

MIT Open Access Articles

Generalized Derivatives of Lexicographic Linear Programs

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

As Published: <https://doi.org/10.1007/s10957-018-1309-2>

Publisher: Springer US

Persistent URL: <https://hdl.handle.net/1721.1/131873>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Generalized Derivatives of Lexicographic Linear Programs

Jose Alberto Gomez¹ · Kai Höffner¹ ·

Kamil A. Khan^{1,2} · Paul I. Barton¹

Communicated by Alexander Martin.

Received: date / Accepted: date

Abstract Lexicographic linear programs are fixed-priority multiobjective linear programs that are a useful model of biological systems using flux balance analysis and for goal-programming problems. The objective function values of a lexicographic linear program as a function of its right-hand side are nonsmooth. This work derives generalized derivative information for lexicographic linear programs using lexicographic directional derivatives to obtain elements of the Bouligand subdifferential (limiting Jacobian). It is shown that elements of the limiting Jacobian can be obtained by solving related linear programs. A nonsmooth equation solving problem is solved to illustrate the

Paul I. Barton (pib@mit.edu) ✉

Jose Alberto Gomez (jagomezr@alum.mit.edu)

Kai Höffner (kai.hoeffner@gmail.com)

Kamil A. Khan (kamilkhan@mcmaster.ca)

¹ Process Systems Engineering Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA, 02139, jagomezr@alum.mit.edu

² Department of Chemical Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada

benefits of using elements of the limiting Jacobian of lexicographic linear programs.

Keywords Linear Programming · Lexicographic Linear Programming · Lexicographic Optimization · Nonsmooth Analysis · Generalized Derivatives · Lexicographic Differentiation · Nonsmooth Equation Solving

Mathematics Subject Classification (2010) 49J52 · 90C05 · 90C29 · 90C31

1 Introduction

A lexicographic linear program (LLP) is a fixed-priority multiobjective linear program (LP) that is a useful model of biological systems using flux balance analysis (FBA) and dynamic FBA (DFBA) [1,2], and for goal-programming problems [3]. These types of models can also be embedded in optimization problems, equation-solving problems, or dynamic systems; therefore, the sensitivity analysis of LLPs is of interest. For example, DFBA can be used to model the dynamics of bioreactors. The optimization of DFBA systems will enable the optimal design of industrial bioprocesses employing microbial communities [4]. Another example comes from bilevel optimization, which is used to model Stackelberg games [5,6], a popular model for markets in economics. When the lower-level optimization problem is a LP, this LP can be expressed as its equivalent KKT conditions to reformulate the bilevel problem as a single-level problem with equilibrium constraints. This reformulation results in bilinear terms which are difficult to handle by optimization algorithms due to

violation of constraint qualifications and nonconvexity. In addition, sensitivity results for nonlinear programs typically require constraint qualifications that would be violated by equilibrium constraints. If all communication between the lower-level LP and upper-level optimization problem can be expressed as a sequence of hierarchical objective functions, a LLP is obtained. The optimal values function of a LLP is inherently more regular than the solution set of a LP as a function of its right-hand side ([1, 2]). However, the objective function values of a LLP are nonsmooth functions of its right-hand side; therefore, the computation of sensitivities is challenging. Generalized derivative information would greatly increase the variety and efficiency of numerical methods that can be applied to problems with LLPs embedded.

In this paper, we develop a theory to compute elements of the B-subdifferential (limiting Jacobian) of LLPs as functions of their right-hand sides. The main tools employed in this development are Nesterov's lexicographic derivatives [7] and the notion of lexicographic directional derivatives (LD-derivatives) [8]. In Section 2, we formally introduce the LLP, present basic theory about LD-derivatives and piecewise affine functions, and derive extensions of the chain rule for directional derivatives. In Section 3, the theory for computing LD-derivatives of the optimal values of a LLP is developed. In Section 4, an example illustrating the relevance of LLPs and the use of LD-derivatives is presented. Finally, conclusions and future work are briefly discussed.

2 Mathematical Preliminaries

Let all norms be the Euclidean norm. Boldface symbols represent vector and matrix-valued quantities. Let V be a subset of a metric space; then $\text{int}(V)$ and $\text{bnd}(V)$ denote the interior and the boundary of V , respectively. Let $L(\mathbb{R}^n; \mathbb{R}^m)$ be the space of linear maps from \mathbb{R}^n to \mathbb{R}^m ; each element of $L(\mathbb{R}^n; \mathbb{R}^m)$ can be identified with an $m \times n$ matrix. The i th column vector of a matrix \mathbf{M} is denoted by \mathbf{m}_i . Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ be the extended real number system. Denote by $GL(n, \mathbb{R})$ the set of all invertible $n \times n$ matrices. Let $\mathbf{0}$ be a vector with all components equal to zero, $\mathbf{1}$ be a vector with all components equal to 1 and \mathbf{e}_i be a vector with all components equal to zero except to the i th component which is equal to one. Let \mathbf{I}_m be the identity matrix with m rows. Let $\mathbf{x}^1, \mathbf{x}^2 \in \mathbb{R}^m$; $\mathbf{x}^1 > \mathbf{x}^2$ if for all $i \in \{1, \dots, m\}$, $x_i^1 > x_i^2$. Let a function f be C^k and PC^k if it is k times continuously and piecewise continuously differentiable in the sense of [9], respectively.

2.1 Definition of LLPs

Consider a LLP parameterized by its right-hand side. For each $i \in \{0, 1, \dots, n_h\}$, let $g^i : \mathbb{R}^{m+i} \rightarrow \overline{\mathbb{R}}$. For $\mathbf{z}_i \in \mathbb{R}^{m+i}$ and $i \in \{0, \dots, n_h\}$:

$$g^0(\mathbf{z}_0) = \inf_{\mathbf{v} \in \mathbb{R}^{n_v}} \mathbf{c}_0^T \mathbf{v}, \quad \text{s.t. } \mathbf{A}\mathbf{v} = \mathbf{z}_0, \quad \mathbf{v} \geq \mathbf{0}, \quad (1)$$

$$g^i(\mathbf{z}_i) = \inf_{\mathbf{v} \in \mathbb{R}^{n_v}} \mathbf{c}_i^T \mathbf{v}, \quad \text{s.t. } \begin{bmatrix} \mathbf{A}^T \mathbf{c}_0 & \dots & \mathbf{c}_{i-1} \end{bmatrix}^T \mathbf{v} = \mathbf{z}_i, \quad \mathbf{v} \geq \mathbf{0}. \quad (2)$$

Assumption 2.1 Let \mathbf{A} be of full row rank. For all $i \in \{0, \dots, n_h\}$, let \mathbf{A} and \mathbf{c}_j , with $j = 0, \dots, i$, be such that $g^i(\mathbf{z}_i) > -\infty$ for all $\mathbf{z}_i \in \mathbb{R}^{m+i}$. In addition for all $i > 0$, let $\left[\mathbf{A}^T \mathbf{c}_0 \dots \mathbf{c}_{i-1} \right]^T$ be full row rank.

The need for \mathbf{A} to be full row rank is not limiting in any sense. Linearly dependent rows imply redundant constraints. Deleting redundant constraints from \mathbf{A} and \mathbf{z} in LP (1) and redundant cost vectors in LP (2) results in LLPs with the same optimal solution set and same objective function values. Let $F := \{\mathbf{z} \in \mathbb{R}^m : -\infty < g^0(\mathbf{z}) < +\infty\}$, $\mathbf{A}^0 := \mathbf{A}$ and $\mathbf{A}^i := \left[\mathbf{A}^T \mathbf{c}_0 \dots \mathbf{c}_{i-1} \right]^T$ for $i \in \{1, \dots, n_h\}$. Only LLPs satisfying Assumption 2.1 will be considered.

Definition 2.1 Let Assumption 2.1 hold. Let $\mathbf{q}^i : F \rightarrow \mathbb{R}^{m+i}$, $\mathbf{q}^0 : \mathbf{z} \mapsto \mathbf{z}$ and for $i \in \{1, \dots, n_h\}$, $\mathbf{q}^i : \mathbf{z} \mapsto \left[\mathbf{z}^T g^0(\mathbf{q}^0(\mathbf{z})) \dots g^{i-1}(\mathbf{q}^{i-1}(\mathbf{z})) \right]^T$. Let $\mathbf{h} : F \rightarrow \mathbb{R}^{n_h+1}$ such that $h_i = g^i \circ \mathbf{q}^i$.

Remark 2.1 The function \mathbf{h} is the vector of optimal values of an LLP parameterized by its right-hand side.

Proposition 2.1 Consider LP (1) and suppose \mathbf{A} and \mathbf{c}_0 are such that for all $\mathbf{z} \in \mathbb{R}^m$, $g^0(\mathbf{z}) > -\infty$. \mathbf{A} is full row rank if and only if $\text{int}(F) \neq \emptyset$.

Proof: The proof for Proposition 2.1 can be found in Appendix A. □

Assumption 2.1 implies F is nonempty by Proposition 2.1. For $i \in \{0, \dots, n_h\}$, let $F_i := \{\mathbf{z}_i \in \mathbb{R}^{m+i} : -\infty < g^i(\mathbf{z}_i) < +\infty\}$. Under Assumption 2.1, all sets F_i are closed [1] and convex [10]. The functions g^i are convex on F_i [10]. Proposition 2.2 shows that if F is nonempty, all sets F_i are nonempty.

Proposition 2.2 *Let Assumption 2.1 hold and let $\mathbf{b} \in F$. Then, for all $i \geq 1$, $\mathbf{q}^i(\mathbf{b}) \in F_i$.*

Proof: By Assumption 2.1, $g^i(\mathbf{z}_i) > -\infty$ for any $\mathbf{z}_i \in \mathbb{R}^{m+i}$, $i \in \{0, \dots, n_h\}$. Then, we only need to show that $g^i(\mathbf{q}^i(\mathbf{b})) < +\infty$, which is equal to LP (2) being feasible for $i = \{1, \dots, n_h\}$. Assume $\mathbf{q}^{i-1}(\mathbf{b}) \in F_{i-1}$ for $i \in \{1, \dots, n_h\}$. Then, there exists $\mathbf{v}_0 \geq \mathbf{0}$ such that $\mathbf{c}_{i-1}^T \mathbf{v}_0 = h_{i-1}(\mathbf{b})$ and $\mathbf{A}^{i-1} \mathbf{v}_0 = \mathbf{q}^{i-1}(\mathbf{b})$. Then, $\mathbf{A}^i \mathbf{v}_0 = \mathbf{q}^i(\mathbf{b})$, $g^i(\mathbf{q}^i(\mathbf{b})) = h_i(\mathbf{b}) \leq \mathbf{c}_i^T \mathbf{v}_0 < +\infty$, and $\mathbf{q}^i(\mathbf{b}) \in F_i$. Since $\mathbf{q}^0(\mathbf{b}) = \mathbf{b} \in F_0$, the proof follows by induction and $\mathbf{q}^i : F \rightarrow F_i$ for all i . \square

Remark 2.2 Let Assumption 2.1 hold and let $\mathbf{z} \in F$. Then $\mathbf{h}(\mathbf{z}) \in \mathbb{R}^{n_h+1}$.

2.2 Lexicographically Smooth Functions

Definition 2.2 Let $X \subset \mathbb{R}^m$ and $\boldsymbol{\rho} : X \rightarrow \mathbb{R}^n$. Let $\mathbf{x} \in X$ and $\mathbf{d} \in \mathbb{R}^m$ be such that there exists $\delta_{\mathbf{d}} > 0$ such that for all $\epsilon \in]0, \delta_{\mathbf{d}}[$, $\mathbf{x} + \epsilon \mathbf{d} \in X$. The (one-sided) *directional derivative* of $\boldsymbol{\rho}$ at $\mathbf{x} \in X$ in the direction $\mathbf{d} \in \mathbb{R}^m$ is given by $\boldsymbol{\rho}'(\mathbf{x}; \mathbf{d}) := \lim_{\tau \downarrow 0} \frac{\boldsymbol{\rho}(\mathbf{x} + \tau \mathbf{d}) - \boldsymbol{\rho}(\mathbf{x})}{\tau}$, if the limit exists in \mathbb{R}^n . If at \mathbf{x} , the limit exists in \mathbb{R}^n for any $\mathbf{d} \in \mathbb{R}^m$, then, $\boldsymbol{\rho}$ is *directionally differentiable* at \mathbf{x} .

Note that in Definition 2.2 the set X is not required to be open. If $\mathbf{x} \in \text{int}(X)$, Definition 2.2 reduces to the standard definition of directional derivative.

Definition 2.3 [7]. Let $X \subset \mathbb{R}^m$ be open and $\boldsymbol{\rho} : X \rightarrow \mathbb{R}^n$ be Lipschitz near $\mathbf{x} \in X$ and directionally differentiable. $\boldsymbol{\rho}$ is *lexicographically smooth* (or *l-smooth*) at \mathbf{x} if for any $q \in \mathbb{N}$ and any matrix $\mathbf{M} = [\mathbf{m}_1 \dots \mathbf{m}_q] \in \mathbb{R}^{m \times q}$ the

following functions are well defined:

$$\begin{aligned}\rho_{\mathbf{x},\mathbf{M}}^{(0)} : \mathbb{R}^m &\rightarrow \mathbb{R}^n : \mathbf{d} \mapsto \rho'(\mathbf{x}; \mathbf{d}), \\ \rho_{\mathbf{x},\mathbf{M}}^{(j)} : \mathbb{R}^m &\rightarrow \mathbb{R}^n : \mathbf{d} \mapsto [\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \mathbf{d}), \quad \forall j \in \{1, \dots, q\}.\end{aligned}\tag{3}$$

From Lemma 3 in [7],

$$\begin{aligned}\rho_{\mathbf{x},\mathbf{M}}^{(k)}(\tau \mathbf{d}) &= \tau \rho_{\mathbf{x},\mathbf{M}}^{(k)}(\mathbf{d}), \quad \forall \mathbf{d} \in \mathbb{R}^m, \forall \tau \geq 0, \\ \rho_{\mathbf{x},\mathbf{M}}^{(k)}(\mathbf{d} + \tau \mathbf{y}) &= \rho_{\mathbf{x},\mathbf{M}}^{(k)}(\mathbf{d}) + \tau \rho_{\mathbf{x},\mathbf{M}}^{(k)}(\mathbf{y}), \quad \forall \mathbf{d} \in \mathbb{R}^m, \forall \mathbf{y} \in \text{span}\{\mathbf{m}_1, \dots, \mathbf{m}_k\},\end{aligned}\tag{4}$$

$\forall \tau \in \mathbb{R}$, and all $k \in \{0, \dots, q\}$. Also, for all $\mathbf{d} \in \text{span}\{\mathbf{m}_1, \dots, \mathbf{m}_k\}$ and all $k \in \{1, \dots, q-1\}$, $\rho_{\mathbf{x},\mathbf{M}}^{(k)}(\mathbf{d}) = \rho_{\mathbf{x},\mathbf{M}}^{(k+1)}(\mathbf{d}) = \dots = \rho_{\mathbf{x},\mathbf{M}}^{(q)}(\mathbf{d})$. These relations imply that $\rho_{\mathbf{x},\mathbf{M}}^{(k)}$ is linear on $\text{span}\{\mathbf{m}_1, \dots, \mathbf{m}_k\}$. From Lemma 2.1 in [11]:

$$\rho_{\mathbf{x},\mathbf{M}}^{(k-1)}(\mathbf{m}_k) = \rho_{\mathbf{x},\mathbf{M}}^{(k)}(\mathbf{m}_k) = \dots = \rho_{\mathbf{x},\mathbf{M}}^{(q)}(\mathbf{m}_k), \quad \forall k \in \{1, \dots, q\}.\tag{5}$$

Definition 2.4 [7]. If $\mathbf{M} \in \mathbb{R}^{m \times k}$ has full row rank and ρ is l -smooth at \mathbf{x} , then the mapping $\rho_{\mathbf{x},\mathbf{M}}^{(k)} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is linear and the matrix $\mathbf{J}_L \rho(\mathbf{x}; \mathbf{M}) = \mathbf{J} \rho_{\mathbf{x},\mathbf{M}}^{(k)}(\mathbf{0}) \in \mathbb{R}^{n \times m}$ is called the *lexicographic derivative* (l -derivative) of ρ at \mathbf{x} in the directions \mathbf{M} .

