the sum of the series in (5.14) is identical to the entire series of \( G(., \tilde{\alpha}) \) at \( u = 0 \). There, we have the simple identification of the coefficients: \( p_r = G_1^{(2r)}/(2r)! \) and \( q_r = G_2^{(2r+1)}/(2r + 1)! \). (Note that the two entire series at \( u = \zeta^{1/2} \) and \( u = -\zeta^{1/2} \) merge.) Undoubtedly numerical troubles will arise as \( \zeta^{1/2} \) gets smaller, and another method must be found for values of \( \zeta^{1/2} \) in a disk around 0. These troubles are only numerical as the functions \( p_r \) and \( q_r \) are holomorphic in \( \tilde{\alpha} \) (or of certain other parameters which will be defined subsequently) in this same neighborhood of \( \tilde{\alpha}_0, D_1 \).

2. The power series method

We now propose a complementary procedure to compute the coefficients \( p_r \) and \( q_r \), to be applied for \( \zeta^{1/2} \) in a disk \( D_\sigma(0, \sigma) \).

Consider (5.14). The equality must hold for \( u = -\zeta^{1/2} \) and \( u = \zeta^{1/2} \), and this over two disks centered at \( -\zeta^{1/2} \) and \( \zeta^{1/2} \). Hence we have the two equations:

\[
\sum_{k=0}^{\infty} \frac{G_1^{(k)}}{k!} (u - \zeta^{1/2})^k = \sum_{r=0}^{\infty} p_r(\tilde{\alpha}) (u^2 - \zeta)^r + \sum_{r=0}^{\infty} q_r(\tilde{\alpha}) u(u^2 - \zeta)^r \quad (5.18)
\]
and:

\[
\sum_{k=0}^{\infty} \frac{G_2^{(k)}}{k!} (u + \zeta^{1/2})^k = \sum_{r=0}^{\infty} p_r(\bar{\alpha})(u^2 - \zeta)^r + \sum_{r=0}^{\infty} q_r(\bar{\alpha}) u(u^2 - \zeta)^r
\]  

(5.19)

As each series is unique, we identify the corresponding terms on both sides of the equations. These identifications can be formalized as Euclidean divisions (identical to the procedure of Bleistein) of the polynomials \((u - \zeta^{1/2})^k\) or \((u + \zeta^{1/2})^k\) by the quartic polynomial \((u^2 - \zeta)\). The resulting expressions appear then as infinite power series of \(\zeta^{1/2}\) with coefficients taken from \(G_1^{(k)}\) and \(G_2^{(k)}\). Hence, we have the set of coupled equations:

\[
\begin{align*}
\begin{cases}
p_0 &= G_1^{(0)} - G_1^{(2)} \zeta^{1/2} + G_1^{(4)} \zeta - \frac{2}{3} G_1^{(3)} \zeta^{3/2} + \frac{1}{3} G_1^{(5)} \zeta^2 - \frac{2}{15} G_1^{(6)} \zeta^3 + \ldots \\
p_0 &= G_2^{(0)} + G_2^{(2)} \zeta^{1/2} + G_2^{(4)} \zeta + \frac{2}{3} G_2^{(3)} \zeta^{3/2} + \frac{1}{3} G_2^{(5)} \zeta^2 + \frac{2}{15} G_2^{(6)} \zeta^3 + \ldots \\
q_0 &= G_1^{(1)} - G_1^{(2)} \zeta^{1/2} + \frac{2}{3} G_1^{(3)} \zeta - \frac{2}{3} G_1^{(4)} \zeta^{3/2} + \frac{2}{15} G_1^{(5)} \zeta^2 - \frac{2}{45} G_1^{(6)} \zeta^3 + \ldots \\
q_0 &= G_2^{(1)} + G_2^{(2)} \zeta^{1/2} + \frac{2}{3} G_2^{(3)} \zeta + \frac{2}{3} G_2^{(4)} \zeta^{3/2} + \frac{2}{15} G_2^{(5)} \zeta^2 + \frac{2}{45} G_2^{(6)} \zeta^3 + \ldots \\
p_1 &= \frac{G_1^{(2)}}{2} - \frac{G_1^{(3)}}{2} \zeta^{1/2} + \frac{G_1^{(4)}}{3} \zeta - \frac{G_1^{(5)}}{6} \zeta^{3/2} + \frac{G_1^{(6)}}{15} \zeta^2 - \frac{G_1^{(7)}}{45} \zeta^3 + \ldots \\
p_1 &= \frac{G_2^{(2)}}{2} + \frac{G_2^{(3)}}{2} \zeta^{1/2} + \frac{G_2^{(4)}}{3} \zeta + \frac{G_2^{(5)}}{6} \zeta^{3/2} + \frac{G_2^{(6)}}{15} \zeta^2 + \frac{G_2^{(7)}}{45} \zeta^3 + \ldots \\
q_1 &= \frac{G_1^{(3)}}{6} - \frac{G_1^{(4)}}{6} \zeta^{1/2} + \frac{G_1^{(5)}}{10} \zeta - \frac{2}{45} G_1^{(6)} \zeta^{3/2} + \frac{G_1^{(7)}}{63} \zeta^2 - \frac{G_1^{(8)}}{210} \zeta^3 + \ldots \\
q_1 &= \frac{G_2^{(3)}}{6} + \frac{G_2^{(4)}}{6} \zeta^{1/2} + \frac{G_2^{(5)}}{10} \zeta + \frac{2}{45} G_2^{(6)} \zeta^{3/2} + \frac{G_2^{(7)}}{63} \zeta^2 + \frac{G_2^{(8)}}{210} \zeta^3 + \ldots \\
\vdots
\end{cases}
\end{align*}
\]

(5.20)
Each pair of equations is in fact twice the same equation: they correspond to the entire series of some derivative of $G(\cdot, \xi)$ at $\zeta^{1/2}$ and $-\zeta^{1/2}$, taken at $u = 0$. However, we cannot use directly such formulas since only a finite number of the terms $G_{1}^{(k)}$ and $G_{2}^{(k)}$ are computable.

The constraint of knowing only a finite number of the terms $G_{1}^{(k)}$ and $G_{2}^{(k)}$ implies the truncation of the series in the above formulas. Then the coupled equations are no longer identical, and we need to decide what combination should be used. It is easy to show that by considering half the sum of each pair of equations, the truncation errors are optimal. Hence, if the sequences $(G_{k})$ are truncated at $k = K$, we obtain the following finite sums:

\[
\begin{align*}
\phi_{0} &= \frac{G_{1}^{(0)} + G_{2}^{(0)}}{2} - \frac{G_{1}^{(1)} - G_{2}^{(1)}}{2} \zeta^{\frac{1}{2}} + \ldots + O(\zeta^{K_{\phi_{0}}}) \\
\psi_{0} &= \frac{G_{1}^{(1)} + G_{2}^{(1)}}{2} - \frac{G_{1}^{(2)} - G_{2}^{(2)}}{2} \zeta^{\frac{1}{2}} + \ldots + O(\zeta^{K_{\psi_{0}}}) \\
\phi_{1} &= \frac{G_{1}^{(2)} + G_{2}^{(2)}}{4} - \frac{G_{1}^{(3)} - G_{2}^{(3)}}{4} \zeta^{\frac{1}{2}} + \ldots + O(\zeta^{K_{\phi_{1}}}) \\
\psi_{1} &= \frac{G_{1}^{(3)} + G_{2}^{(3)}}{12} - \frac{G_{1}^{(4)} - G_{2}^{(4)}}{12} \zeta^{\frac{1}{2}} + \ldots + O(\zeta^{K_{\psi_{1}}}) \\
&\vdots
\end{align*}
\]

with $K_{\phi_{r}} = |E(-(K-2r)/2)|$ and $K_{\psi_{r}} = |E(-(K-2r-1)/2)|$. (Note that the bound on the index of truncation is a function of $r$.)

In fact these series are in powers of $\zeta$ rather than $\zeta^{1/2}$, and can be rewritten in
such a manner. Thus (5.21) is replaced by:

\[
\begin{align*}
p_0 &= \frac{G_1^{(0)} + G_2^{(0)}}{2} \\
q_0 &= \frac{G_1^{(1)} + G_2^{(1)}}{2} - \frac{G_1^{(3)} + G_2^{(3)}}{6} \zeta + \frac{G_1^{(5)} + G_2^{(5)}}{15} \zeta^2 + \ldots + O(\zeta^{K_w}) \\
p_1 &= \frac{G_1^{(2)} + G_2^{(2)}}{4} - \frac{G_1^{(4)} + G_2^{(4)}}{12} \zeta + \frac{G_1^{(6)} + G_2^{(6)}}{30} \zeta^2 + \ldots + O(\zeta^{K_{r_1}}) \\
q_1 &= \frac{G_1^{(3)} + G_2^{(3)}}{12} - \frac{G_1^{(5)} + G_2^{(5)}}{30} \zeta + \frac{17 (G_1^{(7)} + G_2^{(7)})}{1260} \zeta^2 + \ldots + O(\zeta^{K_{r_1}}) \\
&\vdots
\end{align*}
\]

(5.22)

A form much more appropriate to numerical evaluations, as the differences

\((G_2^{(k)} - G_1^{(k)})\) would have induced cancellation errors as \(|\zeta| \to 0|\).

For \(K\) given, how far in \(r\) should we use these formulas, is a matter of imposing

the maximum permissible error in the computations of the \(p_r\) and \(q_r\). By example,

we use the exact formula for \(p_0\). Assuming that we want a truncation error for \(p_r - q_r\) in \(O(\zeta^{K_{r_1}})\) respectively in \(O(\zeta^{K_{q_r}})\) – the computations of \(G_1^{(k)}\) have to be done

up to \(K = \max\{2r + 2K_{p_r}, 2r + 1 + 2K_{q_r}\}\). The finite sequences \((K_{p_r})\) and \((K_{q_r})\) need not be constant. This aspect, as well as the determination of the various radii \(\sigma_r\) of the disks \(D_\sigma(0, \sigma_r)\) within which these formulas should be used, will be discussed later (see Section 5.5.4).

We are now reduced to a similar problem as the ones treated in Chapter 4, that is the computation of a finite sequence of derivatives of \(G(., \vec{a})\) taken at the images of the cols, or \((G_1^{(k)})\) and \((G_2^{(k)})\).
5.3.3 The computation of \( (G_1^{(k)}) \) and \( (G_2^{(k)}) \)

Without any further modification, the procedure in Section 4.1.2 can be applied at each col. Unlike the case of an isolated col where only \( G_2^{(2k)} \) were needed, here all of the derivatives have to be computed at both cols. Therefore the formulas are given as in (4.10) with \( z_1 \) and \( z_2 \) in place of \( e \), thus giving two sets of equations for \( 0 \leq k \leq K \).

Thus we have broken down the task to the computation of two finite sequences of derivatives of the mapping \( z \) taken at \( u = \zeta^{1/2} \) and \( u = -\zeta^{1/2} \), or the sequences \( (z_1^{(k)}) \) and \( (z_2^{(k)}) \). This problem is identical to its analogues treated in Chapter 4, at least for a fixed value of the parameter \( \bar{\alpha} \). However this simplicity is spurious. The least we can say is that by looking for the knowledge of the mapping \( z \), the inverse of \( u \), for \( \bar{\alpha} \) in \( D_1 \), we are asking for quantitative information whereas the existing mathematical theory is entirely qualitative. Hence, more than ever before, we will face the gap between theory and numerical (or quantitative) applications. The difficulties encountered in the next section on the unfolding of the fold catastrophe, give good indications of the kind of problems that would need to be addressed if we were dealing with higher order catastrophes.

5.3.4 The unfolding of the fold catastrophe

The name of fold singularity was first used by Whitney in a 1955 article which laid the foundation of singularity theory. The term cusp singularity, for the singularity "immediately after" the fold, was also introduced. These were later renamed fold and cusp catastrophes when catastrophe theory emerged. Singularity theory deals principally with catastrophes of smooth \((k\text{-times differentiable})\) real mappings, although some results exist for complex mappings (see Arnold (1986) for an overview and references). Dealing with holomorphic (i.e. everywhere analytic), thus infinitely differentiable, functions simplifies significantly the theory and consequently numerical applications. Thus the results presented hereafter are specific to the cases of holomor-
phic functions of several variables in much the same way as the results of Chapter 3.

1. The fold map

The fold map is defined as the parametric mapping \( u \mapsto w \) given by the polynomial \( p_3 \) of Table 1.1, or:

\[
p_3(u, \xi_1) = \frac{u^3}{3} + \xi_1 u = w
\]

In writing the equality \( f(z, \bar{\alpha}) = p_3(u, -\zeta(\bar{\alpha})) + A(\bar{\alpha}) \), in other words the relation which defines the change of variables, we imply that locally, for \( \bar{\alpha} \) in a neighborhood of \((z_0, \bar{\alpha}_0)\), the function of the variable \((z, \bar{\alpha})\), \( f(z, \bar{\alpha}) - A(\bar{\alpha}) \), behaves like the function of the variable \((u, \bar{\alpha})\), \( p_3(u, -\zeta(\bar{\alpha})) \). The definition of the Levinson's transformation, that is the mapping \( u : z \mapsto u \), regular and bijective for \((z, \bar{\alpha})\) in \( W_1 \times D_1 \) and given by:

\[
f(z, \bar{\alpha}) = \frac{u^3}{3} - \zeta(\bar{\alpha})u + A(\bar{\alpha})
\]

is simply the formalization of this remark.

As before the exact knowledge of the inverse mapping \( z = u^{-1} \) is not as much needed as the values of its derivatives at \( u(z_1) \) and \( u(z_2) \). (The knowledge of a finite number of these values correspond in fact to a local approximation of this mapping.) In all of the examples about the coalescence of two saddle points, treated in various articles – Chester et al. (1957), Ursell (1960, 1965) –, either an approximation of \( z \), or the mapping itself are explicitly known. Such cases are possible only when the cols are explicitly known functions of \( \bar{\alpha} \). A constraint of this kind is too restrictive for our purpose, therefore once again we use the idea of a point inversion.
2. The point inversion of the Levinson's transformation

The procedure is globally the same as the ones used in Chapter 4: with $\zeta(\bar{\alpha})$ defined by the relation:

\[ \zeta^{3/2}(\bar{\alpha}) = \frac{3}{4}(f(z_2, \bar{\alpha}) - f(z_1, \bar{\alpha})) \quad (5.24) \]

we repeatedly differentiate (5.23) with respect to $u$ and thus obtain a system of function relations:

\[
\begin{align*}
\frac{\partial f}{\partial z} \frac{\partial z}{\partial u} &= u^2 - \zeta \\
\frac{\partial f}{\partial z} \frac{\partial^2 z}{\partial u^2} + \frac{\partial^2 f}{\partial z^2} \frac{\partial z}{\partial u} &= 2u \\
\frac{\partial f}{\partial z} \frac{\partial^3 z}{\partial u^3} + 3 \frac{\partial^2 f}{\partial z^2} \frac{\partial z}{\partial u^2} + \frac{\partial^3 f}{\partial z^3} \frac{\partial z}{\partial u} &= 2 \\
\vdots 
\end{align*}
\]

valid for $(z, \bar{\alpha})$ in $W_1 \times D_1$ (correspondingly $(u, \bar{\alpha})$ in $U_1 \times D_1$).

Upon using the correspondences $z_1 \leftrightarrow \zeta^{1/2}$, $z_2 \leftrightarrow -\zeta^{1/2}$, we obtain two sets of equations:

\[
\begin{align*}
&f_i^{(2)} z_i^{(1)^2} = (-1)^{i+1} 2 \zeta^{1/2} \\
&3 f_i^{(2)} z_i^{(1)} z_i^{(2)} + f_i^{(3)} z_i^{(1)^3} = 2 \\
&4 f_i^{(2)} z_i^{(1)} z_i^{(3)} + 3 f_i^{(2)} z_i^{(2)^2} + 6 f_i^{(3)} z_i^{(1)^2} z_i^{(2)} + f_i^{(4)} z_i^{(1)^4} = 0 \\
&\vdots 
\end{align*}
\]

(5.26)
for $i = 1, 2$ and with the notations:

$$ f_i^{(n)} = \frac{\partial f}{\partial z}(z, \bar{\alpha}) \bigg|_{z = z_i} $$

$$ z_i^{(n)} = \frac{\partial z}{\partial u}(u, \bar{\alpha}) \bigg|_{u = (-1)^{i+1} \zeta^{1/2}} $$

These sets are triangular and non-linear in the unknowns $z_i^{(n)}$. Each system is solved independently for $\zeta \neq 0$, and gives the following explicit expressions:

$$ z_i^{(1)} = 2^{1/2} \left( \frac{(-1)^{i+1} \zeta^{1/2}}{f_i^{(2)}} \right)^{1/2} $$

$$ z_i^{(2)} = \frac{2 - f_i^{(3)} z_i^{(1)^3}}{3 f_i^{(2)} z_i^{(1)}} $$

$$ z_i^{(3)} = -\frac{3 f_i^{(2)} z_i^{(2)^2} + 6 f_i^{(3)} z_i^{(1)^2} z_i^{(2)} + f_i^{(4)} z_i^{(1)^4}}{4 f_i^{(2)} z_i^{(1)}} $$

(5.27)

for $i = 1, 2$.

Apart from the first two equations, these formulas are identical to (4.27). The only difference lies in the fact that now $f_i^{(2)}$ is allowed to tend to 0. Naturally, such expressions are not valid for $\bar{\alpha} = \bar{\alpha}_0$ and the corresponding equations have to be solved.
Thus for $\zeta = 0, \, z_1 = z_2 = z_0$, we have the system, degenerated form of (5.26):

$$
\begin{align*}
&f_0^{(3)} z_0^{(1)^3} = 2 \\
&6 f_0^{(3)} z_0^{(1)^2} z_0^{(2)} + f_0^{(4)} z_0^{(1)^4} = 0 \\
&10 f_0^{(3)} z_0^{(1)^2} z_0^{(3)} + 15 f_0^{(3)} z_0^{(1)} z_0^{(2)^2} + 10 f_0^{(4)} z_0^{(1)^3} z_0^{(2)} + f_0^{(5)} z_0^{(1)^5} = 0
\end{align*}
$$

(5.28)

which once solved, gives the expressions:

$$
\begin{align*}
z_0^{(1)} &= \left( \frac{2}{f_0^{(3)}} \right)^{\frac{1}{3}} \\
z_0^{(2)} &= -\frac{f_0^{(4)} z_0^{(1)^2}}{6 f_0^{(3)}} \\
z_0^{(3)} &= -\frac{5 f_0^{(3)} z_0^{(2)^2} + 10 f_0^{(4)} z_0^{(1)^2} z_0^{(2)} + f_0^{(5)} z_0^{(1)^4}}{10 f_0^{(3)} z_0^{(1)}} \\
\vdots
\end{align*}
$$

(5.29)

Each of the formulas is obtained as the limit of its analogues in (5.27) by applying L'Hospital rule. However such expressions are of no numerical utility at all!

Numerically the systems (5.27) become ill-conditioned as $|\zeta|$ gets smaller. However the regularity of the functions of $\alpha, z_i^{(n)}$, indicates that since $f_i^{(2)}$ is $O(\zeta^{1/2})$ as $|\zeta^{1/2}| \to 0$, the numerators must be $O(\zeta^{1/2})$. Thus cancellations will occur in the numerators, which is acceptable, but followed by a division by a term of order $\zeta^{1/2}$ with a result of order 1, which is unacceptable. Of course the theory guarantees the regularity of these functions but does not give any other practical method to compute them.
An approach for removing this obstacle, consists in expanding each term in the numerators and the denominators in power series of \((z_2 - z_1)\). Then \(\zeta^{1/2}\) is shown to be \(O((z_2 - z_1))\), and similarly for \(f_i^{(2)}\). The necessary conditions for the derivatives \(z_i^{(n)}, n \geq 2\), to exist is then that the constant term in each denominator, obtained after expanding and summing the various products, be zero. The whole process is rather complex but can be carried out using symbolic computation. The details are omitted but the coefficients would be functions of \(\alpha\). Thus the resulting formulas would have to be evaluated at each new value of \(\alpha\) rendering the computational effort particularly heavy.

This last point requires some explanations: however close the local behaviors of the functions \(p_3(u, \alpha) + A(\alpha)\) and \(f(z, \alpha)\) are, \(f\) cannot be reduced, in general, to the behavior of a power series in \((z_2 - z_1)\) or \(\zeta^{1/2}\). However a particular choice for the parameter \(\alpha\) under the constraint of taking \(\alpha\) in \(\mathbb{C}\), leads to the results of Chester et al. (1957) where all the quantities of interest are regular series of \(\alpha^{1/2}\). (This was actually used in the original proof of the regularity and bijective property of the mapping \(u\)) A generalization of this result to the case of \(\alpha\) in \(\mathbb{C}^l\) is therefore necessary. Indeed each \(z_i^{(n)}\) is a function of \(\alpha\) and of \(\alpha\) only. Therefore, a natural idea is to look for multiple entire series of these quantities in terms of \(\alpha\).

3. An entire series method

We propose here a complementary method based on the existence of some multiple entire series for \(z_i^{(n)}\). In fact, what matters for our problem is not "simply" the existence of some entire series for \(z_i^{(n)}\), but rather their forms and some practical procedure to obtain them.
Thus, we have the following results (see Appendix B):

Upon a translation on the variable $z$, there exists a choice of certain parameters $\vec{\beta} = (\beta_1, \ldots, \beta_i)$ such that:

$$
\zeta^{1/2} \propto \beta_1 \tau_\zeta(\beta_1^2, \beta_2, \ldots, \beta_i) \tag{5.30}
$$

and for $n \geq 1$:

$$
z_1^{(n)} \propto \tau_n(\beta_1, \beta_2, \ldots, \beta_i) \tag{5.31}
$$
$$
z_2^{(n)} \propto \tau_n(-\beta_1, \beta_2, \ldots, \beta_i)
$$

where the symbol $\tau(\beta_1, \beta_2, \ldots, \beta_i)$ denotes a multiple entire series in $\beta_1, \ldots, \beta_i$ of the form:

$$
1 + \sum_{|J|=1}^{\infty} a(J) \beta_1^{i_1} \ldots \beta_i^{i_i} \tag{5.32}
$$

An eventual procedure to obtain the various coefficients of the corresponding truncated series, consists in following the steps of the proof of (5.31) as given in Appendix B. This requires us to be able to expand $f_i^{(n)}$ in multiple series of $\vec{\beta}$ and necessitates a new local parameterization of $f$.

The advantage of this method lies in the fact that the coefficients of all the series (5.31) are now depending on $\vec{\alpha}_0$ only, and thus can be determined once and for all. Therefore for a given $\vec{\alpha}_0$, the computational task is reduced to the computation of finite multiple series knowing the values of their coefficients. The orders at which the series are truncated as well as the various radii $\tau_n$ defining the disks $|\zeta^{1/2}| \leq \tau_n$ within which the series $\tau_n$ are used to compute $z_i^{(n)}$, have to be determined in accordance with the desired accuracy. As these values depend on the function $f$ considered, this task is performed on a case by case basis.
5.3.5 The schematic procedure

From the previous analysis, we can sketch the overall procedure for computing the coefficients $A_s, B_s, C_s, D_s$.

Given the index $s$ and the type of series considered – i.e. $A, B, C$ or $D$ – it is possible to determine the indices of the various quantities to be computed: $r_p$ of $p_r$, $r_q$ of $q_r$, and consequently $k_{r_p,r_q}$ of $G_i^{(k)}$ and $n_{r_p,r_q}$ of $z_i^{(n)}$.

**The procedure**

For each new value of $\tilde{\alpha}$, the following steps have to be performed on the desired ranges of $s$:

1. the computation of $z_i^{(n)}$, for $1 \leq n \leq n_{r_p,r_q}$, using the equations of (5.27) or the power series of (5.31), depending on the value of $|\zeta^{1/2}|$ with respect to $\tau_{n_{r_p,r_q}}$,

2. the computation of $G_i^{(k)}$, for $0 \leq k \leq k_{r_p,r_q}$, using the adapted formulas of (4.23),

3. the computation of $p_r$, for $0 \leq r \leq r_p$, and $q_r$, for $0 \leq r \leq r_q$, using the expressions of (5.17) or the power series of (5.22) depending again on the value of $|\zeta^{1/2}|$ with respect to $\sigma_{r_p}$ and $\sigma_{r_q}$,

4. finally the computation of $A_s, B_s, C_s, or D_s$ using the recursion (5.12), (5.13).

(where $i$ is taking the values $1, 2$).

The final expansion is then obtained by computing the four partial sums $\sum A_s/N^{2s}$, $\sum B_s/N^{2s}$, $\sum C_s/N^{2s}$ and $\sum D_s/N^{2s}$, and multiplying them by the generic functions based on the Airy function or its derivative.

This procedure even in this schematic form, is obviously more complicated than the ones described in Chapter 4. Since for the computation of $z_i^{(n)}$ or the $G_i^{(k)}$ two
complementary procedures are available, we have four possible combinations of algorithms. The total accuracy depends on this choice in an intricate manner. More precisely, the dependence of the errors, at a given step, on errors inherited from previous calculations, is much stronger than in the procedures of Chapter 4. This is due to the use of expressions which are nearly singular. By example, the misuse of the "direct" procedure given by (5.27) in computing $z^{(n)}$ is detrimental to the computation of $p_r$ and $q_r$, and this independently from the choice of the procedure at that stage. A careful determination of the radii $\tau_{r_r}, r_q$ and $\sigma_{r_r}, \sigma_{r_q}$ is therefore crucial for the robustness of the results.

