CONVERGENCE AND ERROR BOUND FOR
PERTURBATION OF LINEAR PROGRAMS[1]

Dedicated to Professor Olvi Mangasarian on the occasion of his 65th birthday, whose boundless energy and creativity continue to inspire.

June 22, 1997 (revised May 7, 1998)

PAUL TSENG
Department of Mathematics, University of Washington
Seattle, Washington 98195, U.S.A. (tseng@math.washington.edu)

ABSTRACT

In various penalty/smoothing approaches to solving a linear program, one regularizes the problem by adding to the linear cost function a separable nonlinear function multiplied by a small positive parameter. Popular choices of this nonlinear function include the quadratic function, the logarithm function, and the $x\ln(x)$-entropy function. Furthermore, the solutions generated by such approaches may satisfy the linear constraints only inexactly and thus are optimal solutions of the regularized problem with a perturbed right-hand side. We give a general condition for such an optimal solution to converge to an optimal solution of the original problem as the perturbation parameter tends to zero. In the case where the nonlinear function is strictly convex, we further derive a local (error) bound on the distance from such an optimal solution to the limiting optimal solution of the original problem, expressed in terms of the perturbation parameter.

1 Introduction

The two main topics of this paper, error bound and perturbation of linear programs, are ones on which Olvi Mangasarian has left indelible marks, as he has done in so many areas of mathematical programming.

Consider the linear program (P):

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b,\ x \geq 0, \end{array}$$

where $c \in \Re^n, A \in \Re^{m\times n}, b \in \Re^m$. We denote by S the set of optimal solutions of (P), which is polyhedral and assumed to be nonempty, and we denote by X the set of feasible solutions of (P), i.e., $X := \{x \in \Re^m : Ax \leq b, x \geq 0\}$. In various

[1]This research is supported by National Science Foundation Grant CCR-9311621.

penalty/smoothing approaches to solving this classical optimization problem, one adds to the objective a separable nonlinear function multiplied by a small positive scalar ϵ and solves the resulting problem instead. Such a function has the form

$$f(x_1, ..., x_n) = f_1(x_1) + \cdots + f_n(x_n),$$

where each $f_j : \Re \mapsto \Re \cup \{\infty\}$ is either continuous on $[0, \infty)$ or continuous on $(0, \infty)$ with $\lim_{\xi \downarrow 0} f_j(\xi) = \infty$. The corresponding perturbed problem, denoted (P^ϵ), has the form:

$$\begin{array}{ll} \text{minimize} & c^T x + \epsilon f(x) \\ \text{subject to} & Ax \leq b^\epsilon, \ x \geq 0, \end{array}$$

where $b^\epsilon \in \Re^m$ satisfies $b^\epsilon \to b$ as $\epsilon \to 0$. We note that previous studies focus on the case where $b^\epsilon = b$ for all ϵ, but the general case is also of interest since in practice the constraints $Ax \leq b$ may be satisfied only inexactly. One popular choice of f_j is the quadratic function

$$f_j(x_j) = (x_j)^2/2 \tag{1}$$

which has been studied by Karlin [Kar59, p. 238], Mangasarian [Man81, Man84, Man86] and others [MNP96, Pin96]. In particular, it was shown in [Man84] that, for all ϵ sufficiently small (and with $b^\epsilon = b$), the unique solution of (P^ϵ) equals the least 2-norm solution of (P). [A weaker version of this result was rediscovered in [MNP96].] The equivalence between the solution set of (P^ϵ) and the solution set of (P) for small ϵ is further studied in [FeM91, MaR79] for arbitrary real-valued function f that is either continuously differentiable around S or Lipschitz continuous around S or convex around X. Another popular choice of f_j is the logarithm function

$$f_j(x_j) = -\ln(x_j) \tag{2}$$

which has been much studied in the context of interior penalty methods and Karmarkar-type interior-point methods (see [FiM68, McL80, Meg89, Wri92] and references therein). A third choice is the entropy function