Definition 2.5 [7]. Let the function $\rho : X \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$ be l -smooth at $\mathbf{x} \in X$. The set $\partial_L \rho(\mathbf{x}) := \{\mathbf{J}_L \rho(\mathbf{x}; \mathbf{M}) \in \mathbb{R}^{n \times m} : \mathbf{M} \in GL(m, \mathbb{R})\}$ is called the *lexicographic subdifferential* of ρ at \mathbf{x} .

Nesterov shows that l -derivatives exist whenever $\boldsymbol{\rho}$ is l -smooth at \mathbf{x} [7]. If $\boldsymbol{\rho}$ is (Fréchet) differentiable at \mathbf{x} , then $\boldsymbol{\rho}_{\mathbf{x},\mathbf{M}}^{(k)}(\mathbf{d}) = \mathbf{J}\boldsymbol{\rho}(\mathbf{x})\mathbf{d}$, for $k = 0, \dots, q$ and for any $\mathbf{M} \in \mathbb{R}^{m \times q}$. For a scalar function ρ , it has been shown in [7] that $\partial_L \rho(\mathbf{x})$ is a subset of Clarke's generalized Jacobian ($\partial \rho(\mathbf{x})$), hence for any $\mathbf{M} \in GL(m, \mathbb{R})$ we have that $\mathbf{J}\boldsymbol{\rho}_{\mathbf{x},\mathbf{M}}^{(m)}(\mathbf{0}) \in \partial \rho(\mathbf{x})$. The *lexicographic directional derivative* of $\boldsymbol{\rho}$ (or *LD-derivative*) [8] at $\mathbf{x} \in X$ in the directions $\mathbf{M} \in \mathbb{R}^{m \times q}$ is $\boldsymbol{\rho}'(\mathbf{x}; \mathbf{M}) := \left[\boldsymbol{\rho}_{\mathbf{x},\mathbf{M}}^{(0)}(\mathbf{m}_1) \cdots \boldsymbol{\rho}_{\mathbf{x},\mathbf{M}}^{(q-1)}(\mathbf{m}_q) \right] = \left[\boldsymbol{\rho}_{\mathbf{x},\mathbf{M}}^{(q)}(\mathbf{m}_1) \cdots \boldsymbol{\rho}_{\mathbf{x},\mathbf{M}}^{(q)}(\mathbf{m}_q) \right]$, with the equality following from (5). This definition is particularly useful since first, for $\mathbf{M} \in GL(m, \mathbb{R})$ the LD-derivative and the l -derivative are related by $\boldsymbol{\rho}'(\mathbf{x}; \mathbf{M}) = \mathbf{J}_L \boldsymbol{\rho}(\mathbf{x}; \mathbf{M})\mathbf{M}$. Second, the chain rule for LD-derivatives has a simple and intuitive structure. Let $p \in \mathbb{N}$ and Y be an open subset of \mathbb{R}^p , let $\boldsymbol{\zeta} : X \subset \mathbb{R}^m \rightarrow Y$ and $\boldsymbol{\rho} : Y \rightarrow \mathbb{R}^n$ be l -smooth at $\mathbf{x} \in X$ and $\boldsymbol{\zeta}(\mathbf{x})$, respectively. The LD-derivative of the l -smooth composition of $\boldsymbol{\rho} \circ \boldsymbol{\zeta}$ at $\mathbf{x} \in X$ is given by:

$$[\boldsymbol{\rho} \circ \boldsymbol{\zeta}]'(\mathbf{x}; \mathbf{M}) = \boldsymbol{\rho}'(\boldsymbol{\zeta}(\mathbf{x}); \boldsymbol{\zeta}'(\mathbf{x}; \mathbf{M})). \quad (6)$$

All differentiable functions and convex functions are l -smooth [7]. Also, piecewise differentiable functions in the sense of Scholtes [9] are l -smooth and their l -derivatives are elements of the B-subdifferential [8]. If $\mathbf{M} = \mathbf{I}_m$, the LD-derivative is an element of the B-subdifferential from $\boldsymbol{\rho}'(\mathbf{x}; \mathbf{M}) = \mathbf{J}_L \boldsymbol{\rho}(\mathbf{x}; \mathbf{M})\mathbf{M}$. Detailed examples that illustrate LD-derivatives can be found in [12].

2.3 Piecewise Linear and Piecewise Affine Functions

Definition 2.6 [9]. A continuous function $\boldsymbol{\rho} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ is called *piecewise linear* (*affine*) if a finite set of linear (affine) functions $\boldsymbol{\rho}_i : \mathbb{R}^m \rightarrow \mathbb{R}^n$ exists such that the inclusion $\boldsymbol{\rho}(\mathbf{x}) \in \{\boldsymbol{\rho}_1(\mathbf{x}), \dots, \boldsymbol{\rho}_k(\mathbf{x})\}$ holds for every $\mathbf{x} \in \mathbb{R}^m$. The linear (affine) functions $\boldsymbol{\rho}_i$ are called *selection functions*.

Since linear functions are also affine, piecewise linear functions are also piecewise affine. Every piecewise affine function is Lipschitz continuous (Proposition 2.2.7 in [9]). Piecewise affine functions are closed under composition [9]. In fact, given two piecewise affine functions $\boldsymbol{\rho} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $\boldsymbol{\zeta} : \mathbb{R}^n \rightarrow \mathbb{R}^p$, with affine selection functions $\{\boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_k\}$ and $\{\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_l\}$, the selection functions of $\boldsymbol{\zeta} \circ \boldsymbol{\rho}$ are in the set that considers all possible compositions $\boldsymbol{\zeta}_i \circ \boldsymbol{\rho}_j$ for $i \in \{1, \dots, l\}$ and $j \in \{1, \dots, k\}$ [9]. Therefore, the composition of two piecewise linear functions is piecewise linear as all selection functions are linear functions and linear functions are closed under composition.

Lemma 2.1 [9]. Let $\boldsymbol{\rho} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ be a piecewise affine function. Then for any point $\mathbf{x}_0 \in \mathbb{R}^m$, there exists $\delta > 0$ such that for any \mathbf{x} such that $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, $\boldsymbol{\rho}(\mathbf{x}) = \boldsymbol{\rho}(\mathbf{x}_0) + \boldsymbol{\rho}'(\mathbf{x}_0; \mathbf{x} - \mathbf{x}_0)$.

Proof: The proof of this lemma is in Section 2.2.2 in [9]. □

Proposition 2.3 Consider a piecewise affine function $\boldsymbol{\theta} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ and a function $\boldsymbol{\rho} : S \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$. Suppose that $\boldsymbol{\rho}(\mathbf{x}) = \boldsymbol{\theta}(\mathbf{x})$ for all $\mathbf{x} \in S$. Given $\mathbf{d} \in \mathbb{R}^m$ and $\mathbf{x} \in S$ suppose that there exists $\delta_{\mathbf{d}}^*$ such that for $\tau \in]0, \delta_{\mathbf{d}}^*[$, $\mathbf{x} + \tau \mathbf{d} \in S$. Then, $\boldsymbol{\rho}$ has a directional derivative at \mathbf{x} in the direction \mathbf{d} and

there exists $\delta_{\mathbf{d}} \in]0, \delta_{\mathbf{d}}^*]$ such that for $0 \leq \epsilon < \delta_{\mathbf{d}}$, $\boldsymbol{\rho}(\mathbf{x} + \epsilon \mathbf{d}) = \boldsymbol{\rho}(\mathbf{x}) + \boldsymbol{\rho}'(\mathbf{x}; \epsilon \mathbf{d})$.

Proof: The proof can be found in Appendix A. \square

From now on we shall call functions piecewise linear (affine) if they coincide with a piecewise linear (affine) function on their domain of definition.

Corollary 2.1 *Let $\boldsymbol{\rho} : S \subset \mathbb{R}^m \rightarrow Z \subset \mathbb{R}^n$ be a piecewise affine function and let $\mathbf{d} \in \mathbb{R}^m$, $\mathbf{x} \in S$. If there exists $\delta_{\mathbf{d}}^* > 0$ such that $\mathbf{x} + \tau \mathbf{d} \in S$ for all $\tau \in]0, \delta_{\mathbf{d}}^*]$, then there exists $\delta_{\mathbf{d}}$ such that $\boldsymbol{\rho}(\mathbf{x}) + \tau \boldsymbol{\rho}'(\mathbf{x}; \mathbf{d}) \in Z$ for all $\tau \in]0, \delta_{\mathbf{d}}[$.*

Proof: From Proposition 2.3 there exists $\delta_{\mathbf{d}} \in]0, \delta_{\mathbf{d}}^*]$ such that for $\tau \in [0, \delta_{\mathbf{d}}[$, $\boldsymbol{\rho}(\mathbf{x} + \tau \mathbf{d}) = \boldsymbol{\rho}(\mathbf{x}) + \boldsymbol{\rho}'(\mathbf{x}; \tau \mathbf{d}) = \boldsymbol{\rho}(\mathbf{x}) + \tau \boldsymbol{\rho}'(\mathbf{x}; \mathbf{d})$ (Eqn. (4)). Since $\boldsymbol{\rho}(\mathbf{x} + \tau \mathbf{d}) \in Z$ for all $\tau \in]0, \delta_{\mathbf{d}}[$, then $\boldsymbol{\rho}(\mathbf{x}) + \tau \boldsymbol{\rho}'(\mathbf{x}; \mathbf{d}) \in Z$. \square

Proposition 2.4 *Let $X \subset \mathbb{R}^m$ be open, $\boldsymbol{\rho} : X \rightarrow \mathbb{R}^n$ be PC^1 (piecewise continuously differentiable), $\mathbf{x} \in X$, and $\mathbf{M} \in \mathbb{R}^{m \times q}$. Then $\boldsymbol{\rho}$ is l -smooth at \mathbf{x} and all functions in the sequence in (3) are piecewise linear.*

Proof: $\boldsymbol{\rho}$ is l -smooth at \mathbf{x} since it is piecewise differentiable [8]. Let $\{\boldsymbol{\rho}^1, \dots, \boldsymbol{\rho}^k\}$ be a set of selection functions for $\boldsymbol{\rho}$ at \mathbf{x} with $k \in \mathbb{N}$. From Proposition 4.1.3 in [9] for $i \in \{1, \dots, n\}$, $\mathbf{d} \in \mathbb{R}^m$, $\rho_i'(\mathbf{x}; \mathbf{d}) \in \{\nabla \rho_i^1(\mathbf{x})^T \mathbf{d}, \dots, \nabla \rho_i^k(\mathbf{x})^T \mathbf{d}\}$. From Theorem 3.1.2 in [9], $\boldsymbol{\rho}'(\mathbf{x}; \cdot)$ is globally Lipschitz continuous. It is clear then that all components of the function $\boldsymbol{\rho}'(\mathbf{x}; \cdot)$ are piecewise linear, and therefore $\boldsymbol{\rho}_{\mathbf{x}, \mathbf{M}}^{(0)}$ is piecewise linear and PC^1 . It follows by induction that for $j \in \{1, \dots, q\}$, $\boldsymbol{\rho}_{\mathbf{x}, \mathbf{M}}^{(j)}$ are piecewise linear. \square

2.4 Extensions of Directional Derivatives

Assumption 2.2 Let $X \subset \mathbb{R}^m$ be open, $Z \subset \mathbb{R}^n$, $\rho : X \rightarrow Z$, $\zeta : Z \rightarrow \mathbb{R}^p$ and $\sigma : X \rightarrow \mathbb{R}^p$, where $\sigma := \zeta \circ \rho$. Assume ρ is l -smooth at $\mathbf{x} \in X$ and ζ is locally Lipschitz continuous. Note that Z is not required to be open.

Definition 2.7 Let Assumption 2.2 hold, let $\mathbf{K} \in \mathbb{R}^{m \times q}$, $\mathbf{M} = \rho'(\mathbf{x}; \mathbf{K})$ and consider the sets $G_{\mathbf{x}}^{(i)} := \{\rho_{\mathbf{x}, \mathbf{Z}}^{(i)}(\mathbf{z}_{i+1}) : \mathbf{Z} \in \mathbb{R}^{m \times q}, \mathbf{z}_{q+1} \in \mathbb{R}^m\} \subset \mathbb{R}^n$ for $i \in \{0, \dots, q\}$. ζ is said to be ρ -weakly l -smooth at \mathbf{x} if for any matrix \mathbf{K} the following sequence is well-defined:

$$\begin{aligned} \zeta_{\rho(\mathbf{x}), \mathbf{M}}^{(0)} : G_{\mathbf{x}}^{(0)} &\rightarrow \mathbb{R}^p : \mathbf{d} \mapsto \zeta'(\rho(\mathbf{x}); \mathbf{d}), \\ \zeta_{\rho(\mathbf{x}), \mathbf{M}}^{(j)} : G_{\mathbf{x}}^{(j)} &\rightarrow \mathbb{R}^p : \mathbf{d} \mapsto [\zeta_{\rho(\mathbf{x}), \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \mathbf{d}), \forall j \in \{1, \dots, q\}. \end{aligned} \quad (7)$$

Remark 2.3 Let Assumption 2.2 hold, let $\mathbf{K} \in \mathbb{R}^{m \times q}$, $\mathbf{M} = \rho'(\mathbf{x}; \mathbf{K})$, and let ζ be l -smooth at $\rho(\mathbf{x})$. Then, ζ is ρ -weakly l -smooth at \mathbf{x} .