However, the difficulties of the uniform treatment of the coalescence of two cols are not reduced to the above. In our analysis we have overlooked two essential "details" which, if not correctly treated, would irremediably result in the failure of the method.

5.4 Some inevitable choices

In fact, the term of "details" is rather inappropriate and we shall use "choices" instead. We have, voluntarily, passed over two steps in Section 5.3, which correspond effectively to choices of root branches in the complex plane. They are essential as they are the very definition of the Levinson's transformation for $f$.

Choosing the branch of a root in the complex plane always raises problems. In general, theory is satisfied with the existence of such a choice: "one can choose". However, in practice a choice has to be made, and difficulties arise there. In our case, we know from Chester et al. (1957), that the Levinson's transformation is uniquely defined. Therefore the correct choice has to be made, for which no criterion exists apart from the existence of the transformation. Moreover, our applications impose that these choices be uniform in $\tilde{a}$, thus bringing additional difficulties in an implementation. Finally, the entanglement of the numerical consequences of these
choices with the numerical troubles described earlier, worsens the situation.

These choices are the essence of the uniform treatment of the coalescence of two saddle points. In any application, whether analytical or numerical, they would have to be faced. Of all the steps in the method of steepest descents, they are among the most arduous ones to systematize. Similar, in a way, to the problem of determining the relevant cols, the difficulty lies in the quantitative representations of qualitative features.

5.4.1 The definition of $\zeta^{1/2}$

The first of these choices is encountered in the definition of $\zeta^{1/2}$. As we have seen in the various formulas of Section 5.3, $\zeta^{1/2}$ plays a determining role. However, it is only defined through its cubic power through equation (5.24):

$$\zeta^{3/2}(\tilde{\alpha}) = \frac{3}{4}(f(z_2, \tilde{\alpha}) - f(z_1, \tilde{\alpha}))$$

This choice of $\zeta^{1/2}$ is in fact related to the choice of the path $\Gamma$ for the functions $\mathcal{F}_r(\zeta, N, \Gamma)$ and $\mathcal{G}_r(\zeta, N, \Gamma)$, and we now detail the process sketched in Section 5.3.1.

As $\tilde{\alpha}$ describes the neighborhood of $\tilde{\alpha}_0$, $W_1$, the subset of relevant cols in $\{z_1, z_2\}$ changes each time $\tilde{\alpha}$ crosses a Stokes surface (i.e. a component of $R$). In fact it is easier to consider the new parameter $\tilde{\beta}$. This local parameterization of $f$ is such that as $\beta_1$ describes a circle around 0, the set of relevant cols in $\{z_1, z_2\}$ is modified 3 times, say in the sequence: $\{z_1\}$, $\{z_1, z_2\}$, $\{z_2\}$, $\{z_1\}$. In the $u$-plane, this must correspond to the exact same pattern with the cols $\zeta^{1/2}$ and $-\zeta^{1/2}$ of $p_3$. Indeed, as $\beta_1$ describes a circle, $\zeta^{1/2}$ describes a contour around 0 (see (B.10)). Therefore, we should have the corresponding sequence: $\{\zeta^{1/2}\}$, $\{\zeta^{1/2}, -\zeta^{1/2}\}$, $\{-\zeta^{1/2}\}$, $\{\zeta^{1/2}\}$.  

171
In both spaces, these sequences depend on the paths of integrations: respectively $C$ for $I$, and $\Gamma$ for $\mathcal{F}_r(\zeta, N, \Gamma)$ and $\mathcal{G}_r(\zeta, N, \Gamma)$. If $C$, and thus the sequence in the $z$-plane, are given, we have the latitude of choosing $\Gamma$ and $\zeta^{1/2}$. Given the correspondences $z_1 \leftrightarrow \zeta^{1/2}$, $z_2 \leftrightarrow -\zeta^{1/2}$, for a non-critical value $\beta_{1,\text{ref}}$ of $\beta_1$, we have three possible choices for $\arg \zeta^{1/2}$, corresponding to $\zeta^{1/2}$ being in one of the sectors: $[0, 2\pi/3[$, $[2\pi/3, 4\pi/3[$, $[4\pi/3, 2\pi[$. (Note that we also have three choices for $\Gamma$: $C_1$, $C_2$ or $C_3$.)

Let us choose a value of $\arg \zeta^{1/2}$, and denote it $\arg \zeta^{1/2}_{\text{ref}}$. Then, there exists a unique choice for $\Gamma$, say $\Gamma_{\text{ref}}$ for which, as $\arg \beta_1$ describes $[\arg \beta_{1,\text{ref}}, \arg \beta_{1,\text{ref}} + 2\pi]$, we recover the sequence $\{\zeta^{1/2}\}, \{-\zeta^{1/2}\}, \{\zeta^{1/2}\}$ in the $u$-plane.

The pair $\left(\beta_{1,\text{ref}}, \arg \zeta^{1/2}_{\text{ref}}\right)$ constitutes then a reference value for all other computations for $\bar{\alpha}$ in $D_1$.

### 5.4.2 The choice of the correct branch

The second choice arises in the local inversion of the Levinson's transformation in the form of the first equation of (5.27), namely:

$$z^{(1)}_i = 2^{1/2} \left( \frac{(-1)^{i+1} \zeta^{1/2}}{f^{(2)}_i} \right)^{1/2}$$  \hspace{1cm} (5.33)

and corresponds to choosing one of the branches of the square root.

This choice is in fact, part of the definition of the Levinson's transformation. Chester, Friedman and Ursell showed that there exists only one branch of the transformation which is uniformly regular and bijective. This is precisely this branch that we must select when choosing the branch of the square root.

The consequences of a wrong choice are the definition of a mapping $z$ which is singular at $\bar{\alpha}_0$, denying any validity to the results. Numerically, this translates in values of $|z^{(n)}_i|$ which increase faster with $n$ than the non-singular branch, this for $\bar{\alpha} \neq \bar{\alpha}_0$. Such a behavior can be mistaken for the numerical failure of the formulas of (5.27),
especially since as $\bar{\alpha} \to \bar{\alpha}_0$ these values increase without bounds. It is therefore necessary to find some criterion that will enable us to choose the appropriate branch of this square root. This can be done systematically for small enough values of $|\beta|$, and extended to larger values using connectedness arguments.

Consider $\bar{\alpha}$ in a neighborhood of $\bar{\alpha}_0$, small enough so that the multiple series of $z_i^{(n)}$ are converging. The constraint of having regular series implies, as stated earlier, that:

\textit{the constant term in the multiple series of the numerators, in the various equations of (5.27), should be 0.}

This necessary condition can be applied in particular to the numerator of the second equation of (5.27) or:

$$2 - f_i^{(3)} z_i^{(1)^3}$$  \hspace{1cm} (5.34)

Hence we deduce:

\textit{Criterion 1}

$$a_{i(3)}(\bar{0}) a_{1}(\bar{0})^3 = 2$$  \hspace{1cm} (5.35)

As the definition of $f_i^{(3)}$ does not suffer from any indetermination, this condition only bears on $z_i^{(1)^3}$. It suffices then to change eventually the sign of $z_i^{(1)^3}$ in order to ensure that the right branch has been selected. Since this problem of choosing the correct branch only appears in (5.33), the computation of the series for $z_i^{(d)}$, $d \geq 2$, does not present additional problems.

This criterion is of course only valid for the multiple series method. When using the direct method we can only formulate a condition which might not be uniformly valid in $W_1$:

\textit{Criterion 2}

\textit{The sign of $z_i^{(1)}$ is chosen such that: $|2 - f_i^{(3)} z_i^{(1)^3}|$ is minimal.}
This simple criterion is satisfactory as long as $\bar{\alpha}$ is such that, for one of the branches:

$$\Re \left\{ f^{(3)}_1 z^{(1)^3}_1 \right\} > \delta > 0 \quad (5.36)$$

and this uniformly in the region described by $\bar{\alpha}$.

- It is interesting to note that the determination of $\zeta^{1/2}$ is of no consequence in the choice of the appropriate branch of the Levinson’s transformation since the cubic power of $z^{(n)}_1$ is involved in (5.35).

The extension of this choice to larger regions of $D_1$ is done by ensuring its uniformity along paths in $D_1$. This requires the knowledge of the branch for some reference value and its extension along a continuous path in $D_1$ (see Appendix C).

Hence we conclude our considerations of the “delicate” choices in the uniform treatment of the fold catastrophe, and our analysis of the procedure for computing the expansion. It remains now to focus on the practical aspects of an implementation, following the steps given in Section 5.3.5. The resulting algorithms do possess some common numerical features inherited from the methods we proposed.

### 5.5 Numerical characteristics

The “direct” methods of Section 5.3 present the same numerical properties as those for the cases of an end-point and an isolated col. Namely the same question of floating point number representation, and more important, the same problems of cancellations have to be addressed. If the remarks and remedies of Sections 4.1.2 and 4.2.2 regarding number representations are still valid, cancellations have to be treated in a more radical way. They necessitate the use of complementary methods: the multiple entire series (5.31) for the computation of $z^{(n)}_1$, and the power series (5.22) for the sequences $(p_r)$ and $(q_r)$. 174
• Cancellation errors

Cancellations occur in the same manners as seen in Chapter 4, and in addition in the computation of the sequences \((p_r)\) and \((q_r)\). A general statement on their occurrence is again impossible as both extremes – total cancellation or absence of cancellation – might take place. The resulting errors however cannot be reduced by extending the floating point number representations as in Chapter 4, since the algorithms for the local inversion of the Levinson’s transformation and for the computation of the sequences \((p_r)\) and \((q_r)\) become numerically ill-behaved. This motivated the search for complementary methods for these particular steps. The computation of the sequences \(G_i^{(k)}\) however did not require such alternative algorithms since the current method stays well-behaved. Nevertheless cancellation errors may occur at this stage and the final form of the power series method for the sequences \((p_r)\) and \((q_r)\) was devised to minimize their influence on the final result.

The regimes in which cancellation errors affect the results the most, are undoubtedly at the boundaries of the regions of validity of the complementary methods. Their importance there depends entirely on the local properties of the functions \(f\) and \(g\), or more precisely:

— on the occurrence or not of cancellations

— on the sizes of the domains of convergence of the series

The use of extended number representations is the only systematic means to minimize them, but we cannot guarantee that they will not exceed a given bound for all of the cases. The uniformity of the numerical evaluations will then have to be assessed when encoding the algorithms for a specific case. The influence of such errors on the final result depends of course as in Chapter 4 on the values of \(N\) and on the truncation indices of the asymptotic series. In this respect their importance might be contained, but this can only be asserted when the final result is available numerically.
- **Particular cases**

Perhaps more than for the cases of the ordinary method of steepest descents, it is of interest to test the methods for each stage of the whole procedure on simple but characteristic examples. This is especially relevant to the inversion of the Levinson's transformation upon which depend the results.

The comparison of numerical results from the inversion of polynomials and rational functions highlights the influence of the local behavior of \( f \) on the properties of the sequences \( (z_i^{(n)}) \).

### 5.6 The building of an implementation

Quite differently to Chapter 4 where we presented what the algorithms could be, here we only describe *how an algorithm should be built* for the case of the fold catastrophe. This lack of *determination* is inherent to this case: some of the methods detailed in Sections 5.3 and 5.4 are obviously case dependent (e.g. the multiple entire series). Therefore an implementation of the method of steepest descents for the fold catastrophe is necessarily case specific. This is apparent in the procedures we propose which are *combinations of complementary methods* requiring *numerical explorations* of their regions of validity.

Our claim is now that for a given case which presents a fold catastrophe, the algorithm for its uniform evaluation by the method of steepest descents, can be built in a *systematic manner* by using the methods given in Section 5.3.

Both for the inversion of the Levinson's transformation and for the computation of the sequences \((p,),(q,)\), we need to introduce uniform errors associated to the quantities \(z_i^{(n)}, p, q\). We name the assessment of these errors, their control and the investigation of the corresponding regions of validity of the different procedures, "the building of the uniform algorithms". The total algorithm is then constructed by
assembling these uniform algorithms in the schematic procedure given in 5.3.5.

The first step is thus to build the uniform algorithm for the inversion of the Levinson's transformation, the cornerstone of the whole procedure.

5.6.1 The numerical inversion of the Levinson's transformation

The task here is multiple. It consists in:

(i) fixing the reference value \((\beta_{\text{ref}}, \arg \zeta^{1/2}_{\text{ref}})\) for the definition of \(\zeta^{1/2}\), and determining the three regions of \(\tilde{O}_2\) where a given branch of (5.24) is uniformly valid.

(ii) determining the regions of the \(\tilde{\alpha}\)-space where a given branch of (5.33) is uniformly valid.

(iii) finding for each \(z_i^{(n)}\) the radius \(\tau_n\) of the disk, \(|\zeta^{1/2}| < \tau_n\), inside which the multiple entire series are to be used, the multi-index of truncation \(J_n\), and the bound on the error \(\varepsilon_{z_i^{(n)}}\).

Steps (i) and (ii) correspond to the choices described in Section 5.4. They involve the "numerical" continuation of branches in the complex plane, uniformly in \(\tilde{\alpha}\), and its practical aspects are treated in Appendix C. The determination of the regions of \(\tilde{O}_2\) requires numerical "explorations" of a given neighborhood of \(\tilde{\alpha}_0\), unless the boundaries of these domains are simply defined. In addition, (ii) involves the computation of the first coefficient of the multiple entire series for \(z_i^{(1)}\) and \(f_i^{(3)}\). In fact, at this stage we include the computation of the entire series for each \(z_i^{(n)}\). (A method for the numerical computation of these coefficients is given in Appendix D.)

In (iii) we assess the numerical properties of the combination multiple series/direct formulas, in terms of their regions of validity and the uniform minimal errors they
guarantee. We choose to characterize these regions in the $u$-plane (i.e. in terms of half the distance between the images by $u$ of the cols, or $\zeta^{1/2}$), since they can be quite well approximated by disks. This is justified by the form of the multiple entire series where $\beta_1$ plays a central role, and the fact that $\zeta^{1/2} \sim \beta_1$ in a neighborhood of $\alpha_0$. Consequently the resulting tests are much simpler than the characterization of regions of the $\alpha$ or $\beta$-spaces. Nevertheless we cannot rule out the necessity of delimiting more finely the geometry of these regions in certain cases.

The finite sequence of the three determinant parameters, or $((\tau_n, J_n, \epsilon_{z_i^{(n)}}))$, completely characterizes the uniform numerical inversion of the Levinson's transformation. Since the series are convergent series, the triplet $(\tau_n, J_n, \epsilon_{z_i^{(n)}})$ is such that:

- the series are convergent for $|\zeta^{1/2}| < \tau_n$ (upper bound on $\tau_n$),
- $J_n$ ensures a relative truncation error well below $\epsilon_{z_i^{(n)}}$,
- $\tau_n$ is defined by the minimal relative difference between the series and the direct formula results. The value of this difference gives $\epsilon_{z_i^{(n)}}$.

This description is enhanced by the following comments:

- For this investigation and for the final encoding, the use of large mantissas (16 digits if a final result of at least 5-6 digits is desired) is necessary as in most cases the errors will be dominated by cancellation errors.

- Relative differences are used, since the $z_i^{(n)}$ are usually increasing sequences of $n$.

- The number of derivatives $z_i^{(n)}$ to be computed, or $n_{\text{max}}$, has to be fixed a priori.

- The sequence $\tau_n$ is increasing with $n$ (due to the worsening of cancellation errors with $n$).

Once this “exploration” is completed, the different information (coefficients, region definitions, formulas) can be encoded in their final form to constitute the algorithm for the inversion of the Levinson’s transformation. (This encoding should include an
eventual optimization of the formulas.) Then, and only then, the process of defining the algorithm for the computation of the sequences \((p_r), (q_r)\) can be considered.

5.6.2 The computation of \((G_i^{(k)})\)

The direct algorithm based on the adapted formulation of (4.10) can be used without any further modifications. Its generality is preserved, therefore its encoding can be done as in Chapter 4, and will be valid for all cases.

5.6.3 The uniform computation of the sequences \((p_r), (q_r)\)

The aim here is much simpler than for the computation of \(z_i^{(n)}\). It consists simply in the equivalent of (iii) above. The same methodology as in Section 5.6.1 is employed since the power series of (5.22) are also convergent. This results in the definition of \((\sigma_{p_r}, K_{p_r}, \epsilon_{p_r})\) for each \(p_r\) (and similarly for each \(q_r\) where:

- \(\sigma_{p_r}\) is the radius of the disk \(|\zeta^{1/2}| < \sigma_{p_r}\) in which the series for \(p_r\) in (5.22) is used.

- \(K_{p_r}\) is the order at which this series is truncated.

- \(\epsilon_{p_r}\) is the upper bound on the error, given by the relative difference between the results from the series and the direct formulas on \(|\zeta^{1/2}| = \sigma_{p_r}\).

Remarks similar to those of Section 5.6.1 can be made. In addition we need to relate the maximum order \(r_{\text{max}}\) at which the sequences \((p_r)\) and \((q_r)\) are computed, to the truncation orders of the series, and to \(n_{\text{max}}\). In particular \(r_{\text{max}}\) must satisfy the constraint: \(\max(\{2r + K_{p_r}\}, \{2r + 1 + K_{q_r}\}) + 1 \leq n_{\text{max}}\). Hence, for example, computing \(p_r\), for \(0 \leq r \leq 7\), using series truncated at the order 3 in \(\zeta\), requires the computation of the first 21 terms of \((z_i^{(n)})\).

Again, once this task is completed, the corresponding algorithm can be encoded in its final form. The next step in building the implementation corresponds to the
arrangement of the three algorithms evaluating \( z_i^{(n)} \), \( G_i^{(k)} \), and the sequences \((p_r), (q_r)\) toward the computation of the four asymptotic series.

### 5.6.4 The truncations of the asymptotic series

The recursion relations (5.12) which give the sequences \((A_s/N^{2s}), (B_s/N^{2s}), (C_s/N^{2s})\) and \((D_s/N^{2s})\), are not case specific. Therefore, they can be encoded in a fixed form, valid for any integral \(I\). The only information that might change, is the first terms of the recursions which depend on \(\Gamma_{\text{ref}}\).

#### 1. The recursions

The recursion relations, as noted earlier, tend to decrease in sets of four, for moderate values of \(r\), typically: \(2r \leq N\). This can be seen by rewriting (5.12) as:

\[
\begin{align*}
    \mathcal{F}_r &= \left(\frac{2r}{N}\right)^2 \left(1 - \frac{1}{r}\right) \left[ \left(1 - \frac{5}{2r}\right) \mathcal{F}_{r-3} + \left(1 - \frac{3}{r}\right) \zeta \mathcal{F}_{r-4} \right] \\
    \mathcal{G}_r &= \left(\frac{2r}{N}\right)^2 \left[ \left(1 - \frac{1}{2r}\right) \left(1 - \frac{2}{r}\right) \mathcal{G}_{r-3} + \left(1 - \frac{1}{r}\right) \left(1 - \frac{3}{r}\right) \mathcal{G}_{r-4} \right]
\end{align*}
\]

(5.37)

However, we are interested in the sequences \((p_r \mathcal{F}_r)\) and \((q_r \mathcal{G}_r)\). Since \((p_r)\) and \((q_r)\) are (hopefully) most of the time decreasing, these product sequences are much better behaved than the above recursion relations would indicate. Of course ultimately the increasing trend will dominate, giving to the series this so characteristic feature of being divergent, but asymptotic.

The encoding of these recursions does not present any numerical difficulty. Their independence should be preserved in the algorithm, since the truncation tests are performed independently on each of the four series.
2. The truncation tests

The truncation tests are exactly the same as in Section 4.3.2. They have to be performed independently on each of the four series, with:

\[
\text{mod}_s = \left| \frac{A_s}{N^{2s}} \right|, \quad \left| \frac{B_s}{N^{2s}} \right|, \quad \left| \frac{C_s}{N^{2s}} \right| \quad \text{or} \quad \left| \frac{D_s}{N^{2s}} \right| \quad (5.38)
\]

This complicates significantly the overall algorithm in the arrangement of the different procedures. Indeed, this independence is only apparent since the four series share some common information. The main effort is then to combine these various algorithms in order to make the evaluation of the series as efficient as possible.

3. The global arrangement

The recursion relations indicate that on one side, \((p, F_r)\) generates \((A_s/N^{2s})\) and \((B_s/N^{2s})\), and on the other side, \((q, F_r)\) generates \((C_s/N^{2s})\) and \((D_s/N^{2s})\). Hence we can draw the diagram of Figure 5-1 in terms of information flow. It appears then that the feedback information issued from the truncation tests has to be distributed at each stage of the computation, and depends on the series being considered. This information is not easily accessible but can be determined once and for all, and encoded as in Table 5.1.

The final stage consists in assembling the four series multiplied by the appropriate generic functions.

5.6.5 Assembling the final expansion

This last step involves the computation of the Airy function and its derivative of a complex argument. It also corresponds to the last stage in the computation of the
Figure 5-1: Information flow

Table 5.1: Feedback information for each series

<table>
<thead>
<tr>
<th>$s_A$</th>
<th>$r_p$</th>
<th>$s_B$</th>
<th>$r_p$</th>
<th>$s_C$</th>
<th>$r_q$</th>
<th>$s_D$</th>
<th>$r_q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>6</td>
<td>11</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>7</td>
<td>12</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>—</td>
<td>—</td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>—</td>
<td>—</td>
<td>9</td>
<td>12</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>10</td>
<td>12</td>
</tr>
</tbody>
</table>
expansion, once the asymptotic series have been truncated.

Efficient and accurate algorithms for the evaluation of the complex Airy function exist, and can be found, by example, in Schulten, Anderson and Gordon (1979). Care is required, taking into account the behavior of the Airy function and the details of these algorithms. Such considerations are fundamental for the correctness of the numerical results.

The main concern lies in the evaluation of the Airy function of large complex argument. Indeed, as $N$ gets large and since $\zeta$ takes values a priori in a disk centered at the origin, the argument of the Airy function, or $N^{2/3} \zeta$, can become large as well. In this case it is appropriate to compute a “scaled” form of the Airy function, rather than the function itself, that is:

$$
\text{Ai}(N^{2/3} \zeta) = \exp\left[-\frac{2}{3} N \zeta^{3/2}\right] \overline{\text{Ai}}(N^{2/3} \zeta)
$$

where $\overline{\text{Ai}}$ stands for the “scaled” Airy function.

Scaled results are available from such algorithms, and must be used at all times since the final result is multiplied by the exponential term $\exp[N A(\bar{\alpha})]$ which depends on $N$. In fact, both exponents are added, so that we have:

$$
A(\bar{\alpha}) - \frac{2}{3} \zeta^{3/2} = \begin{cases} 
  f(z_1, \bar{\alpha}) \\
  f(z_2, \bar{\alpha})
\end{cases}
$$

depending on the choice of the branch for $\zeta^{1/2}$ which is made in these algorithms. Finally, the exponential term, including the scaling of the Airy function, becomes:

$$
\exp[N f(z_i, \bar{\alpha})] \quad \text{for } i = 1 \text{ or } i = 2
$$
At this stage the building of an algorithm which ensures a uniform numerical evaluation of $I$ over $\mathbb{O}^2$ is completed. In this presentation we have emphasized the description of the rules in order to obtain an algorithm which will produce the uniform numerical evaluation of a given integral $I$. Other aspects such as the optimization of the various formulas, and the role of symbolic computations are the same as in Chapter 4. In fact, the importance of such an environment is even more acute as some of the algorithms are used in their critical regimes. The possibility of using arbitrary number representations is particularly useful when attempting to isolate the consequences of a wrong choice of a root branch from the numerical deterioration of the direct methods.

5.7 Conclusion

We can only emphasize the contrast between the procedures presented in Chapter 4 and those described in this chapter on the fold catastrophe. The main striking feature of the uniform treatment of this case is the complexity of the procedures it involves. Exactly as suspected, the mathematical complexity of the fold catastrophe is mirrored in the procedures for the numerical evaluation of the expansion, and in particular in the point inversion of the Levinson's transformation.

This complexity has an obvious consequence: the apparent impossibility of providing an algorithm which will ensure the uniform numerical evaluation of any integral $I$ presenting a fold catastrophe. Nevertheless we can infer general rules regarding the building of a robust algorithm for any particular integral $I$. Some of these rules are given under the forms of tools for constructing such an algorithm. They consist in procedures of symbolic computation for the systematic generation of the formulas present in the direct methods and in the power series method for the sequences $(p_r)$, $(q_r)$. They allow us to focus our efforts towards the fundamental problems of the uniform evaluation of $I$: the determination of the regions of the $\bar{\alpha}$-space in which a given method, or choice of a root branch, is numerically uniformly valid. This
task, which involves numerical explorations exceeds the capacity of systematization provided by symbolic computations. However this can be carried out following the guidelines given in Sections 5.6.1 and 5.6.3, along with the method for obtaining the multiple entire series in the inversion of the Levinson’s transformation.

1. Robustness

The robustness of the resulting algorithm depends in a crucial manner on the construction of uniform algorithms for each stage of the whole procedure. For this purpose the guidelines we present, furnish a suitable frame.

The main obstacle in obtaining a minimal fixed error in the numerical results, is the critical influence of cancellations in the direct methods for the inversion of the Levinson’s transformation, and the computation of the sequences \((p_r), (q_r)\). From the analysis of these phenomena carried out in Chapter 4, their occurrence cannot be ruled out, while the only systematic remedy is the use of extended number representations in the direct methods. The final influence of cancellation errors can be estimated only when assessing the regions of validity of the complementary methods based on series expansions around \(\bar{\alpha}_0\).