$$f_j(x_j) = x_j \ln(x_j), \tag{3}$$

which was studied by Fang et al. [Fan92, FaT93, RaF92] and, from a dual exponential penalty view, by Cominetti et al. [CoD94, CoS94] (see [FRT97, FaT96] for further discussions). Such a perturbation of (P) also arises in the asymptotic analysis of a certain perturbed entropy minimization problem, as communicated to the author by A. Lewis at the University of Waterloo. In [CoS94], it was shown that, as $\epsilon \to 0$ (and with $b^\epsilon = b$), the unique solution of (P^ϵ) approaches the least $x\ln(x)$-entropy solution of (P). This result was generalized in a recent work of Auslender et al. [ACH97, Theorem 3.4] to the case where each f_j is the conjugate of a certain kind of strictly convex differentiable function, namely, a function satisfying the assumptions (H_0), (H_2), (H_3) in [ACH97] (see the end of Section 2 for further discussions of this).

For arbitrary f that is bounded on X (and assuming $b^\epsilon = b$ for all ϵ), it is easily shown that the difference in the optimal objective value of (P^ϵ) and of (P) is within a constant factor of ϵ (see the proof of Proposition 2.1(c) in [Tse95]).

Let S^ϵ denote the set of optimal solutions of (P^ϵ) and let

$$S^* := \arg\min\{\sum_{j\in J^*} f_j(x_j) : x \in S\},$$

where $J^* := \{j \in \{1, ..., n\} : x_j > 0 \text{ for some } x \in S\}$. We will show that, as $\epsilon \to 0$, S^ϵ converges S^* in a certain sense (see Proposition 1) and, when each f_j is strictly convex (so S^ϵ and S^*, if nonempty, are singletons), we will give an estimate of the *rate* of convergence (see Proposition 2). The latter estimate is given as a local (error) bound on the distance from S^ϵ to S^*, expressed in terms of the perturbation parameter ϵ and $b^\epsilon - b$. To our knowledge, the only previous bounds of this kind were those obtained by Mangasarian et al. [Man84, MaR79] for the case $b^\epsilon = b$ and f being Lipschitz continuous around S^* (not necessarily separable), for which this distance equals zero for all ϵ sufficiently small. When f is not Lipschitz continuous around S^*, such as when each f_j is the logarithm function or the entropy function, our bound appears to be new, even for the case $b^\epsilon = b$.

In our notation, all vectors are column vectors, $\Re^n$ denotes the space of n–dimensional real column vectors, and ^{T} denotes transpose. For any vector $x \in \Re^n$, we denote by x_i the ith component of x and, for any $I \subseteq \{1, ..., n\}$, by x_I the vector obtained after removing from x those x_i with $i \notin I$. We also denote by $\|x\|$ the Euclidean norm of x, i.e., $\|x\| = \sqrt{x^T x}$. For any $B \in \Re^{m\times n}$ and any $I \subset \{1, ..., m\}$ and $J \subset \{1, ..., n\}$, we denote by B_I the submatrix of B obtained by removing all rows of B with indices outside of I and by B_{IJ} the submatrix of B_I obtained by removing all columns of B_I with indices outside of J. We also denote by $|I|$ the cardinality of I and denote $I^c := \{1, ..., m\}\backslash I$.

2 Convergence of Solutions of Perturbed Problem

Below we derive a general condition for S^ϵ to converge to S^* in a certain sense.

Proposition 1 *Assume there exists an $x^* \in S^*$ and an $\hat{x} \in X$ with $f(\hat{x}) < \infty$. Then, for any sequence of positive scalars $\Upsilon = \{\epsilon^1, \epsilon^2, ...\}$ tending to zero and any sequence of vectors $x^\epsilon \in S^\epsilon$, $\epsilon \in \Upsilon$, converging to some x^∞, we have $x^\infty \in S^*$.*