Proposition 2.5 Let $\zeta : Z \subset \mathbb{R}^n \rightarrow \mathbb{R}^p$ be Lipschitz near $\mathbf{z} \in Z$. Let the set $G_{\mathbf{z}}$ contain any $\mathbf{d} \in \mathbb{R}^n$ such that $\zeta'(\mathbf{z}; \mathbf{d})$ exists. Then $\zeta'(\mathbf{z}; \cdot)$ is globally Lipschitz on $G_{\mathbf{z}}$.

Proof: The proof is very similar to that of Theorem 3.1.2 in [9]. Let $\mathbf{d}_1, \mathbf{d}_2$ be in $G_{\mathbf{z}}$. First, $\|\zeta'(\mathbf{z}; \mathbf{d}_1) - \zeta'(\mathbf{z}; \mathbf{d}_2)\| = \lim_{\tau \downarrow 0} \frac{\|\zeta(\mathbf{z} + \tau \mathbf{d}_1) - \zeta(\mathbf{z} + \tau \mathbf{d}_2)\|}{\tau}$, and then $\lim_{\tau \downarrow 0} \frac{\|\zeta(\mathbf{z} + \tau \mathbf{d}_1) - \zeta(\mathbf{z} + \tau \mathbf{d}_2)\|}{\tau} \leq \lim_{\tau \downarrow 0} L \frac{\|\tau \mathbf{d}_1 - \tau \mathbf{d}_2\|}{\tau} = L \|\mathbf{d}_1 - \mathbf{d}_2\|$, where L is the Lipschitz constant for ζ near \mathbf{z} . \square

Corollary 2.2 *Let Assumption 2.2 hold and let ζ be ρ -weakly 1-smooth at \mathbf{x} .*

Then the functions in the sequence (7) are globally Lipschitz continuous.

Proof: Let us assume that for $j \in \{1, \dots, q\}$, $\zeta_{\rho(\mathbf{x}), \mathbf{M}}^{(j-1)}$ is globally Lipschitz continuous on $G_{\mathbf{x}}^{(j-1)}$ with constant L . By Proposition 2.5, $\zeta_{\rho(\mathbf{x}), \mathbf{M}}^{(j)}$ is globally Lipschitz continuous. Since $\zeta_{\rho(\mathbf{x}), \mathbf{M}}^{(0)}$ is globally Lipschitz continuous from Proposition 2.5, the proof follows by induction. \square

Next, a generalization of the chain rule for l -smooth functions is presented.

Theorem 2.1 *Let Assumption 2.2 hold, $\mathbf{M} \in \mathbb{R}^{m \times q}$, and ζ be ρ -weakly 1-smooth at $\mathbf{x} \in X$. Then σ is 1-smooth at \mathbf{x} and $\sigma'(\mathbf{x}; \mathbf{M}) = \zeta'(\rho(\mathbf{x}); \rho'(\mathbf{x}; \mathbf{M}))$.*

Proof: The proof can be found in Appendix A. \square

Remark 2.4 Theorem 2.1 provides a chain rule when $\rho(\mathbf{x}) \in \text{bnd}(Z)$ which is not possible in the classical theory.

Theorem 2.2 *Let Assumption 2.2 hold. Suppose that σ is 1-smooth at \mathbf{x} .*

Let $\rho(\mathbf{x}) \in \text{bnd}(Z)$. Let $G_{\mathbf{x}}^{(j)}$ be the sets described in Definition 2.7. Assume that for all $\mathbf{d} \in \mathbb{R}^m$ there exists $\delta_{\mathbf{d}} > 0$ such that $\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}) \in Z$ for any $\tau \in]0, \delta_{\mathbf{d}}[$. Assume that for any $j \in \{0, \dots, q-1\}$, $\mathbf{m}_{j+1} \in \mathbb{R}^m$, and for $\mathbf{M} \in \mathbb{R}^{m \times q}$ there exists $\delta_{\mathbf{m}_{j+2}} > 0$ such that for all $\tau \in]0, \delta_{\mathbf{m}_{j+2}}[$, $\rho_{\mathbf{x}, \mathbf{M}}^{(j)}(\mathbf{m}_{j+1}) + \tau [\rho_{\mathbf{x}, \mathbf{M}}^{(j)}]'(\mathbf{m}_{j+1}; \mathbf{m}_{j+2}) \in G_{\mathbf{x}}^{(j)}$. Then, ζ is ρ -weakly 1-smooth at \mathbf{x} and for $\mathbf{M} \in \mathbb{R}^{m \times q}$, $\sigma'(\mathbf{x}; \mathbf{M}) = \zeta'(\rho(\mathbf{x}); \rho'(\mathbf{x}; \mathbf{M}))$.

Proof: The proof to Theorem 2.2 can be found in Appendix A and is quite similar to the proof for Theorem 2.1. \square

Remark 2.5 The assumption that for any $\mathbf{m}_{q+1} \in \mathbb{R}^m$, $\mathbf{M} \in \mathbb{R}^{m \times q}$ and $j \in \{0, \dots, q-1\}$, there exists $\delta_{\mathbf{m}_{j+2}} > 0$ such that for all $\tau \in (0, \delta_{\mathbf{m}_{j+2}})$, $\boldsymbol{\rho}_{\mathbf{x}, \mathbf{M}}^{(j)}(\mathbf{m}_{j+1}) + \tau[\boldsymbol{\rho}_{\mathbf{x}, \mathbf{M}}^{(j)}]'(\mathbf{m}_{j+1}; \mathbf{m}_{j+2}) \in G_{\mathbf{x}}^{(j)}$ may seem difficult to verify at first glance. However, if Assumption 2.2 holds and in addition $\boldsymbol{\rho}$ is a PC^1 function, then this assumption follows from Proposition 2.4 and Corollary 2.1.

Theorems 2.1 and 2.2 are extensions of Theorem 3.1.1 in [9] and the chain rule (6) from [7] under weaker assumptions. In our Theorems, ζ is not required to be directionally differentiable; its directional derivatives need to exist in \mathbb{R}^p only in certain directions. The chain rule in [13] cannot be applied if $\boldsymbol{\rho}(\mathbf{x}) \in \text{bnd}(Z)$; however, Theorems 2.1 and 2.2 can deal with this situation.

3 LD-Derivatives of Lexicographic Linear Programs

3.1 Computation of LD-Derivatives of LLPs

In this section, we derive the LD-derivatives of a LLP as a function of its right-hand side. We are aware that the literature suggests approaches such as Proposition 4.12 in [14]. An explanation of why this Proposition is not applicable to our case can be found in Appendix B. Here we use Theorems 2.1 and 2.2 and apply them to piecewise linear functions defined on closed sets. Then, we extend the results in [15] and apply them to LLPs. Next, we obtain the LD-derivatives of a LLP as a function of some components of its right-hand side. Finally, we use a LP that is always feasible to obtain an extended system [16]. This extended system provides a way of dealing with LLPs becoming

infeasible due to numerical inaccuracies in the implementation of optimization and equation solving algorithms.

Assumption 3.1 Let g^i, \mathbf{q}^i and \mathbf{h} for $i \in \{0, \dots, n_h\}$ be defined as in Section 2 and let Assumption 2.1 hold. Assume that $\mathbf{b}_0 \in \text{int}(F)$. Notice that $\text{int}(F)$ is nonempty by Proposition 2.1.

Proposition 3.1 *Let Assumption 3.1 hold. Then \mathbf{h} and \mathbf{q}^i are l -smooth at \mathbf{b}_0 , and \mathbf{h} and \mathbf{q}^i are piecewise linear functions on F for all $i \in \{0, \dots, n_h\}$.*

Proof: From Lemma 3.1 in [8], piecewise differentiable functions in the sense of Scholtes (see Section 4.1 in [9]) are l -smooth on the interior of their domains. Piecewise linear functions are piecewise differentiable functions. Therefore, the proof shows that for $i \in \{0, \dots, n_h\}$, h_i is piecewise linear. All functions g^i are piecewise linear and convex on their respective domains F_i [10]. \mathbf{q}^0 is a linear function on F , therefore, h_0 is piecewise linear on F as it is the composition of a linear function with a piecewise linear function. Now assume that for $i \in \{1, \dots, n_h\}$, h_{i-1} and \mathbf{q}_{i-1} are piecewise linear on F . \mathbf{q}^i is a piecewise linear function on F because both h_{i-1} and \mathbf{q}_{i-1} are piecewise linear on F . Then, h_i is piecewise linear on F because it results from the composition of piecewise linear functions. Since h_0 and \mathbf{q}^0 are piecewise linear on F , it follows by induction that h_i and \mathbf{q}^i are piecewise linear on F for all i . Since $\mathbf{b}_0 \in \text{int}(F)$, both \mathbf{h} and \mathbf{q}^i are l -smooth at \mathbf{b}_0 for $i \in \{0, \dots, n_h\}$. \square

The following proposition assumes \mathbf{q}^i is piecewise affine to obtain a more general result; however, Proposition 3.2 applies to piecewise linear \mathbf{q}^i .

Proposition 3.2 *For $i > 0$ let $F \subset \mathbb{R}^m$ and $F_i \subset \mathbb{R}^{m+i}$ be closed, $g^i : F_i \rightarrow \mathbb{R}$ be piecewise linear and $\mathbf{q}^i : F \rightarrow \text{bnd}(F_i)$ be piecewise affine. Let $h_i = g^i \circ \mathbf{q}^i$. Then for $\mathbf{b}_0 \in \text{int}(F)$, g^i is \mathbf{q}^i -weakly l -smooth at \mathbf{b}_0 and for $\mathbf{M} \in \mathbb{R}^{m \times q}$, $h'_i(\mathbf{b}_0; \mathbf{M}) = [g^i]'(\mathbf{q}^i(\mathbf{b}_0); [\mathbf{q}^i]'(\mathbf{b}_0; \mathbf{M}))$.*

Proof: h_i is piecewise affine as it results from the composition of a piecewise affine function with a piecewise linear function. Then, h_i and \mathbf{q}^i are both l -smooth at \mathbf{b}_0 because they are PC^1 functions. To apply Theorem 2.2, for any $\mathbf{d} \in \mathbb{R}^m$ we need to find $\delta_{\mathbf{d}} > 0$ such that for any $\tau \in]0, \delta_{\mathbf{d}}[$, $\mathbf{q}^i(\mathbf{b}_0) + \tau[\mathbf{q}^i]'(\mathbf{b}_0; \mathbf{d}) \in F_i$. By Remark 2.5, the rest of the assumptions are satisfied because \mathbf{q}^i are PC^1 functions for all i . From Proposition 2.3, since \mathbf{q}^i is piecewise affine there exists $\delta_{\mathbf{d}}^*$ such that $\mathbf{q}^i(\mathbf{b}_0 + \epsilon \mathbf{d}) = \mathbf{q}^i(\mathbf{b}_0) + \epsilon[\mathbf{q}^i]'(\mathbf{b}_0; \mathbf{d})$ for any $\epsilon \in [0, \delta_{\mathbf{d}}^*[$. Since $\mathbf{q}^i(\mathbf{b}_0 + \epsilon \mathbf{d}) \in F_i$ for all ϵ such that $\mathbf{b}_0 + \epsilon \mathbf{d} \in F$, we can set $\delta_{\mathbf{d}} = \delta_{\mathbf{d}}^*$. Then, $h'_i(\mathbf{b}_0; \mathbf{M}) = [g^i]'(\mathbf{q}^i(\mathbf{b}_0); [\mathbf{q}^i]'(\mathbf{b}_0; \mathbf{M}))$ by Theorem 2.2. In addition, Theorem 2.2 establishes that g^i is \mathbf{q}^i -weakly l -smooth at \mathbf{b}_0 . \square

Proposition 3.3 *Let Assumption 2.1 hold and let $g : F \rightarrow \mathbb{R} : \mathbf{z} \mapsto g^0(\mathbf{z})$. Let $\mathbf{z}_0 \in F$ and $\mathbf{d} \in \mathbb{R}^m$ be such that there exists $\delta > 0$ such that for all $\epsilon \in [0, \delta[$, $\mathbf{z}_0 + \epsilon \mathbf{d} \in F$. Then,*

$$g'(\mathbf{z}_0; \mathbf{d}) = \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \mathbf{d}^T \boldsymbol{\lambda}, \quad \text{s.t. } \mathbf{A}^T \boldsymbol{\lambda} \leq \mathbf{c}_0, -\mathbf{z}_0^T \boldsymbol{\lambda} \leq -g(\mathbf{z}_0). \quad (8)$$

Proof: The proof of this proposition can be found in Appendix A. \square

Remark 3.1 Let Assumption 2.1 hold and let $g : F \rightarrow \mathbb{R} : \mathbf{z} \mapsto g^0(\mathbf{z})$. Let $\mathbf{z}_0 \in F$ and $\mathbf{d} \in \mathbb{R}^m$ be such that there exists $\delta > 0$ such that for all $\epsilon \in [0, \delta[$,

$\mathbf{z}_0 + \epsilon \mathbf{d} \in F$. Let $\Lambda^*(\mathbf{z}) \subset \mathbb{R}^m$ be the dual optimal solution set of LP (1). Then,

LP (8) is equivalent to $g'(\mathbf{z}_0; \mathbf{d}) = \max_{\boldsymbol{\lambda} \in \Lambda^*(\mathbf{z}_0)} \mathbf{d}^T \boldsymbol{\lambda}$.

