TAIWANESE JOURNAL OF MATHEMATICS
Vol. 5, No. 3, pp. 603-608, September 2001
This paper is available online at http://www.math.nthu.edu.tw/tjm/

PERTURBATIONS AND APPROXIMATE MINIMUM IN CONSTRAINED OPTIMIZATION

B. D. Craven

Abstract. An approximate minimum, for the minimization of a function f over a feasible set S, is a point ξ such that $f(x) \geq f(\xi) - \epsilon$ for all feasible x near the minimum point p of f on S. This concept is relevant when the problem data, or the computation, are approximate. Under regularity assumptions, an approximate minimum is a local minimum of a perturbation of the given problem. This depends on the property of a strict local minimum, that a small perturbation moves the minimum point only by a small amount.

1. Introduction and Definitions

Suppose that the constrained minimization problem:

(1) $$\text{MIN } J(x) \text{ subject to } g(x) \leq 0,$$

reaches a local minimum at $x = \bar{x}$, where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are continuous functions. If the data for the problem, of the computation, are approximate, one may wish to consider *approximate minima*, namely, those points ξ in a neighbourhood of $\bar{x}$ for which $f(\xi) \leq f(\bar{x}) + \epsilon$. Since $\bar{x}$ is a minimum, $f(\bar{x}) \leq f(\xi)$.

Note that an unconstrained approximate minimum point is not necessarily one where the gradient is small; there are counterexamples [1, 2].

These approximate minima may be related to exact minima of suitably perturbed problems. Consider the perturbed problem:

(2) $$\text{MIN}_x \, f(x, q) \text{ subject to } g(x, q) \leq 0,$$

Received February 29, 2000; revised July 31, 2000.
Communicated by P. Y. Wu.
2000 *Mathematics Subject Classification*: 90C30.
Key words and phrases: Perturbation, approximate minimum, strict local minimum.

in which q is a perturbation parameter, and $f(x,0) = f(x), g(x,0) = g(x)$, and in particular the linearly perturbed problem:

$$\text{MIN}_x\, f(x) + b^T(x - \bar{x}) \text{ subject to } g(x) \leq r, \tag{3}$$

in which the vectors b and r comprise the perturbation parameter q, with $||q||$ assumed to be sufficiently small. In (3), the gradient of the objective and the level of the constraint are each perturbed by a small amount. If (3) is minimized at a point $\hat{x}(q)$, denote $\Phi(q) := f(\hat{x}(q), q)$. Under some regularity conditions (see, e.g., Craven [5]),

$$\Phi'(0) = f_q(\bar{x}, 0) + \bar{\lambda} g_q(\bar{x}, 0), \tag{4}$$

where $\bar{\lambda}$ is the Lagrange multiplier for the minimum of (1), and f_q and g_q denote partial derivatives with respect to q. However, the linear approximation:

$$\epsilon \geq f(\xi) - f(\bar{x}) = \Phi(q) - \Phi(0) \approx \Phi'(0)q$$

may not be sufficient; quadratic terms may be needed.

The results depend on the following definitions and theorem.

Definition 1. A local minimum of (1) is a *strict local minimum* if for all sufficiently small $\rho > 0$, there exists positive ξ such that $f(x) \geq f(x) + \xi$ whenever x is feasible and $||x - \bar{x}|| = \rho$.

Theorem 1. Perturbation of strict local minimum (Craven [6, Theorem 4.7.1]). *For problem* (2), *assume that*

(i) *the unperturbed problem* (*with* $q = 0$) *reaches a strict local minimum at* $x = \bar{x}$,

(ii) *for each* $q, g(\bar{x}, q) = g(\bar{x})$,

(iii) *the functions* $f(.,.)$ *and* $g(.,.)$ *are uniformly continuous on bounded sets*,

(iv) *when* $q \neq 0$, $f(., q)$ *reaches a minimum on each closed bounded set.*

Then, *whenever* $||q||$ *is sufficiently small*, *the perturbed problem* (2) *reaches a local minimum at a point* $\bar{x}(q)$, *where* $\bar{x}(q) \to 0$ *as* $q \to 0$.