2. Efficiency

The efficiency of the resulting algorithm relies, as for the other cases, on the preservation of the asymptotic nature of the expansions in their numerical forms. An algorithm built along the guidelines of Section 5.6 will satisfy this condition. The remarks on the influence of the range of values of \(N\) made in Section 4.4, are then relevant. However, the total computational cost is here much higher in terms of numbers of operations, even after optimization. This higher cost is imputed to the larger number of stages in the global process: four instead of two; to an increased complexity in the inversion of the Levinson’s transformation, and to the use of an Airy function.
In particular applications, where speed is a main priority, alternative methods for obtaining more efficiently the coefficients of the series should be considered. Such methods will be based on approximations and interpolations of numerical values obtained from an algorithm of this kind. This option is of particular interest since it allows the use of extended number representations for the computation of these values without any consequences on the run time of the final algorithm.

3. Generality

If the principles given for constructing uniform algorithms are by essence of general use, the algorithms are necessarily case specific. In addition, practical restrictions intervene in the process of obtaining the multiple entire series (see Appendix D). If those restrictions are acceptable for isolated fold catastrophes, they would become extremely constraining for extended fold catastrophes such as in the cases of higher catastrophes. Thus the practical extension of the methods for the fold catastrophe, to other catastrophes, is highly questionable. We must also note that the sequences of Chester et al. is particular to the fold catastrophe. For the treatment of other catastrophes, algorithms for the computation of the Bleistein sequences are still to be devised.

Our work would be incomplete if we did not investigate the overall numerical properties of such a fold catastrophe expansion. To our knowledge, the procedures described in this chapter and the previous one constitute the first attempt to numerically obtain, in a systematic manner, uniform steepest descent expansions up to an arbitrary order. Hence, we propose as an ultimate presentation of all the tools and algorithms presented earlier, to treat in full a particular example.
Chapter 6

The Kelvin wave source potential

We present here our first and main example of a numerical application of the method of steepest descents.

The Kelvin wave source potential is particularly suitable for such an application. Indeed this function has played a persistently stimulating role in the development of asymptotic methods for approximations of integrals. Hence shortly after formulating his principle of stationary phase (Kelvin, 1887a), Lord Kelvin revised his previous article ‘On Ship Waves’ (Kelvin, 1887b) by applying this very method to the potential of a translating source. Following the formulation by Debye of the method of steepest descents, Hogner (1923) then Peters (1949) studied the Kelvin potential using this new tool. However these studies failed to provide treatments of the vicinity of the rays $\theta = \pm 19^\circ 28'$ in the Kelvin wave pattern. This obstacle, characteristic in fact of a fundamental failure of the ordinary method of steepest descents, motivated the work of Chester, Friedman and Ursell (1957). Not surprisingly, Ursell applied this latest improvement to the Kelvin potential (Ursell, 1960).

The application to the Kelvin potential of the algorithms described in Chapters 4 and 5 can be viewed as a numerical extension of the treatment found in Ursell (1960). Thus we extend the use of the method of steepest descents to the whole domain where it can possibly be applied: namely, for any submergence of the pair source-field point away from a neighborhood of the axis $\rho = 0$. This contrasts with previous theoretical
works concerned only with the free-surface elevation when both the source and the field-point are on the free-surface (with the exception of Chung and Lim, 1991) and always at leading order, or numerical works persistently considering the source to be submerged. The other and more interesting aspect from the standpoint of applied mathematics, is the extension of the results in terms of accuracy, or the ability of obtaining numerically the asymptotic series not only at leading order but at any order. As a whole this particular example presents a complete numerical implementation of the method of steepest descents using the algorithms devised in previous chapters.

In the following presentation, we want to demonstrate how the various tools and algorithms could be applied to a given integral. Thus we start from the analytical expression of the integral, and perform the necessary analysis to obtain a form amenable to the use of these algorithms. Obviously, some of the problems encountered in the implementation of the method of steepest descent are greatly simplified by the form of the Kelvin potential.

6.1 The “single integral” part of the Kelvin wave source potential

The overall procedure for treating the Kelvin potential was outlined in Clarisse (1990). We now give the corresponding details in this chapter. As already mentioned in Example 1 of Section 2.2, the application of the method of steepest descents corresponds precisely to the last portion of the physical space not yet treated (see Clarisse, 1989 & 1990). Hence there exists several complementary expressions for the potential:

— a convergent Neumann series obtained by Bessho (Bessho, 1964, eq. 4.2) and valid in a region extending from the near field to the far field, away from the axis $\rho = 0$ (see Section 2.2).
— an asymptotic series originally found by Bessho (Bessho, 1964, eq. 4.3) but in an incomplete form, and later fully derived by Ursell (1988), valid in the neighborhood of \( \rho = 0 \) but away from the origin. (The results of Ursell (1988) are stated for \( z \) bounded, but can be extended in their forms of Theorems 1 and 2 to unbounded \( z \) and bounded \( \rho \).) This expression is numerically complementary to the above in a region \(-16.75 \leq X \leq -1\), corresponding to a six digit absolute accuracy.

— an asymptotic series in the same spirit as the above (Clarisse, 1989), valid in the neighborhood of the origin and complementary to both series, with again six digit absolute accuracy.

They furnish efficient and robust means of obtaining numerical values of the potential and its derivatives, or integrals. They also provide analytical formulas indispensable to the analysis of the potential in its most critical regime (Clarisse, 1991). The goal is here to provide a method for the complementary region - i.e. \( X \leq -16.75 \) and away from \( \rho = 0 \) - which would present the same characteristics of robustness and accuracy, as well as providing corresponding analytical formulas. Once completed, this collection of algorithms would form a complete set of algorithms for the efficient and robust numerical evaluation of the potential, anywhere in the fluid domain. Such a set can be considered as a numerical and analytical reference for this function, especially in conjunction with the algorithms for the near-field disturbance developed in Newman (1987a).

The wave like disturbance of the Kelvin wave source potential is also often called "single integral" part of the Kelvin potential. There exists various expressions for this function (see Noblesse, 1981, by example), but we retain the form of (2.1), that is:

\[
    w(X, Y, Z) = \int_{0}^{\infty} \sin X \sqrt{1 + t^2} \cos Y t \sqrt{1 + t^2} \exp[-Z(1 + t^2)] \, dt
\]

for \( X < 0 \).
Its mathematical properties are studied extensively in Euvrard (1983). By symmetry of its domain of definition, we are reduced to consider the eighth of space: \( X < 0, Y \geq 0, Z \geq 0, \rho \neq 0 \). Our goal is to find an expression of (6.1) which allows its numerical evaluation in a region complementary to the domains of validity of the expressions earlier mentioned. After numerical exploration of these domains, this complementary region appears to be bounded by: \( X \leq -16.75 \) and \( \rho \geq 1/2 \) (based on a six digit absolute accuracy). Therefore, we can assume that the radial distance from the origin, \( R \), is large. The application of the method of steepest descents as done by Hogner (1923), Peters (1949) or Ursell (1960), is then legitimate. We use here a form of \( w \) which differs slightly from Ursell (1960, eq. (3.2)).

We consider \( w \) under the form (Ursell, 1988, eq. (2.2)):

\[
w(X, Y, Z) = \frac{e^{-\frac{X}{2}}}{2} R \left\{ \frac{I^+ + I^-}{2} \right\}
\]  

(6.2)

with:

\[
I^\pm(X, \rho, \varphi) = \int_{-\infty + i\varphi/2}^{+\infty + i\varphi/2} \exp \left[ -\frac{\rho}{2} \cosh(2u - i\varphi) \pm iX \cosh u \right] \cosh u du
\]

(6.3)

where \((X, \rho, \varphi)\) are the cylindrical coordinates of axis \( X \). The path of integration ensures the uniform convergence of the integral with respect to its parameters.

The relation between \( I^+ \) and \( I^- \):

\[
I^-(X, \rho, \varphi) = \overline{I^+(X, \rho, \varphi)}
\]

reduces the problem to the treatment of \( I^+ \). Upon performing the change of variable \( s = \exp[-u + i\varphi/2] \), we obtain \( I^+ \) in the generic form of Section 1.1.1, namely:

\[
I^+(R, \theta, \varphi) = I(N, \tilde{\alpha}) = \int_C g(z, \tilde{\alpha}) \exp[N f(z, \tilde{\alpha})] dz
\]

(6.4)
where \((R, \theta, \varphi)\) are now the spherical coordinates associated to \((X, \rho, \varphi)\), and with:

- \(\tilde{a} \in \mathbb{C}^2\), with the relations: \(\alpha_1 = e^{-i\varphi/2}\) and \(\alpha_2 = i \cot \theta\)
- \(g(z, \tilde{a}) = \frac{1}{2}(\alpha_1 z + \alpha_1^{-1} z^{-1})\)
- \(f(z, \tilde{a}) = -\frac{1}{4} \left( z^2 - 2 \alpha_2 \alpha_1 z - 2 \alpha_2 \alpha_1^{-1} z^{-1} + z^{-2} \right)\)
- \(N = R \sin \theta = \rho\)
- and \(C\) being a path joining \(0e^{i\theta}\) to \(\infty e^{i\theta}\).

This definition of \(I\) draws some remarks:

- We have the freedom of choice for the parameterization of \(f\) and \(g\) in \(\tilde{a}\). The only constraint is that this dependence must be holomorphic on some domains of the \(\tilde{a}\)-space.

- The form of (6.4) is much more suitable to a steepest descent analysis than (6.3), mainly from the fact that the functions \(f\) and \(g\) are now rational functions. Thus we are avoiding the considerations of periodic functions with an infinite number of basins and cols.

- It appears from the definition of \(N\) that our expansions will be in powers of \(\rho\) instead of \(R\) as in Ursell (1960). This difference is only apparent, and our results will be entirely equivalent to those of Ursell.

### 6.2 The domains of definition

Most of the subsequent analysis, including the determination of the relevant cols, depends on the domains \(W\) of the complex \(z\)-plane, and \(D\) of the \(\tilde{a}\)-space, in which \(f\) and \(g\) will be considered.
According to the assumptions of Section 1.1.1, \( z = 0 \) must be excluded from \( W \). For convenience, we in fact exclude a steepest path \( S_0^* \) issued from \( z = 0 \) and ending at infinity, and thus define:

\[
W = C \setminus S_0^*
\]  

(6.5)

a simply connected domain excluding the pole \( z = 0 \) (see Figure 6-4). This choice allows the use of Cauchy's theorem freely in \( W \).

Regarding the dependence on \( \bar{\alpha} \), both \( f \) and \( g \) have a pole of order 1 at \( \alpha_1 = 0 \). We therefore define:

\[
D = C^2 \setminus \{ (\alpha_1, \alpha_2) : \arg(\alpha_1) = -\pi \}
\]  

(6.6)

which is simply connected. Our region of interest is in fact restricted to a subset, \( D' \), made of the portion of the cylinder \( |\alpha_1| = 1 \), defined by: \( 0 \leq \arg \alpha_1 \leq \pi/4 \), \( \Re \alpha_2 = 0 \), \( \Im \alpha_2 > 0 \). In all the computations we will consider \( \bar{\alpha} \) in \( D' \), except when otherwise specified.

By definition \( W \times D' \) is such that \( f \) and \( g \) satisfy the assumptions of Section 1.1.1. From there, we can apply the method of steepest descents and at first determine the relevant cols.

### 6.3 The relevant col problem

A priori, we need to locate all the saddle points of \( f \), or find the solutions of the quartic equation:

\[
z^4 - \alpha_2 \alpha_1 z^3 + \alpha_2 \alpha_1^{-1} z - 1 = 0
\]  

(6.7)

for \( \bar{\alpha} \) in \( D \). However, by restricting \( \bar{\alpha} \) to \( D' \), this equation takes the form of:

\[
z^4 + a z^3 + \bar{\alpha} z - 1 = 0
\]  

(6.8)

where: \( a = -\alpha_1 \alpha_2 = -i \cot \theta e^{-i\varphi/2} \).
1. Locating the cols

Obtaining numerically the solutions of (6.7) does not raise any specific difficulties, since analytical expressions for the roots of a quartic polynomial provide an efficient and robust numerical method. These formulas are preferred here, as pure numerical methods are likely to fail in a neighborhood of the coalescence of the cols. Equation (6.8) presents some symmetry properties which simplifies further the problem. Hence if \( z_1 \) and \( z_2 \) are two roots of (6.8), we have:

\[
(6.8) \quad \Leftrightarrow \quad (z - z_1)(z - z_2)(z + \frac{1}{z_1})(z + \frac{1}{z_2}) = 0 \quad (6.9)
\]

As \( a \) describes an annular strip in the sector \([\pi/2, 3\pi/4]\), it appears that two of the roots are bounded to be in the first quadrant, whereas the two others are in the third quadrant of the complex \( z \)-plane. Since the end points of the path \( C \) are located in the first quadrant, it seems reasonable to assume that only the two roots located in the first quadrant will matter. If this assumption is confirmed subsequently, its veracity is not obvious. The two roots in the first quadrant describes two subsets, 1: \( \{z_1(\bar{a}) : \bar{a} \in D'\} \) and 2: \( \{z_2(\bar{a}) : \bar{a} \in D'\} \), of the \( z \)-plane as shown in Figure 6-1. The intersection of these subsets is a single point, and corresponds to the coalescence of the four roots in pairs. Double roots can be shown to occur only for \((\theta_0, \varphi_0) = (\sin^{-1}(1/3), \pi/2)\).

2. Indexing the cols

As discussed in Chapter 3, an important task is to be able to index and follow the saddle points as \( \bar{a} \) describes \( D' \). Here, this problem is simplified by the geometry of the sets, and by the fact that the cols are coalescing by pairs. Although this is possible, it is not necessary to use multiple entire series expansions as mentioned in Section 3.3.3 in order to follow the cols in the vicinity of the double roots.
Figure 6-1: The sets described by the roots in the first quadrant, for $4^\circ \leq \theta \leq 50^\circ$, $0^\circ \leq \varphi \leq 90^\circ$. The curves inside the sets are curves of constant depth of the pair source-field-point; or trajectories of the cols for a constant depth as $(\theta, \varphi)$ varies.

3. Determining the relevant cols

The problem of determining the relevant saddle points is here particularly a propos. In the classical stationary phase analysis of the wave elevation, when the source is on the free-surface (i.e. $\varphi_0 = \pi/2$), three regimes are easily distinguished as $\theta$ varies (see Tuck et al. (1971) by example):

- the presence of two points of stationary phase, for $\theta < \theta_0$
- the coalescence of the two stationary phase points, for $\theta = \theta_0$
- the absence of stationary phase point, for $\theta > \theta_0$

However, this analysis fails as soon as $\varphi \neq \pi/2$. For such cases, no point of stationary phase appear in the range of integration, and the behavior of the integral cannot be reduced to some local contributions as before. By studying the images, by $f$, of
the sets $1$ and $2$ — that is the sets described by $w_1 = f(z_1, \alpha)$ and $w_2 = f(z_2, \alpha)$ respectively, as $\alpha$ is in $D'$ — we can infer that the value of the integral is given by the contributions from the relevant saddle points among $\{z_1, z_2\}$. For any value of $\alpha$ there will be at least one relevant saddle point, denoted $z_I$.

Hence for $\varphi = \varphi_0$, the above results are therefore modified as follows:

— for $\theta < \theta_0$, the cols $z_I$ and $z_{II}$ are relevant

— for $\theta = \theta_0$, both cols are coalescent at $z_0 = \sqrt{2 + \sqrt{3}} e^{i \pi / 4}$, or in other words, $f$ presents a fold catastrophe

— for $\theta > \theta_0$, only $z_I$ is relevant

This is illustrated in Figure 6-2, under the form of the images, in the $w$-plane, of the sets 1 and 2 of Figure 6-1.

---

**Figure 6-2:** The images by $f$ of the sets 1 and 2 of Figure 6-1

The curves are again curves of constant depth of the pair source-field-point. The right view is an enlargement of the neighborhood of $w_0 = f(z_0, \alpha_0)$.
These two image sets are overlapping, and in fact we have the strict inclusion, in terms of images: \( 1 \subset 2 \). To distinguish which of these two sets correspond to \( \{f(z_1, \bar{a})\} \) is most difficult. By chance, we can invoke physical arguments to solve this problem. From observations of ship wakes, we know that no waves are present in a sector \( \theta > \theta_0 \).

This implies that \( \Re\{f(z_1, \bar{a})\} \) must be negative in this region, since the factor in front of the expansion of \( I \) is \( \exp[N f(z_1, \bar{a})] \). Hence we conclude that \( z_1 \) is \( z_i \).

For other values of \( \varphi \), a Stokes surface (i.e. a component of the set \( \mathcal{R}_R \)) defined in terms of \((\theta, \varphi)\) (in fact in terms of \( \bar{a} \)), divides \( D' \) in two regions:

- one, where both \( z_1 \) and \( z_2 \) are relevant, characterized by:
  \[
  7\pi/2 \leq \arg(w_2 - w_1) < 4\pi
  \]

- the other, where only \( z_1 \) is relevant, given by: \( 4\pi < \arg(w_2 - w_1) \)

The difference between the stationary phase analysis and the steepest descent analysis for a non zero submergence of the source, is seen in an enlargement of the vicinity of the line \( \Re w = 0 \), as in Figure 6-3.

This line corresponds to the trajectories of \( w_1 \) and \( w_2 \), as \( \theta \) varies in \([0, \theta_0]\) while \( \varphi \) remains constant at \( \varphi_0 \). The method of stationary phase applied to \( I \), is nothing else but an integration on this line: by definition of the method, \( I \) must be such that \( \Re\{f(z, \bar{a})\} = 0 \), which is precisely the line \( \Re w = 0 \). The results of the stationary phase analysis appear then clearly, and the results of Ursell (1960) should be recovered in the neighborhood of \( \theta_0 \).

When the source submergence is non-zero, the saddle points follow trajectories similar to those of Figure 6-3, for a constant depth of the field-point. Consequently, none of the cols will cross the line of integration. (Each set is in fact on two separate sheets of a Riemann surface which are connected at \( w_0 \) along the half-line \( \Re w = 0 \), \( \Im w \geq \Im w_0 \). When \( w_2 \) crosses the line \( \Re w = 0 \), it does not lie on the same sheet as the line of integration.) The limitations of the stationary phase analysis are then obvious: as soon as \( \varphi \neq \varphi_0 \), \( I \) cannot be reduced anymore to the local contributions of some points of \( \Re w = 0 \). This drawback can be in part avoided if one uses physical
Figure 6-3: The vicinity of \( \Re w = 0 \)

Trajectories, passed a certain submergence, are not shown. Lines are referenced in pairs (i.e. one line per col) by the submergence of the pair source-field-point.

cylindrical coordinates of axis \( Z \), with the radial distance as large parameter. (This approach is given in Wong, 1989, and used in Chung and Lim, 1991.) The formulation of the single integral is then such that points of stationary phase will always be present within the range of integration inside the Kelvin angle, while none are present outside this angle. This situation is acceptable for obtaining leading behaviors, but unsatisfactory when accurate results are needed for moderate values of \( R \).

From this discussion, the problem of determining the relevant cols seems fairly simple. Of course the long history of articles on the subject and the use of physical observations are of precious help. Solving this problem in its mathematical context would lead to apply the formalism of Chapter 3. As an illustration, we reproduce in Figure 6-4 the representation of the basins of \( f \) for \((\theta, \varphi) = (10^\circ, 80^\circ)\) given in Figure 3-8.
Figure 6-4: The basins of $f$ for $(\theta, \varphi) = (10^\circ, 80^\circ)$

The domain in $z$ is given by $-2 \leq \Re z \leq 5$, $-2 \leq \Im z \leq 5$, the conventions being the same as in Chapter 3. The four saddle points are given as well as the paths of steepest ascent issued from them. The cut $S_0^*$ is represented by the dashed line originating from the origin. The end-point at infinity of the initial path of integration, $C$, is symbolized by the arrow at the right of the plot.

The application of the formalism of Chapter 3 yields directly the result in the form of the appropriate path in $G_f$ (see Figure 3-8). Hence only the two saddle points located in the first quadrant matter. We distinguish three subdomains of $D$ corresponding to three different expressions for $I$:

- $O_1^1$ where $I$ is given by an isolated col expansion from $z_1$

- $O_2^1$ where $I$ is given by the sum of the isolated col expansions from $z_1$ and $z_2$

- $O_0^2$, a neighborhood of $\tilde{\alpha}_0$, where $I$ is given by a fold catastrophe expansion.

The task is then to obtain numerically the expansions for each of these domains, and to ensure that they provide a uniform numerical approximation of $I$ on a collection $\{\tilde{O}_1^1, \tilde{O}_2^1, \tilde{O}_0^2\}$ of subsets of $D'$. 
6.4 The isolated col expansions

Since the path of integration $C$ are such that $\Re f = -\infty$ at its end-points, the contribution to the integral comes solely from the saddle points. Hence the expansion is formally written:

$$I \sim \frac{\sqrt{2\pi}}{N^{1/2}} \sum_{z_i \in Z_R} \exp[N f(z_i, \bar{a})] \sum_{k=0}^{\infty} \frac{G_{z_i}^{(2k)}}{k!} \frac{1}{(2N)^k}$$

(6.10)

where $Z_R \subset \{z_1, z_2\}$.

The algorithm presented in Chapter 4 is applied in its original form. The only delicate point is the uniform determination of the square root present in the inversion of the mapping $u$ (see Section 4.2.2). For this particular example, a choice of the branch is uniform for $(\theta, \varphi)$ in $\tilde{O}_1 \cup \tilde{O}_2$ as shown in Figure 6-5: the surfaces described by the real and imaginary parts of $z_{z_i}^{(1)}$ are smooth everywhere in the domains of relevance of the respective cols.

Regarding these figures, some comments are necessary:

- All computations are in 16 digit mantissa (double precision in FORTRAN 77).
- The domain of computation used for this figure and the following, is a patch $[1^\circ, 50^\circ] \times [0^\circ, 90^\circ]$ in terms of $(\theta, \varphi)$, from which is excluded a neighborhood of the fold catastrophe at $(\theta_0, \varphi_0)$ since all the quantities are singular there.
- The real and imaginary parts of both $z_{z_1}^{(1)}$ and $z_{z_2}^{(1)}$ increase in magnitude in the vicinity of $(\theta_0, \varphi_0)$, characterizing the failure of the quadratic mapping $u$, and consequently of the asymptotic expansions.
- The discontinuities in the surfaces for $z_{z_2}^{(1)}$ correspond to the trace of the component of $R_R$ in this patch: the expansion relative to $z_2$ is not computed outside $\tilde{O}_2$ and is set there to 0.
The computations of the coefficients of the series are carried up to a maximum order of $K = 5$. This value will appear to be sufficient for the domain of interest and for the desired absolute accuracy of six significant digits, but is in no way restrictive.

1. Coefficients of the expansion

Along with the numerical values of the terms $G_{2k}^{(2k)}/k! (2N)^k$, for any value of $N$, we have also access to intermediate results. This allows the study of the different stages of the algorithm, especially in terms of cancellations. However in our case, the final results do not show cancellation errors exceeding the desired minimal error.

The limitations of the validity of the expansions when $(\theta, \varphi)$ is in a neighborhood of the critical value, are clearly apparent in the values of $|G_{2k}^{(2k)}|/k! (2N)^k$, under the form of a volcano cone shape. This is illustrated in Figures 6-6 and 6-7.
Figure 6-6: Values of $|G_{zz}^{(2k)}|/k!(2N)^k$ for $k = 0$ and $k = 3$, at $R = 16$
The values are scaled by $10$ and $10^3$ respectively.

Figure 6-7: Values of $|G_{zz}^{(2k)}|/k!(2N)^k$ for $k = 0$ and $k = 5$, at $R = 16$
The values are scaled by $10$ and $2 \times 10^4$ respectively.
The radii of the ridges are increasing with the order \( k \) as a consequence of applying the truncation test: as terms get larger, they are not incorporated in the sum, hence the value 0 is attributed. Each coefficient inherits its features from the terms \( z_{2i}^{(n)} \) it involves. We therefore find the same characteristics, modulated of course by the factors \( 1/k! (2N)^k \).

The domain in \((\theta, \varphi)\) we have chosen for displaying these results does not necessarily correspond to the actual domain in which the expansions will be used. This remark is particularly valid for the vicinity of the fold catastrophe, where the practical boundary can only be assessed by numerical comparison with the fold catastrophe expansion.

2. Region of validity

Once the numerical values of the expansions are available, the outer boundary of \( \bar{O}_1 \cup \bar{O}_2 \) has to be determined with respect to a certain error \( \epsilon \). This determination is performed by comparing the results from the isolated col expansions, with the already existing expressions for the potential. Using the notations of Clarisse (1989), we denote:

— (CS1), the convergent Neumann series valid in a region delimited by the surface \( M \leq 21 \) for \( X \leq -16 \), with \( M = X^2/4\rho \)

— (AS1), the Bessho-Ursell asymptotic series, complementary of (CS1) for \( -16 \leq X \leq -1 \). Its domain of validity for large \(|X|\) can be assessed to be of the form \( \rho < 1 \), but necessitates an increasing number of terms (of the order of \( M \)).

and in addition:

— (SD1), the steepest descent expansion given by (6.10), valid in the region \( \bar{O}_1 \cup \bar{O}_2 \), which is to be determined.
The constraint on the region of validity of (AS1) for large $|X|$ has a symmetry of revolution about the $X$-axis. The boundary of $\tilde{O}_1 \cup \tilde{O}_2$ is also expected to present a symmetry of revolution in this region. The traces of the intersection between the two domains, on portions of spheres of constant $R$ give the results of Table 6.1 in terms of $R$ and $\theta$, for $\varphi$ in $[0^\circ, 90^\circ]$.