Proposition 3.4 *Let Assumption 3.1 hold. Then for $i \in \{0, \dots, n_h\}$ and $j \in \{0, \dots, q\}$ and $\mathbf{M} \in \mathbb{R}^{m \times q}$, the LD-derivatives of \mathbf{h} at \mathbf{b}_0 are given by*

$$[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(j)}(\mathbf{d}) = \max_{\boldsymbol{\lambda} \in \mathbb{R}^{m+i}} \left[[\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(j)}(\mathbf{d}) \right]^T \boldsymbol{\lambda}, \quad (9)$$

$$s.t. \quad \begin{bmatrix} [\mathbf{A}^i]^T \\ -\mathbf{q}^i(\mathbf{b}_0)^T \\ -[\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1)^T \\ \vdots \\ -[\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(j-1)}(\mathbf{m}_j)^T \end{bmatrix} \boldsymbol{\lambda} \leq \begin{bmatrix} \mathbf{c}_i \\ -h_i(\mathbf{b}_0) \\ -[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1) \\ \vdots \\ -[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) \end{bmatrix}.$$

Proof: The case for $i = 0$ is established in Theorem 3.3 in [15]. From Proposition 3.1, \mathbf{h} and \mathbf{q}^i are piecewise linear and l -smooth at \mathbf{b}_0 for all i . By Propositions 3.1 and 3.2, $h'_i(\mathbf{b}_0; \mathbf{M}) = [g^i]'(\mathbf{q}^i(\mathbf{b}_0); [\mathbf{q}^i]'(\mathbf{b}_0; \mathbf{M}))$ for $i \in \{1, \dots, n_h\}$. The case for all i and $j = 0$ is established by strong duality of LPs [10] and Proposition 3.3; just use F_i as the domain of the function g in Proposition 3.3, \mathbf{A}^i as the technology matrix, and \mathbf{c}_i as the cost vector. Then for all i , $[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{d})$ is given by LP (9). Now assume that $[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(j-1)}(\mathbf{d})$ is given by LP (9) for all i and for $j \in \{1, \dots, q\}$. Since for all i , g^i is \mathbf{q}^i -weakly l -smooth at \mathbf{b}_0 , the assumptions of Proposition 3.3 are satisfied and $[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(j)}(\mathbf{d})$ is given by LP (9); use technology matrix $\left[\mathbf{A}^i - \mathbf{q}^i(\mathbf{b}_0) - [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1) \cdots - [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) \right]$, cost vector $\left[\mathbf{c}_i^T - h_i(\mathbf{b}_0) - [h_i]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1) \cdots - [h_i]_{\mathbf{b}_0, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) \right]$, and $(G_i)_{\mathbf{b}_0}^{(j)}$ is

the domain, where $(G_i)_{\mathbf{b}_0}^{(j)}$ corresponds to the sets in Definition 2.7. Since the case for $j = 0$ is established, the proof follows by induction. \square

To calculate the LD-derivatives of LLPs with LP (9), we optimize over the dual optimal solution set of LPs (1) and (2). Just apply Remark 3.1 recursively.

Definition 3.1 Let Assumption 3.1 hold, let $\mathbf{M} \in \mathbb{R}^{m \times q}$ and let $S_0^i(\mathbf{b}_0, \mathbf{M})$ be the solution set of LPs (1) and (2) and $S_k^i(\mathbf{b}_0, \mathbf{M})$ be the solution set of LP (9) with $j = k - 1$ in (9) and $k \in \{1, \dots, q + 1\}$.

Remark 3.2 Let Assumption 3.1 hold. Given that \mathbf{h} is l -smooth at \mathbf{b}_0 , for any $\mathbf{d} \in \mathbb{R}^m$, $\mathbf{M} \in \mathbb{R}^{m \times q}$, and $j \in \{0, \dots, q\}$, $-\infty < \mathbf{h}_{\mathbf{b}_0, \mathbf{M}}^{(j)}(\mathbf{d}) < +\infty$ and for $i \in \{0, \dots, n_h\}$, $[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(j)}(\mathbf{d})$ is given by the primal version of LP (9):

$$\begin{aligned}
 [h_i]_{\mathbf{b}_0, \mathbf{M}}^{(j)}(\mathbf{d}) = & \\
 \min_{\mathbf{v} \in \mathbb{R}^{n_v + j + 1}} & \left[\mathbf{c}_i^T - h_i(\mathbf{b}_0) - [h_i]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1) \cdots - [h_i]_{\mathbf{b}_0, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) \right] \mathbf{v}, \\
 \text{s.t.} & \left[\mathbf{A}^i - \mathbf{q}^i(\mathbf{b}_0) - [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1) \cdots - [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) \right] \mathbf{v} = [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(j)}(\mathbf{d}), \\
 & \mathbf{v} \geq \mathbf{0}.
 \end{aligned} \tag{10}$$

Given a fixed $j \in \{0, \dots, q\}$, LPs (10) for all i form a LLP.

Corollary 3.1 If Assumption 3.1 holds, the LD-derivative of \mathbf{h} at \mathbf{b}_0 in the directions $\mathbf{M} \in \mathbb{R}^{m \times q}$ is given by

$$\mathbf{h}'(\mathbf{b}_0; \mathbf{M}) =$$

$$\begin{bmatrix} \boldsymbol{\lambda}_0^T \mathbf{m}_1 & \boldsymbol{\lambda}_0^T \mathbf{m}_2 & \cdots & \boldsymbol{\lambda}_0^T \mathbf{m}_q \\ \boldsymbol{\lambda}_1^T [\mathbf{q}^1]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1) & \boldsymbol{\lambda}_1^T [\mathbf{q}^1]_{\mathbf{b}_0, \mathbf{M}}^{(1)}(\mathbf{m}_2) & \cdots & \boldsymbol{\lambda}_1^T [\mathbf{q}^1]_{\mathbf{b}_0, \mathbf{M}}^{(q-1)}(\mathbf{m}_q) \\ \vdots & & & \\ \boldsymbol{\lambda}_{n_h}^T [\mathbf{q}^{n_h}]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1) & \boldsymbol{\lambda}_{n_h}^T [\mathbf{q}^{n_h}]_{\mathbf{b}_0, \mathbf{M}}^{(1)}(\mathbf{m}_2) & \cdots & \boldsymbol{\lambda}_{n_h}^T [\mathbf{q}^{n_h}]_{\mathbf{b}_0, \mathbf{M}}^{(q-1)}(\mathbf{m}_q) \end{bmatrix}$$

with $\boldsymbol{\lambda}_i \in S_q^i(\mathbf{b}_0, \mathbf{M})$ for $0 \leq i \leq n_h$.

Proof: The definition of LD-derivative is

$$\begin{aligned} h'_i(\mathbf{b}_0; \mathbf{M}) &= \left[[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(q)}(\mathbf{m}_1) \ [h_i]_{\mathbf{b}_0, \mathbf{M}}^{(q)}(\mathbf{m}_2) \ \cdots \ [h_i]_{\mathbf{b}_0, \mathbf{M}}^{(q)}(\mathbf{m}_q) \right], \\ &= \left[[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1) \ [h_i]_{\mathbf{b}_0, \mathbf{M}}^{(1)}(\mathbf{m}_2) \ \cdots \ [h_i]_{\mathbf{b}_0, \mathbf{M}}^{(q-1)}(\mathbf{m}_q) \right]. \end{aligned}$$

The second equality follows from Equation (5). For any $\boldsymbol{\lambda}_i \in S_q^i(\mathbf{b}_0, \mathbf{M})$,

$[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(q-1)}(\mathbf{m}_q) = \boldsymbol{\lambda}_i^T [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(q-1)}(\mathbf{m}_q)$. Moreover, $\boldsymbol{\lambda}_i \in S_q^i(\mathbf{b}_0, \mathbf{M}) \subset S_{q-1}^i(\mathbf{b}_0, \mathbf{M})$

and then $[h_i]_{\mathbf{b}_0, \mathbf{M}}^{(q-2)}(\mathbf{m}_{q-1}) = \boldsymbol{\lambda}_i^T [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(q-2)}(\mathbf{m}_{q-1})$. Following this argument,

$$h'_i(\mathbf{b}_0; \mathbf{M}) = \left[\boldsymbol{\lambda}_i^T [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(0)}(\mathbf{m}_1) \ \boldsymbol{\lambda}_i^T [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(1)}(\mathbf{m}_2) \ \cdots \ \boldsymbol{\lambda}_i^T [\mathbf{q}^i]_{\mathbf{b}_0, \mathbf{M}}^{(q-1)}(\mathbf{m}_q) \right].$$

□

A simple example illustrating the calculation of LD-derivatives for LLPs can be found in Appendix C. So far we have considered the optimal values of the LLP to be a function of all right-hand sides of the equality constraints. In practice, we might be interested in the optimal value as a function of a small number of components of the right-hand side. Let us suppose that only the

first k components of the right-hand side are variable. For some $k < m$, $k \in \mathbb{N}$, $\mathbf{B} \in \mathbb{R}^{m \times k}$ full column rank, and $\mathbf{b}_0 \in \mathbb{R}^m$ let $\tilde{\mathbf{b}} : \mathbb{R}^k \rightarrow \mathbb{R}^m : \mathbf{u} \mapsto \mathbf{B}\mathbf{u} + \mathbf{b}_0$ and consider the functions $\tilde{\mathbf{q}}^i := \mathbf{q}^i \circ \tilde{\mathbf{b}}$ and $\tilde{h}_i := g^i \circ \tilde{\mathbf{q}}^i$ for $i \in \{0, \dots, n_h\}$. Their domains are given by $\tilde{F} := \{\mathbf{u} \in \mathbb{R}^k : -\infty < \tilde{\mathbf{h}}(\mathbf{u}) < +\infty\}$, which means all components of $\tilde{\mathbf{h}}(\mathbf{u})$ take values in \mathbb{R} . Therefore $\tilde{\mathbf{b}}(\tilde{F}) \subset F$ and for $i \in \{0, \dots, n_h\}$, $\tilde{\mathbf{q}}^i(\tilde{F}) \subset F_i$.

Assumption 3.2 Let Assumption 2.1 hold. Suppose that $\text{int}(\tilde{F})$ is nonempty, and that $\mathbf{u}_0 \in \text{int}(\tilde{F})$.

Computing LD-derivatives of $\tilde{\mathbf{h}}$ can be challenging when $\mathbf{B}\mathbf{u} + \mathbf{b}_0$ is in the boundary of F . For example, consider LP (18) and let $\mathbf{B} = [0 \ 1]^T$ and $\mathbf{b}_0 = \mathbf{0}$. For such $\tilde{\mathbf{b}}$, $\tilde{F} = \{u : u \geq 0\}$, and $\forall u \in \tilde{F}$, $\tilde{\mathbf{b}}(u) \in \text{bnd}(F)$. Therefore the chain rule (6) is not applicable. However, the extensions of LD-derivatives presented in Theorems 2.1 and 2.2 can help us compute the LD-derivatives for this case. We shall now show how to compute directional and LD-derivatives of $\tilde{\mathbf{h}}$.

Proposition 3.5 *Let Assumption 3.2 hold. Then for $i \in \{0, \dots, n_h\}$, $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{q}}^i$ are 1-smooth at \mathbf{u}_0 and piecewise affine on \tilde{F} .*

Proof: From Proposition 3.1, \mathbf{h} and \mathbf{q}^i are piecewise linear on F for all i . Since $\tilde{\mathbf{b}}$ is piecewise affine on \mathbb{R}^k , $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{q}}^i$ are piecewise affine on \tilde{F} for all i . Hence, they are piecewise differentiable functions and l -smooth at \mathbf{u}_0 [8]. \square

The following Remark is analogous to Proposition 3.4. The chain rule (6) can be applied when $\tilde{\mathbf{b}}(\mathbf{u}_0) \in \text{int}(F)$. If $\tilde{\mathbf{b}}(\mathbf{u}_0) \in \text{bnd}(F)$, then it follows from Propositions 3.5, 3.2 and 3.3 and the proof of Proposition 3.4.

Remark 3.3 Let Assumption 3.2 hold at $\mathbf{u} = \mathbf{u}_0$. For $i \in \{0, \dots, n_h\}$ and any $\mathbf{M} \in \mathbb{R}^{k \times q}$, $\tilde{h}'_i(\mathbf{u}; \mathbf{M}) = g^i(\tilde{\mathbf{q}}^i(\mathbf{u}); [g^i]'(\tilde{\mathbf{q}}^i(\mathbf{u}); [\tilde{\mathbf{q}}^i]'(\mathbf{u}; \mathbf{M}))$. Hence, $\tilde{\mathbf{h}}'(\mathbf{u}; \mathbf{M})$ can be computed using LP (9) with $\tilde{\mathbf{q}}$ instead of \mathbf{q} and $\tilde{\mathbf{h}}$ instead of \mathbf{h} .

3.2 Phase I LP as an Extended System

Definition 3.2 Consider the LP (1). A Phase I LP of (1) is given by [10]:

$$h_{-1}^E(\mathbf{z}) = \min_{\mathbf{v} \in \mathbb{R}^{nv}, \mathbf{s}_+, \mathbf{s}_- \in \mathbb{R}^m} \sum_{i=1}^m s_{+i} + s_{-i}, \quad \text{s.t. } \mathbf{A}\mathbf{v} + \mathbf{s}_+ - \mathbf{s}_- = \mathbf{z}, \quad (11)$$

$$\mathbf{v} \geq \mathbf{0}, \mathbf{s}_+ \geq \mathbf{0}, \mathbf{s}_- \geq \mathbf{0}.$$

Notice that LP (11) is always feasible for any $\mathbf{z} \in \mathbb{R}^m$ and its objective function value is equal to zero if and only if (1) is feasible and positive otherwise [10].

When LLPs become infeasible, DFBA simulations, optimization algorithms, or nonsmooth equation solving methods fail. The Phase I LP of the Simplex algorithm can be used to extend the domain of \mathbf{h} because it provides an alternative LLP that is always feasible [16]. In particular, when the LLP presented in (1) and (2) is feasible, the extended system given by the Phase I LP and the original system coincide. Otherwise, the extended system is still defined and provides a penalty function [16], which can be used as a constraint in optimization problems or as an equation in nonsmooth equation solving problems.