Proof. If the claim is false, then either (i) $x^\infty \notin S$ or (ii) $x^\infty \in S$ but $x^\infty \notin S^*$. In case (i), since $x^* \in S$ and x^ϵ converges to x^∞ so that $x^\infty \in X$, there is some constant $\rho > 0$ such that $c^T x^\epsilon \geq c^T x^* + \rho$ for all sufficiently small $\epsilon \in \Upsilon$. Now, x^* satisfies $Ax^* \leq b$ and $x^* \geq 0$ and, for each $\epsilon \in \Upsilon$, $Ax \leq b^\epsilon$ and $x \geq 0$ has a solution (namely x^ϵ) so, by a lemma of Hoffman [Hof52], there exists a solution y^ϵ satisfying $\|y^\epsilon - x^*\| \leq \kappa\|b^\epsilon - b\|$ for some constant $\kappa > 0$ depending on A only. Similarly, there exists a solution $\hat{y}^\epsilon$ satisfying $\|\hat{y}^\epsilon - \hat{x}\| \leq \kappa\|b^\epsilon - b\|$. Fix any $\theta \in (0, 1)$ satisfying $\theta(c^T\hat{x} - c^T x^*) < \rho$. Then,

$y^\epsilon \to x^*$ and $\hat{y}^\epsilon \to \hat{x}$ as $\epsilon \to 0$, implying $w^\epsilon = (1-\theta)y^\epsilon + \theta\hat{y}^\epsilon \to (1-\theta)x^* + \theta\hat{x}$ and (since f is continuous at $(1-\theta)x^* + \theta\hat{x}$)

$$c^T w^\epsilon + \epsilon f(w^\epsilon) \;\to\; c^T x^* + \theta(c^T\hat{x} - c^T x^*) \;<\; c^T x^* + \rho \qquad \text{as } \epsilon \to 0.$$

Moreover, w^ϵ is a feasible solution of (P^ϵ). Now, the convergence of x^ϵ implies $f(x^\epsilon)$ is bounded from below, so

$$\liminf_{\epsilon\in\Upsilon}\{c^T x^\epsilon + \epsilon f(x^\epsilon)\} \;\geq\; c^T x^* + \rho \;>\; \lim_{\epsilon\in\Upsilon,\epsilon\to 0}\{c^T w^\epsilon + \epsilon f(w^\epsilon)\},$$

contradicting $x^\epsilon \in S^\epsilon$ for all $\epsilon \in \Upsilon$. In case (ii), we have $\sum_{j\in J^*} f_j(x_j^\infty) > \sum_{j\in J^*} f_j(x_j^*)$. Let $d := x^* - x^\infty$. Then $c^T d = 0$, $d_j = 0$ for all $j \notin J^*$, and $A_I d \leq 0$, where $I := \{i \in \{1,...,m\} : A_i x^\infty = b_i\}$, and for any $\alpha \in (0,1)$ sufficiently near 1, we have $\sum_{j\in J^*} f_j(x_j^\infty + \alpha d_j) < \sum_{j\in J^*} f_j(x_j^\infty)$ and $A_{I^c}(x^\infty + \alpha d) < b_{I^c}$. Then, for all sufficiently small $\epsilon \in \Upsilon$, the vector $z^\epsilon := x^\epsilon + \alpha d$ satisfies $c^T z^\epsilon = c^T x^\epsilon$, $\sum_{j\notin J^*} f_j(z_j^\epsilon) = \sum_{j\notin J^*} f_j(x_j^\epsilon)$, $\sum_{j\in J^*} f_j(z_j^\epsilon) < \sum_{j\in J^*} f_j(x_j^\epsilon)$ and $Az^\epsilon \leq b^\epsilon$, $z^\epsilon \geq 0$, contradicting $x^\epsilon \in S^\epsilon$. ■

The assumptions in Proposition 1 are quite mild and, in particular, if S is nonempty bounded and $\inf_{x\in[0,\infty)^n} f(x) > -\infty$, then S^ϵ is nonempty whenever (P^ϵ) has a feasible solution. If we do not assume x^ϵ converges but do assume it is bounded, then Proposition 1 implies each cluster point of x^ϵ is in S^*. If we do not even assume x^ϵ is bounded, then, while we can still show that $\min_{s\in S} \|x^\epsilon - s\| \to 0$ as $\epsilon \to 0$, we do not know of an analogous result on the convergence of x^ϵ to S^*. In the cases where $b^\epsilon = b$ and each f_j is Lipschitz continuous on $[0,\infty)$ or is convex real-valued on an open interval containing $[0,\infty)$, Proposition 1 can also be deduced from the results of Mangasarian and Meyer [MaR79] (also see [FeM91]). In the case where $b^\epsilon = b$ and f_j is the logarithm function (2), Proposition 1 corresponds to a well-known result about interior-point methods, namely, the convergence of the central path to the analytic center of the optimal face [McL80, Theorem 9]. In the case where $b^\epsilon = b$ and f_j is the entropy function (3), Proposition 1 corresponds to [CoS94, Proposition 4.1]. In the case where $f_j(x_j) = \sup_{y_j\in\Re}(x_j y_j - g_j(y_j))$ [Roc70, Chapter 12] for some convex $g_j : \Re \mapsto \Re \cup \{\infty\}$ satisfying $\lim_{y_j\to-\infty} g_j(y_j)/y_j = 0$ and $\lim_{y_j\to\infty} g_j(y_j)/y_j = \infty$, it can be shown by a standard duality argument that a dual of (P^ϵ) is