Remarks. If $\bar{x}$ is a strict minimum, then there is no feasible curve $x = \omega(\alpha)(\alpha \geq 0)$ starting at $\bar{x}$, with f constant along the curve. In the proof of Theorem 1, the strong feasibility assumption (ii) is used only to ensure that $g(\bar{x}, q) \leq 0$, so the latter may be assumed instead. Assumption (iii) follows from continuity in finite dimensions. A *strict minimum* is not enough to ensure that the Lagrange multiplier $\hat{\lambda}$ corresponding to $\hat{x}$ converges to the multiplier $\bar{\lambda}$ corresponding to $\hat{x}$.

Definition 2. Problem (2) (or (3)) is called *locally unique* if for all sufficiently small $||q||$ (or $||(b,r)||$), at most one point ξ in a neighbourhood of $\bar{x}$ satisfies (for some multiplier λ) the condition:

(5) $$f_x(\xi,q) + \lambda^T g_x(\xi,q) = 0, \lambda^T g(\xi,q) = 0$$

(6) $$(\text{or } f'(\xi) + b^T + \lambda^T g'(\xi) = 0, \lambda^T(g(\xi) - r) = 0).$$

Remarks. Such points will be called *KKT points*. Note that (4), together with $\lambda \geq 0$, is the necessary Karush-Kuhn-Tucker condition for a minimum of (2) at ξ. Condition (4), with $\lambda \geq 0$, is also necessary and sufficient for a *quasimin* of (2) at ξ (see [4]), namely,

(7) $$f(x) - f(\xi) + \mathbf{o}(||x - \xi||) \geq 0 \text{ for feasible } x \to \xi.$$

If $f(.,q)$ and $g(.,q)$ are C^2, and all constraints are assumed active (thus $g(\bar{x}) = 0, g(\xi,q) = 0$), then *locally unique* holds if the matrix

(8) $$\begin{bmatrix} f_{xx}(\bar{x},0) & g_x(\bar{x},0)^T \\ g_x(\bar{x},0) & 0 \end{bmatrix}$$

is nonsingular, for then $f_x(\xi,q)+\lambda^T g_x(\xi,q) = 0$, $g(\xi,q) = 0$ can be solved locally for ξ and λ. (It suffices if $f_{xx}(\bar{x},0)$ is nonsingular and $g_x(\bar{x},0)$ has full rank.) Less restrictively, it suffices, using Clarke's implicit function theorem [3], if $f_x(.,q)$ and $g_x(.,q)$ are Lipschitz functions, and the matrix,

(9) $$\begin{bmatrix} A & K^T \\ K & 0 \end{bmatrix}$$

is nonsingular for each A in the generalized Jacobian $\partial f_x(\bar{x},0)$ and $K \in g_x(\bar{x},0)$. The matrix is nonsingular if each A is nonsingular and each K has full rank (using, the partitioned inverse matrix theorem).

Example 0. Let $f(x) = |x|$, $x \in \mathbb{R}$. Then f reaches an unconstrained strict minimum at 0. A linear perturbation to $|x| + qx$ with $|q| < 1$ does not move the minimum away from 0. A perturbation (with $q > 0$) to

(10) $$f(x,q) = -x(x \leq q), x - 2q(x > q),$$

moves the minimum to q. Note that, for $0 < x < q$, $f(x,q) - f(x,0) = -2x$, but the coefficient -2 is not *sufficiently small*.

Example 1. Let $f(x) := (1/2)x^TAx(x \in \mathbb{R}^n)$, where A is a positive definite matrix; then $\bar{x} = 0$, and ξ is an (unconstrained) approximate minimum when

$$\frac{1}{2}x^TAx \geq \frac{1}{2}\xi^TA\xi - \epsilon, \forall x \tag{11}$$

thus when $(1/2)\xi^TA\xi \leq \epsilon$. Let $f(x,q) := (1/2)x^TAx + q^Tx$; then $f(.,q)$ is minimized at $x = \hat{x} : -A^{-1}b$, and $f(\hat{x},q) = -(1/2)q^TA^{-1}q$. Now $\hat{x}$ is an approximate minimum of $f(.)$ exactly when q lies in the ellipsoid $(1/2)q^TA^{-1}q \leq \epsilon$.

