May 5, 2003

APPENDIX F: The Lagrange multiplier technique.

Let f be a real-valued function defined on an open set $U \subset \mathbb{R}^k$ such that f has a gradient $\nabla f = (\partial f/\partial x_1, \ldots, \partial f/\partial x_k)$ at each point of U. Then at any local maximum or minimum x of f, $\nabla f(x) = 0$. The classical Lagrange multiplier technique extends this necessary condition on extrema to extrema subject to a constraint $g(x) = c$. Namely, let λ be a new real variable called a Lagrange multiplier. For $\lambda \in \mathbb{R}$ and $x \in U$ let $F(x, \lambda) := f(x) + \lambda[g(x) - c]$. Set $\nabla_{x,\lambda} F = 0$ where $\nabla_{x,\lambda} = (\partial/\partial x_1, \ldots, \partial/\partial x_k, \partial/\partial\lambda)$. Solutions will satisfy the constraint since $\partial F/\partial\lambda = g(x) - c$. As will be shown, under some conditions, extrema of f under the constraint will be among those x with $\nabla_{x,\lambda} F(x, \lambda) = 0$ for some λ. (But, with or without constraints, f need not in general have any extrema, even though there are points where the gradient is 0, as in Solari's example in Section 3.2.)

The following Theorem F.1 includes the assumption that $\nabla g(x) \neq 0$ at all x where $g(x) = c$. If $\nabla g(x) = 0$, then $\nabla_{x,\lambda} F(x, \lambda) = (\nabla_x f(x), g(x) - c) = 0$ implies $\nabla_x f(x) = 0$, as for an unconstrained extremum, which will often be incompatible with the constraint. Here is a specific example. Let $k = 2$, let $f(x, y) := (x-1)^2 + y^2$, and let the constraint be $g(x, y) := x^2 = c := 0$, Here g is a "bad" function in that its gradient is 0 when $g = c$. To minimize f subject to $g = c$, we have $f(0, y) = 1 + y^2$, minimized at $x = y = 0$. But with a Lagrange multiplier we get $F((x, y), \lambda) = (x - 1)^2 + y^2 + \lambda x^2$. Setting $\nabla_{x,\lambda} F = 0$ gives $\partial F/\partial y = 2y = 0$ (correct), $\partial F/\partial\lambda = x^2 = 0$ (correct), but $\partial F/\partial x = 2(x - 1) + 2\lambda x = 0$ which with $x = 0$ gives $-2 = 0$, a contradiction.

The theorem stated next says that under its hypotheses, minima of f subject to $g = c$ exist and can be found via Lagrange multipliers. Hypothesis (b) follows from conditions in Proposition F.2.

F.1 Theorem. (a) Let f and g be C^1 functions from an open set $U \subset \mathbb{R}^k$ into $\mathbb{R}$, where $k \geq 2$. Let $c \in \mathbb{R}$, suppose $g(x) = c$ for some $x \in U$ and that the gradient $\nabla g(x) \neq 0$ for all such x.

(b) Also suppose that for some compact $C \subset U$ with $C_1 := \{x \in C : g(x) = c\} \neq \emptyset$,

$$K := \inf\{f(x) : x \in C_1\} < \inf\{f(x) : x \in U \setminus C\}.$$

Let $C_2 := \{x \in C_1 : f(x) = K\}$. Then C_2 is non-empty and at every point x of C_2, $\nabla_{x,\lambda} F(x, \lambda) = 0$ for some λ.

Proof. C_1 is compact and non-empty, so $C_2 \neq \emptyset$ since a continuous real function on a compact set attains its infimum. At every point y of the boundary of C, $f(y) \geq \inf\{f(x) : x \in U \setminus C\}$. Thus C_2 is included in the interior of C.

The equation $\partial F(x, \lambda)/\partial\lambda = g(x) - c = 0$ holds for all $x \in C_1 \supset C_2$ and all λ.

Let $x \in C_2$. Then $\nabla g(x) \neq 0$ by hypothesis. Thus $\partial g(x)/\partial x_j \neq 0$ for some j, where we can assume $j = 1$. In a neighborhood of x, we can make a C^1 change of coordinates $G(x_1, x_2, ..., x_k) := (g(x), x_2, ..., x_k)$ which has a non-zero Jacobian determinant equal to $\partial g/\partial x_1$ and so G has a C^1 inverse. Thus we can assume $g(x) \equiv x_1$ in a neighborhood of x.

The vector equation $0 = \nabla_x F(x, \lambda) = \nabla_x f(x) + \lambda \nabla_x g(x)$ is equivalent to $\nabla_x f(x) = -\lambda \nabla_x g(x) = (-\lambda, 0, ..., 0)$. Thus $\lambda = -\partial f(x)/\partial x_1$, which can be solved trivially by taking it as the definition of λ at x. We get the further equations $\partial f(x)/\partial x_j = 0$ for $j = 2, ..., k$, and these are, indeed, necessary for f to have a minimum at x in the hyperplane $x_1 = c$. The Theorem is proved. $\square$

F.2 Proposition. If hypotheses (a) of Theorem F.1 hold and $f(x) \to +\infty$ as x approaches the boundary of U or $|x| \to +\infty$, $x \in U$, then hypotheses (b) hold.

Proof. Choose $x_0 \in U$ with $g(x_0) = c$. Every point x of U has a neighborhood U_x with compact closure included in U. Since U is separable, countably many such neighborhoods $U_{x(n)}$ form a base of the topology of U (RAP, Proposition 2.1.4), so $U = \bigcup_n U_{x(n)}$. Let C_n be the closure of $\bigcup_{j=1}^n U_{x(j)}$. Then C_n is compact. If $y_n \in U \setminus C_n$ for all n and for some $M < \infty$, $f(y_n) \leq M$ for all n, then $|y_n|$ must be bounded, so y_n has a subsequence converging to some $y \in \mathbb{R}^k$. By the definitions, y must be in the boundary of U, a contradiction. So, taking $M := f(x_0) + 1$, there is a compact $C \subset U$ such that $f(x) > f(x_0) + 1$ for all $x \in U \setminus C$, with $C = C_n$ for some n. The Proposition follows. $\square$

Remark. In the example given just before Theorem F.1, f and g are C^1 (in fact C^∞), $U = \mathbb{R}^2$, $c = 0$, and $f(x, y) \to +\infty$ as $|(x, y)| = \sqrt{x^2 + y^2} \to +\infty$. So all the hypotheses of Theorem F.1 and Proposition F.2 hold except that $\nabla g(x, y) = 0$ when $g = c$.