$$\text{minimize} \quad (b^\epsilon)^T\lambda + \epsilon\sum_{j=1}^n g_j(-(c_j + (A^T)_j\lambda)/\epsilon) \quad \text{subject to} \quad \lambda \geq 0,$$

so Proposition 1 is closely related to a dual convergence result given in [ACH97, Theorem 3.4] which also assumes $b^\epsilon = b$ and $g_1, ..., g_n$ are identical and strictly convex differentiable. The same reference also studied primal convergence and extensions to convex programming. The weaker conclusion that $x^\infty \in S$ is related to the classical theory for penalty methods [FiM68, Theorem 25]. Otherwise, Proposition 1 appears to be a new result.

3 A Local Error Bound for Solutions of Perturbed Problem

In this section, we consider the case where each f_j is strictly convex on $[0, \infty)$, so that S^ϵ and S^*, if nonempty, have single element x^ϵ and x^*. We derive a bound on the distance between x^ϵ and x^* in terms of ϵ and $b^\epsilon - b$. The proof of this uses Proposition 1, the optimality conditions for (P^ϵ) and (P), as well as a result of Hoffman [Hof52] relating to an upper Lipschitzian property of the solution set of a linear system with respect to right-hand side perturbation [WaW69].

Proposition 2 *Assume each f_j is strictly convex and continuously differentiable on $(0, \infty)$ and define $\nabla f_j(0) := \lim_{\xi \downarrow 0} \nabla f_j(\xi)$. Assume there exists an $x^* \in S^*$ and an $\hat{x} \in X$ with $\nabla f_j(\hat{x}_j) > -\infty$ for all $j = 1, ..., n$. Then, for any sequence of positive scalars $\Upsilon = \{\epsilon^1, \epsilon^2, ...\}$ tending to zero and any bounded sequence of vectors $x^\epsilon \in S^\epsilon$, $\epsilon \in \Upsilon$, there exist positive constants $\bar{\epsilon}$, ρ and τ such that, for all $\epsilon \in \Upsilon$ with $\epsilon \leq \bar{\epsilon}$, we have*

$$x_j^\epsilon \leq \nabla f_j^{-1}(-\rho/\epsilon) \;\; \forall j \in K, \qquad \sum_{j \in J} \phi_j(x_j^\epsilon, x_j^*) + \|x_L^\epsilon\| \leq \tau \left(\|b^\epsilon - b\| + \|x_K^\epsilon\|\right),$$

where $J := \{j \in \{1, ..., n\} : \nabla f_j(x_j^*) > -\infty\}$, K is some subset of $J^c := \{1, ..., n\} \backslash J$ and $L = J^c \backslash K$ (K may depend on ϵ), and $\phi_j(\xi, \psi) := (\xi - \psi)(\nabla f_j(\xi) - \nabla f_j(\psi))$ for each $j \in J$.