2. Approximate Unconstrained Minimum

Proposition 1. *Let the C^1 function $f : \mathbb{R}^n \to \mathbb{R}$ reach a strict local minimum at $\bar{x}$. Let the linearly perturbed problem,*

$$\text{MIN}_x\, \bar{f}(x) := f(x) + q^T(x - \bar{x}), \tag{12}$$

be locally unique. Then, for sufficiently small $\epsilon > 0$, ξ is an approximate minimum of $f(.)$ exactly when ξ is a local minimum of $\hat{f}$ for some constant vector q.

Proof. Choose an approximate minimum ξ satisfying $f(\xi) = f(\bar{x}) + \epsilon$ for some $\epsilon > 0$; then $||\xi - \bar{x}||$ is small if ϵ is small. Now ξ is a stationary point of $\hat{f}(.)$ if q is chosen as $-f'(\xi)^T$. Since f is C^1, $q \to 0$ as $\epsilon \to 0$. From Theorem 1, if $||q||$ is sufficiently small, $\bar{f}(.)$ reaches a local minimum at a point $\hat{x}$, where $\hat{x} \to 0$ as $q \to 0$, and thus $f'(\hat{x}) = -q^T$. By the *locally unique* assumption, $\hat{x} = \xi$; thus ξ is a local minimum of $\bar{f}(.)$.

Remarks. If, in particular, f is C^2, and $f''(\bar{x})$ is nonsingular, and b is given, then $f'(\bar{x}) = 0$ and $f'(\xi) = -q^T$ give, for each component i, that $-q_i = (f')'_i(\hat{\zeta}_i)(\xi - \bar{x})$ for intermediate points ζ_i. Construct a matrix M with rows $(f')'_i(\zeta_i)$; then M is nonsingular since f"$(\bar{x})$ is, for $||q||$ small, so ξ is determined uniquely. Less stringently, suppose that $f'(.)$ is Lipschitz, and every element of the Clarke generalized Jacobian $\partial f'(\bar{x})$ is nonsingular; then ξ is determined uniquely.

The conclusion of Proposition 1 does not hold if f is not differentiable (see Example 0).

3. Approximate Constrained Minimum

This linear-quadratic example serves to approximate smooth problems.

Example 2.

$$\text{MIN } f(x) := \frac{1}{2}x^TAx + a^Tx \text{ subject to } K_x \leq k, \tag{13}$$

where now the matrix A need not be positive definite. The Karush-Kuhn-Tucker conditions require that $A\bar{x} + a + K^T\bar{\lambda} = 0$, $\bar{\lambda} \geq 0$. If the origin is shifted to make the solution $\bar{x} = 0$, $f(\bar{x}) = 0$, then the approximate minimum points ξ must satisfy $f(\xi) \leq \epsilon + f(x)$ whenever $Kx \leq k$, and hence $f(\xi) \leq \epsilon$.

Consider a perturbed problem:

$$\text{(14)} \qquad \text{MIN } \frac{1}{2}x^T Ax + a^T x + b^T x \text{ subject to } K_x \leq k + r,$$

where b and r are small (vector) parameters. If inactive constraints are omitted for the unconstrained problem, and if the perturbation does not change the list of active constraints, and (14) reaches a minimum at $\hat{x}$, then KKT requires, for some multiplier $\bar{\lambda}$, that

$$\text{(15)} \qquad \begin{bmatrix} A & K^T \\ K & 0 \end{bmatrix} \begin{bmatrix} \hat{x} - \bar{x} \\ \hat{\lambda} - \bar{\lambda} \end{bmatrix} = \begin{bmatrix} -b \\ r \end{bmatrix}.$$

So the optimum $\hat{x}$ is a linear function of b and r, and is unique under conditions stated above for (9).