<table>
<thead>
<tr>
<th>$R$</th>
<th>$\theta$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>$[6^\circ, 6.5^\circ]$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>20</td>
<td>$[4^\circ, 7.8^\circ]$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>30</td>
<td>$[4^\circ, 5.8^\circ]$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>40</td>
<td>$[1^\circ, 4^\circ]$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>100</td>
<td>$[0.5^\circ, 1.8^\circ]$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>1000</td>
<td>$[0.1^\circ, 0.2^\circ]$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
</tbody>
</table>

Table 6.1: Explorations of the intersection between the domains of validity of (AS1) and (SD1)

These explorations are not maximal in any respect, but guarantees the agreement of these expressions within $\epsilon$, over a connected region.

The absolute error within which the numerical values compare is more than satisfactory, especially since the magnitude of $I^+$ can be of order 1 in these regions. It should be noted that comparisons are performed in terms of $I^+$, i.e. without including the factor $\exp[-Z/2]$ of (6.2). The asymptotic properties of (SD1) appears to be extremely good with respect to the values of $\rho$ which stay near 1. It must be noted that the task for computing (AS1) at $R = 1000$ is extremely heavy, and that another asymptotic expansion is desirable to complement (AS1) and (SD1) in a sector of small angle $\theta$. Such expansion may be obtained by following the treatment of the singular axis given in Ursell (1960).
The comparison with (CS1) is simpler in terms of geometry. The surfaces at constant $M$ are of parabolic sections in any plane of constant $\varphi$. Therefore for large values of $|X|$, the region of validity of (CS1) corresponds to a domain where the magnitude of $I^+$ is well below the required error tolerance. Consequently the intersection between the two domains of validity is of finite extent in $R$. For large values of $R$, the boundary of $\tilde{O}_1 \cup \tilde{O}_2$ can be determined by fixing a lower bound on $|I^+|$. Such a definition produces a surface having the cone of angle $\varphi_0$, as asymptote when $R \to +\infty$.

As above, the region of agreement can be probed by intersecting it with portions of spheres. Such results are given in Table 6-2. The higher error observed for $R = 16$, indicates that (SD1) may not be used in this region (for the values of $\theta$ given above).

<table>
<thead>
<tr>
<th>$R$</th>
<th>$\theta$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>$[38^\circ, 40^\circ]$</td>
<td>$&lt; 10^{-6}$</td>
</tr>
<tr>
<td>20</td>
<td>$[38^\circ, 40^\circ]$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>30</td>
<td>$[33^\circ, 40^\circ]$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>40</td>
<td>$[55^\circ, 60^\circ]$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>100</td>
<td>$32.5^\circ$</td>
<td>$</td>
</tr>
<tr>
<td>1000</td>
<td>$21.6^\circ$</td>
<td>$</td>
</tr>
</tbody>
</table>

Table 6.2: Explorations of the intersection between the domains of validity of (CS1) and (SD1)

We emphasized, in the presentation of the algorithms (Chapters 4 and 5), the importance of a uniform numerical approximation of an integral. Here, in a sense, this uniformity is not sufficient. The constraint of providing numerical results in the complementary region of validity of (AS1) and (CS1), requires us to define domains,
\( \tilde{\Omega}_1 \cup \tilde{\Omega}_1 \), whose shapes depend on \( R \) (and therefore on \( N \)), at least for our choice of the parameters. A specific choice of the parameters might allow the definition of a fixed region of the \( \tilde{\sigma} \)-space where uniformity is guaranteed. By example, \((\rho, \varphi)\) seems to describe appropriately the boundary with \((AS1)\), however this choice is not suitable for the boundary with \((CS1)\). In addition an asymptotic expansion whose region of validity would fill a cylindrical wedge between the regions of validity of \((AS1)\) and \((SD1)\), appears to be necessary. This would allow the consideration of \((SD1)\) in a bounded region in \( \theta \), a feature of great computational interest as we will later see. The definition of \( \tilde{\Omega}_1 \cup \tilde{\Omega}_1 \) would be incomplete without the comparison between \((CS1)\) and the expansion for the fold catastrophe.

6.5 The fold catastrophe

The algorithm for the uniform numerical approximation of \( I^+ \) near its fold catastrophe is not readily provided by our analysis of Chapter 5. However, we claimed that, upon following the guidelines of Section 5.6, we could build such an algorithm. It is now appropriate to demonstrate how these directives actually result in the corresponding algorithm for \( I^+ \).

6.5.1 The building of the algorithm

This is the most arduous part of the treatment of \( I^+ \). However this particular integral does present some features which facilitate the task. This is especially true for the inversion of the Levinson's transformation which is at the foundation of the procedure leading to a fold catastrophe expansion.
1. Inversion of the Levinson's transformation

For the building of the corresponding algorithm, we follow exactly the steps (i), (ii) and (iii) of Section 5.6.1. First, it is necessary to parameterize \( f(z, \bar{\alpha}) \) into \( f(z, \bar{\beta}) \) and thus define the new local parameter \( \bar{\beta} = (\beta_1, \beta_2) \) as follows:

\[
\begin{align*}
\beta_1 &= \left[ 2 \left( \alpha_2 \alpha_1 - a_0 + \frac{\alpha_2 \alpha_1^{-1} - \bar{a}_0}{z_0^2} \right) \right]^{1/2} \\
\beta_2 &= \alpha_2 \alpha_1^{-1} - \bar{a}_0 \\
\bar{a}_0 &= 1 + i = \alpha_{z_0} \alpha_{1_0}
\end{align*}
\]  

(6.11)

and the derivative of \( f \) in a neighborhood of \((z_0, \bar{0})\) is then in the form:

\[
f'(z, \bar{\beta}) = -\beta_1^2 + \frac{4\beta_2}{z_0^3} (z-z_0) - \frac{6}{z_0^2} (\beta_2 + \bar{a}_0) z_0 - 2 \] \((z-z_0)^2 + (z-z_0)^3 \phi(z, \bar{\beta}) \)  

(6.12)

which, we note, is different from the form of (B.8).

(i) The definition of \( \zeta^{1/2} \)

The process of defining \( \zeta^{1/2} \) is greatly simplified by the fact that there exists a choice of the branch of (5.24) which is uniform over \( D' \). Hence \( \arg(\zeta^{1/2}) \) is chosen to be in \([7\pi/6, 5\pi/3]\), and its uniformity is illustrated, for \((\theta, \varphi)\) in \([8^\circ, 40^\circ] \times [40^\circ, 90^\circ]\), in Figure 6-8. The uniformity of this choice does not require the determination of \( \beta_{1_{ref}} \).

However when \( \bar{\beta} \) is allowed to describe a whole neighborhood of \( \bar{0} \) as in the generation of the set of points of Appendix D, \( \beta_{1_{ref}} \) is determined by imposing \( \arg(\zeta^{1/2}_{ref}) = 7\pi/6 \).

As a consequence of the choice for the argument of \( \zeta^{1/2} \), the path \( C_2 \) is to be used in the definition of \( \mathcal{F}_r \) and \( \mathcal{G}_r \) (see Section 5.2.1).

(ii) Choice of the branch of the transformation

Selecting the right branch of the transformation is done in two parts. We first compute the double entire series \((\beta_1, \beta_2)\) for \( z_i^{(n)} \). The relations between \( \bar{\beta} \) and \( \bar{a} \) are simple enough so that the set of points characterized in Appendix D is easily
generated. Then the Fast Fourier Transform of the values of $z_{1}^{(n)}$ given by the formulas of (5.27) are performed while $\overline{\beta}$ describes this set. Criterion 1 of Section 5.4.2 is used to determine the right branch. In practice, radii of 1, and 32 points were taken in both variables, defining a set of $2^{10}$ points. This allowed us to obtain the first 31 coefficients of the Fourier series for each $z_{1}^{(n)}$, the coefficients for $z_{2}^{(n)}$ being obtained by replacing $\beta_{1}$ by $-\beta_{1}$.

Under the definition of $\overline{\beta}$ as given by (6.11), the Fast Fourier Transforms contain positive and negative components. This is explained by the non generic form of (6.12): the definition of $\beta_{1}$ is not the rigorous one. Exact expansions of $z_{1}^{(n)}$ as obtained from symbolic computations, contain negative powers of $\beta_{1}$ and $\beta_{2}$, confirming the results of the Fourier transform. By keeping only the positive components of the Fourier transform, we obtain the coefficients of double entire series which are approximating the exact series. This approximation is justified by practical reasons: thus the value of $z_{0}(\alpha)$ (see Appendix B) does not need to be computed. We shall see that its consequences on the numerical results are marginal. The forms of the series as given in Appendix B are recovered. The numerical values of the coefficients thus obtained are
particularly stable with respect to the radii $|\beta_1|$ and $|\beta_2|$: double precision results are unchanged for the $|\beta_i|$'s varying in $[0.1, 1]$. The choice of $|\beta_1| = |\beta_2| = 1$, corresponds to values of $|\zeta^{1/2}|$ of order 1. For this regime the direct algorithm is numerically well behaved and the influence of cancellation errors are thus rendered negligible. The consistency of the results for various values of the radii, confirms that the Levinson's transformation is valid for the set of points considered.

The second part of this process consists in extending this choice to the portion of the computational domain $\tilde{D}_0^3$ where the direct algorithm is used. Here again a single choice of the branch of (5.33) is uniform in $D'$. This is confirmed in Figure 6-9 where $f_i^{(3)} z_i^{(1)^3}$, $i = 1, 2$ is shown to describe a connected, with respect to some discretization, domain in a univalent manner, as $(\theta, \varphi)$ is in $[80^\circ, 400^\circ] \times [400^\circ, 900^\circ]$.

![Figure 6-9: $f_1^{(3)} z_1^{(1)^3}$ and $f_2^{(3)} z_2^{(1)^3}$, in $8^\circ \leq \theta \leq 40^\circ$, $40^\circ \leq \varphi \leq 90^\circ$](image)

In both graphs, $\times$ corresponds to the values of $f_1^{(3)} z_1^{(1)^3}$ [left] and $f_2^{(3)} z_2^{(1)^3}$ [right] at $\zeta = 0$.

(iii) Determination of the radii $\tau_n$

We apply the rules relative to (iii) of Section 5.6.1 to our case of $I^+$. For this purpose the first 13 derivatives of $z$ are computed. This maximum value for the order

---

208
of the derivatives was fixed \emph{a priori}. The determination of the radii $\tau_n$ is done by comparing the results from the entire series with the results from (5.27). The entire series are truncated at the order 10 in terms of $(j_1 + 2j_2)$, where $J = (j_1, j_2)$ is the power multi-index of $\vec{\beta}$. Thus we obtain the results of Table 6.3 in terms of $n$, $\tau_n$ and the relative error $\epsilon_n$. The deterioration of $\epsilon_n$ with $n$ is due to the increasing influence of cancellation errors with $n$, for the small values of $|\zeta^{1/2}|$. No other remedies are available, except for the use of extended mantissas in the computation of $z_i^{(n)}$ for small $|\zeta^{1/2}|$. The relatively high errors for $n \leq 3$ are explained by the error made on $\beta_1$ and $\beta_2$ by using (6.12) instead of the exact form of (B.8). This error is also present for larger values of $n$ as the increase in $|\zeta^{1/2}|$ accentuates the difference between the exact and the approximated solution. The regions defined by the various radii are confined to a small neighborhood of $(\theta, \varphi)$, typically inside the rectangular domain $[18.7^\circ, 20.3^\circ] \times [87^\circ, 90^\circ]$. We will have a full understanding of the consequences of such approximations when analyzing the final results.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\tau_n$</th>
<th>$\epsilon_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$3.8 \times 10^{-2}$</td>
<td>$&lt; 10^{-6}$</td>
</tr>
<tr>
<td>2</td>
<td>$6 \times 10^{-2}$</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>3</td>
<td>$6 \times 10^{-2}$</td>
<td>$&lt; 10^{-6}$</td>
</tr>
<tr>
<td>4</td>
<td>$6.16 \times 10^{-2}$</td>
<td>$&lt; 10^{-5}$</td>
</tr>
<tr>
<td>5</td>
<td>$1.15 \times 10^{-1}$</td>
<td>$&lt; 10^{-5}$</td>
</tr>
<tr>
<td>6</td>
<td>$1.15 \times 10^{-1}$</td>
<td>$&lt; 10^{-5}$</td>
</tr>
<tr>
<td>7</td>
<td>$1.15 \times 10^{-1}$</td>
<td>$&lt; 10^{-4}$</td>
</tr>
<tr>
<td>8</td>
<td>$1.7 \times 10^{-1}$</td>
<td>$&lt; 10^{-4}$</td>
</tr>
<tr>
<td>9</td>
<td>$1.65 \times 10^{-1}$</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>10</td>
<td>$2.05 \times 10^{-1}$</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>11</td>
<td>$2.5 \times 10^{-1}$</td>
<td>$&lt; 10^{-3}$</td>
</tr>
<tr>
<td>12</td>
<td>$2.5 \times 10^{-1}$</td>
<td>$&lt; 10^{-2}$</td>
</tr>
<tr>
<td>13</td>
<td>$2.5 \times 10^{-1}$</td>
<td>$&lt; 10^{-2}$</td>
</tr>
</tbody>
</table>

Table 6.3: Determination of $(\tau_n)$

The algorithm for the computation of $G_i^{(k)}$ is available in a general form and no additional work is necessary. This allows us to consider the last \emph{composite algorithm}, or the one concerned with the sequences $(p_r)$ and $(q_r)$.
2. Computation of the sequences \((p_r),(q_r)\)

Following Section 5.6.3, the central task is the determination of the radii \(\sigma_{p_r}\) and \(\sigma_{q_r}\). We here face a constraining limitation in the accuracy of the power series imposed by the choice of the maximum value of \(n\). If the sequences \((p_r)\) and \((q_r)\) are available up to \(r = 13\) by using the direct method, because of the maximum value of \(n\) we chose, i.e. \(n = 13\), \((p_r)\) and \((q_r)\) are limited to the first 7 and 6 terms, respectively, when using the power series method. Moreover these values are obtained with increasing error due to the forced truncation of the series. This situation is illustrated in Table 6.4, in terms of \(r\) and of the order of the first neglected term: \(\zeta^{K_{p_r}}\) or \(\zeta^{K_{q_r}}\).

![Table 6.4: Truncations of the power series for \((p_r)\) and \((q_r)\)](image)

For higher values of \(r\), no complementary expressions in power series are available since \(G_i^{(k)}\) is available only for \(k \leq 13\). The values of the radii \(\sigma_{p_r}\) and \(\sigma_{q_r}\) have then to be defined as additional boundaries of the domain of computation for each term of the sequences. Their choice is not well defined as no comparison is possible with other results. They should be chosen small enough in order to obtain the maximum number of terms in each sequence, but large enough to avoid the introduction of spurious values in the asymptotic series. It is advisable not to include them in the final algorithm although the corresponding radii could be chosen in a suitable manner (see Appendix E). Hence we obtain the determinations of \((\sigma_{p_r})\) and \((\sigma_{q_r})\) given in Table 6.5.
Table 6.5: Determination of the radii \((\sigma_{p_r})\) and \((\sigma_{q_r})\)

The comparison of the absolute errors with the orders of magnitude of the terms of the series indicates a strong correlation between the loss in accuracy and the lowering of the order of truncation in the power series. We note that for both series the radii stay constant until \(r = 4\), which corresponds to a truncated term in \(O(\zeta^2)\). For such values of the radii, the order of this term can be assessed to be \(O(10^{-8})\). Hence the loss in accuracy of the last terms are directly related to the truncation of the series.

The influence of the errors made in the approximation of \(z_i^{(n)}\), could be significant at this stage. However they appear to be negligible in comparison with the truncation errors. At this stage, we can state that the uniform computation of both sequences \((p_r)\) and \((q_r)\) is ensured throughout the computational domain with an absolute accuracy, as a function of \(r\), given by the values of \(\epsilon\) of Table 6.5. The consequences of these errors on the final result depend then entirely on the properties of the recursion relations of (5.37) in terms of error propagation.

4. Computation of the series \((A_s), (B_s), (C_s), (D_s)\)

The recursion relations in the form of (5.37) provide the corresponding algorithm. Consequently to the limitation of the order at which \(p_r\) and \(q_r\) can be computed, we obtain the numerical values of the coefficients of the series only for the following orders:

- \(A_s\) for \(0 \leq s \leq 3\)
- \( B_s \) for \( 0 \leq s \leq 2 \\
- \( C_s \) for \( 0 \leq s \leq 2 \\
- \( D_s \) for \( 0 \leq s \leq 3 \\

By grouping the terms of the four series in two sums we get:

\[
I^+ \approx \exp[N f(z_1, \tilde{\alpha})] \left[ \frac{\tilde{A}_i(N^{2/3} e^{2i\pi/3} \zeta)}{N^{1/3}} \sum_{s \leq 7} \frac{a_s(\tilde{\alpha})}{N^s} \right. \\
+ \left. \frac{\tilde{A}_i'(N^{2/3} e^{2i\pi/3} \zeta)}{N^{2/3}} \sum_{s \leq 5} \frac{b_s(\tilde{\alpha})}{N^s} \right] 
\]

(6.13)

where \( a_s \) and \( b_s \) are given by the relations:

\[
a_{2s} = A_s \\
a_{2s+1} = D_s \\
b_{2s} = C_s \\
b_{2s+1} = B_s 
\]

(6.14)

and where \( \tilde{A}_i \) and \( \tilde{A}_i' \) are the scaled values of the Airy functions. The factor \( e^{2i\pi/3} \) in the argument of the Airy functions is due to the use of \( C_2 \) in the definition of \( F_r \) and \( G_r \).

The algorithm, once built, allows the computation of the terms in the sums for any values of \((\theta, \varphi)\) where this expansion is valid. The variations of the moduli of these quantities as functions of \((\theta, \varphi)\) for a constant \( N \) are shown in Figures 6-10, 6-11, 6-12 and 6-13, for the computational domain \( 8^\circ \leq \theta \leq 40^\circ, 40^\circ \leq \varphi \leq 90^\circ \) in \((\theta, \varphi)\). We choose to compute these values for \( R = 16 \) which corresponds to the lowest value of \( R \) for which the fold catastrophe expansion will be needed. Graphs for greater values of \( R \) are simply obtained by an appropriate scaling in \( N \). The magnitudes of the moduli are scaled in order to magnify their variations. In all figures the scale of the vertical coordinate is the same as those for the horizontal coordinates. These surfaces were obtained by computing 100 equidistant points in each variable.
Figure 6-10: $|A_s|/N^{2s}$, $s \leq 3$, at $R = 16$ for $8^\circ \leq \theta \leq 40^\circ$, $40^\circ \leq \varphi \leq 90^\circ$

In this graph and the following, $(\theta_0, \varphi_0)$ is signaled by a vertical segment on the $\theta$ axis.

The topographies of all these functions appear to be smooth everywhere except for $A_3$ and $D_3$ in the vicinity of $(\theta_0, \varphi_0)$. Otherwise, some common trends can be observed. All functions are slowly increasing (with the exception of $D_3$) in $\varphi$. The most pronounced variations are along the $\theta$ direction, and, except when localized near $(\theta_0, \varphi_0)$, are present qualitatively for all values of $\varphi$. The most important feature of these functions is that they are slowly-varying: the real and imaginary parts of these quantities present the same features. This contrasts with the behavior of $I^+$ which is known to present oscillations for $\theta$ smaller than or near $\theta_0$.

One can observe that the coefficients are decreasing in magnitude as $s$ increases. As the recursion relations only induce decreasing sequences in sets of four terms, this
property is inherited from the values of $p_r$ and $q_r$, that is from the convergence of the expansion of $G(u, \tilde{\alpha})$ in the form of (5.14).

In order to perceive the limitations of the asymptotic properties of these series, further terms have to be considered (see Appendix E). Additional terms improve the accuracy of the series especially for values of $(\theta, \varphi)$ away from $(\theta_0, \varphi_0)$. From the graphs of Appendix E, it appears that the application of a truncation test will not result in a truncation of any of the series uniformly in $(\theta, \varphi)$. Nevertheless for all terms, we observe an increase in magnitude as $\theta$ and $\varphi$ are moving further away from the critical values. This phenomenon characterizes the limitations of the validity of the expansion in $(\theta, \varphi)$, or the limited validity of the Levinson's transformation – i.e. of the unfolding of the fold catastrophe.
6.5.2 Numerical characteristics

The reliability of the algorithm for the computation of the coefficients depends on the control and the consequences of the errors introduced at the different stages of the overall procedure. We are now able to discern, in this example, the influences of the various approximations made principally during the inversion of the Levinson’s transformation and the computation of the sequences \((p_r)\) and \((q_r)\).

Regarding the inversion of the Levinson’s transformation, it appears that the approximation made when using the definition of the new parameter \(\tilde{\beta}\) given by (6.11), is of no perceivable consequence in subsequent steps of the computation. In a similar manner, the tolerances within which the results from the multiple entire
series and the direct algorithm compare are of consequences well below the minimal accuracy of six significant digits. These assertions can be verified by inspecting the various quantities $z_i^{(n)}$ and $G_i^{(k)}$ in a close neighborhood of the critical value $(\theta_0, \varphi_0)$. This procedure is also useful for the understanding of the disturbed regions in the graphs of $A_3$ (Figure 6-10) and $D_3$ (Figure 6-13).

Hence by computing values of $|A_3|/N^6$ and $|D_3|/N^6$ at $100 \times 100$ points in the rectangular domain $[18.7^\circ, 20.3^\circ] \times [87^\circ, 90^\circ]$, we obtain Figures 6-14 and 6-15. From these graphs, three main features are noticeable:

1. the inclined planar ($A_3$) or parabolic ($D_3$) surface centered at $(\theta_0, \varphi_0)$ of approximately elliptic projection on the $(\theta, \varphi)$-plane. This surface is given by the truncated power series for $p_6$ and $q_5$ as shown on the right parts of the figures.
Figure 6-14: $|A_3|/N^6$ and $|p_6|$, at $R = 16$ for $18.7^\circ \leq \theta \leq 20.3^\circ$, $87^\circ \leq \varphi \leq 90^\circ$

Figure 6-15: $|D_3|/N^6$ and $|q_5|$, at $R = 16$ for $18.7^\circ \leq \theta \leq 20.3^\circ$, $87^\circ \leq \varphi \leq 90^\circ$
The distinction between the two kind of surfaces is consistent with the order of truncation of the series.

2. the boundary region of the rectangular domain away from \( \theta_0 \), nearly flat and undisturbed. This corresponds to the results from the direct formulas for \( p_6 \) and \( q_5 \) in a regime where these are still well behaved numerically.

3. the disturbed region at the boundaries of \( |\zeta^{1/2}| = \sigma_{p_6} \) and \( |\zeta^{1/2}| = \sigma_{q_5} \), respectively, characterized by many spikes and oscillations. This reflects the breakdown of the direct formulas as \( |\zeta^{1/2}| \) gets smaller. Spikes are much more pronounced in the graph of \( |p_6| \) due to the additional power of \( \zeta^{1/2} \) in the denominator compared to the formula for \( q_5 \).

The limitations of the direct algorithm are clearly visible. Undoubtedly, the power series truncated at order 1 and 2 in \( \zeta^{1/2} \) provide poor approximants of the functions \( p_6(\theta, \varphi) \) and \( q_5(\theta, \varphi) \).

However, the consideration of more complete series might not furnish the necessary increase in \( \sigma \) which would entirely cover the regions of ill behavior of the direct formulas. This can be seen in the increase in \( \sigma \) necessary for the elimination of numerical errors in the values of higher order terms as shown in Figures E-1, E-2, E-3 and E-4 of Appendix E. Hence the limitations of the direct algorithms in double precision have been reached. The search for higher accuracy would necessitate the use of extended mantissas.

The behavior of these functions in the outer regions suggest that interpolation methods might be more suitable than low order power series. However the results for lower order coefficients \( (s \leq 2) \) are totally satisfactory, and such considerations are not necessary. It must be emphasized that such errors are bounded by \( 3 \times 10^{-6} \) for \( R = 16 \), for \( s \leq 3 \). Consequently higher minimum values of \( R \) will have to be used in order to guarantee higher accuracies.
6.5.3 Domain of validity

Finally we need to compare the result from this last expansion, or (SD2), with the results from other expressions, namely (AS1), (CS1) and (SD1). The region of validity of (SD2) is such that simultaneous comparisons with (AS1), (CS1) and (SD1) have to be performed. The localized nature of the critical value at \((\theta_0, \varphi_0)\) requires us to adapt the parameterization of \(I^+\). Hence, instead of presenting the results in terms of \((\theta, \varphi)\), we use \(|\zeta^{1/2}|\), where it is understood that comparisons are achieved within \(\epsilon\) over the whole range of \(\arg \zeta^{1/2}\), i.e. \([7\pi/6, 5\pi/3]\).