Proposition 3.6 *Let Assumption 2.1 hold. Now let $\mathbf{h}^E : \mathbb{R}^m \rightarrow \mathbb{R}^{n_h+2}$ where $h_{-1}^E(\mathbf{z})$ is given by LP (11) and for $i \in \{0, \dots, n_h\}$*

$$\begin{aligned}
 h_i^E(\mathbf{z}) = \min_{\mathbf{v} \in \mathbb{R}^{n_v}, \mathbf{s}_+, \mathbf{s}_- \in \mathbb{R}^m} \mathbf{c}_i^T \mathbf{v}, \quad \text{s.t.} \quad \mathbf{A}\mathbf{v} + \mathbf{s}_+ - \mathbf{s}_- = \mathbf{z}, \quad (12) \\
 \sum_{i=1}^m s_{+i} + s_{-i} = h_{-1}^E(\mathbf{z}), \\
 \begin{bmatrix} \mathbf{c}_0^T \\ \vdots \\ \mathbf{c}_{i-1}^T \end{bmatrix} \mathbf{v} = \begin{bmatrix} h_0^E(\mathbf{z}) \\ \vdots \\ h_{i-1}^E(\mathbf{z}) \end{bmatrix}, \\
 \mathbf{v} \geq \mathbf{0}, \mathbf{s}_+ \geq \mathbf{0}, \mathbf{s}_- \geq \mathbf{0}.
 \end{aligned}$$

Then \mathbf{h}^E is 1-smooth on \mathbb{R}^m . If $h_{-1}^E(\mathbf{z}) = 0$, then LPs (1) and (2) are feasible and $h_i^E(\mathbf{z}) = h_i(\mathbf{z})$ for $i \in \{0, \dots, n_h\}$.

Proof: The proof of this proposition can be found in Appendix A. □

Remark 3.4 Let $\tilde{\mathbf{h}}^E = \mathbf{h}^E \circ \tilde{\mathbf{b}}$. Since $\tilde{h}_i^E(\mathbf{u}) = \tilde{h}_i(\mathbf{u})$ for $i \in \{0, \dots, n_h\}$ and $\mathbf{u} \in \tilde{F}$, the LD-derivatives of \tilde{h}_i^E and \tilde{h}_i for $i \in \{0, \dots, n_h\}$ coincide on $\text{int}(\tilde{F})$. Then, $\tilde{\mathbf{h}}^E$ can be used to calculate the LD-derivatives of $\tilde{\mathbf{h}}$.

4 Example Solving Nonsmooth System of Equations

All running times reported are for a 3.20 GHz Intel®; Xeon®; CPU in MATLAB 7.12 (R2011a), Windows 7 64-bit operating system using 4 processors for computations in parallel. The LP solvers were CPLEX [17] and Gurobi [18].

Example 4.1 This example is taken from [19, 20]. In these papers, fermentation of synthesis gas to ethanol and acetate takes place in a bubble column bioreactor with syngas fermenting bacterium *Clostridium Ijungdahlii*. This is a new technology that is being considered for production of biofuels from natural gas. This bubble column bioreactor can be modeled using the partial differential equation (PDE) system described in Appendix D. The spatial dimension of the bubble column was discretized using the finite volume strategy; 100 nodes were considered. The state vector for each finite volume has nine variables. To obtain the growth rate μ and the exchange flux rates v_G , DFBA is used. Given a stoichiometry matrix $\mathbf{S} \in \mathbb{R}^{m \times n_v}$:

$$\mu = \max_{\mathbf{v} \in \mathbb{R}^{n_v}} v_{growth}, \text{ s.t. } \mathbf{S}\mathbf{v} = \mathbf{0}, \mathbf{v}^{LB}(\mathbf{x}) \leq \mathbf{v} \leq \mathbf{v}^{UB}. \quad (13)$$

LP (13) can be put into standard form. Then, following the strategy in [2], LP (13) can be transformed into an LLP using a hierarchy of objectives: (a) Minimize slacks in the Phase I feasibility LP; (b) Maximize growth; (c) Maximize CO uptake; (d) Maximize H₂ uptake; (e) Minimize CO₂ production; (f) Minimize acetate production; (g) Minimize ethanol production. The LLP associated to *Clostridium Ijungdahlii* satisfies Assumption 3.1. The uptake kinetics for CO, H₂ and CO₂ are described by $v_G = \frac{v_{max,G}G}{K_{m,G} + G} \frac{1}{1 + \frac{E_L + A_L}{K_I}}$, and provide some upper bounds to the exchange flux rates in (13). All parameter values are reported in Appendix D. The goal is to compute the steady state of this system. One way is to run the dynamic simulation for a long time. Alternatively, a nonsmooth system of equations can be solved by setting all time

derivatives to zero. This system comprises 901 equations (the last equation corresponds to the penalty), 100 LLPs each one with 682 equality constraints, 1715 variables, and 7 objective functions. Three different strategies were used to obtain sensitivity information: (a) LD-derivatives (LD) in the directions \mathbf{I} ; (b) Directional derivatives (DD) in the coordinate directions; note that these are not guaranteed to be B-subdifferential elements; and (c) Finite differences (FD). Notice that whereas for the LD method, LP (9) is being solved for all i and all j , in the DD method LP (9) is solved only for $j = 0$ (only directional derivatives are computed). Therefore, the LPs being solved to find the LD-derivatives have smaller feasible sets than the ones being solved to find the directional derivative in each coordinate direction. At nonsmooth points, this can result in the DD method not returning an element of the B-subdifferential. The semismooth Newton method [21] was used to solve this system. Two different starting points were considered and the method converged to two different solution points: washout and the non-trivial solution. All finite volumes used the same starting point; therefore, only the starting vector for a single finite volume is reported. Table 1 presents the number of iterations, the 2-norm of the residual vector and the total time for each method. Finding the steady state using the dynamic simulation takes 1462.0 seconds from the first start point and 6629.8 seconds from the second start point. Both dynamic simulations converge to the non-trivial steady state that supports growth and fermentation of the syngas (as opposed to the washout steady state). It can be noticed that finding the analytical derivatives results

in less iterations than finite differences. Also, given that the LPs solved by LD-derivatives have smaller feasible sets than the ones solved by finding the directional derivatives in each coordinate direction, the computation of the steady state using LD-derivatives is faster. When the LD and DD iterations differ, a nondifferentiable point has been encountered.

Table 1 Number of iterations and 2-norm reported for each method.

Start point: [0.1, 1.6421, 80.6372, 0.9032, 53.7581, 0, 0, 0, 0]			
Iteration	LD	DD	FD
1	71.9045	71.9045	71.9045
2	3.6064×10^{-10}	3.6819×10^{-10}	0.0025
3			3.6032×10^{-5}
4			6.3648×10^{-9}
Time(s)	242.6	374.72	319.9
Result	Washout	Washout	Washout
Start point: [20, 0.5, 50, 0, 50, 0, 0, 10, 50]			
Iteration	LD, 2-norm	DD, 2-norm	FD, 2-norm
1	1.448×10^5	1.448×10^5	1.448×10^5
2	1.074×10^3	1.074×10^3	1.074×10^3
3	567.5	567.5	567.5
4	251.1	251.1	251.1
5	12.47	12.47	12.47
6	1.305	1.305	1.306
7	0.0114	0.0114	0.0115
8	8.416×10^{-7}	8.416×10^{-7}	1.098×10^{-5}
9			1.283×10^{-7}
Time(s)	2873.9	4115.7	783.0
Result	Non-trivial solution	Non-trivial solution	Non-trivial solution

5 Conclusions

Lexicographic linear programs are useful to model bioprocesses using flux balance analysis and dynamic flux balance analysis or business decisions involving goal-programming. In both cases, nonsmooth optimization and equation-solving problems embedding lexicographic linear programs can be formulated. To solve these problems, generalized derivative information for lexicographic

linear programs is desirable. This paper obtains elements of the Bouligand subdifferential of the optimal values function of a lexicographic linear program as a function of its right-hand side by solving a number of related linear programs. The two examples presented illustrate how lexicographic directional derivatives can be used to increase the solution speed and reliability of non-smooth equation-solving and optimization problems embedding lexicographic linear programs.

This work opens the possibility of optimizing dynamic flux balance analysis systems. With dynamic flux balance analysis optimization, optimal design of industrial bioprocesses employing microbial communities and parameter estimation will be possible. Initial work on this topic has been presented in [15]. Future work will integrate the work in this paper and that presented in [11, 15] to derive the sensitivities of ordinary differential equation systems with lexicographic linear programs embedded, coming closer to making the systematic optimization of dynamic flux balance analysis systems possible.

Appendices

Appendix A

Proof of Proposition 2.1

First we will show that if $\text{int}(F) \neq \emptyset$, then \mathbf{A} is full row rank. Assume \mathbf{A} is not full row rank and $\text{int}(F) \neq \emptyset$. Then, there exists at least one row that is linearly dependent on the other rows. Without loss of generality, let us assume that the last row of \mathbf{A} is linearly dependent. Let $\mathbf{b}_0 \in \text{int}(F)$. Then, any small perturbation of the last component of \mathbf{b}_0 while

keeping all other components of \mathbf{b}_0 constant results in LP (1) becoming infeasible. Then, $\mathbf{b}_0 \notin \text{int}(F)$ and there is a contradiction.

For the next part assume \mathbf{A} is full row rank. Let $\mathbf{v} = \mathbf{1}$ and $\mathbf{b} = \mathbf{A}\mathbf{v}$. By hypothesis, this $\mathbf{b} \in F$. Let $\mathbf{d} \in \mathbb{R}^m$ and $\epsilon > 0$. Without loss of generality, let $\|\mathbf{d}\| = 1$. Let us find a $\Delta\mathbf{v}$ that satisfies, $\mathbf{A}(\mathbf{v} + \Delta\mathbf{v}) = \mathbf{b} + \epsilon\mathbf{d}$. Since $\mathbf{A}\mathbf{v} = \mathbf{b}$, we need $\mathbf{A}\Delta\mathbf{v} = \epsilon\mathbf{d}$. Since \mathbf{A} is full row rank, the columns of \mathbf{A} span \mathbb{R}^m . Consider all possible collections of m columns that form a basis for the column space of \mathbf{A} . Take such a basis $\tilde{\mathbf{A}}$ and let $\Delta\tilde{\mathbf{v}} = \epsilon\tilde{\mathbf{A}}^{-1}\mathbf{d}$. Without loss of generality, let the columns of $\tilde{\mathbf{A}}$ be the first m columns of \mathbf{A} . Let $\Delta v_j = \Delta\tilde{v}_j$ if $j \in \{1, \dots, m\}$ and 0 otherwise. It is clear that $\Delta\mathbf{v}$ constructed in this form satisfies $\mathbf{A}(\mathbf{v} + \Delta\mathbf{v}) = \mathbf{b} + \epsilon\mathbf{d}$. Moreover, $\|\Delta\mathbf{v}\| = \|\Delta\tilde{\mathbf{v}}\|$. Therefore, $\|\Delta\mathbf{v}\| = \|\Delta\tilde{\mathbf{v}}\| \leq \epsilon\|\tilde{\mathbf{A}}^{-1}\|\|\mathbf{d}\| = \epsilon\|\tilde{\mathbf{A}}^{-1}\|$ and for small enough $\epsilon > 0$, $\|\Delta\mathbf{v}\| < 1$. Then all components $|\Delta v_j| < 1$ and $\mathbf{v} + \Delta\mathbf{v} > \mathbf{0}$, and $\mathbf{b} + \epsilon\mathbf{d} \in F$. By hypothesis $F = \{\mathbf{A}\mathbf{v} : \mathbf{v} \geq \mathbf{0}\}$ [10], so F is a nonempty convex set and $\mathbf{b} + \delta\mathbf{d} \in F$ for all $\delta \in]0, \epsilon[$. Since \mathbf{d} was arbitrary, $\mathbf{b} \in \text{int}(F)$ and $\text{int}(F) \neq \emptyset$. \square

Proof of Proposition 2.3.

Notice that $\rho(\mathbf{x} + \epsilon\mathbf{d}) = \theta(\mathbf{x} + \epsilon\mathbf{d})$ holds $\forall \epsilon \in [0, \delta_{\mathbf{d}}^*]$. From Lemma 2.1 there exists $\delta > 0$ such that $\theta(\mathbf{x} + \epsilon\mathbf{d}) = \theta(\mathbf{x}) + \theta'(\mathbf{x}; \epsilon\mathbf{d})$, $\forall \|\epsilon\mathbf{d}\| < \delta$. Let $\delta_{\mathbf{d}} = \min(\delta, \delta_{\mathbf{d}}^*)$. Given that $\rho(\mathbf{x} + \epsilon\mathbf{d}) = \theta(\mathbf{x} + \epsilon\mathbf{d})$ for all $\epsilon \in]0, \delta_{\mathbf{d}}[$, then $\rho(\mathbf{x} + \epsilon\mathbf{d}) = \rho(\mathbf{x}) + \theta'(\mathbf{x}; \epsilon\mathbf{d})$ for all $\epsilon \in]0, \delta_{\mathbf{d}}[$. Now let us show that $\rho'(\mathbf{x}; \mathbf{d})$ exists and is equal to $\theta'(\mathbf{x}; \mathbf{d})$. Consider any sequence $\{\tau_i\}_{i=0}^\infty$, $\tau_i \in]0, \delta_{\mathbf{d}}[$, $\tau_i \downarrow 0$. From the definition of directional derivative: $\lim_{i \rightarrow \infty} \frac{\theta(\mathbf{x} + \tau_i\mathbf{d}) - \theta(\mathbf{x})}{\tau_i} = \theta'(\mathbf{x}; \mathbf{d})$. Also, for all τ_i , $\frac{\theta(\mathbf{x} + \tau_i\mathbf{d}) - \theta(\mathbf{x})}{\tau_i} = \frac{\rho(\mathbf{x} + \tau_i\mathbf{d}) - \rho(\mathbf{x})}{\tau_i}$. Therefore, $\rho'(\mathbf{x}; \mathbf{d})$ exists and $\rho'(\mathbf{x}; \mathbf{d}) = \theta'(\mathbf{x}; \mathbf{d})$.

From Eqn. (4), $\rho'(\mathbf{x}; \epsilon\mathbf{d}) = \epsilon\rho'(\mathbf{x}; \mathbf{d}) = \epsilon\theta'(\mathbf{x}; \mathbf{d}) = \theta'(\mathbf{x}; \epsilon\mathbf{d})$. Hence for all $\epsilon \in]0, \delta_{\mathbf{d}}[$, $\rho(\mathbf{x} + \epsilon\mathbf{d}) = \rho(\mathbf{x}) + \rho'(\mathbf{x}; \epsilon\mathbf{d})$. \square

Proof of Theorem 2.1.

For ρ to be l -smooth, it is necessarily Lipschitz continuous near \mathbf{x} . Since ζ is locally Lipschitz continuous, the composition $\sigma := \zeta \circ \rho$ is Lipschitz continuous near \mathbf{x} . First, we show that σ is directionally differentiable at \mathbf{x} .