To each $(\hat{x}, \hat{\lambda})$ in a neighbourhood of $(\bar{x}, \bar{\lambda})$ there correspond perturbation parameters (b, r). Conversely, assume that A is nonsingular and K has full rank; then the matrix in (15) is nonsingular, and (15) determines $(\hat{x}, \hat{\lambda})$ uniquely as a continuous function of (b, r); thus *locally unique* holds.

Proposition 2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be C^1 functions; let $f(x)$ reach a strict local minimum, subject to $g(x) \leq 0$, at $x = \bar{x}$; and let a constraint qualification hold. Assume that the perturbed problem* (3) *is locally unique, and that the list of active constraints does not change with a small perturbation. Assume that the constraint $g(\bar{x}) = r$ is feasible, for sufficiently small $\|r\|$, and $f(x) + b^T x$ reaches a minimum on each closed bounded set. Then, when $\epsilon > 0$ is sufficiently small, ξ is an approximate minimum of the given problem exactly when ξ is a local minimum of the perturbed problem, for some suitable b and r of sufficiently small norm.*

Proof. Inactive constraints have no effect; therefore omit them, thus assuming $g(\bar{x}) = 0$. Choose an approximate minimum ξ of the given problem, satisfying, for some $\epsilon > 0$, $f(\xi) = f(\bar{x}) + r$. Choose $r = g(\xi)$. Now ξ will satisfy the Karush-Kuhn-Tucker conditions for (3) if b is chosen as $-[f'(\xi) + \lambda^T g'(\xi)]^T$. Here λ is chosen with $\|\lambda - \bar{\lambda}\|$ sufficiently small, so that $\|b\|$ and $\|r\|$ are sufficiently small that Theorem 1 applies.

Given this b and r, Theorem 1, applied to the strict minimum, shows that the perturbed problem has a local minimum at $x = \hat{x}$, where $\hat{x} \to \bar{x}$ as $\|b\| \to 0$,

$\|r\| \to 0$. Denote by $\hat{\lambda}$ the Lagrange multiplier corresponding to $\hat{x}$. Thus KKT conditions (6) hold, with the same (b, r), both for (ξ, λ) and for $(\hat{x}, \hat{\lambda})$. From the *locally unique* hypothesis, $\hat{x} = \xi$, hence also $\hat{\lambda} = \lambda$. ■

4. Discussion and Applications

If the data for the given optimization problem (1) are somewhat fuzzy, then a more descriptive formulation might replace (1) by a family of perturbed problems (2) or (3), with the perturbation parameters required to be small, in some sense. There is then the possibility of a second optimization, over the perturbation parameters in a specified region. The objective for the second optimization could be the original objective, or a different secondary objetive. Many optimization problems are by nature multi-objective, and a choice of a single objective is then rather arbitrary.

Consider, in particular, the auxiliary objective $c^T q$, with $c = \phi'(0)$ from 4), and a constraint $q^T Q q \leq \delta$, specifying a small region for q. Then $c^T q$ is bounded by $\pm(\delta c^T Q^{-1} c)^{1/2}$, giving a tolerance for the objective value for the given problem (1).

Acknowledgement

Thanks are given to a referee for correcting various misprints.

References

1. A. Auslender and J. P. Crouzeix, Well behaved asymptotical convex functions, *Analyse Non-Linéaire* (1989), 101-122.
2. S. Bolintineanu, Vector variational principles towards asymptotically well behaved vector convex functions, University of Pau, France, Preprint, 1999.
3. F. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
4. B. D. Craven, Lagrangean conditions and quasiduality, *Bull. Austral. Math. Soc.* **16** (1977), 325-339.
5. B. D. Craven, Lagrangian conditions for a minimax, *Proc. Centre Math. Appl. Austral. Nat. Univ.* (1998), 24-33.
6. B. D. Craven, *Control and Optimization*, Chapman & Hall, London, 1995.

Department of Mathematics & Statistics,
University of Melbourne, Victoria 3010, Australia