We first assess the validity of (SD2) at \(R = 16\) by comparison with (AS1) and (CS1) which are complementary. For \((\theta, \varphi)\) in \([8^\circ, 30^\circ] \times [30^\circ, 90^\circ]\), an error smaller than \(10^{-5}\) is ensured everywhere, while errors smaller than \(10^{-6}\) are achieved in the range \([30^\circ, 40^\circ]\). We pursue this investigation by including the results from (SD1) in a region defined by: \(\theta \geq \theta_{\text{min}}\) as given from Table 6.1, \(|\zeta^{1/2}| > 5 \times 10^{-1}\), \(\theta \leq \theta_{\text{max}}\) as given by Table 6.2 and \(\varphi \geq 40^\circ\). We know that for values of \(R\) greater than 16, the region covered by (AS1), (SD1) and (CS1) is connected and ensures a maximum error of \(10^{-6}\) \((10^{-7} \text{ for } R \geq 20)\). Thus by considering (AS1) - (SD1) - (CS1) as a composite expression, the task is reduced to the comparison with (SD2). (The arrangement of the various domains is sketched in Figure 6.16.)

Similarly to Section 6.4 we obtain the results of Table 6.4 by intersection of these regions with portions of spheres of various radii. The convergence properties of (SD2), similarly to (SD1), are surprisingly good as \(N\) varies from 2 to 10 at \(R = 16\). The overall error bound of \(10^{-5}\) indicates a limitation in the accuracy of (SD2) for lower values of \(N\). This tendency is confirmed by the improvement of the accuracy for larger values of \(\theta\), i.e. larger values of \(N\). From this exploration, it appears that the volume of the region of validity of (SD2) is significant. This observation is consistent with the results of Ursell (1965).

These properties of accuracy and the features of the coefficients of the expansion leads to further remarks on the numerical and mathematical characteristics of such
Figure 6-16: Arrangement of the regions of validity of (AS1), (SD1), (SD2) and (CS1) in terms of $\theta$ and $\varphi$

| $R$  | $|\zeta^{1/2}|$       | $\epsilon$  |
|------|----------------------|-------------|
| 20   | [0.5, 0.8]           | $< 10^{-5}$ |
| 30   | [1.2, 1.25]          | $< 10^{-6}$ |
| 40   | [1., 1.25]           | $< 10^{-6}$ |
| 100  | [0.75, 1.25]         | $< 10^{-7}$ |
| 1000 | [0.5, 0.8]           | $< 10^{-7}$ |

Table 6.6: Explorations of the intersection between the domains of validity of (AS1) - (SD1) - (CS1) and (SD2)
6.5.4 Fundamental properties of catastrophe expansions

We emphasized in Section 6.5.2 the slowly varying behavior of the coefficients of the expansion (6.13) as functions of $(\theta, \varphi)$. We now describe the decomposition which is implied in writing (6.13). We illustrate our assertions by considering $I^+$ for $\varphi = 90^\circ$ which corresponds to the case where both the source and the field-point are on the free-surface. The computational domain chosen is a sector $8^\circ \leq \theta \leq 30^\circ$ and $16 \leq R \leq 40$. We know then that the accuracy of the results given by (6.13), is greater than 5 significant digits. The potential $w$ appears as in Figure 6-17.

Let us decompose (6.13) in two sums: one associated with the Airy function (Figure 6-18), the other associated with its derivative (Figure 6-19). This second component is of weaker contribution than the first one, and consequently we will only consider the latter. However the decomposition we carry on is identical for both components.
Figure 6-18: The component in $A_i$ of $w$ at $Z = 0$ for $8^\circ \leq \theta \leq 30^\circ$ and $16 \leq R \leq 40$

Figure 6-19: The component in $A_{i'}$ of $w$ at $Z = 0$ for $8^\circ \leq \theta \leq 30^\circ$ and $16 \leq R \leq 40$
We now break down the component in \( \text{Ai} \) in two factors: one involving the sum \( \sum_{s \leq r} a_s / N' \) (Figure 6-20), the other the product \( \exp[N f(z_1, \bar{a})] \hat{\text{Ai}} \) (Figure 6-21).

Figure 6-20: The real part of \( \sum_{s \leq r} a_s (\theta, \varphi) / N' \) at \( \varphi = 90^\circ \) for \( 8^\circ \leq \theta \leq 30^\circ \) and \( 16 \leq R \leq 40 \).

No variations are visible here.

Figure 6-21: The real part of \( \exp[N f(z_1, \theta, \varphi)] \hat{\text{Ai}}(N^{2/3} e^{2i\pi/3} \zeta) \) at \( \varphi = 90^\circ \) for \( 8^\circ \leq \theta \leq 30^\circ \) and \( 16 \leq R \leq 40 \).
It is then obvious that all of the oscillatory behavior of this component is contained in the term \( \exp[N f(z_1, \alpha)] \) \( \frac{\partial}{\partial \alpha} \). We go further in isolating the scaled Airy function (Figure 6-22) from its factor (Figure 6-23).

Figure 6-22: The real part of \( \frac{\partial}{\partial \alpha} (N^{2/3} e^{2i\pi/3} \zeta) \) for \( \varphi = 90^\circ \), \( 8^\circ \leq \theta \leq 30^\circ \) and \( 16 \leq R \leq 40 \)

Figure 6-23: The real part of \( \exp[N f(z_1, \theta, \varphi)] \) for \( \varphi = 90^\circ \), \( 8^\circ \leq \theta \leq 30^\circ \) and \( 16 \leq R \leq 40 \)
The purpose of this decomposition appears clearly: the "transverse wave" part is entirely contained in the factor \( \exp[N f(z_1, \theta, \varphi)] \) while the "diverging wave" part is embedded in \( \overline{A_i}(N^{2/3} e^{2ix/3} \zeta) \). The complex interferences observed on the boundary and inside the Kelvin wedge is thus accounted for by the *modulation* of these two components and not anymore by their *sum*. This feature is similar to the transformation of a sum of sine functions into a product. But more profound is the total decomposition of the qualitative and quantitative properties of \( w \) onto the Airy function, its derivative and the factor \( \exp[N A(\bar{\alpha})] \), with the series \( \sum a_i/N \) and \( \sum b_i/N \) playing the role of coordinates.

**A universal property**

Perhaps for the first time are we able to perceive in its entirety the power a catastrophe expansion. This power is mirrored in the features displayed by our example of the Kelvin potential, namely:

- a significant region of validity in terms of the parameter \( \bar{\alpha} \) (or \( \theta \) and \( \varphi \)) characteristic of the uniformity of the expansion,

- a *complete* decomposition of a function (our integral \( I^+ \)) on a set functions (here \( \exp[N A(\bar{\alpha})] Ai \) and \( \exp[N A(\bar{\alpha})] Ai' \)) which contains all the oscillatory information, leaving slowly varying (as functions of \( N \) and \( \bar{\alpha} \)) factors. (We use here the term *complete* in the sense "up to negligible terms").

These two characteristics are of course embedded in the procedure used to obtain the expansion: the cubic transformation and the sequences \( (p_\tau), (q_\tau) \) of Chester *et al.* However this decomposition is not unique: the corresponding Bleistein sequence would yield similar results in terms of validity and *decomposition property*.

This power is in no way restricted to fold catastrophe expansions but is common to *any* catastrophe expansion. The generality of this feature is inherited from the *universality* of the classification of catastrophes and their unfoldings, and the generality
of the Bleistein procedure. This remarkable characteristic has been acknowledged in quantum mechanics for the past 20 years at the theoretical level. However until now, the sole application of catastrophe expansions to hydrodynamics is the 1960 article of Ursell... Our numerical results not only confirm the validity of such approach but also open new prospects for numerical applications.

A powerful numerical tool

We saw the promising properties of accuracy and robustness permitted by generalized steepest descent expansions. The use of the method of steepest descents, or rather of properties of holomorphic functions, is here crucial in its ability to reduce a global behavior to local contributions. We have demonstrated through the treatment of the Kelvin potential, that the algorithm devised in Chapter 4 for an ordinary steepest descent expansion conveys this accuracy and this robustness to numerical evaluations. More importantly and for the first time, we have shown that by following the guidelines of Section 5.6, a robust algorithm can be built which preserves the accuracy and uniformity of a fold catastrophe expansion. However in numerical applications these qualities are paid by a low efficiency: typically 10 times lower than the ones permitted by series like (AS1) or (CS1). (It must be noted that the latter are of the order of a 100 times faster than a direct numerical integration of \( w \) to the same accuracy.)

In fact what we called the decomposition property of a catastrophe expansion provides us with means of improving this low efficiency. The major task in evaluating a fold catastrophe expansion is spent in computing the coefficients \( a_s(\bar{\alpha}) \), \( b_s(\bar{\alpha}) \). Moreover the accuracy of the numerical values of these coefficients worsens as \( s \) increases, requiring extended mantissas if higher order terms were needed. But since these coefficients are slowly varying functions over bounded regions in \( \bar{\alpha} \), we could replace their costly evaluations by some appropriate approximations. Then the task would be divided into two parts:

1. high accuracy computations of the coefficients in the bounded region in \( \bar{\alpha} \), using
an algorithm built along the guidelines of Section 5.6 regardless of its efficiency,

2. building of appropriate approximations of the functions $a_*(\alpha), b_*(\alpha)$ in this domain.

Finally the real time evaluation would consist in computing the Airy function, its derivative and the coefficients using these approximations. Gains in efficiency of a factor of 10 could be expected.

There exists no theoretical limitation in applying such scheme to other catastrophes or even ordinary steepest descent expansions. However, we have seen the practical limitations of the procedures used for the fold catastrophe with respect to their extensions to non-isolated catastrophes. It is only through future applications that the practicality of such extensions will be assessed.

Before concluding this presentation of the Kelvin potential we propose a "new" physical interpretation for the interferences present in Kelvin's ship wave pattern.

6.6 Some results on Kelvin's ship wave pattern

The obvious consequence of the algorithms for the computation of (SD1) and (SD2) is to provide complementary expressions to those already existing: (AS1), (CS1) and (AS3). Rigorously we cannot state that the physical domain is treated in its entirety since we noted that an expression is missing to fill the region between (AS1) and (SD1) for large $R$ (typically a finite sector of mean $\theta$ between 1° and 3°). Values of $R$ up to 1000 are however uniformly covered but at a high price in the computation of (AS1). The absolute accuracy provided by the current expansions is of at least 5 significant digits in the quarter of sphere $R \leq 1000$, $X \leq 0$, and higher (6, 7 digits and higher) away from the region $R \approx 16$. 

227
1. The "classical" wave analysis

The form of the isolated col expansion (6.10) is a typical example of an ordinary steepest descent analysis. It is then usual to attribute to each saddle point a part of the wave system inside the wedge $|\theta| < \theta_0$. Hence the transverse wave system is identified to $z_1$ while the diverging waves are associated to $z_2$. In such analysis the graphs of Figure 6-2 and 6-3 have a particular significance: $\Re\{w_i\}$ and $\Im\{w_i\}$ give respectively the growth rate and the frequency of the corresponding wave systems for large $N$'s. The decomposition of the wave pattern as a sum of diverging and transverse wave systems is classical in hydrodynamics. Contour lines of such a decomposition at leading order for the case of the free-surface elevation, when both the field point and the source point are on the free surface, can be found in Euvrard (1983, Fig. VIII.1, after A. Gamst).

In fact the main information about this expansion, concerns its region of validity: (SD1) is characterized by a high accuracy despite the low values of $N = \rho$, typically as small as 1, at which it might be used (see Table 6.1). More interest lies in the study and interpretation of the wave field closer to the critical line $\theta = \theta_0$.

2. A "non-classical" analysis

We first note that the region of validity of (SD2) covers an important part of the wave field behind the disturbance: more than half of the Kelvin wedge is accurately described by (SD2). The numerical values for the leading order given in Ursell (1960) for $10^\circ \leq \theta \leq 20^\circ$, as well as the statement in Ursell (1965) about the extensions of the validity of (6.13) to larger regions, were totally a propos. Contrary to the usual opinion among hydrodynamicists, the expansion in Airy functions is valid in a much wider region than the near vicinity of the critical line, and this for values of the radial distance $R$ as small as 16. This is clearly visible in the portion of sector $8^\circ \leq \theta \leq 23^\circ$, $95 \leq R \leq 135$, shown in Figure 6.24. Many diverging waves are present in this patch which describes fairly well what a ship wave pattern resembles. We may even wonder...
if the expansion (6.10) is of any numerical use in view of its small region validity (between 2° and 6°)!

\[ \frac{X}{Y} / Z \]

Figure 6-24: The potential \( w \) for \( \varphi = 90^\circ, 8^\circ \leq \theta \leq 23^\circ, 95 \leq R \leq 135 \)

An interpretation of the transition region around the line \( \theta = \theta_0 \) when both the source and the field-point are on the free-surface, is clearly suggested by the decomposition carried out in Section 6.5.4. A \textit{rectilinear} wave front along \( \theta = \theta_0 \), of amplitude decaying as the radial distance \( R \) increases through the factors in \( N^{-1/3} \), propagates like a \textit{plane wave system}. This wave system which corresponds in fact to the diverging waves, is \textit{modulated} by the transverse wave system. Hence we depart from the “summation of contributions” assumption of the classical wave analysis. This interpretation is in fact \textit{far from being new}, at least in optics and quantum mechanics, and is common to \textit{any} fold catastrophe in asymptotics of integrals! We should note that the product \( \exp[N f(z_1, \bar{a})] \widetilde{Ai}(N^{1/3} e^{2i\pi/3} \zeta) \) comes originally from the products \( \exp[N (f(z_1, \bar{a}) + f(z_2, \bar{a}))/2] \text{Ai}(N^{1/3} e^{2i\pi/3} \zeta) \), and thus that the modulations due to the diverging wave system present in both parts of this last product are canceling.
out, leaving only the modulation from the transverse wave system. It must be noted that only the diverging wave system presents a wave front. The transverse waves do extend beyond the critical lines although with a vanishing behavior driven by \( f(z_1, \bar{a}) \).

Identical interpretations can be made for a \textit{submerged source}. Naturally the Airy function does not present anymore a pronounced transition as for the case of zero submergence. The argument of \( \text{Ai} \), or \( N^{2/3} e^{2i\pi/3} \), is no more purely real: in other words the fold catastrophe \textit{only} takes place when \( Z = 0 \). However, for small submergence of the pair source-field-point, the influence of the fold catastrophe is still significant. Such examples are given for the potential over the domain \( 4 \leq Y \leq 15, -16 \leq X \leq -40 \), for various submergences \( d \) of the source: \( d = 0 \) (Figure 6-25), \( d = 2\pi/100 \) (Figure 6-26), \( d = 2\pi/10 \) (Figure 6-27) and \( d = 4\pi/10 \) (Figure 6-28), where \( 2\pi \) is the wavelength of the transverse wave system.

![Figure 6-25: The potential \( w \) at \( d = 0 \) over the domain \( 4 \leq Y \leq 15, -16 \leq X \leq -40 \)](image)
Figure 6-26: The potential $w$ scaled by $\exp[d/2]$ at $d = 2\pi/100$ over the domain $4 \leq Y \leq 15$, $-16 \leq X \leq -40$

Figure 6-27: The potential $w$ scaled by $\exp[d/2]$ at $d = 2\pi/10$ over the domain $4 \leq Y \leq 15$, $-16 \leq X \leq -40$
As the submergence increases, the magnitude of the potential decreases (and this independently from the factor $\exp[-d/2]$). The diverging waves which are predominant in the structure of the wave pattern for $d = 0$ and even $d = 2\pi/100$, are much weaker for $d = 2\pi/10$, and have nearly totally disappeared for $d = 4\pi/10$. This phenomenon is again accounted by the Airy function whose arguments is of increasing modulus as the submergence increases. Thus only the transverse wave modulation associated to $\exp[N f(z_1, \alpha)]$ remains: $\Re\{f(z_1, \alpha)\}$ is a slowly decreasing function of $z$ as can be seen in Figures 6-2 and 6-3.

Such computations are not limited to this domain. In any case the same features would be observed in any portions of the sector $8^\circ \leq \theta \leq 40^\circ$. The presentation of a full wave pattern, i.e. extending to small but non-zero values of $\rho$ does present some practical problem relative to the number of grid points. The patch of Figure 6-23 by example, required 100 points in the radial direction and 400 points in the $\theta$ direction. The constraint of representing the diverging waves without distorting them, necessitates the use of adapted grids when treating the vicinity of the $x$-axis.
All the results presented here, are in terms of the potential. However, quantities like the wave elevation, or any derivatives of the potential can be obtained with minimal additional effort. The necessary modifications take place at the stage of the computation of $G_i^{(k)}$: the function $g(z, \tilde{\alpha})$ must be replaced by its corresponding equivalent. In addition it is necessary to redefine the sequences of radii $(\sigma_{p_r})$ and $(\sigma_{q_r})$, but no further modifications are necessary. Integrals of the potential can also be treated in the same manner, as long as the integrand, or $g(z, \tilde{\alpha}) \exp[N f(z, \tilde{\alpha})]$ is integrable in closed form with respect to the parameter $\tilde{\alpha}$.

This concludes our example of the application of our algorithms to the Kelvin wave source potential.
Chapter 7

The impulse response Green function in finite depth

We present here the second example of the method of steepest descents applied to the numerical evaluation of an integral. This example pertains again to the field of free-surface hydrodynamics and is of interest in several aspects. The difficulties associated with this new integral differ from those encountered in Chapter 6 with the Kelvin wave source potential. The main problem is here to derive an analysis which would result in a uniform expansion: indeed, the form of the integral we are considering places the problem of its uniform treatment beyond the theoretical frame of Section 1.2.

Several attempts of treatment of this integral (or related functions) using the method of steepest descents have been made: Kajiura (1963), Newman (1990) and Ursell (1991). All these results reduce to the same interpretation, namely: a fold catastrophe along with a coalescing singularity in the integrand function $g$. The nature of this singularity – a logarithmic branch point – and the presence of a large parameter in the argument of $g$ render these analyses difficult to justify. In addition, as emphasized by Ursell, the validity of the resulting approximation is extremely limited. Nevertheless we follow this approach and present an asymptotic expansion for the impulse Green function. In the process, we deal with integrals for which the
Levinson's transformations are trivial, thus avoiding most of the difficulties proper to the fold catastrophe and detailed in Chapter 5. Consequently, we have the luxury of having a complete knowledge of the functions $G(u, \tilde{\alpha})$, enabling us to devise simple, robust, but specific algorithms for obtaining the coefficients of the expansions. A numerical implementation shows that if the final expansion is well behaved, its approximation of the Green function is unsatisfactory. The reason of this failure, far from being obvious, seems to deny any possible uniform treatment through the method of steepest descents. Consequently, we outline a new analysis based on asymptotic approximations of two-dimensional integrals, which should yield a uniform expansion of the impulse response Green function. The complete justification of this analysis lies beyond the scope of this work.

7.1 A mathematical problem

By the beginning of the XIX-th century, the problem of the evolution of the free surface after an initial perturbation in the theory of linear free-surface waves was the subject of a prize problem issued by the French Academy of Sciences ("Académie des Sciences"). Both Cauchy in 1815 and Poisson in 1816, independently, answered the question in the case of infinite depth. Subsequent studies of other variants of this problem – known now as the Cauchy-Poisson problem – were made by other hydrodynamicists. Among those, was Lord Kelvin who formulated his principle of stationary phase (Kelvin, 1887a) in order to infer the behavior, for large times and distances away from the initial perturbation, of the three-dimensional solution in infinite depth. The study of the two-dimensional case was undertaken by Lord Rayleigh (1909) both in finite and infinite depth. The interest in first finding these solutions and then studying their behavior was legitimate in regard of their general character both for unsteady and steady problems. Hence the approach of Kelvin for ship wave patterns (Kelvin, 1887b), was to consider an infinite sequence in space and time of impulse disturbances. Knowing the solution for an impulse, the wave elevation was
obtained by its indefinite integration thus yielding the steady solution.

For the problem of the impulse response in three dimensions, it is worth having in mind the case of infinite depth as Lord Rayleigh did:

- In infinite depth, the phase velocity of waves as function of their wave length is unbounded. This corresponds to the stretching of the disturbance in space as time increases: the longer the wave length is, the higher its phase velocity is.

- In finite depth however, the upper bound (\( \sqrt{g/h} \), if \( h \) denotes the depth, and \( g \) the acceleration of gravity) on the phase velocity results in the presence of a transition region corresponding to a local, but moving, accumulation of energy. This transition — a wave front — moving at the velocity \( \sqrt{g/h} \) clearly separates two regions: one ahead of the front for radial distances \( x >> t \sqrt{g/h} \) characterized by the absence of waves (or rather the presence of only evanescent modes), the other behind the front \( (x << t \sqrt{g/h}) \) where waves are clearly visible.

This distinction between finite and infinite depth is also valid for the two-dimensional cases. It is also interesting to relate the elementary solution for the three-dimensional case in infinite depth — with no wave front — to its equivalent with steady motion, the Kelvin wave source potential, which presents wave fronts along the critical lines. This calls for caution in any attempt of generalization of the features of a given case to another. In the presence of a wave front in finite depth we can expect to identify a catastrophe in the corresponding mathematical formulation of the solution. This is indeed the case, as highlighted by Whitham (1974), for the impulse response wave elevation in two dimensions (although without the formalism of catastrophe theory!). An Airy function (characteristic of the fold catastrophe) appears quite obviously in the asymptotic behavior for large times and distances close to the wave front. The argument of this Airy function is proportional to the non-dimensional distance \( (1 - X/T) \) (where \( X = x/h \) and \( T = t \sqrt{g/h} \)). We could immediately conclude that the three-dimensional case would be similar and thus expect again the presence of an Airy function. Rigorously, however, since we know that the three-
dimensional case does present a wave front, we should only presume the presence of a catastrophe in the mathematical solution. The determination of the type of catastrophe is then left to the analysis of the solution.

The three-dimensional case has been in fact treated by Kajiura (1963) as a model for the behavior of an idealized tsunami. The initial condition was then of an impulsive displacement of the bottom in a radius about the origin at \( t = 0 \). Kajiura found that the leading order involved the integral:

\[
\int_0^{+\infty} u^4 \exp \left[ i \frac{u^6}{6} + i p u^2 \right] du
\]

where the parameter \( p \) is again proportional to \((1 - X/T)\) (for details see also Mei, 1983). As noted by Kajiura, this integral is in fact:

\[
\frac{d^2}{dp^2} \int_0^{+\infty} \exp \left[ i \frac{u^6}{6} + i p u^2 \right] du
\]

We recognize here the incomplete generalized Airy function \( \mathcal{U}_6(\xi, \Gamma_0) \) (using the notations of Section 1.2.3) with \( \xi = (0, -p, 0, 0) \) and \( \Gamma_0 \) joining 0 to \( \infty e^{-i\pi/4} \). In the formalism of catastrophe theory, this integral is related to the partial derivative of the butterfly integral:

\[
\mathcal{U}_5(\xi, \Gamma) = \int_{\Gamma} \exp \left[ \frac{u^6}{6} + \sum_{l=1}^{4} \xi_l u^l \right] du
\]

on the real axis \( \xi_1 = \xi_3 = \xi_4 = 0 \) in \( \mathbb{C}^4 \). Hence the catastrophe associated to the wave front in this case is the butterfly catastrophe which denotes, a priori, a much more complex behavior than for the two-dimensional case.

When looking for the time solution potentials for large objects of complex shapes such as offshore structures or ships, a slightly different integral than the ones previously mentioned is involved: namely the initial condition is given as a Dirac at \( t = 0 \) in terms of source strength. Solving the corresponding numerical problem after discretization of the body involves the numerical evaluation of this function at each point.
in space and time. As already seen in Chapter 2, the requirements for robustness and efficiency are such that numerical integration techniques are inappropriate for large times and distances. The consideration of the finite depth case is particularly relevant to offshore structure problems, and as noted by Newman (1990) lacks the attention that the infinite depth case was given.

In what follows we will be concerned with the time dependent part of the impulse response Green function which can be written (Newman, 1990, eq. (23) & (25)):

$$\Phi(X, Z, Z', T) = g^{1/2} h^{-3/2} \left[ \delta(T) F_1(X, Z, Z') + F_2(X, Z, Z', T) \right]$$  \hspace{1cm} (7.3)

for $T > 0$, and where $F_2(X, Z, Z', T) = F(X, Z - Z', T) + F(X, 2 - Z - Z', T)$ is the time dependent part, with:

$$F(X, V, T) = 2 \int_0^{+\infty} \frac{\omega(k) \cosh(Vk)}{\sinh(2k)} \sin(T\omega(k)) J_0(Xk) \, dk$$  \hspace{1cm} (7.4)

The function $\omega$ represents the dispersion relation: $\omega(k) = \sqrt{k} \tanh k$ based on the non-dimensional quantities:

- $X$, the horizontal distance between the source point and the field point, non-dimensionalized by $h$, the water depth
- $Z$ and $Z'$, the non-dimensional altitude of the source point and field point, with respect to the plane of the free-surface
- $T$, the non-dimensional time $t \sqrt{g/h}$

Subsequently we will only consider the integral $F$ of (7.4) which contains all the information regarding the propagation in time of the initial disturbance. This function only differs from the two-dimensional case by the presence of the Bessel function $J_0$ in place of a cosine function of the same argument. Although both functions have very similar behaviors for large positive and real arguments, the Bessel function $J_0$ complicates significantly the matter. However Kajiura in his particular case, by relating $J_0$
to the Hankel function of the first kind \(H_0^{(1)}\) and using the asymptotic expansion for large arguments of the latter, obtained his result involving the integral (7.1). A similar approach was followed by Newman (1990) for \(F\), although independently from the earlier work of Kajiura, and resulted in an asymptotic behavior involving the square of an Airy function. Both of these results consisted in the leading order only of the corresponding functions. A complete asymptotic expansion was later given by Ursell (1991) for the Cauchy-Poisson problem. We now give an overview of this last approach before applying it to the function \(F\) of (7.4).