We need to show that $\lim_{\tau \downarrow 0} \frac{\sigma(\mathbf{x} + \tau\mathbf{d}) - \sigma(\mathbf{x})}{\tau} = \lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x} + \tau\mathbf{d})) - \zeta(\rho(\mathbf{x}))}{\tau} \in \mathbb{R}^p$ for all $\mathbf{d} \in \mathbb{R}^m$. First, $\lim_{\tau \downarrow 0} \frac{\rho(\mathbf{x} + \tau\mathbf{d}) - \rho(\mathbf{x}) - \tau\rho'(\mathbf{x}; \mathbf{d})}{\tau} = \mathbf{0}$ from the definition of the direc-

tional derivative. Next given $\mathbf{d} \in \mathbb{R}^m$, $\zeta'(\rho(\mathbf{x}); \rho'(\mathbf{x}; \mathbf{d})) = \lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d})) - \zeta(\rho(\mathbf{x}))}{\tau} \in \mathbb{R}^p$ by assumption. We show that $\lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d})) - \zeta(\rho(\mathbf{x}))}{\tau} = \lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x} + \tau \mathbf{d})) - \zeta(\rho(\mathbf{x}))}{\tau}$, which is equivalent to showing that $\lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d})) - \zeta(\rho(\mathbf{x} + \tau \mathbf{d}))}{\tau} = \mathbf{0}$. From the existence of the directional derivative $\zeta'(\rho(\mathbf{x}); \rho'(\mathbf{x}; \mathbf{d}))$ there exists $\epsilon_1 > 0$ such that $\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}) \in Z$, for all $\tau \in]0, \epsilon_1[$. Since ζ is locally Lipschitz continuous there exists $\epsilon_2 > 0$ and $K > 0$ such that,

$$\|\zeta(\rho(\mathbf{x} + \tau \mathbf{d})) - \zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}))\| \leq K \|\rho(\mathbf{x} + \tau \mathbf{d}) - (\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}))\|,$$

$\forall \tau \in]0, \epsilon_1[$ such that $\|\rho(\mathbf{x} + \tau \mathbf{d}) - (\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}))\| < \epsilon_2$. This is equal to

$$\left\| \frac{\zeta(\rho(\mathbf{x} + \tau \mathbf{d})) - \zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}))}{\tau} \right\| \leq K \left\| \frac{\rho(\mathbf{x} + \tau \mathbf{d}) - (\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}))}{\tau} \right\|,$$

$\forall \tau \in]0, \epsilon_1[$ such that $\|\rho(\mathbf{x} + \tau \mathbf{d}) - (\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}))\| < \epsilon_2$. By the existence of $\rho'(\mathbf{x}; \mathbf{d})$ it follows that $\lim_{\tau \downarrow 0} \left\| \frac{\rho(\mathbf{x} + \tau \mathbf{d}) - (\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}))}{\tau} \right\| = 0$, and then

$$\lim_{\tau \downarrow 0} \left\| \frac{\zeta(\rho(\mathbf{x} + \tau \mathbf{d})) - \zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}))}{\tau} \right\| = 0.$$

Since $\mathbf{d} \in \mathbb{R}^m$ was arbitrary, σ is directionally differentiable at \mathbf{x} and

$$\sigma_{\mathbf{x}, \mathbf{M}}^{(0)}(\mathbf{d}) = \zeta_{\rho(\mathbf{x}), \rho'(\mathbf{x}; \mathbf{M})}^{(0)}(\rho_{\mathbf{x}, \mathbf{M}}^{(0)}(\mathbf{d})).$$

Now let us assume $\sigma_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) = \zeta_{\rho(\mathbf{x}), \rho'(\mathbf{x}; \mathbf{M})}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j))$ for $j \in \{1, \dots, q\}$, for all $\mathbf{M} \in \mathbb{R}^{m \times q}$ and $\mathbf{m}_{q+1} \in \mathbb{R}^m$. Let $\mathbf{y} := \rho(\mathbf{x})$, $\mathbf{Y} := \rho'(\mathbf{x}; \mathbf{M})$, and $\hat{\mathbf{m}} := \mathbf{m}_{j+1}$. We need to show that the following limit exists in \mathbb{R}^p for all $\mathbf{M} \in \mathbb{R}^{m \times q}$ and $\mathbf{m}_{q+1} \in \mathbb{R}^m$:

$$\begin{aligned} \lim_{\tau \downarrow 0} \frac{\sigma_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau \hat{\mathbf{m}}) - \sigma_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j)}{\tau} = \\ \lim_{\tau \downarrow 0} \frac{\zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau \hat{\mathbf{m}})) - \zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j))}{\tau}. \end{aligned}$$

Given $\mathbf{M} \in \mathbb{R}^{m \times q}$ and $\mathbf{m}_{q+1} \in \mathbb{R}^m$, since ζ is ρ -weakly l -smooth at \mathbf{x} , we know the following limit exists in \mathbb{R}^p :

$$\begin{aligned} & [\zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}]'(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j), [\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})) = \\ & \lim_{\tau \downarrow 0} \frac{\zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})) - \zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j))}{\tau}, \end{aligned}$$

and $\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}}) \in G_{\mathbf{x}}^{(j-1)}$ for $\tau > 0$ small enough. We also know from the definition of l -smoothness that:

$$\lim_{\tau \downarrow 0} \frac{\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau \hat{\mathbf{m}}) - \rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) - \tau[\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})}{\tau} = \mathbf{0}.$$

We will show that

$$\begin{aligned} & \lim_{\tau \downarrow 0} \frac{\zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})) - \zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j))}{\tau} = \\ & \lim_{\tau \downarrow 0} \frac{\zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau \hat{\mathbf{m}})) - \zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j))}{\tau}, \end{aligned}$$

which is equivalent to

$$\lim_{\tau \downarrow 0} \frac{\zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})) - \zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau \hat{\mathbf{m}}))}{\tau} = \mathbf{0}.$$

From Corollary 2.2, $\zeta_{\mathbf{y}, \mathbf{Y}}^{(k)}$ is globally Lipschitz continuous for all $k \in \{0, \dots, q\}$. Then, there exists $\delta > 0$ such that $\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}}) \in G_{\mathbf{x}}^{(j-1)}$ if $\tau \in [0, \delta[$, and there exists $K > 0$ and $\tau \in [0, \delta[$ such that

$$\begin{aligned} & K \left\| \frac{\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}}) - \rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau \hat{\mathbf{m}})}{\tau} \right\| \geq \\ & \left\| \frac{\zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})) - \zeta_{\mathbf{y}, \mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x}, \mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau \hat{\mathbf{m}}))}{\tau} \right\|. \end{aligned}$$

By the existence of the directional derivative $[\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})$,

$$\lim_{\tau \downarrow 0} \left\| \frac{\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau [\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}}) - \rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau \hat{\mathbf{m}})}{\tau} \right\| = 0,$$

and then

$$\lim_{\tau \downarrow 0} \left\| \frac{\zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau [\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})) - \zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau \hat{\mathbf{m}}))}{\tau} \right\| = 0.$$

Since $\mathbf{M} \in \mathbb{R}^{m \times q}$ and $\mathbf{m}_{q+1} \in \mathbb{R}^m$ were arbitrary, $\sigma_{\mathbf{x},\mathbf{M}}^{(j-1)}$ is directionally differentiable at \mathbf{m}_j and $\sigma_{\mathbf{x},\mathbf{M}}^{(j)}(\hat{\mathbf{m}}) = \zeta_{\mathbf{y},\mathbf{Y}}^{(j)}(\rho_{\mathbf{x},\mathbf{M}}^{(j)}(\hat{\mathbf{m}}))$ for all $\mathbf{M} \in \mathbb{R}^{m \times q}$ and $\mathbf{m}_{q+1} \in \mathbb{R}^m$. Given that $\sigma_{\mathbf{x},\mathbf{M}}^{(0)}(\mathbf{m}_1) = \zeta_{\mathbf{y},\mathbf{Y}}^{(0)}(\rho_{\mathbf{x},\mathbf{M}}^{(0)}(\mathbf{m}_1))$ for any $\mathbf{M} \in \mathbb{R}^{m \times q}$, the rest of the LD-derivatives follow by induction. \square

Proof of Theorem 2.2

First we show the limit for the directional derivative. Given $\mathbf{d} \in \mathbb{R}^m$ we want to show that the following limit exists in \mathbb{R}^p :

$$\zeta'(\rho(\mathbf{x}); \rho'(\mathbf{x}; \mathbf{d})) = \lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d})) - \zeta(\rho(\mathbf{x}))}{\tau}.$$

We know that $\lim_{\tau \downarrow 0} \frac{\sigma(\mathbf{x} + \tau \mathbf{d}) - \sigma(\mathbf{x})}{\tau} = \lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x} + \tau \mathbf{d})) - \zeta(\rho(\mathbf{x}))}{\tau}$ exists in \mathbb{R}^p .

If $\lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d})) - \zeta(\rho(\mathbf{x}))}{\tau} = \lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x} + \tau \mathbf{d})) - \zeta(\rho(\mathbf{x}))}{\tau}$, which is equivalent to showing that

$$\lim_{\tau \downarrow 0} \frac{\zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d})) - \zeta(\rho(\mathbf{x} + \tau \mathbf{d}))}{\tau} = 0, \quad (14)$$

the proof is complete. Notice that $\zeta(\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}))$ is well defined by the assumption of $\rho(\mathbf{x}) + \tau \rho'(\mathbf{x}; \mathbf{d}) \in Z$ for all $\tau \in]0, \delta_{\mathbf{d}}[$. Since ζ is locally Lipschitz, the proof of Theorem 2.1 establishes (14). Also notice that

$$\sigma'(\mathbf{x}; \mathbf{d}) = \zeta'(\rho(\mathbf{x}); \rho'(\mathbf{x}; \mathbf{d})) = \sigma_{\mathbf{x},\mathbf{M}}^{(0)}(\mathbf{d}) = \zeta_{\rho(\mathbf{x}), \rho'(\mathbf{x}; \mathbf{M})}^{(0)}(\rho_{\mathbf{x},\mathbf{M}}^{(0)}(\mathbf{d})).$$

Given that $\mathbf{d} \in \mathbb{R}^m$ was arbitrary, the directional derivative of ζ at $\rho(\mathbf{x})$ exists in all directions $\rho_{\mathbf{x},\mathbf{M}}^{(0)}(\mathbf{d})$ and the first function of sequence (7) is well-defined with its corresponding domain $G_{\mathbf{x}}^{(0)} = \{\rho_{\mathbf{x},\mathbf{Z}}^{(0)}(\mathbf{z}_1) : \mathbf{Z} \in \mathbb{R}^{m \times q}, \mathbf{z}_{q+1} \in \mathbb{R}^m\}$. By Proposition 2.5, $\zeta_{\rho(\mathbf{x}),\rho'(\mathbf{x};\mathbf{M})}^{(0)}$ is globally Lipschitz on $G_{\mathbf{x}}^{(0)}$.

Now let us assume that for $j \in \{1, \dots, q\}$ and for all $\mathbf{M} \in \mathbb{R}^{m \times q}$ and $\mathbf{m}_{q+1} \in \mathbb{R}^m$, $\sigma_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j) = \zeta_{\rho(\mathbf{x}),\rho'(\mathbf{x};\mathbf{M})}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j))$, and that the function $\zeta_{\rho(\mathbf{x}),\rho'(\mathbf{x};\mathbf{M})}^{(j-1)}$ is well-defined in the sense of (7) and is globally Lipschitz on its domain

$$G_{\mathbf{x}}^{(j-1)} = \{\rho_{\mathbf{x},\mathbf{Z}}^{(j-1)}(\mathbf{z}_j) : \mathbf{Z} \in \mathbb{R}^{m \times q}, \mathbf{z}_{q+1} \in \mathbb{R}^m\}.$$

For brevity, let $\mathbf{y} := \rho(\mathbf{x})$ and $\mathbf{Y} := \rho'(\mathbf{x};\mathbf{M})$. Also, let $\hat{\mathbf{m}} := \mathbf{m}_{j+1}$. We want to show that

$$\begin{aligned} & [\zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}]'(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j), [\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})) = \\ & \lim_{\tau \downarrow 0} \frac{\zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})) - \zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j))}{\tau} \end{aligned}$$

exists in \mathbb{R}^p . We know that the following limits exist in \mathbb{R}^p :

$$\begin{aligned} & \lim_{\tau \downarrow 0} \frac{\sigma_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau\hat{\mathbf{m}}) - \sigma_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j)}{\tau} = \\ & \lim_{\tau \downarrow 0} \frac{\zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau\hat{\mathbf{m}})) - \zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j))}{\tau}. \end{aligned}$$

Analogous to the proof in Theorem 2.1, the proof is complete if we show that

$$\lim_{\tau \downarrow 0} \frac{\zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}})) - \zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j + \tau\hat{\mathbf{m}}))}{\tau} = \mathbf{0}. \quad (15)$$

For $\tau > 0$ small enough, $\mathbf{m}_{q+1} \in \mathbb{R}^m$, and $\mathbf{M} \in \mathbb{R}^{m \times q}$, $\zeta_{\mathbf{y},\mathbf{Y}}^{(j-1)}(\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}(\mathbf{m}_j) + \tau[\rho_{\mathbf{x},\mathbf{M}}^{(j-1)}]'(\mathbf{m}_j; \hat{\mathbf{m}}))$ is well-determined by assumption. Then, the proof of Theorem 2.1 establishes (15) and $\sigma_{\mathbf{x},\mathbf{M}}^{(j)}(\hat{\mathbf{m}}) = \zeta_{\mathbf{y},\mathbf{Y}}^{(j)}(\rho_{\mathbf{x},\mathbf{M}}^{(j)}(\hat{\mathbf{m}}))$. Since \mathbf{M} and \mathbf{m}_{q+1} are arbitrary, the domain of $\zeta_{\mathbf{y},\mathbf{Y}}^{(j)}$ is $G_{\mathbf{x}}^{(j)} = \{\rho_{\mathbf{x},\mathbf{Z}}^{(j)}(\mathbf{z}_{j+1}) : \mathbf{Z} \in \mathbb{R}^{m \times q}, \mathbf{z}_{q+1} \in \mathbb{R}^m\}$, and by Proposition 2.5 $\zeta_{\mathbf{y},\mathbf{Y}}^{(j)}$ is globally Lipschitz on $G_{\mathbf{x}}^{(j)}$. Since the case for $j = 0$ was established, the proof follows by induction. \square

Proof of Proposition 3.3.