7.2 The result of Ursell

Ursell considers the solution to the Cauchy-Poisson problem, that is the response to an impulse disturbance of the free-surface elevation at \(t = 0\). In particular the integral (if written in non-dimensional coordinates):

\[
\Psi(X, T) = \int_0^{+\infty} J_0(Xk) \cos(T\omega(k)) \, k \, dk
\]

(7.5)
is treated.

The first obstacle to a steepest descent analysis is the dependence of \(\Psi\) on two large parameters \(X\) and \(T\). The form of the integral strongly suggests a stationary phase approach, but in order to apply the method of steepest descents, it is necessary to extend the integrand to an holomorphic function of its variables. More precisely the integrand should be in a form:

\[
g(Xk) \, e^{Xr(k)} \, e^{T\omega(k)}
\]

where \(g\) would be such that \(k \mapsto |g(k)| \exp[-a \, k], \forall a \in \mathbb{IR}^{+}\) is bounded in parts of the complex plane.
Ursell does so by introducing the Hankel function of the first kind, yielding:

\[ \Psi(X, T) = \frac{1}{2} \int_{\Gamma_1} H_0^{(1)}(Xk) \cos(T\omega(k)) \, k \, dk \]  

(7.6)

where \( \Gamma_1 \) joins \(-\infty\) to \(+\infty\) passing above the logarithmic branch point at \( k = 0 \). The advantage is then that \( H_0^{(1)}(Xk) \exp[-iXk] \) is bounded in the half plane \( \Re k > 0 \).

From there \( \Psi \) is rewritten as:

\[ \Psi = \frac{1}{4} (\Psi^+ + \Psi^-) \]

with:

\[ \Psi^\pm = \int_{\Gamma_1} H_0^{(1)}(Xk) e^{-iXk} \exp[i(Xk \pm T\omega)] \, dk \]  

(7.7)

In writing the exponent function as \( T(kX/T \pm \omega) \), it can be seen that \( \Psi^- \) accounts for the main contribution to \( \Psi \). Under the assumption that \( X/T \) stays in a small enough neighborhood of 1, the saddle points of \( f(k, X/T) = kX/T - \omega \) are in a neighborhood of 0. \( f \) can then be approximated by the polynomial:

\[ f(k, X/T) \approx \frac{k^3}{6} - (1 - \frac{X}{T}) k \]

and \( \Psi^- \) replaced by the integral:

\[ A = \int_{\infty}^{\infty} H_0^{(1)}(Xu) e^{-iXu} \exp \left[ i \frac{T}{2} \left( \frac{u^3}{3} + \eta \, u \right) \right] u \, du \]  

(7.8)

with \( \eta = 2(X/T - 1) \). The canonical form of a function with two coalescing critical points clearly appear in the exponent function: the Levinson's transformation is here simply the identity. It remains nevertheless to expand in a series in \( u \), the function \( H_0^{(1)}(Xu) \exp[-iXu] \). The difficulty is here that \( X \) is large while \( u \) is in a neighborhood of 0, and no local series expansion can be used legitimately. Ursell, by means of the inverse Mellin transform, expresses \( H_0^{(1)}(z) \, \exp[-iz] \) as:

\[ H_0^{(1)}(z) \, \exp[-iz] = -\frac{1}{\pi^2} \int_{1/4 - i\infty}^{1/4 + i\infty} z^{-s} \, e^{i\frac{\pi}{2} s} \, \kappa(s) \, ds \]
with:
\[ \kappa(s) = 2^{-s} \frac{\Gamma^2(s) \Gamma(\frac{1}{2} - s)}{\Gamma(\frac{1}{2})} \]

This leads to:
\[
A = -\frac{i}{\pi^2} \int_{-\infty}^{\infty} \exp \left[ i \frac{T}{2} p_3(u, \eta) \right] \int_{1/4 - i\infty}^{1/4 + i\infty} X^{-s} u^{1-s} e^{i\pi s/2} \kappa(s) \, ds \, du
\]
\[
= -\frac{i}{\pi^2} \int_{1/4 - i\infty}^{1/4 + i\infty} X^{-s} e^{i\pi s/2} \kappa(s) \int_{-\infty}^{\infty} \exp \left[ i \frac{T}{2} p_3(u, \eta) \right] \, du \, ds
\]

with the notations of Chapter 1, and where the interchange in the order of integration is legitimate.

Obtaining the asymptotic expansion consists then in shifting the contour of integration in \( s \) to the right. In doing so, the integral in \( s \) is transformed into a sum of residues over the poles of \( \kappa(s) \): \( s = m + 1/2 \), for \( m \) in \( \mathbb{N} \). Hence the expansion is:

\[
A \sim \sum_{m=0}^{\infty} \frac{d_m}{(2X)^{m-1/2}} \int_{-\infty}^{\infty} \exp \left[ i \frac{T}{2} p_3(u, \eta) \right] \frac{du}{(ue^{-i\pi s/2})^{m-1/2}} \quad (7.9)
\]

with:
\[ d_m = \frac{(-1)^m}{\pi^{3/2}} \frac{\Gamma^2(m + \frac{1}{2})}{\Gamma(m + 1)} \]

If the integrals in (7.9) are denoted \( B(m - 1/2, \eta, T/2) \), Ursell proves the relation:

\[
\zeta B(s - 1, \eta, \frac{T}{2}) - B(s - 3, \eta, \frac{T}{2}) = -2 \frac{(s - 1)}{T} B(s, \eta, \frac{T}{2}) \quad (7.10)
\]

and the identity:
\[
B(1/2, \eta, 1) = 2^{5/3} \eta^{3/2} Ai^2(2^{-2/3} \eta) \quad (7.11)
\]

Relation (7.10) signifies that (7.9) can be expressed in terms of \( B(1/2, \eta, T/2) \), \( B(-1/2, \eta, T/2) \) and \( B(-3/2, \eta, T/2) \), where the last two functions are proportional to the first and second derivatives in \( \eta \) of \( B(1/2, \eta, T/2) \). This canonical function is not simply the Airy function as in the two-dimensional case, but a generalized Weber
function in the sense of Section 1.2.3. Although the asymptotic expansion of \( H_0^{(1)} \) for large arguments has not been used in the process, Ursell notes that using this expansion would bear an identical result.

Identity (7.11) relates the leading order found by Newman (1990), to the result of Ursell. Moreover, if we perform the change of variable \( u = v^2 \), we recover a derivative of the butterfly integral, and thus a form equivalent to the result of Kajiura. This same identity furnishes also means of evaluating numerically, and efficiently, the canonical function (and its derivatives), instead of having to devise methods to compute the butterfly integral.

The form of (7.9) requires some comments. The integrals involved present square root branch points at \( u = 0 \) from which the dependence in \( X \) has been extracted. The series in (7.9) appears then as an attempt of uniformly decomposing the behavior of \( H_0^{(1)}(Xu) \) on a sequence of rational functions. However this decomposition is non-uniformly valid in a non-vanishing neighborhood of \( \eta = 0 \). This comes from the fact that the behavior of \( H_0^{(1)}(Xu) \) is not uniform with respect to \( X \) in a fixed neighborhood of \( u = 0 \). Hence the expanded form of (7.9), obtained by repeatedly using recursion (7.10), has decreasing terms only for small values of \( |\eta| \) (of the order of \( O(1/\sqrt{T}) \)) as remarked by Ursell. However this expansion might still be of numerical interest, since we are concerned with a fixed accuracy, and in any case is worth considering from the algorithmic point of view.

7.3 An asymptotic expansion in terms of the butterfly integral

We have seen that it was sufficient to treat the function \( F \) of (7.4). Following the same approach as Ursell, we rewrite \( F \) by introducing \( H_0^{(1)} \).
Hence:

\[ F = \int_{\Gamma_1} \frac{\omega(k) \cosh(Vk)}{\sinh 2k} H_0^{(1)}(Xk) \sin(T\omega(k)) \, dk \]

where \( \Gamma_1 \) is defined as in Section 7.2. For convenience we perform the change of variables \( k = iu \), so that:

\[ F = \int_{\Gamma_2} g(u, V) h(Xu) e^{-Xu} \sinh(T\bar{\omega}(u)) \, du \]

where \( \Gamma_2 \) now joins \( \infty e^{-iu/2} \) to \( \infty e^{iu/2} \), passing to the right of the origin, and with:

\[
\begin{align*}
  &g(u, V) = \frac{\bar{\omega}(u) \cos Vu}{\sin 2u} \\
  &\bar{\omega}(u) = \sqrt{u \tan u} \\
  &h(Xu) = H_0^{(1)}(iXu) e^{Xu}
\end{align*}
\]

\( F \) is then expressed as: \( F = (I^+ - I^-)/2 \) where:

\[ I^\pm = \int_{\Gamma_2} g(u, V) h(Xu) e^{-Xu} \exp \left[ -X \, u \pm T \, \bar{\omega}(u) \right] \, du \]

A priori both integrals \( I^+ \) and \( I^- \) contribute (through their real part) to \( F \), however only \( I^+ \) is of importance. This last assertion is directly related to the problem of finding the relevant cols.

### 7.3.1 The relevant cols

Finding the set of relevant cols necessitates a somewhat detail study of the exponent functions.

Let \( f^\pm(u, \alpha) \) denote the functions \(-\alpha \, u \pm \bar{\omega}(u)\). \( f^\pm \) inherit the same branch points as \( \bar{\omega} \), namely the points \( u = (2k + 1) \pi/2, \, k \in \mathbb{Z} \). In addition \( \bar{\omega} \) is defined such that \( \bar{\omega}(e^{i\pi} \, u) = e^{i\pi} \, \bar{\omega}(u) \) resulting in \( \bar{\omega} \) being holomorphic in the disc \( D_o(0, \pi/2) \). Thus \( \bar{\omega} \)
is an odd function of \( u \), holomorphic in the region:

\[
\mathbb{C} \ \setminus \ \bigcup_{n=0}^{\infty} \left\{ \left[ (2n + 1) \pi/2, (2n + 2) \pi/2 \right] \cup \left[ -(2n + 2) \pi/2, -(2n + 1) \pi/2 \right] \right\}
\]

where the omitted segments represent the necessary branch cuts. (This is illustrated in Figure 7-1 for the real and imaginary parts of \( \tilde{\omega}' \).) However we have seen the importance of defining functions on simply connected domains. Hence we define \( \tilde{\omega} \) on the domain \( D_{\tilde{\omega}} = \mathbb{C} \ \setminus \ \{ -\infty, -\pi/2 \} \cup \{ \pi/2, +\infty \} \).

Figure 7-1: Real and imaginary parts of \( \tilde{\omega}' \) in the strip \( 0 \leq \Re u \leq 10, -0.5 \leq \Im u \leq 0.5 \) (A good example of the algorithm of root extraction described in Appendix D: the reference values are taken along the imaginary axis.)
Finding the cols of $f^\pm$ is then equivalent to finding the solutions of:

$$\tilde{\omega}'(u) = \pm x \quad \text{for } x \text{ real and } x \text{ in a neighborhood of } 1, u \in D_{\tilde{\omega}}$$  (7.12)

Since $\tilde{\omega}'$ has singularities in $(u - (2k + 1)\pi/2)^{-3/2}$ at the branch points $(2k + 1)\pi/2$, and in $(u - (2k + 1)\pi/2)^{-1/2}$ at the branch points $k\pi, k \neq 0$, equation (7.12) has obviously at least one solution in some neighborhoods of these points. Using the additional property that $\tilde{\omega}(u) = \overline{\tilde{\omega}(u)}$, also true for its derivatives, we deduce that there exists in fact two solutions, conjugate one of the other, in such neighborhoods. Denoting these solutions $u^+_{k}$ and $u^-_{k}$ for $f^\pm$, we have a countable infinity of solutions given by:

$$\bigcup_{n=1}^{\infty} \{u^+_n, u^-_n, -u^+_n, -u^-_n\}$$  (7.13)

where $u^+_n$ is in a neighborhood of $n\pi$, while $u^-_n$ is in a neighborhood of $(2n - 1)\pi/2$. In addition $f^+$ has two additional saddle points in a neighborhood of 0 denoted $u_0$ and $-u_0$ where $u_0$ is real if $x < 1$, imaginary if $x > 1$, and 0 if $x = 1$.

It remains of course to determine the relevant saddle points or finding the appropriate paths in the graphs $G_{f^+}$ and $G_{f^-}$. The practical task of building the graphs of $f^+$ and $f^-$ is rendered difficult by an infinite number of cols. In this process, the formalism of Chapter 3 and the properties of the graphs $G_{f^+}$ and $G_{f^-}$ are extremely helpful. We refer here to Figures 3-9 and 3-10 of Chapter 3, where the basins of $f^+$ and $f^-$ are given in the domain $[-10,10] \times [-2,2]$ for $\alpha = 1.2$ for $f^+$, and $\alpha = 1.0$ for $f^-$. Having noticed that $\Re f^\pm(u) = \Re f^\pm(-\overline{u})$, we can infer that:

- For $f^+$, only $u_0$ and $-u_0$ are possibly relevant for any value of $\alpha$. More precisely, only $u_0$ (assuming $u_0 > 0$) is relevant for $\alpha > 1$, both $u_0$ and $-u_0$ being relevant for $\alpha < 1$. The values of $\alpha > 1$ are ramification values for $u_0$ and $-u_0$ only, while any $\alpha < 1$ is non-critical for $f^+$. This comes from the relations:

$$\forall \alpha \quad 0 \leq \Re f^+(u_0) < \Re f^+(u_1^+) < \ldots < \Re f^+(u_n^+) < \Re f^+(u_{n+1}^+) < \ldots$$
• For $f^-$, we have the following relations:

$$\forall \alpha \quad 0 > \Re f^-(u^-_i) > \ldots > \Re f^-(u^-_n) > \Re f^-(u^-_{n+1}) > \ldots$$

and similarly for the real parts. In addition $\alpha$ is always a ramification value for any pair $(u^-_n, -u^-_n)$. The above relations ensure that the graph of $f^-$, $G_{f^-}$, is identical for any two distinct values of $\alpha$. Since for $\alpha = 1.0$ all of the cols $u^-_n, u^-_n$ are relevant, they will be relevant for any other value of $\alpha$. (The cols $-u^-_n, -u^-_n$ are also relevant but their contributions to $I^-$ are without effect since $\alpha$ is a ramification value for $f^-$.)

Hence the contribution of $I^-$ is in fact an infinite sum over $n$ of the contributions from the cols $u^-_n, u^-_n$! From the above relations and since $|\exp[Tf^+(\pm u_0, \alpha)]| >> |\exp[Tf^-(u^-_1, \alpha)]|$ for $T >> 1$, uniformly for $\alpha$ in a neighborhood of 1, the contribution from $I^-$ is in fact exponentially small and we only need to consider $I^+$.

We now impose $\alpha$ to be in a small enough neighborhood of 1, so that we may approximate $f^+(u, \alpha)$ by the polynomial $p_3(u, \zeta)/2$ with $\zeta = 2(1 - X/T)$, under the condition that $|\zeta|^{5/2} << 1$. The extension of the region of validity of the expansion to larger values of $|\zeta|$ (larger domain around the wave front) would then require to consider the exact Levinson's transformation $u \mapsto v$ defined by: $f^+(u, \alpha) = p_3(v, \xi)/2$ requiring to use the procedures described in Chapter 5. However we choose the above approximation since much stronger limitations bear on the validity of this expansion, as we have seen in Section 7.2.

### 7.3.2 The butterfly catastrophe expansion

We are thus reduced to consider the integral:

$$\tilde{I}^+ = \int_{\Gamma_3} g(u, V) h(Xu) \exp \left[ \frac{T}{2} p_3(u, \zeta) \right] du \quad (7.14)$$

where $\Gamma_3$ is now joining $\infty e^{-i\pi/3}$ to $\infty e^{i\pi/3}$, and for which we follow the procedure of Ursell outlined in Section 7.2. There is no
obstacle in doing so since the change in the order of integration, and the deformation of the path of integration in $s$ (see Section 7.2) are legitimate. Thus, this results in the asymptotic expansion:

$$\bar{I}^+ \approx -i \frac{\sqrt{2}}{\pi^{3/2} X^{1/2}} \sum_{n=0}^{\infty} (-1)^n \frac{\Gamma^2(n + \frac{1}{2})}{\Gamma(n + 1)} \frac{1}{(2X)^n} I_n(T, V, \zeta)$$

(7.15)

where:

$$I_n(T, V, \zeta) = \int_{\Gamma_3} \frac{g(u, V)}{u^{n+1/2}} \exp \left[ \frac{T}{2} p_3(u, \zeta) \right] du$$

(7.16)

Similarly to the case considered by Ursell, we are now left with integrals $I_n$ presenting square root branch points at $u = 0$. If in our case the functions in factor of the exponential $g(u, V)/u^{n+1/2}$ are not of simple form, the integrals $I_n$ are suitable for an application of the generalized method of steepest descents. Hence, by applying directly the procedure of Bleistein corresponding to the coalescence of two saddle points with an algebraic singularity, we obtain (see eq. (1.34) of Section 1.2.2):

$$I_n \sim \frac{T}{2} \left( \frac{T}{2} \right)^{-p/3} \sum_{k=0}^{\infty} \left( \frac{T}{2} \right)^{-k} \sum_{p=0}^{2} \left( \frac{T}{2} \right)^{-p/3} P_{n,k}^{(p)}(V, \zeta) U_{2,1,n}(W, 0, \Gamma_3)$$

(7.17)

with, according to the notations of Section 1.2.2:

$$U_{2,1,n}(W, 0, \Gamma_3) = \int_{\Gamma_3} u^{p-n-1/2} \exp[p_3(u, W)] du \quad , \quad W = \zeta \left( \frac{T}{2} \right)^{2/3}$$

247
and with \( P_{n,k}^{(p)} \) defined through – see (1.32) and (1.33):

\[
\begin{align*}
G_0(u, V) &= g(u, V) \\
G_{n,k}(u, V, \zeta) &= R_{n,k}(u, V, \zeta) + \Psi(u, \zeta) H_{n,k}(u, V, \zeta) \\
\Psi(u, \zeta) &= u^3 + \zeta u = u p_3'(u, \zeta)
\end{align*}
\]

where

\[
R_{n,k}(u, V, \zeta) = \sum_{p=0}^{2} P_{n,k}^{(p)}(V, \zeta) u^p
\]

\[
G_{n,k+1}(u, V, \zeta) = -u \frac{\partial}{\partial u} H_{n,k}(u, V, \zeta) + \left(n - \frac{1}{2}\right) H_{n,k}(u, V, \zeta)
\]

The series (7.17) is asymptotic for large values of \( T \), and obtaining the full expansion at the wave front is just a matter of replacing each \( I_n \) by its asymptotic expansion in (7.15). However we rearrange (7.17) as follows:

\[
I_n \sim 2 T^{n-1/6} \sum_{k=0}^{\infty} \left(\frac{T}{2}\right)^{-k} \sum_{p=0}^{2} T^{-p} P_{n,k}^{(p)}(V, \zeta) \bar{U}_{n-p}
\]

where \( \bar{U}_n \) satisfies the recursion:

\[
\bar{U}_{n+1} = \frac{1}{2n+1} \left[ T^{-2} \bar{U}_{n-2} + \zeta \bar{U}_n \right]
\]

and:

\[
\begin{align*}
\bar{U}_0 &= \mathcal{U}_5(\bar{W}', \Gamma_4) \\
\bar{U}_1 &= T^{2/3} \frac{\partial}{\partial \xi_2} \mathcal{U}_5(\bar{W}', \Gamma_4) = T^{2/3} \frac{\partial^2}{\partial \xi_1^2} \mathcal{U}_5(\bar{W}', \Gamma_4) \\
\bar{U}_2 &= T^{4/3} \frac{\partial^2}{\partial \xi_2^2} \mathcal{U}_5(\bar{W}', \Gamma_4) = T^{4/3} \frac{\partial^4}{\partial \xi_1^4} \mathcal{U}_5(\bar{W}', \Gamma_4)
\end{align*}
\]

with \( \mathcal{U}_5 \) defined by (7.2), \( \Gamma_4 \) a path joining \( \infty e^{-i \pi/6} \) to \( \infty e^{i \pi/6} \), and \( \bar{W}' \) being \((0, \zeta, T^{2/3}/2, 0, 0)\).
It is not necessary to express the canonical function $U_{2,1}$ in terms of the butterfly integral $U_5$ in order to obtain the recursion (7.20): the relation (7.10) is sufficient. The point is here to interpret the combination of a fold catastrophe and of a coalescing branch point at 0, as a catastrophe itself. This is possible here, and the catastrophe we are concerned with, is the butterfly catastrophe.

By chance we are only dealing with a trace of the butterfly catastrophe on the real line $\xi_1 = \xi_3 = \xi_4 = 0$, instead of the complete space of parameters $C^4 (\approx R^8)$. For this reason only the even derivatives of $U_5$ in $\xi_1$ are present in the expansion (instead of the four derivatives $\partial U_5 / \partial \xi_1, \ldots, \partial^4 U_5 / \partial \xi_1^4$ for the general case).

The interpretation we propose is obvious for $I_0$ which is in fact the leading behavior in (7.15) as $X$ and $T$ are large, and $\zeta$ in a neighborhood of 0. For the other integrals $I_n$, $n \geq 1$, the decomposition given by (7.17) is not so clear. However the recursion relation (7.20) indicates that the expansions are in no way different in essence. Simply, the coefficients of these expansions, once they are in canonical form (i.e. in terms of $\tilde{U}_0, \tilde{U}_{-1}, \tilde{U}_{-2}$), are not directly given by the Bleistein sequence $(R_{nk})$.

Aside from such considerations, remains the problem of evaluating these coefficients, or of devising the appropriate algorithms. We have seen in Chapter 4 and 5 that the procedure of Bleistein, at least in its contour integral formulation, was of no practical use. We even preferred the method of Chester et al. for the treatment of the fold catastrophe. But here, in presence of a coalescing branch point, we are bound to find a numerical equivalent to (7.18).

### 7.3.3 The algorithms

The algorithms we present here, are entirely case specific. We are thus departing from the constraints we fixed in Chapter 2. We could certainly devise algorithms for the case of the coalescence of two cols with an algebraic singularity in the same spirit as Chapter 5. However, their application to the impulse Green function would be
awkward in view of the analysis of Sections 7.3.1 and 7.3.2. In other words by doing so, we would deny the knowledge we have about the integral $I_n$, and thus overlook important simplifications. This knowledge – the analytical expression of the function $g(u, V)$ – provides us with the entire series of $g(u, V)$ at $u = 0$. Hence we have access, in theory, to the analytical formulas of these series coefficients as functions of $V$.

**A general algorithm**

Let us, for the moment, assume that we have a similar knowledge of $G(u, \alpha)$ of (1.29) (see Chapter 1) and that $\tilde{\beta}(\alpha)$ and $\tilde{\xi}(\alpha)$ are known functions of $\alpha$. In fact we only need the ability of computing the numerical values of these quantities for any $\alpha$. (This requirement might seem weaker but in practice is as strong as the analytical knowledge of the functions themselves.) Starting then from an approximant of $G_0$ in the form of its truncated entire series at $u = 0$ – a polynomial in $u$ of a certain degree – it is easy to derive, at the level of *symbolic computation*, a procedure for the computation of the coefficients of the various polynomials $R_k$, $H_k$ and $G_k$. Such a procedure results in explicit relations (functions of $k$) between the coefficients of these polynomials or, if fully evaluated, in complete analytical expressions for the coefficients $P_k^{(p)}$ in terms of $\tilde{\beta}(\alpha)$, $\tilde{\xi}(\alpha)$ and the coefficients of $G_0(u, \alpha)$. These explicit relations can in turn be used as algorithms for the numerical evaluation of the coefficients $P_k^{(p)}$. The generality of these algorithms is then only restricted by the assumption we made, i.e. the knowledge of the entire series of $G(u, \alpha)$. However we need to keep in mind that, for the results to be valid, all of the points $\beta_\mu(\alpha)$ and cols $u_n(\alpha)$ have to be within the disk of convergence of this series. This might significantly restrict the domain of validity in the $\alpha$-space of the above algorithms in certain cases. (These limitations are consistent in our case with the approximations made in Section 7.3.2 and do not entail any further the validity of the final results.)
A specific case

By being more specific in our assumptions, i.e. by imposing \((r_\mu)\) (see (1.28)) and \(m\) to be fixed, it is then possible, in reasonable cases (for moderate values of \(m\) and \(q\)), to derive simple relations corresponding to each of the equations of (1.32). This corresponds to the lowest level of abstraction, level suitable to an implementation in any programming language.

Naturally we chose this last approach here, and detail the various steps implied in (7.18).

a) Computation of the polynomial \(G_0(u,V)\)

Since we know the expression of \(g(u,V)\) (see Section 7.3), we can obtain, through symbolic computation, the coefficients of the truncated entire series as functions of \(V\). This entire series is even in \(u\) and its coefficients are even polynomials in \(V\). Thus we have:

\[
G_0(u,V) \approx \sum_{p=0}^{p_{G_0}} G_0^{(p)}(V^2) u^p
\]

where \(G_0^{(p)} = 0\) for \(p\) odd, and the degree of \(G_0^{(p)}\) in \(V^2\) is \(p/2\).

These polynomials are uniformly valid in \(V\) and their coefficients are computed once \(p_{G_0}\) is fixed. The numerical values of \(G_0^{(p)}(V^2)\), \(p\) even, are then evaluated for each new value of \(V\).

In practice, since we know that \(V\) lies in the interval \([0,2]\), we compute the coefficients of \(G_0^{(p)}(4v)\) where now \(v\) is in \([0,1]\) (\(v = V^2/4\)). This allows the elimination of coefficients of magnitudes smaller than a certain tolerance, resulting in a significant gain in number of FLOPs (from 30% to 70% at this stage).
b) Euclidean division of $G_k$ by $\Psi$

Here we take advantage of the fact that $\Psi$ is of simple form. We have then the relations for $k \leq E(p_{G_0}/3)$:

\[
\begin{align*}
H_k^{(p_{G_k} - 3)} &= G_k^{(p_{G_k})} \\
H_k^{(p_{G_k} - 4)} &= G_k^{(p_{G_k} - 1)} \\
H_k^{(p)} &= G_k^{(p + 3)} - \zeta H_k^{(p + 2)} , \quad 0 \leq p \leq p_{G_k} - 5
\end{align*}
\]

and

\[
\begin{align*}
P_k^{(2)} &= G_k^{(2)} - \zeta H_k^{(1)} \\
P_k^{(1)} &= G_k^{(1)} - \zeta H_k^{(0)} \\
P_k^{(0)} &= G_k^{(0)}
\end{align*}
\]

with the notations: $G_k = \sum_{p=0}^{p_{G_k}} G_k^{(p)} u^p$, etc... and where we have dropped the dependence in $n$.

Such relations do not raise any difficulty whether in terms of numerical properties (a priori $\zeta$ is in $]-1,1[$, and the above relations are stable) or implementation.

c) Obtaining $G_{k+1}$ from $H_k$

The differential equation giving $G_{k+1}$ from $H_k$ results in the following linear relation in terms of polynomial coefficients:

\[
G_{k+1}^{(p)} = (n - p - \frac{1}{2}) H_k^{(p)} , \quad 0 \leq p \leq p_{H_k} = p_{G_k} - 3
\]

Again, its implementation in floating point numbers does not raise any difficulty. A variant may include at this stage the factor $2/T$ in order to avoid overflow for large values of $n$ and $p$. 

252
d) The asymptotic expansion of $I_n$

The combination of the three above procedures, allows us to obtain the values of $P_{n,k}^{(p)}$, $0 \leq p \leq 2$, for any $n$ and $k$, $k \leq E(pG_0/3)$. It remains to perform the summation in $p$ of (7.19) involving $\bar{U}_n$, $\bar{U}_{n-1}$, $\bar{U}_{n-2}$ expressed by their projections on the basis $\bar{U}_0$, $\bar{U}_{-1}$, $\bar{U}_{-2}$. This is done through the use of recursion (7.18) with the initial terms given by $\bar{U}_0$, $\bar{U}_{-1}$, $\bar{U}_{-2}$. The form of (7.18) was chosen so that $(\bar{U}_n)$ is decreasing in sequence of 3 as soon as $T > 1$, and $|\zeta| < 1$. This induced the coefficient in factor of (7.19) which is included when summing the series for $\bar{I}^+$.

The indices at which each of the sequences $(P_{n,k}^{(p)})$, $0 \leq p \leq 2$, are truncated, are bounded by $E(pG_0/3)$, where $pG_0$ is the degree of the approximation of $G_0$. In fact the actual number of non-zero terms might be smaller than this bound by a half (this is the case for $n = 0$).

The asymptotic summation in (7.19) is performed in fact on three truncated series, one for each of the three component of the basis $(\bar{U}_0$, $\bar{U}_{-1}$, $\bar{U}_{-2}$). Hence a truncation test is applied independently on each of the three finite series. This test is based on an alternative of the type (4.33) of § 4.3.1, along with an enforcement of the bound on the indices.

e) The asymptotic expansion of $\bar{I}^+$

The contributions from each $I_n$, $n \geq 0$, are incorporated in (7.15) under the form:

$$
\bar{I}^+ \sim -i \frac{23/2}{\pi^{1/2}} \left( \frac{T}{X} \right)^{1/2} T^{-2/3} \sum_{n=0}^{\infty} \alpha(n) \left( \frac{T}{2X} \right)^n \sum_{k=0}^{\infty} \left( \frac{2}{T} \right)^k \sum_{p=0}^{2} T^{-p} P_{n,k}^{(p)} \bar{U}_{n-p} \tag{7.21}
$$

with:

$$
\begin{align*}
\alpha(0) &= 1 \\
\alpha(n) &= (1 - n - \frac{1}{4n}) \alpha(n-1) , \quad n \geq 1
\end{align*}
$$
The truncation bears again on each of the three series in \( n \), defined by the generic terms \( \alpha(n) \left(\frac{T}{2X}\right)^n \left(\sum_{k=0}^{K(i,n)} \cdots \right) \cdot \bar{U}_{i} \), \( 0 \leq i \leq 2 \), where the indices \( K(i,n) \) are given by the truncations of the asymptotic expansion of \( I_n \). Since the series in \( n \) are a priori infinite, an arbitrary upper bound on \( n \) is a precaution against unnecessary computations for small values of \( T \) and \( X \) or large values of \( \zeta \).

The asymptotic properties of (7.21) are more easily perceived if \( T/X \) is replaced by \( 2/(2 - \zeta) \). The behavior for varying \( k \), all other variable being held fixed, is typical of an asymptotic expansion as the magnitude of the coefficients are increasing past a certain row, this again for values of \( \zeta \) in a neighborhood of 0. For varying \( n \) however, the magnitudes of all terms are decreasing if \( \zeta < 1 \). This last property emphasize the importance of having a truncation test based on a fixed tolerance and upper bounds on the indices.

f) The computation of \( \bar{U}_0, \bar{U}_{-1}, \bar{U}_{-2} \)

The final step is the computation of the canonical function and its derivatives. Although we have expressed the expansion (7.15) in terms of the butterfly integral, we use the identity (after (7.11)):

\[
\mathcal{U}_5((0, \xi_2, 0, 0), \Gamma_4) = i^{5/6} \pi^{3/2} \text{Ai}^2(-2^{-1/3} \xi_2)
\]

to compute \( \mathcal{U}_5 \) and its first two derivatives in \( \xi_2 \). These derivatives are then expressed in terms of \( \xi_2 \), \( \text{Ai} \) and \( \text{Ai}' \) only.

We conclude here the presentation of our algorithm for the impulse response Green function in finite depth, whose global arrangement is given in Figure 7.2. We can only emphasize the ease with which we devised the various components. The difference with the procedure presented in Chapter 5 for the simpler case of the fold catastrophe, both in terms of complexity and numerical properties, is striking. The simplicity of this last example, apart from the analysis of Sections 7.3.1 and 7.3.2,
comes entirely from the triviality of the Levinson's transformation. Since \( g(u, V) \) is known analytically, we have the additional assumption on the knowledge of the entire series of \( G_0(u, V) \) at \( u = 0 \) which permits the derivation of the above procedures. This simplicity is counterbalanced by a smaller region of validity than the one permitted by a more general algorithm in the spirit of Chapter 5. However before considering the inversion of a Levinson's transformation and its complications, it is of interest to assess the numerical validity of (7.21).

\[ \sum_{k_0}^{k_2} \sum_{n_0}^{n_2} \]

\[ H_k \quad R_k \]

\[ G_{k+1} \]

\[ \tilde{U}_n \quad \tilde{U}_{n-1} \quad \tilde{U}_{n-2} \]

\[ a(n) \]

\[ G_0 \]

Figure 7-2: Global arrangement of the algorithm

### 7.3.4 Numerical results

The chosen implementation is based on a truncation of the entire series of \( G(u, V) \) at the order 48 in \( u \). This enables the computation of, between 8 and 17 terms for each of the series in \( k \), while the upper limit in \( n \) is fixed to 20. Such choices are motivated more by an exploration of the properties of the asymptotic expansion than by practical computations. In practice an order of truncation of \( p_G = 24 \) is sufficient.
Tolerance for the truncation criterion is chosen to be $10^{-8}$ for an implementation in double precision in FORTRAN 77. The numerical validity of the truncated expansion is assessed by comparison with direct numerical integration as described in Newman (1990), for the range $20 \leq T \leq 80$ and $|\zeta| < 0.4$. Higher values of $T$ induce insufficient convergence in the Romberg part of the numerical integration algorithm. This range of $T$ is indeed sufficient to appreciate the usefulness of the expansion. We present $F(X,T,0)$ for $T = 20, 40, 80, |\zeta| < 0.4$ evaluated both by the truncated asymptotic expansion and by numerical integration (see Figures 7-3, 4, 5).

Figure 7-3: $F(X,T,0)$ for $T = 20, |\zeta| < 0.4$: numerical integration versus asymptotic series

For these three cases it clearly appears that although the truncation errors in the asymptotic expansion are small, the absolute errors between the expansion and the numerical integration are significant as soon as $\zeta$ is away from 0. Furthermore the increase in the values of $T$ does not bring any drastic improvement in the absolute errors as it could be expected from a truly uniform asymptotic expansion. This denies any practical use for this expansion.
Figure 7-4: $F(X, V, T)$ for $T = 40$, $|\zeta| < 0.4$: numerical integration versus asymptotic series

Figure 7-5: $F(X, V, T)$ for $T = 80$, $|\zeta| < 0.4$: numerical integration versus asymptotic series
7.4 A uniform asymptotic expansion in terms of the hyperbolic umbilic integral

In view of the failure of the approach of Sections 7.2 and 7.3, we now outline a different analysis. This new approach yields an asymptotic approximation which should be uniformly valid in a region of constant size in terms of $(1 - X/T)$.

We start from the expression of $F$ given in Section 7.3, or:

$$F(X, V, T) = \int_{\Gamma_2} g(z_1, V) H_0^{(1)}(iXz_1) \sinh(T \tilde{\omega}(z_1)) dz_1$$  \hspace{1cm} (7.22)

where $g$ and $\Gamma_2$ are defined as before and with $\tilde{\omega}$ and $g$ defined in $D_\tilde{\omega}$ (see Section 7.3.1).

Since we are dealing with two large parameters, $X$ and $T$, it is natural to seek not to favor one or the other. This is effectively done by introducing the integral expression of $H_0^{(1)}$:

$$H_0^{(1)}(z) = \frac{1}{i\pi} \int_{-\infty+\pi/2}^{+\infty+\pi/2} \exp[z \sinh t] dt \quad \text{for} \quad 0 < \arg z < \pi$$

(Note the bounds of integration.)

This leads, by letting $\alpha_1 = V$, to:

$$F = \frac{1}{i\pi} \int_{\Gamma_2} g(z_1, \alpha_1) \int_{-\infty}^{+\infty} \exp[-Xz_1 \cosh z_2] dz_2 \sinh(T \tilde{\omega}(z_1)) dz_1$$

Both integrals are uniformly convergent with respect to the other variable of integration given the path $\Gamma_2$ in $z_1$ and the real straight line in $z_2$. We can thus rewrite $F$ as:

$$F = \frac{1}{i\pi} \int_{\Omega} g(z_1, \alpha_1) \exp[-Xz_1 \cosh z_2] \sinh(T \tilde{\omega}(z_1)) dz_1 dz_2$$

where the integration is understood to be a double integration over the domain, or surface of $\mathbb{C}^2$, $\Omega = \{(z_1, z_2) : z_1 \in \Gamma_2, z_2 \in \mathbb{R}\}$.
Hence we consider the double integrals over $\Omega$:

$$
\mathcal{I}^\pm(T, \bar{\alpha}) = \int_\Omega g(z_1, \alpha_1) \exp[T f^\pm(z_1, z_2, \alpha_2)] \, dz_1 \, dz_2
$$

(7.23)

where: $f^\pm(z_1, z_2, \alpha_2) = \pm \tilde{\omega}(z_1) - \alpha_2 z_1 \cosh z_2$, so that:

$$
\mathcal{F}(X, V, T) = \frac{1}{2i\pi} \left( \mathcal{I}^+(T, \bar{\alpha}) - \mathcal{I}^-(T, \bar{\alpha}) \right)
$$

with $\bar{\alpha} = (\alpha_1, \alpha_2) = (V, X/T)$.

The problem is now reduced to the finding of uniform asymptotic expansions for the integrals $\mathcal{I}^\pm$ as $T$ gets large, while $\bar{\alpha}$ remains in $[0, 2] \times [1-\delta, 1+\delta]$ with $0 < \delta < 1$. The results for the asymptotic approximation of multiple integrals (see Wong, Chap. VIII and IX, 1989) are concerned mainly with stationary phase or Laplace type integrals. However we have here two complex double integrals and the question of deforming the initial surface is far from being trivial. The problem is nevertheless the same as in the method of steepest descents and consists in finding the critical points of an integral. Such points are eventually: points on the boundaries of the surface, saddle points of the exponent functions, and of course singular points of the function $g$ when coalescing with other critical points. Naturally these points can form curves on the surface of integration. The principles for deforming the surface of integration are the same as for one-dimensional complex integrals (mainly Cauchy's theorem for holomorphic functions of several variables) with the constraint that any deformation must be uniform in both variables.

The relevant saddle points

The functions of $(z_1, z_2)$, $f^+$ and $f^-$, are defined in the domain (of $C^2$), $D_{f^\pm} = D_{\omega} \times C$. Each function being periodic of period $(0, 2i\pi)$, it is sufficient to consider the domain $D'_{f^\pm} = D_{\omega} \times ([\mathbb{R} \times] - i\pi, i\pi])$. The domain of definition of $g$, $D_g$, coincides exactly with $D_{f^\pm}$, therefore the singular points of $G$ are of no matter here. We are thus left
with finding only the saddle points of \( f^\pm \), since \( \Omega \) is unbounded. Such points satisfy the conditions:

\[
\begin{cases}
\frac{\partial f^\pm}{\partial z_1} = 0 \\
\frac{\partial f^\pm}{\partial z_2} = 0
\end{cases} \quad \iff \quad \begin{cases}
\pm \bar{\omega}'(z_1) - \alpha_2 \cosh z_2 = 0 \\
-\alpha_2 z_1 \sinh z_2 = 0
\end{cases}
\]

Using the same conventions as in Section 7.3.1 for the zeros of \( \bar{\omega}'(z_1) = \alpha_2 \cosh z_2 \), the set of saddle points of \( f^+ \), \( Z(f^+) \), is:

\[
(u_+^n, i \pi), (u_+^n, i l \pi), (-u_+^n, i \pi), (-u_+^n, i l \pi) \quad n \geq 1, \ l \in \mathbb{Z}
\]

\[
(u_0, i 2l \pi), (-u_0, i 2l \pi)
\]

\[
\begin{cases}
(0, \text{arcosh } 1/\alpha_2 + i 2l \pi), (0, -\text{arcosh } 1/\alpha_2 + i 2l \pi) & \text{if } \alpha_2 \leq 1 \\
(0, i \text{arccos } 1/\alpha_2 + i 2l \pi), (0, -i \text{arccos } 1/\alpha_2 + i 2l \pi) & \text{if } \alpha_2 \geq 1
\end{cases}
\]

(with coalescence of the last two pairs of sets onto \((0, i 2l \pi)\) if and only if \( \alpha_2 = 1 \))

while the set \( Z(f^-) \), is:

\[
(u_-^n, i \pi), (u_-^n, i l \pi), (-u_-^n, i \pi), (-u_-^n, i l \pi) \quad n \geq 1, \ l \in \mathbb{Z}
\]

\[
(u_0, i (2l + 1) \pi), (-u_0, i (2l + 1) \pi)
\]

\[
\begin{cases}
(0, \text{arcosh } 1/\alpha_2 + i (2l + 1) \pi), (0, -\text{arcosh } 1/\alpha_2 + i (2l + 1) \pi) & \text{if } \alpha_2 \leq 1 \\
(0, i \text{arccos } 1/\alpha_2 + i (2l + 1) \pi), (0, -i \text{arccos } 1/\alpha_2 + i (2l + 1) \pi) & \text{if } \alpha_2 \geq 1
\end{cases}
\]

(with the same remark as above).

Thus the coalescence of four saddle points (for a fixed \( l \)) occurs at \( \alpha_2 = 1 \) both for \( f^+ \) and \( f^- \). For any path in \( z_1 \) homotopic to \( \Gamma_2 \) in the right half-plane of \( D_{\bar{\omega}} \), the path in \( z_2 \) (i.e. the real line) ensures the uniform convergence of (7.23). In other words, any homotopic deformation of \( \Omega \) can be such that its projection on the \( z_2 \)-plane is
the real straight line $\mathbb{R}$. This restricts the set of possible saddle points to:

$$Z(f^+) \quad \text{with } l = 0 \quad \text{for } f^+$$

$$(u_n^-, 0), (u_n^+, 0), (-u_n^-, 0), (-u_n^+, 0) \quad n \geq 1 \quad \text{for } f^-$$

and hence the problem of determining the relevant saddle points is the same as in Section 7.3.1.

**The unfolding of the catastrophe at $\alpha_2 = 1$**

According to the above results and those of Section 7.3.1, we only need to consider $f^+$. Since $\mathbb{H}^+$ is a double integral, the co-rank $k$ of the catastrophe at $\alpha_2 = 1$ might be 1 or 2. The determination of this co-rank necessitates to consider the differential forms of degree two and three of $f^+$, in particular:

$$d^2 f^+(z_1, z_2) = \begin{pmatrix}
\bar{w}'(z_1) & -\alpha_2 \sinh z_2 \\
-\alpha_2 \sinh z_2 & -\alpha_2 z_1 \cosh z_2
\end{pmatrix}$$

which is of rank 2 except at $(0, 0)$ for $\alpha_2 = 1$ where its rank is 0. The 3-form $d^3 f^+$ is however non null at this point. We are thus in presence of a hyperbolic umbilic catastrophe (co-rank 2, co-dimension 3); see Thom (1977).

We know (singularity theory) that there exists a diffeomorphism $u : (z_1, z_2) \mapsto (u_1, u_2)$ from a neighborhood of $(0, 0)$ in $D_z$ to a neighborhood of $(0, 0)$ in $\mathbb{C}^2$ such that:

$$f^+(z_1, z_2, \alpha_2) = u_1^3 + u_2^3 + \zeta_3 u_1 u_2 + \zeta_2 u_2 + \zeta_1 u_1$$

$$= \mathcal{F}_5(u_1, u_2, \zeta)$$

(The right hand side is the canonical form of the hyperbolic umbilic singularity.)

Hence we have:

$$\mathbb{H}^+ \approx \int_\Delta G(u_1, u_2, \bar{\alpha}) \exp[T \mathcal{F}_5(u_1, u_2, \zeta)] du_1 du_2$$
where \( G(u_1, u_2, \alpha') = -g(z_1, \alpha_1)/J[u] \) with \( J[u] \) the Jacobian of \( u \), and where \( \Delta \) is the domain of \( \mathbb{C}^2 \), \( \Delta = \{(u_1, u_2) \mid u_1 \in \Gamma_3, \ u_2 \in \Gamma_3\} \) (\( \Gamma_3 \) being defined as in Section 7.3.1).

The asymptotic expansion of \( \mathcal{I}_+ \) is then given by (see Connor, 1976):

\[
\mathcal{I}_+ \sim T^{-2/3} \sum_{k=0}^{\infty} T^{-k} \sum_{(p,q) = (0,0)}^{(1,1)} T^{-(p+q)/3} P_k^{(p,q)} \int_{\Delta} u_1^p u_2^q \exp[F_5(u_1, u_2, \bar{X})] \, du_1 \, du_2
\]

(7.24)

where \( \bar{X} = T^{2/3} \zeta \), \( (P_k^{(p,q)}) \) being defined through the recursion:

\[
\begin{align*}
G_0(u_1, u_2, \alpha') &= G(u_1, u_2, \alpha) \\
G_k(u_1, u_2, \alpha') &= R_k(u_1, u_2, \alpha') + \bar{H}_k(u_1, u_2, \alpha') \cdot \nabla F_5(u_1, u_2, \bar{\zeta}) \\
\text{where} \quad R_k(u_1, u_2, \alpha') &= \sum_{(p,q) = (0,0)}^{(1,1)} P_k^{(p,q)} u_1^p u_2^q \\
G_{k+1}(u_1, u_2, \alpha') &= -\nabla \cdot \bar{H}_k(u_1, u_2, \alpha')
\end{align*}
\]

(7.25)

Upon approximating, as in Section 7.3.1, \( f^+ \) by its germ, \( u \) is given by:

\[
\begin{pmatrix}
  u_1 \\
  u_2
\end{pmatrix} = \left(\frac{2}{3}\right)^{1/3} \begin{pmatrix}
  1 & -i \alpha_2^{1/2} \\
  1 & i \alpha_2^{1/2}
\end{pmatrix} \begin{pmatrix}
  z_1 \\
  z_2
\end{pmatrix}
\]

and \( \bar{\zeta} = ((3/2)^{1/3}(1 - \alpha_2), (3/2)^{1/3}(1 - \alpha_2), 0) \).

The canonical functions are then the traces of the hyperbolic umbilic integral and its derivatives taken on the real line \( (x, x, 0) \) in \( \mathbb{C}^3 \) (\( \simeq \mathbb{R}^6 \)), or:

\[
\mathcal{H}^{(p,q,0)}(\bar{X}) = \int_{\Delta} u_1^p u_2^q \exp[u_1^2 + u_2^2 + X_2 u_2 + X_1 u_1] \, du_1 \, du_2
\]
with:

\[ \mathcal{H}^{(p,q,0)} = \frac{\partial^{p+q} \mathcal{H}^{(0,0,0)}}{\partial X_1^p \partial X_2^q} \]

which are easily expressible in terms of the Airy function and its derivatives since:

\[ \mathcal{H}^{(0,0,0)}(X_1, X_2, 0) = 3^{-2/3} \text{Ai}(-3^{-1/3} X_1) \text{Ai}(-3^{-1/3} X_2) \]

Hence we recover the same behavior for the canonical functions present in (7.24) as for those in (7.21), as well as the same dependence in powers of \( T^{-1/3} \). However the coefficients are different in essence. Thus the resulting expansion, (7.24), is a single series of standard form for a steepest descent expansion, where only coalescing saddle points are involved. In addition we expect that (7.24) will provide overlapping regions with the classical steepest descent expansions (\(|1 - X/T| \) away from 0) of significant size.

For completeness, the steps underlying this analysis should be detailed. On a more practical level remain also the questions of obtaining the coefficients of the expansion, and especially of the inversion of the mapping \( u \). This last point is of importance since the form of (7.24) strongly suggests that the size of its region of validity in terms of \( \alpha_2 \) is significant.

### 7.5 Conclusion

In view of this last result, the failure of expansion (7.21) to furnish a uniform approximation of \( F \) at the wave front becomes clearer. Whether decomposed as a fold catastrophe with a coalescing logarithmic branch point, or as a butterfly catastrophe, the result obtained from the method of steepest descents (i.e. expansion (7.21)) is incomplete. The true behavior of \( F \) at the wave front is in fact given, as implied in (7.24), by a hyperbolic umbilic catastrophe, whose nature is intrinsically different.
from any catastrophe of co-rank 1.

This emphasizes the importance of a correct analysis (the main difficulty of this last example) when in presence of a catastrophe. On this analysis rely the identification of the catastrophe and consequently the whole nature of the resulting asymptotic expansion. The fundamental differences between the various catastrophes are of important consequences for numerical applications: there exists no alternative but to deal with the full complexity, i.e. exact identity, of a given case. This implies not only the consideration of catastrophes of higher co-dimensions such as the cusp, swallow-tail or butterfly in the method of steepest descents, but also of catastrophes of higher co-ranks such as the hyperbolic or elliptic umbilics in methods for asymptotic approximation of multi-dimensional integrals. The problem of devising practical methods for the numerical evaluation of expansions other than the fold catastrophe type is still open.
Conclusion

We would like to view this work as contributing to two different areas. The first and most important contribution in our mind concerns systematic implementations and uses of the generalized method of steepest descents for numerical computations. The analytical power of this method, especially after the emergence of catastrophe theory, is well established but we wish to extend it to numerical applications as well.

Following other contributions toward this aim, we were able to focus our efforts on two practical problems.

Hence we considered at first the untreated problem of determining the relevant saddle points, and proposed a rigorous formulation of this problem in the cases of holomorphic functions. This formalism appears promising both for theoretical and practical considerations. It remains nevertheless to extend it to the cases of functions with singular points as well as to propose an implementation. A more fundamental work is required to investigate a practical characterization of basins on which depends this problem, and to prove the conjecture stated in Chapter 3.

The second practical obstacle in an implementation consists in obtaining numerically the coefficients of steepest descent expansions. In the simple cases of the ordinary method of steepest descents, the corresponding natural algorithms appear to preserve the robustness of the method. The main numerical characteristic of these procedures is the eventuality of cancellations whose effects can be limited by the use of sufficiently large mantissas. In this study which can be viewed as a numerical complement of the corresponding work done by Dingle (Dingle, 1973), the role played by symbolic
computation is crucial. In addition, symbolic computation furnishes a frame for implementations which preserve entirely the generality of the method and which are capable of arbitrary precision and exact results.