Notice that \mathbf{z}_0 is not required to be in the interior of F . g is a convex and piecewise linear function on F [10] of the form $g(\mathbf{z}) = \max_{\boldsymbol{\lambda} \in \Lambda} \mathbf{z}^T \boldsymbol{\lambda}$ where Λ is the finite set that contains all extreme points of the polyhedron $\mathbf{A}^T \boldsymbol{\lambda} \leq \mathbf{c}_0$. Λ is nonempty because \mathbf{A} is full row rank (Theorem 2.6 in [10]). Let $\tilde{g} : \mathbb{R}^m \rightarrow \mathbb{R} : \mathbf{z} \mapsto \max_{\boldsymbol{\lambda} \in \Lambda} \mathbf{z}^T \boldsymbol{\lambda}$ and let $\mathbf{z} \in \mathbb{R}^m$. From Proposition 2.2.7 in [13] since \tilde{g} is a convex function that is Lipschitz near \mathbf{z} , the generalized gradient of \tilde{g} at \mathbf{z} ($\partial \tilde{g}(\mathbf{z})$) coincides with the subdifferential at \mathbf{z} and for all $\mathbf{d} \in \mathbb{R}^m$, $\tilde{g}'(\mathbf{z}; \mathbf{d}) = \tilde{g}^\circ(\mathbf{z}; \mathbf{d})$ (where the second quantity is the generalized directional derivative defined in Section 2 of [13]). Let $\tilde{J}(\mathbf{z}) := \{\boldsymbol{\lambda} \in \Lambda : \tilde{g}(\mathbf{z}) = \mathbf{z}^T \boldsymbol{\lambda}\}$ and $J(\mathbf{z}) := \{\boldsymbol{\lambda} \in \Lambda : g(\mathbf{z}) = \mathbf{z}^T \boldsymbol{\lambda}\}$; it is clear that $\tilde{J}(\mathbf{z}) = J(\mathbf{z})$, $\forall \mathbf{z} \in F$. Given that \tilde{g} is the pointwise maximum of convex differentiable functions, $\partial \tilde{g}(\mathbf{z}) = \text{co}(\tilde{J}(\mathbf{z}))$ (section 3.4 in [22]) and for $\mathbf{z} \in F$,

$$\text{co}(\tilde{J}(\mathbf{z})) = \text{co}(J(\mathbf{z})) = \{\boldsymbol{\lambda} : \mathbf{A}^T \boldsymbol{\lambda} \leq \mathbf{c}_0, \mathbf{z}^T \boldsymbol{\lambda} = g(\mathbf{z})\}.$$

Finally, $\{\boldsymbol{\lambda} : \mathbf{A}^T \boldsymbol{\lambda} \leq \mathbf{c}_0, \mathbf{z}^T \boldsymbol{\lambda} = g(\mathbf{z})\} = \{\boldsymbol{\lambda} : \mathbf{A}^T \boldsymbol{\lambda} \leq \mathbf{c}_0, -\mathbf{z}^T \boldsymbol{\lambda} \leq -g(\mathbf{z})\}$. For $\mathbf{z} \in F$, $\tilde{g}'(\mathbf{z}; \mathbf{d}) = \tilde{g}^\circ(\mathbf{z}; \mathbf{d}) = \max\{\mathbf{d}^T \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \partial \tilde{g}(\mathbf{z})\}$ by Proposition 2.1.2 in [13], and therefore $\max\{\mathbf{d}^T \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \partial \tilde{g}(\mathbf{z})\} = \max\{\mathbf{d}^T \boldsymbol{\lambda} : \mathbf{A}^T \boldsymbol{\lambda} \leq \mathbf{c}_0, -\mathbf{z}^T \boldsymbol{\lambda} \leq -g(\mathbf{z})\}$. For $\mathbf{z} \in F$ and $\mathbf{d} \in \mathbb{R}^m$ such that there exists $\delta > 0$ such that for all $\epsilon \in [0, \delta]$, $\mathbf{z} + \epsilon \mathbf{d} \in F$ and $\tilde{g}'(\mathbf{z}; \mathbf{d}) = g'(\mathbf{z}; \mathbf{d})$. Then, $g'(\mathbf{z}_0; \mathbf{d})$ is given by LP (8). \square

Proof of Proposition 3.6.

Under Assumption 2.1 F is nonempty. If $h_{-1}^E(\mathbf{z}) = 0$, then $\mathbf{z} \in F$ because there exists $\mathbf{v} \geq \mathbf{0}$ such that $\mathbf{A}\mathbf{v} = \mathbf{z}$. Then LP (1) is feasible and by Proposition (2.2) LP (2) is also feasible for all i . If $h_{-1}^E(\mathbf{z}) = 0$, then the variables \mathbf{s}_+ , $\mathbf{s}_- = \mathbf{0}$ and they can be removed together with the constraint $\sum_{i=1}^m s_{+i} + s_{-i} = h_{-1}^E(\mathbf{z})$ from LP (12) resulting in $h_i^E(\mathbf{z}) = h_i(\mathbf{z})$ for all $i \in \{0, \dots, n_h\}$. By Proposition 2.1, F is nonempty and for $\mathbf{z} \in F$, $\mathbf{h}^E(\mathbf{z}) \in \mathbb{R}^{n_h+2}$. This implies that dual LPs of (11) and (12) are always feasible. Now let us show that (11) and (12) satisfy Assumption 3.1. Let $n_h^p := n_h + 1$, $\mathbf{A}^p := \begin{bmatrix} \mathbf{A} & \mathbf{I}_m & -\mathbf{I}_m \end{bmatrix}$. Let $(\mathbf{c}_0^p)^T := \begin{bmatrix} \mathbf{0}^T & \mathbf{1}^T & \mathbf{1}^T \end{bmatrix}$ and for $i \in \{0, \dots, n_h\}$, let $(\mathbf{c}_{i+1}^p)^T := \begin{bmatrix} \mathbf{c}_i^T & \mathbf{0}^T & \mathbf{0}^T \end{bmatrix}$. Then LPs (11) and (12) can be expressed in the format of LPs (1) and (2). Since the dual LPs of (11)

and (12) are always feasible, \mathbf{A}^p and \mathbf{c}_i^p for $i \in \{0, \dots, n_h^p\}$ are such that $\mathbf{h}^E(\mathbf{z}) > -\infty$ for all $\mathbf{z} \in \mathbb{R}^m$. Since \mathbf{A} is full row rank, \mathbf{A}_p is full row rank. Then LPs (11) and (12) satisfy Assumption 3.1. Then by Proposition 3.1, \mathbf{h}^E is l -smooth for any $\mathbf{z} \in \mathbb{R}^m$. \square

Appendix B

Proof of Proposition 4.12 in [14] not being applicable.

This proposition refers to optimization problems of the form

$$\min_{\mathbf{x} \in X} f(\mathbf{x}, \mathbf{u}), \quad \text{s.t. } \mathbf{x} \in \Phi, \quad (16)$$

where $\mathbf{u} \in U$ is a parameter vector and Φ is nonempty and closed. Under certain conditions, this proposition calculates the directional derivative of the optimal value function. To apply Proposition 4.12 in [14], we need to consider the duals of each LP in (1) and (2):

$$h_i(\mathbf{z}) = \max_{\boldsymbol{\lambda} \in \mathbb{R}^{m+i}} [\mathbf{q}^i(\mathbf{z})]^\top \boldsymbol{\lambda}, \quad \text{s.t. } [\mathbf{A}^i]^\top \boldsymbol{\lambda} \leq \mathbf{c}_i. \quad (17)$$

For $i \in \{0, \dots, n_h\}$, let $\mathcal{D}^i \subset \mathbb{R}^{m+i}$ be the feasible set of LP (17). Since Proposition 4.12 in [14] considers a minimization problem, $f^i(\boldsymbol{\lambda}, \mathbf{z}) = -[\mathbf{q}^i(\mathbf{z})]^\top \boldsymbol{\lambda}$. Notice that the feasible set of (17) is independent of \mathbf{z} and nonempty under Assumption 2.1 for $i \in \{0, \dots, n_h\}$. The application of Proposition 4.12 in [14] requires that the inf-compactness condition is satisfied at $\mathbf{z} \in F$. The following propositions show that the inf-compactness condition cannot be satisfied by LLPs.

Definition B.1 Consider the optimization problem (16). The inf-compactness condition holds at $\mathbf{u}_0 \in U$ if there exists $\alpha \in \mathbb{R}$ and a compact set $C \subset X$ such that for every \mathbf{u} near \mathbf{u}_0 , the level set $\text{lev}_\alpha f(\cdot, \mathbf{u}) := \{\mathbf{x} \in \Phi : f(\mathbf{x}, \mathbf{u}) \leq \alpha\}$ is nonempty and contained in C .

Proposition B.1 Let Assumption 2.1 hold and consider LP (17). Then for all i , the inf-compactness condition is not satisfied at \mathbf{z}_0 if $\mathbf{q}^i(\mathbf{z}_0) \in \text{bnd}(F_i)$.

Proof: F_i is closed [1]. Let us assume $\mathbf{q}^i(\mathbf{z}_0) \in \text{bnd}(F_i)$. Since $\mathbf{q}^i(\mathbf{z}_0) \in F_i$, the solution set of LP (17) is nonempty, and since it can be described as a polyhedron, it is closed and convex.

Proposition 3.5 in [15] implies that $\mathbf{q}^i(\mathbf{z}_0) \in \text{int}(F_i)$ if and only if solution set of LP (17) is nonempty, convex, and compact. Since $\mathbf{q}^i(\mathbf{z}_0) \in \text{bnd}(F_i)$, the solution set of LP (17) cannot be compact and therefore must be unbounded. Since the optimal level set is unbounded and this is the smallest nonempty level set, there is no nonempty bounded level set at \mathbf{z}_0 and the inf-compactness condition is not satisfied. \square

Proposition B.2 *Let Assumption 2.1 hold. For all $i > 0$, $\mathbf{q}^i(F) \subset \text{bnd}(F_i)$.*

Proof: By Proposition 2.2, $\mathbf{q}^i(F) \subset F_i$. In addition, $q_{m+i}^i(\mathbf{z}) = h_{i-1}(\mathbf{z})$. Let $\mathbf{d} = -\mathbf{e}_{m+i}$. For any $\epsilon > 0$, $\mathbf{q}^i(\mathbf{z}) + \epsilon\mathbf{d} \notin F_i$ because the $(m+i)^{\text{th}}$ component of $\mathbf{q}^i(\mathbf{z})$ is $h_{i-1}(\mathbf{z})$ which is optimal. Hence, any value less than $h_{i-1}(\mathbf{z})$ results in an infeasible LP. Then $\mathbf{q}^i(\mathbf{z}) \in \text{bnd}(F_i)$ and thus $\mathbf{q}^i(F) \subset \text{bnd}(F_i)$. \square

Hence, by Propositions B.1 and B.2, Proposition 4.12 in [14] cannot be applied to LLPs and an alternative method to calculate the LD-derivatives of the optimal values function is required.

Appendix C

Let $\mathbf{h} : \mathbb{R}_+^2 \rightarrow \mathbb{R}^2$ where:

$$\mathbf{h}(\mathbf{b}) = \text{lex} \min_{\mathbf{v} \in \mathbb{R}^2} \mathbf{C}^T \mathbf{v}, \text{ s.t. } v_1 \leq b_1, v_1 + v_2 \leq b_2, v_1, v_2 \geq 0, \quad (18)$$

with $\mathbf{C} = -\mathbf{I}_2$, where first v_1 is maximized, and then v_2 is maximized. LP (18) can be transformed into standard form to obtain the form of LLP (1) and (2):

$$\mathbf{h}(\mathbf{b}) = \text{lex} \min_{\mathbf{v} \in \mathbb{R}^2} \hat{\mathbf{C}}^T \mathbf{v}, \text{ s.t. } \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \mathbf{v} = \mathbf{b}, \mathbf{v} \geq \mathbf{0}, \quad (19)$$

where $\hat{\mathbf{C}} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}$. For LP (18), $F = \{\mathbf{b} : \mathbf{b} \geq \mathbf{0}\}$. Consider $\mathbf{b}_0 = [1 \ 1]^T$. Clearly, $\mathbf{b}_0 \in \text{int}(F)$ and $\mathbf{h}(\mathbf{b}_0) = [-1 \ 0]^T$. \mathbf{h} is not differentiable at \mathbf{b}_0 . Given $\mathbf{M} \in \mathbb{R}^{2 \times 2}$ and LP

(19), $\mathbf{h}'(\mathbf{b}_0; \mathbf{M})$ can be computed with Proposition 3.4. Consider $\mathbf{M}_1 = \mathbf{I}_2$, $\mathbf{M}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $\mathbf{M}_3 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$, and $\mathbf{M}_4 = -\mathbf{I}_2$. The resulting LD-derivatives are: $\mathbf{h}'(\mathbf{b}_0; \mathbf{M}_1) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $\mathbf{h}'(\mathbf{b}_0; \mathbf{M}_2) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\mathbf{h}'(\mathbf{b}_0; \mathbf{M}_3) = \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix}$, and $\mathbf{h}'(\mathbf{b}_0; \mathbf{M}_4) = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$.