In considering the simplest case of a generalized steepest descent expansion, i.e. a fold catastrophe, the main difficulty lies in the inversion of the corresponding Levinson's transformation. This requires complementary algorithms both for this inversion and for the computation of the sequences of Chester, Friedman and Ursell. Despite the lack of determination which results from such composite algorithms, we were able to formulate guidelines for the building of robust algorithms for the fold catastrophe expansion of Chester et al.. The application of these results to the Kelvin wave source potential provided a confirmation of our claim of robustness. More importantly this example allowed us to outline the decomposition property of a fold catastrophe expansion in particular, and of catastrophe expansions in general. This property has important consequences for the numerical evaluation of the coefficients of the series and should induce drastic improvements in efficiencies of real time computations.

Further efforts are needed regarding the practical inversions implied by the unfoldings of mappings and which are the cornerstone of the generalized method. Namely, algorithms should be devised:

1. for fold catastrophes which are non isolated,

2. for catastrophes of higher co-dimensions and co-ranks,

although it is unclear whether the growing complexities of catastrophes will not deny any practical use to these procedures.

Along the same line, algorithms for the computation of the Bleistein sequences should be investigated.

The second contribution concerns two fundamental solutions of hydrodynamics.

In treating the wave part of the Kelvin potential we have furnished robust methods
for evaluating the potential, its derivatives or certain integrals, in the far field and
for any submergence of the pair source-field-point. Yet, the whole physical domain
is not fully treated, and a last expansion based on the ordinary method of steepest
descents should be found for the neighborhood of the singular axis for large distances
\((R > 100)\) away from the source. Perhaps the most remarkable feature of this example
is the extent of the region of validity of the fold catastrophe expansion which confirms
the results of Ursell (1960, 1965). This led us to propose a different interpretation
of the interferences taking place in more than half of the Kelvin wedge. However
new this interpretation might seem to hydrodynamicists, it is in fact a well known
phenomenon in optics and is an intrinsic property of the fold catastrophe.

Finally, our consideration of the three-dimensional impulse response Green function in
finite depth emphasizes the importance of asymptotic methods for multi-dimensional
integrals and of the classification of catastrophes. Previous asymptotic treatments
by the methods of stationary phase or steepest descents resulted in an expansion
valid only in a vanishing neighborhood of the wave front. Although the reason of
this failure is not obvious, the method of steepest descents is inappropriate in this
case. Thus we presented a uniform asymptotic expansion in terms of the hyperbolic
umbilic integral based on results from asymptotics of two-dimensional integrals and
from catastrophe theory. The complete justification of this expansion is yet to be
produced as well as practical methods for computing its coefficients.

Hence by this last example, we demonstrate the importance of catastrophe the-
ory in asymptotics of free-surface waves. From similitude with quantum mechanics,
hydrodynamics should gain new physical understandings of wave fronts, interferences
and other phenomena characteristic of catastrophes (e.g. Trulsen, 1991). The ques-
tion of the practicality of numerical implementations of the method of steepest de-
scents is however still open. In the case of an isolated fold catastrophe as for the
Kelvin potential, the answer is positive. But for catastrophes of higher co-dimensions
or co-ranks, or which are not isolated, answers will only be provided through future
investigations.


- Clarisse, J.-M., 1990, “Thirty years after... the evaluation of the single integral part of the Kelvin wave source potential in the far-field”, Fifth International Workshop on Water Waves and Floating Bodies, Manchester, U.K.


- Kelvin, (Lord), 1887a, "On the waves produced by a single impulse in water of any depth, or in dispersive medium", *Philos. Mag.*, 23, pp. 252-255.


Appendix A

On some properties of the sets $\bar{O}_t$

From the properties of the various domains of $\bar{\alpha}$ defined in Chapter 3, we can be more precise about the topological properties of the sets $\bar{O}_t$. There is a genuine interest in doing so since this information can be used when assessing the intersections of such regions.

We distinguish two kinds of subsets $\bar{O}_t$:

1. A subset $\bar{O}^1_t$ is belonging to the first kind, if it satisfies the condition:

   $$\exists s_t : \bar{O}^1_t \subset \bar{O}^1_{s_t}$$

   This corresponds to a connected subset of $D$ for which $\bar{\alpha}$ is either not element of $K_R, C_R$ or $C_C$. That is we have the inclusion:

   $$\bar{O}^1_t \subset O^1_{s_t} \cup \{ R_R \setminus \{C_R \cup C_C \} \}$$

   The expansion corresponding to $\bar{O}^1_t$ is the same as the one associated to $O^1_{s_t}$.
and we have the following property:

*Two subsets of the first kind can only intersect on $R_R \setminus \{C_R \cup C_C\}$, that is on a portion of a Stokes surfaces.*

Thus the geometry of their intersections are “well defined” since they are parts of Stokes surfaces.

2. A subset $O_i^2$ of the second kind is associated to an expansion of any of the forms given in Section 1.2.2. Such a subset is a neighborhood in $D$ of the points of $K_R \cup C_R \cup C_C$.

Unlike subsets of the first kind, we have the property:

*The intersection of a subset $O_i^2$ with another subset of any kind, if it is not empty, is of non-empty interior in $D$.)*

which is directly inherited from the properties of the domains $O_i^2$ of Section 3.3.2.

These intersections are much more difficult to assess as their shapes are not well defined (although they depend on properties of $f$ and $g$, see Martin, 1974), and they require computational intensive explorations.
Appendix B

On certain multiple entire series

Here we are concerned with the existence and forms of certain multiple entire series for the derivatives of $z$ with respect to $u$, taken at $z = z_i$, or the sequences $(z_i^{(n)})$, $i = 1, 2$.

We assume that $\alpha_0$ is a fold catastrophe of $f$ - or in other words that $f$ has a saddle point of order 2, $z_0(\alpha_0)$, for $\alpha = \alpha_0$ - away from any other critical points of $f$. Using the set notations of Chapter 3, this is written as: $\alpha_0 \in K_1 \setminus (K_2 \cup C_{c} \cup C)$.

Consider a neighborhood of $\alpha_0$ in $D$, say $D_2$, small enough, such that for $\alpha$ in $D_2 \setminus K_1$, there exist two, and only two, distinct cols of order 1 of $f$, $z_1(\alpha)$ and $z_2(\alpha)$. Their limit as $\alpha \rightarrow \alpha_0$ is in $D_2 \setminus K_1$, is $z_0(\alpha_0)$. By translation on the variable $(z, \alpha)$, we are reduced to the case $(z_0, \alpha_0) = (0, \bar{0})$.

If we write the conditions for $z_i(\alpha)$ to be a col of order 1 of $f$, we have:

$$0 = f^{(1)}(0, \alpha) + f^{(2)}(0, \alpha) z_i + f^{(3)}(0, \alpha) \frac{z_i^2}{2!} + \Phi(z_i, \alpha) z_i^3 \quad (B.1)$$
with:
\[
\begin{align*}
  f'(0, 0) &= 0 \\
  f''(0, 0) &= 0 \\
  f'''(0, 0) &\neq 0
\end{align*}
\]  
(B.2)

and \( \Phi \) analytic at \((0, 0)\).

Since the function \( f \) is holomorphic, each of its derivatives with respect to \( z \) taken at \( z = 0 \), can be written as a multiple entire series in \( \alpha_1, \ldots, \alpha_t \), or:
\[
\frac{f^{(n)}(0, \vec{\alpha})}{n!} = \mathcal{A}_n(\vec{\alpha}) = \sum_{|J|=0}^{\infty} a_n(J) \alpha_1^{j_1} \ldots \alpha_t^{j_t}
\]  
(B.3)

where \( J \) denotes the \( l \)-index \((j_1, \ldots, j_t)\) and \( |J| = \sum_k j_k \).

More specifically we know that:
\[
\begin{align*}
  \mathcal{A}_1(\vec{\alpha}) &= \sum_{|J|=1}^{\infty} a_1(J) \alpha_1^{j_1} \ldots \alpha_t^{j_t} \\
  \mathcal{A}_2(\vec{\alpha}) &= \sum_{|J|=1}^{\infty} a_2(J) \alpha_1^{j_1} \ldots \alpha_t^{j_t}
\end{align*}
\]  
(B.4)

and \( a_3(0) \neq 0 \).

Let us write:
\[
\beta_1^2 = \mathcal{A}_1(\vec{\alpha})
\]  
(B.5)

so that when substituting in (B.1) we get:
\[
0 = \beta_1^2 + z_i \sum_{|J|=1}^{\alpha_1^{j_1}} a_2(J) \alpha_1^{j_1} \ldots \alpha_t^{j_t} + z_i^2 \sum_{|J|=0}^{\infty} a_3(J) \alpha_1^{j_1} \ldots \alpha_t^{j_t} + \ldots
\]  
(B.6)

Before trying to express \( z_i \) in a multiple power series, it is necessary to perform an additional translation on the variable \( z \).
There exists $z_0(\bar{\alpha})$ such that $f^{(2)}(z_0(\bar{\alpha}), \bar{\alpha}) = 0$ for $\bar{\alpha}$ in $D_2 \setminus K_1$. Thus if we perform the translation $-z_0(\bar{\alpha})$ on $z$, we obtain, instead of (B.6):

$$0 = \beta_1^2 + \tilde{z}_1^2 \sum_{|J|=0}^{\infty} \tilde{a}_3(J) \alpha_1^{j_1} \ldots \alpha_l^{j_l} + \ldots$$  \hspace{1cm} (B.7)

with $\tilde{z}_i(\bar{\alpha}) = z_i(\bar{\alpha}) - z_0(\bar{\alpha})$.

- Since $z_0$ is a function of $\bar{\alpha}$, we are in fact treating a function other than $f(z, \bar{\alpha})$, namely $\tilde{f}(z, \bar{\alpha}) = f(z + z_0(\bar{\alpha}), \bar{\alpha})$. However, these two functions merge as $\bar{\alpha} \to \bar{\alpha}_0$.

The next step is to define the new parameters that will be used in the multiple series expressions of the $z_i^{(n)}$ (which should rather be denoted $\tilde{z}_i^{(n)}$). Thus, we choose $\beta_1$ as given by (B.5). The other parameters are chosen upon expressing one of the $\alpha_k$'s in terms of $\beta_1$ and of the remaining $\alpha_k$'s. This is done as follows: there exists at least a multi-index $J_0$, such that $\tilde{a}_1(J_0) \neq 0$, and such that $J_0 = (j_{1,0}, 0, \ldots, 0)$ with $j_{1,0}$ minimal. Then $\alpha_1^{j_{1,0}}$ can be expressed as a function of $(\beta_1, \alpha_2, \ldots, \alpha_l)$, analytic at $\tilde{0}$. (This might require a redefinition of $\beta_1$ if $j_{1,0} > 1$, in order to have a series entire in $\beta_1$.) For the other parameters, we simply choose $\beta_2 = \alpha_2, \ldots, \beta_l = \alpha_l$.

Hence, we have:

$$0 = \beta_1^2 + \tilde{z}_1^2 \sum_{|J|=0}^{\infty} \tilde{a}_3(J) \beta_1^{j_1} \ldots \beta_l^{j_l} + \tilde{z}_i^3 \Phi(\tilde{z}_1, \tilde{\beta})$$  \hspace{1cm} (B.8)

from which we deduce that:

$$\tilde{z}_i = (-1)^{i+1} \frac{\beta_1}{\tilde{a}_3(0)} \rho_0((-1)^{i+1} \beta_1, \beta_2, \ldots, \beta_l)$$  \hspace{1cm} (B.9)

where $\rho(\beta_1, \beta_2, \ldots, \beta_l)$ represents a multiple series of the form given in (5.3.4).

It can then be shown, using (5.24) and (5.27), that:

$$\tilde{\xi}^{1/2} \propto \beta_1 \rho_\xi(\beta_1^2, \beta_2, \ldots, \beta_l)$$  \hspace{1cm} (B.10)
and for \( n \geq 1 \):
\[
\tilde{z}_i^{(n)} \propto r_n((-1)^{i+1} \beta_1, \beta_2, \ldots, \beta_i)
\]  
(B.11)

- These results are just the generalization of the results for \( l = 1 \) of Chester, Friedman and Ursell (1957). It is important to notice that the series thus obtained correspond to the inversion of the Levinson’s transformation associated with \( \tilde{f} \).

The relations between \( \tilde{z}_i^{(n)} \) and \( z_i^{(n)} \)

Since we are primarily interested in \( z_i^{(n)} \), we ought to find the relationships between \( \tilde{z}_i^{(n)} \) and \( z_i^{(n)} \). We have already, by definition of \( \tilde{f}(z, \bar{\alpha}) \):
\[
z_i(\bar{\alpha}) = \tilde{z}_i(\bar{\alpha}) + z_0(\bar{\alpha})
\]  
(B.12)

and with no further difficulty, we deduce that:
\[
f_i^{(n)} = \tilde{f}_i^{(n)}
\]  
(B.13)

In addition, it can be seen from (5.24) that \( \zeta \) is identical to \( \zeta \), and thus that \( z_i^{(n)} = \tilde{z}_i^{(n)} \).

Thus everything is equivalent to the local inversion of the Levinson’s transformation for \( f \), except for the values of \( z_i(\bar{\alpha}) \) which are not needed in the subsequent computations.

Finally, if we could obtain the various multiple entire series in the parameter \( \beta \), of the functions \( \zeta^{1/2} \) and \( f_i^{(n)} \), it would be possible to obtain the series for \( z_i^{(n)} \) by using successively the equations of (5.27).
The case of $\bar{\alpha}_0$ describing part of $K_1$

All the above results are based on a fixed value of $\bar{\alpha}_0$ in $K_1 \setminus \{K_2 \cup C_c \cup C\}$. In the most general case, $\bar{\alpha}_0$ is allowed to describe a whole region of $K_1$. This happens in the consideration of any catastrophe of higher co-dimension such as the cusp (i.e. Pearcey’s integral), the swallowtail, etc... There, the fold catastrophe appears along parts of the caustic set, and therefore not only at point values of the parameter $\bar{\alpha}$. The results in entire series are easily extended to such cases simply by considering the coefficients $a(J)$ to be functions of $\bar{\alpha}_0$, as long as, of course, $\bar{\alpha}_0$ stays in $K_1 \setminus \{K_2 \cup C_c \cup C\}$. This implies that the process of obtaining the coefficients of the series would have to be repeated at each new value of $\bar{\alpha}_0$. Moreover, the uniform treatment of values of $\bar{\alpha}_0$ in a neighborhood of $K_1$ would require the expansions of these coefficients in entire series of $\bar{\alpha}_0$. We can perceive here, one aspect of the increasing complexity of catastrophe of higher co-dimension, and henceforth could question the practicality of their uniform treatment.
Appendix C

Root extractions in the complex plane

The methods presented in Chapters 4 and 5 involve in several places the extractions of complex roots. These operations require the definition of certain branch cuts in the complex plane. In an implementation of these methods, artificial discontinuities may then appear along the selected cuts. Such discontinuities are inadmissible and necessitate some refinements in the algorithms for the complex roots.

For this purpose, we simply apply the continuity property of the complex $n$-th root, $\sqrt[n]{\cdot}$, to paths of the complex plane. Thus we know that for any path $\Gamma$, its image by $\sqrt[n]{\cdot}$ must be connected. This translates into the following definition of the complex $n$-th root, based on the finite discrete equivalents of the paths $\Gamma$.

*Given $z$ and its argument $\arg z$, $\sqrt[n]{z}$ is given by: $\sqrt[n]{|z|} \exp[i \arg z/n]$.*

The continuity of the function thus defined relies on the determination of the argument. The principle of this determination is as follows:

*For a given $z$, say $z_{\text{ref}}$, the argument of $z_{\text{ref}}$ is chosen to be its determination in $[0, 2\pi]$.***
For any $z$ other than $z_{\text{ref}}$, we construct a finite sequence:

$$(z_0 = z_{\text{ref}}, z_1, \ldots, z_{j-1}, z_j = z)$$

such that $|\arg z_{i+1} - \arg z_i| < \pi$ ($\leq \pi/2$ is sufficient by example) for one of their determinations. In practice this is always possible if $z_{\text{ref}}$ and $z$ are finite.

Then $\arg z$ is determined from this sequence, using the above constraint to determine all the arguments of the intermediate points $z_i$.

In theory, the whole process would have to be repeated for each new value of $z$. However in our applications, this task can be reduced to the stage of pre-computations. Indeed, it suffices to determine the regions of the $\tilde{\alpha}$-space in which the quantities of interest, say $z$, satisfies the condition $|\arg z - \arg z_{\text{ref}}| < \pi$ for some reference value $z_{\text{ref}}$. Then the determination of the arguments needs only to be performed for these reference values.

This approach guarantees the uniformity in $\tilde{\alpha}$ of the results for the various roots involved in our procedures.
Appendix D

The computation of the coefficients in the multiple entire series

The usefulness of the entire series method for the computation of $z_i^{(n)}$ in the vicinity of $\alpha_0$, relies on the ability of computing the coefficients of these series. Two aspects intervene in this matter:

- the local transformation of $f(z, \alpha)$ into $f(z, \beta)$

- the numerical evaluation of the coefficients

The second aspect depends entirely on the possibility of defining the new parameters $\beta$, but we consider it at first.

1. Numerical evaluation of the coefficients

We assume that the functions $z_i$ and $f_i^{(n)}$ are known as functions of $\beta$ in the sense that for any $\beta$ in a neighborhood of the origin, the numerical values of $z_i(\beta)$ and of the derivatives $f_i^{(n)}$ up to some finite $n$, are available.

Since all the above functions of $\beta$ are analytic at $\tilde{u}$, we have the following formulas
for the coefficients of the entire series:

\[ a[z_1^{(n)}](J) = \frac{1}{(2\pi i)^l} \oint \cdots \oint \frac{z_1^{(n)}}{\beta_1^{i_1+1} \cdots \beta_l^{i_l+1}} \, d\beta_1 \cdots d\beta_l \quad |J| \geq 0 \]  

(D.1)

where, by definition:

\[ z_1^{(n)} = a[z_1^{(n)}](0) + \sum_{|J|=1}^{\infty} a[z_1^{(n)}](J) \beta_1^{i_1} \cdots \beta_l^{i_l} \]  

(D.2)

and where the \( k \)-th contour integral is around 0 in the variable \( \beta_k \). (From now on, we will only consider \( z_1^{(n)} \) since the series for \( z_2^{(n)} \) is obtained by changing \( \beta_1 \) into \(-\beta_1\).)

We emphasize that the contours do not depend on \( \beta \). As a consequence, our objections to the use of contour integral methods in computing the asymptotic series are not relevant here.

In theory it suffices to perform numerically the multiple integrations along closed contours. However, by imposing these contours to be circles in the \( \beta_k \)-planes (i.e. poly-circles), we benefit from the equivalence between the coefficients of the entire series and the coefficients of the Fourier series of \( z_1^{(n)} \) taken as a function of the arguments of the \( \beta_k \)'s. This relationship between the two series is of special interest since the coefficients of the Fourier series can be obtained accurately and efficiently using Fast Fourier Transform algorithms. More precisely, we have the relation:

\[ a[z_1^{(n)}](J) = \frac{1}{(2\pi i)^l} \frac{1}{|\beta_1|^{i_1} \cdots |\beta_l|^{i_l}} c[z_1^{(n)}](J) \]  

(D.3)

where \( c[z_1^{(n)}](J) \) denotes the coefficients of the Fourier series, and \(|\beta_k|\) the radius of the circle in \( \beta_k \).

The computation of the Fourier coefficients is done by using finite sets of points defined to satisfy the following constraints:

1. If \( \max j_k = J_k - 1 \) is the maximum order in \( \beta_k \), the corresponding set is given by: \( \{ \beta_{k,0} \exp[i2\pi p/2J_k] \mid 1 \leq p \leq 2J_k \} \).
2. For all \( k \), the radius \( |\beta_{k,0}| \) is chosen such that the Levinson's transformation exists. (This gives upper bounds for the radii.)

3. For all \( k \), the radius \( |\beta_{k,0}| \) is chosen such that cancellation errors in the formulas of (5.27) have minimal effects on the accuracy of the numerical result of \( z_1^{(n)} \). (This imposes lower bounds on the radii.)

4. The parameters \( \beta_k \) are eventually rescaled to improve numerical accuracy, so that \( |\beta_{k,0}| = 1 \). This is done through a re-definition of \( f(z,\bar{\beta}) \).

With such sets of points the Fourier coefficients are computed from the numerical values of \( z_1^{(n)} \) obtained at these points by applying the formulas of (5.27).

The advantage of this approach is to use the direct method where it is accurate, and of being efficient while staying simple as the number of parameters increases. The drawbacks are its dependence upon the generation of equally spaced points on circles of the \( \bar{\beta} \)-space.

2. The local re-parameterization of \( f \)

The generation of these circles is the practical aspect of the definition of the new parameters described in Appendix B. This ability is not granted a priori by the kind of knowledge of \( f \) that we have assumed so far, and therefore necessitates some additional effort.

For this purpose we need to go back to the derivation of the series. This derivation was based on the consideration of the function:

\[
\tilde{f}(z,\bar{\alpha}) = f(z + z_0(\bar{\alpha}),\bar{\alpha})
\]

with \( z_0(\bar{\alpha}) \) satisfying:

\[
f^{(2)}(z_0(\bar{\alpha}),\bar{\alpha}) = 0
\]
If \( z_0(\tilde{\alpha}) \) is not known a priori, it can be approximated by solving numerically this equation (note that \( f^{(3)}(0, \tilde{\alpha}) \neq 0 \)). Hence, we require to be able to determine not only \( z_1(\tilde{\alpha}) \) and \( z_2(\tilde{\alpha}) \), but also this particular point \( z_0(\tilde{\alpha}) \). However this additional constraint is restricted to a neighborhood of \( \tilde{\alpha}_0 \).

Once this value is available, we recover (B.7) by considering the entire series of \( \tilde{f}(z, \tilde{\alpha}) \) at \( z = 0 \), or:

\[
0 = \beta_1^2 + \tilde{z}_1 \sum_{|J|=0}^{\infty} \tilde{a}_3(J) \alpha_1^j \cdots \alpha_k^j + \ldots
\]

with \( \beta_1^2 = \lambda_1(\tilde{\alpha}) \).

The definition of the new parameters is based on the application of the implicit function theorem to the latter equality. This inversion can be performed symbolically, and the coefficients in the series of \( \lambda_1(\tilde{\alpha}) \) being computed numerically using again FFT algorithms. This allows us to obtain the value of say \( \alpha_1 \) in terms of \( \beta_1 \) and \( \alpha_k \) for \( k \geq 2 \).

Thus finally, we can obtain a finite sequence of correspondences:

\[
(\beta_1, \ldots, \beta_l) \leftrightarrow (\alpha_1, \ldots, \alpha_l)
\]

as each \( \beta_k \) describes a set of equally spaced points on a circle of radius \( |\beta_k| \) and centered at the origin. The generation of such a sequence involves extractions of complex roots (in the definition of \( \beta_1 \) by example) which are treated following the principle described in Appendix C. This set of correspondences is then used directly for the computation of the Fourier coefficients.

The fact that this procedure needs to be performed only once, compensates for its complexity. However, if \( \tilde{\alpha}_0 \) describes a whole subset, that is when \( K_1 \) is a manifold of dimension 1 or higher, the procedure becomes extremely heavy: the whole process would have to be repeated for possibly all values of \( \tilde{\alpha}_0 \) which are of interest! Undoubtedly another method must be sought to treat such cases.
Appendix E

Higher order terms of the series

$A_s$, $B_s$, $C_s$, $D_s$

The results we present here are for the higher order terms of the fold catastrophe expansion of Chapter 6, for which power series approximations in the neighborhood of the fold catastrophe are not available due to the fixed arbitrary bound on $n$: $n \leq 13$. Hence for $|\zeta^{1/2}|$ sufficiently large, we can compute the values of $A_s$, for $4 \leq s \leq 9$, $B_s$, for $3 \leq s \leq 7$, $C_s$, for $3 \leq s \leq 9$ and $D_s$, for $4 \leq s \leq 10$. The moduli of these coefficients divided by the corresponding powers of $N$, over the computational domain $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$ in terms of $(\theta, \varphi)$, are given respectively in Figures E-1, E-2, E-3 and E-4.

Since no complementary approximations are available for the neighborhood of the fold catastrophe where the direct algorithms are ill-behaved numerically, the values are not computed there. In principle, the regions left by the indentations of the computational domains for these coefficients, should correspond to regimes where the direct formulas are well-behaved numerically. A criterion which would allow these truncations to be defined rigorously should be based on upper bounds of the slopes of the moduli. In practice the definition of such criteria is rather difficult, since the bounds depend on the variations of each of these functions. The results shown in the following figures, and which are far from being perfect in view of the spikes present...
in Figures E-1 and E-4, were obtained by arbitrary truncations based on contour plots and upper bounds applied to the moduli themselves. The consideration of these high order coefficients in such extreme regions, is only motivated by the search for maximum accuracy of the expansions, which has a sense only for large values of $N$. 

287
Figure E-1: $|A_s|/N^{2s}$ at $R = 16$, for $4 \leq s \leq 9$ and $(\theta, \varphi)$ in $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$
Figure E-2: $|B_s|/N^{2s}$ at $R = 16$, for $3 \leq s \leq 7$ and $(\theta, \varphi)$ in $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$
Figure E-3: $|C_s|/N^{2s}$ at $R = 16$, for $3 \leq s \leq 9$ and $(\theta, \varphi)$ in $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$
Figure E-4: $|D_s|/N^{2s}$ at $R = 16$, for $4 \leq s \leq 10$ and $(\theta, \varphi)$ in $[8^\circ, 40^\circ] \times [40^\circ, 90^\circ]$