Different elements of the lexicographic subdifferential are obtained from the linear system $\mathbf{J}_L \mathbf{h}(\mathbf{b}; \mathbf{M}) \mathbf{M} = \mathbf{h}'(\mathbf{b}; \mathbf{M})$. The resulting l -derivatives are:

$$\begin{aligned} \mathbf{J}_L \mathbf{h}(\mathbf{b}; \mathbf{M}_1) &= \mathbf{J}_L \mathbf{h}(\mathbf{b}; \mathbf{M}_2) = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}, \\ \mathbf{J}_L \mathbf{h}(\mathbf{b}; \mathbf{M}_3) &= \mathbf{J}_L \mathbf{h}(\mathbf{b}; \mathbf{M}_4) = \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix}. \end{aligned}$$

Notice that \mathbf{M}_1 and \mathbf{M}_2 result in the same l -derivative matrix as well as \mathbf{M}_3 and \mathbf{M}_4 . These two l -derivatives form the B-subdifferential of \mathbf{h} at $\mathbf{b}_0 = [1 \ 1]^T$. Proposition 2.6.2 in [13] shows that for a non-scalar function \mathbf{h} evaluated at \mathbf{b}_0 , Clarke's generalized Jacobian is a subset of the Cartesian product of the generalized gradients of each component of \mathbf{h} . In this example, the Cartesian product of the generalized gradients of h_0 and h_1 at $\mathbf{b}_0 = [1 \ 1]^T$ results in the convex hull of $\left\{ \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix} \right\}$. However, the kinks in the functions h_0 and h_1 are lined up such that the B-subdifferential of \mathbf{h} at $\mathbf{b}_0 = [1 \ 1]^T$ contains only two matrices. The LD-derivatives are guaranteed to find at most these two matrices (Figure 2). The results of this example can be easily verified by expressing \mathbf{h} as:

$$\begin{aligned} h_0(\mathbf{b}) &= -\min\{b_1, b_2\} = \frac{|b_1 - b_2| - b_1 - b_2}{2} \\ h_1(\mathbf{b}) &= -\max\{0, b_2 - b_1\} = \frac{b_1 - b_2 - |b_1 - b_2|}{2}. \end{aligned}$$

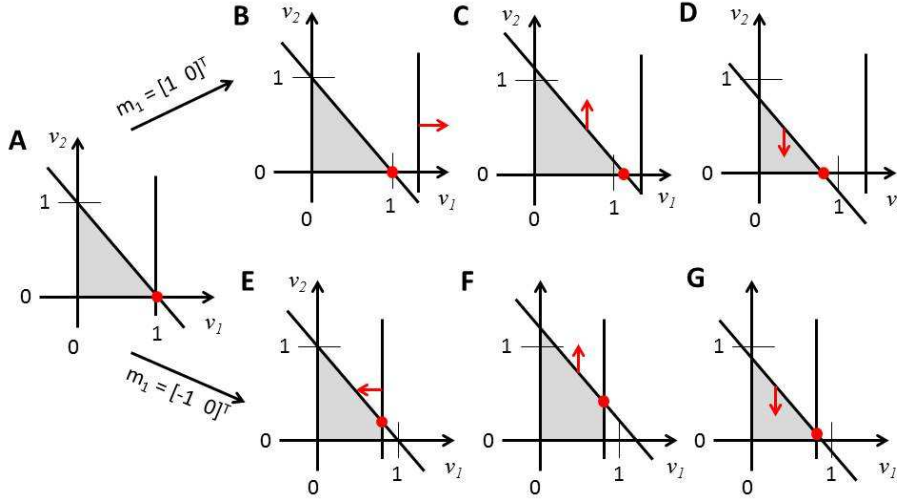


Fig. 1 This figure shows graphically how M_1, M_2, M_3 and M_4 result in different LD-derivatives at $\mathbf{b}_0 = [1 \ 1]^T$. A) Feasible set (gray) and optimal solution point (red dot) for LP (18). If the first column of the matrix of directions is $[1 \ 0]^T$, B is obtained. In B, b_1 is increased and the optimal solution point does not change. Then, the second column of the matrix of directions can be $[0 \ 1]^T$ or $[0 \ -1]^T$ resulting in C and D, respectively. In C, the solution point changes such that h_0 decreases and in D it changes such that h_0 increases. If the first column of the matrix of directions is $[-1 \ 0]^T$, E is obtained. In E, the solution point changes such that h_0 increases and h_1 decreases. Then, the second column of the matrix of directions can be $[0 \ 1]^T$ or $[0 \ -1]^T$ resulting in F and G, respectively. In F, the solution point changes such that h_1 decreases, and in G it changes such that h_1 increases.

Appendix D

Model description of Example 4.1.

(a) Mass balance of biomass of *C. ljungdahlii*:

$$\begin{aligned} \frac{\partial X}{\partial t}(z, t) &= \mu(z, t)X(z, t) - \frac{u_L}{\epsilon_L} \frac{\partial X}{\partial z}(z, t) + D_A \frac{\partial^2 X}{\partial z^2}(z, t), \\ u_L X(0, t) - \epsilon_L D_A \frac{\partial X}{\partial z}(0, t) &= 0, \quad \frac{\partial X}{\partial z}(L, t) = 0, \quad X(z, 0) = X_0, \end{aligned}$$

where X is the concentration of biomass, t is time, μ is the growth rate, u_L is the liquid velocity, z is the spatial position, D_A is the diffusivity, ϵ_L is the liquid volume fraction in the reactor, and X_0 is the initial biomass concentration.

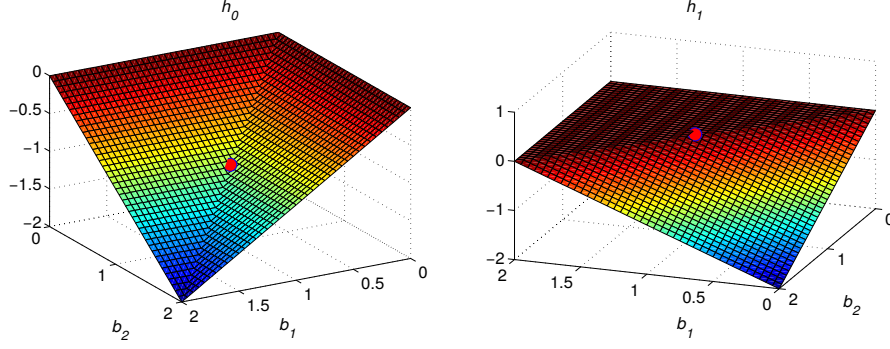


Fig. 2 Surface plots of \mathbf{h} with respect to \mathbf{b} . The red dots indicate the point $\mathbf{b}_0 = [1 \ 1]^T$. Notice that both components of \mathbf{h} can be divided into two regions of differentiability with two different gradients. In particular, $\nabla h_0(\mathbf{b}) = [-1 \ 0]^T$ or $\nabla h_0(\mathbf{b}) = [0 \ -1]^T$ and $\nabla h_1(\mathbf{b}) = [0 \ 0]^T$ or $\nabla h_1(\mathbf{b}) = [1 \ -1]^T$. The four matrices of directions $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3$ and \mathbf{M}_4 probe possible combinations of these gradients at \mathbf{b}_0 , resulting in two different l -derivative matrices. In fact, these two matrices form the B-subdifferential of \mathbf{h} at \mathbf{b}_0 . In this case, the generalized Jacobian of \mathbf{h} at \mathbf{b}_0 is a strict subset of the Cartesian product of the generalized gradients of each component of \mathbf{h} at \mathbf{b}_0 .

(b) Mole balances of liquid-phase CO, H₂, CO₂, ethanol, and acetate:

$$\begin{aligned} \frac{\partial G_L}{\partial t}(z, t) = & v_G(z, t)X(z, t) + \frac{k_{m,G}}{\epsilon_L}(G^*(z, t) - G_L(z, t)) - \frac{u_L}{\epsilon_L} \frac{\partial G_L}{\partial z}(z, t) + D_A \frac{\partial^2 G_L}{\partial z^2}(z, t), \\ u_L G_L(0, t) - \epsilon_L D_A \frac{\partial G_L}{\partial z}(0, t) = u_L G_{gF} H_G, \quad \frac{\partial G_L}{\partial z}(L, t) = 0, \quad G_L(z, 0) = G_{L0}, \end{aligned}$$

where G can be CO, H₂, CO₂, ethanol, and acetate concentrations, v_G refers to the exchange flux rate for species G , $k_{m,G}$ the liquid mass transfer coefficient for species G , G_L the liquid concentration of species G , G^* the liquid concentration in equilibrium with the gas concentration of species G , G_{gF} the gas concentration in the feed of species G , G_{L0} the initial concentration of species G in the liquid, and H_G is Henry's constant for species G .

(c) Mole balances of gas-phase CO, H₂, and CO₂:

$$\begin{aligned} \frac{\partial G_g}{\partial t}(z, t) = & -\frac{k_{m,G}}{\epsilon_g}(G^*(z, t) - G_L(z, t)) - \frac{u_g}{\epsilon_g} \frac{\partial G_g}{\partial z}(z, t), \\ G_g(0, t) = G_{gF}, \quad G_g(z, 0) = G_{g0}, \end{aligned}$$

where G_g is the concentration of species G in the gas, ϵ_g is the gas volume fraction, u_g is the gas velocity, and G_{g0} is the initial gas concentration of species G .

- (d) Column pressure profile: $P(z) = P_L + \rho_L g(L - z)$, where P is the pressure as a function of position in the column, P_L is the pressure at the top of the column, ρ_L is the density of the liquid, L is the size of the column, and g is gravitational acceleration.

The spatial dimension of the bubble column was discretized using the finite volume strategy; 100 nodes were considered. The state vector for each finite volume is:

$$\mathbf{x} = [X, \text{CO}_L, \text{CO}_g, \text{H}_{2g}, \text{H}_{2L}, \text{CO}_{2L}, \text{CO}_{2g}, A, E],$$

where A stands for acetate and E for ethanol. To obtain the growth rate μ and the exchange flux rates v_G , this problem is transformed into a DFBA problem using LP (13).

The resulting nonsmooth system of equations were solved for the following parameter values:

$$\begin{aligned} u_g &= 75 \text{ m/h}, L = 25\text{m}, u_L = 0.25 \text{ m/h}, D_A = 0.25 \text{ m}^2/\text{h}, T = 310.15\text{K}, \\ P_L &= 1 \text{ atm}, \rho_L = 1000 \text{ kg/m}^3, P_{\text{CO}} = 0.6 * P_0, P_{\text{H}_2} = 0.4 * P_0, \\ G_{gF} &= \frac{P_G}{R * T}, H_{\text{CO}} = 8.0 \times 10^{-4} \frac{\text{mol}}{(L * \text{atm})}, H_{\text{H}_2} = 6.6 \times 10^{-4} \frac{\text{mol}}{(L * \text{atm})}, \\ H_{\text{CO}_2} &= 2.5 \times 10^{-2} \frac{\text{mol}}{(L * \text{atm})}, k_{m,\text{CO}_2} = k_{m,\text{CO}} = 80/\text{h}, k_{m,\text{H}_2} = 2.5k_{m,\text{CO}}, \\ k_{m,E} &= k_{m,A} = 0, \epsilon_g = u_g \frac{0.53}{3600(0.15 + u_g/3600)}, \epsilon_L = 1 - \epsilon_g, \\ v_{\text{max},\text{CO}} &= 35 \frac{\text{mmol}}{\text{g} \times \text{h}}, v_{\text{max},\text{CO}} = v_{\text{max},\text{CO}_2} = 35 \frac{\text{mmol}}{\text{g} \times \text{h}}, v_{\text{max},\text{H}_2} = 70 \frac{\text{mmol}}{\text{g} \times \text{h}}, \\ K_{m,\text{CO}} &= K_{m,\text{CO}_2} = 0.02 \text{ mmol/L}, K_{m,\text{H}_2} = 0.02 \text{ mmol/L}, K_I = 10 \text{ mmol/L}, \end{aligned}$$

where P_0 is the pressure at the bottom of the column, P_G is the partial pressure of species G , T is the temperature in Kelvin, and R is the universal gas constant. This system comprises 901 equations (the last equation corresponds to the penalty), 100 LLPs each one with 682 equality constraints, 1715 variables, and 7 objective functions.

Acknowledgements The authors would like to acknowledge Dr. Stuart Harwood for the insightful discussions that took place while preparing the paper.

References

1. Harwood, S.M., Höffner, K., Barton, P.I.: Efficient solution of ordinary differential equations with a parametric lexicographic linear program embedded. *Numerische Mathematik* **133**(4), 623–653 (2016)
2. Höffner, K., Harwood, S.M., Barton, P.I.: A reliable simulator for dynamic flux balance analysis. *Biotechnology and Bioengineering* **110**(3), 792–802 (2013)
3. Ignizio, J.P.: Linear programming in single- & multiple-objective systems. Prentice Hall, Englewoods Cliffs, N.J. 07632 (1982)
4. Höffner, K., Barton, P.I.: Design of microbial consortia for industrial biotechnology. *Computer Aided Chemical Engineering* **34**, 65–74 (2014)
5. Von Stackelberg, H.: The theory of the market economy. Oxford University Press (1952)
6. Ferris, M.C., Pang, J.S.: Engineering and economic applications of complementarity problems. *SIAM Review* **39**(4), 669–713 (1997)
7. Nesterov, Y.: Lexicographic differentiation of nonsmooth functions. *Mathematical Programming* **104**(2-3), 669 (2005)
8. Khan, K.A., Barton, P.I.: A vector forward mode of automatic differentiation for generalized derivative evaluation. *Optimization Methods and Software* **30**(6), 1185–1212 (2015)
9. Scholtes, S.: Introduction to Piecewise Differentiable Equations. Springer, New York (2012)
10. Bertsimas, D., Tsitsiklis, J.N.: Introduction to Linear Optimization. Athena Scientific, Belmont, MA (1997)
11. Khan, K.A., Barton, P.I.: Generalized derivatives for solutions of parametric ordinary differential equations with non-differentiable right-hand sides. *Journal of Optimization Theory and Applications* **163**(2), 355–386 (2014)

12. Barton, P.I., Khan, K.A., Stechlinski, P., Watson, H.A.: Computationally relevant generalized derivatives: theory, evaluation and applications. *Optimization Methods and Software* (In Press)
13. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. SIAM (1990)
14. Bonnans, J.F., Shapiro, A.: *Perturbation analysis of optimization problems*. Springer Science & Business Media, New York (2013)
15. Höffner, K., Khan, K.A., Barton, P.I.: Generalized derivatives of dynamic systems with a linear program embedded. *Automatica* **63**, 198–208 (2016)
16. Gomez, J.A., Höffner, K., Barton, P.I.: DFBAlab: A fast and reliable MATLAB code for dynamic flux balance analysis. *BMC Bioinformatics* **15**(409) (2014)
17. CPLEX IBM ILOG: 12.7 User’s Manual. https://www.ibm.com/support/knowledgecenter/SSSA5P_12.7.0/ilog.odms.studio.help/pdf/u (2017)
18. Gurobi Optimization: Gurobi optimizer reference manual. URL: <http://www.gurobi.com/documentation/> (2017)
19. Chen, J., Gomez, J.A., Höffner, K., Barton, P.I., Henson, M.A.: Metabolic modeling of synthesis gas fermentation in bubble column reactors. *Biotechnology for Biofuels* **8**(89) (2015)
20. Chen, J., Gomez, J.A., Höffner, K., Phalak, P., Barton, P.I., Henson, M.A.: Spatiotemporal modeling of microbial metabolism. *BMC Systems Biology* **10**(21) (2016)
21. Qi, L., Sun, J.: A nonsmooth version of Newton’s method. *Mathematical programming* **58**(1-3), 353–367 (1993)
22. Boyd, S., Vandenberghe, L.: *Subgradients: Notes for EE364b, Stanford University, Winter 2006-07*. https://see.stanford.edu/materials/lsocoe364b/01-subgradients_notes.pdf (2008)