Proof. Since each f_j is strictly convex on $(0, \infty)$, x^* must be the only element of S^*. Since x^ϵ is bounded and hence has convergent subsequences, then Proposition 1 yields $x^\epsilon \to x^*$ as $\epsilon \to 0$. Also, $\hat{x}$ satisfies $A\hat{x} \leq b, \hat{x} \geq 0$ and, for each $\epsilon \in \Upsilon$, $Ax \leq b^\epsilon, x \geq 0$ has a solution (namely x^ϵ), so, by a lemma of Hoffman [Hof52], there exists a solution y^ϵ satisfying $\|y^\epsilon - \hat{x}\| \leq \kappa \|b^\epsilon - b\|$, where $\kappa > 0$ is a constant depending on A only. Thus, $y^\epsilon \to \hat{x}$ as $\epsilon \to 0$, implying (since $\nabla f_j(\hat{x}_j) > -\infty$ for all j) $y_j^\epsilon \to \hat{x}_j > 0$ whenever $\nabla f_j(0) = -\infty$. Moreover, y^ϵ is a feasible solution of (P^ϵ). Since f is convex on $[0, \infty)^n$ and the directional derivative of $x \mapsto c^T x + \epsilon f(x)$ at x^ϵ in the direction $y^\epsilon - x^\epsilon$ is

$$c^T(y^\epsilon - x^\epsilon) + \epsilon \sum_{j=1}^{n} \nabla f_j(x_j^\epsilon)(y_j^\epsilon - x_j^\epsilon),$$

it must be that $\nabla f_j(x_j^\epsilon) > -\infty$ and $\epsilon \nabla f_j(x_j^\epsilon)$ is bounded as $\epsilon \to 0$, for all $j = 1, ..., n$. [Otherwise, because $\nabla f_j(x_j^\epsilon) = -\infty$ implies $x_j^\epsilon = 0$ and $\epsilon \nabla f_j(x_j^\epsilon)$ being unbounded implies $\epsilon \nabla f_j(x_j^\epsilon) \to -\infty$ and $x_j^\epsilon \to 0$ as $\epsilon \to 0$ along some subsequence of Υ, this directional derivative would be negative (possibly $-\infty$) for some $\epsilon \in \Upsilon$, contradicting $x^\epsilon \in S^\epsilon$.]

For each $\epsilon \in \Upsilon$, since x^ϵ is an optimal solution of the convex program (P^ϵ) and $\nabla f_j(x_j^\epsilon) > -\infty$ for $j = 1, ..., n$ (so $x_{J^c}^\epsilon > 0$), it follows from the Karush-Kuhn-Tucker theorem (see [Roc70, Corollary 28.3.1]) that there exist index sets $I \subset \{1,...,m\}$, $M \subset J$, and $\lambda^\epsilon \in \Re^m$ satisfying

$$\epsilon \nabla f_j(x_j^\epsilon) + c_j + (A^T)_j \lambda^\epsilon \begin{cases} = 0 & j \in M \cup J^c \\ \geq 0 & j \in J \backslash M \end{cases}, \quad \lambda_I^\epsilon \geq 0, \; \lambda_{I^c}^\epsilon = 0, \tag{4}$$

$$x_{J\backslash M}^\epsilon = 0, \quad A_I x^\epsilon = b_I^\epsilon, \quad A_{I^c} x^\epsilon < b_{I^c}^\epsilon. \tag{5}$$

The number of such index sets I and M is finite and independent of ϵ, so, by passing into a subsequence if necessary, we can assume I and M are fixed for all $\epsilon \in \Upsilon$. Also, since $\epsilon \nabla f(x^\epsilon)$ is bounded as $\epsilon \to 0$, it follows from a lemma of Hoffman [Hof52] that there exists λ^ϵ satisfying (4) that is bounded as $\epsilon \to 0$. Thus, we can assume that λ^ϵ is bounded.

Then, for sufficiently small $\epsilon \in \Upsilon$ so that $\nabla f_j(x_j^\epsilon) \leq 0$ for all $j \in J^c$ (since $\nabla f_j(x_j^\epsilon) \to -\infty$ as $\epsilon \to 0$), we have from (4) that λ^ϵ satisfies

$$\epsilon \nabla f_j(x_j^\epsilon) + c_j + (A^T)_j \lambda^\epsilon \begin{cases} = 0 & j \in M \\ \geq 0 & j \in J \backslash M \end{cases}, \quad c_{J^c} + (A^T)_{J^c} \lambda^\epsilon \geq 0, \quad \lambda_I^\epsilon \geq 0, \; \lambda_{I^c}^\epsilon = 0.$$

Since $\epsilon \nabla f_j(x_j^\epsilon) \to 0$ as $\epsilon \to 0$, for all $j \in J$, so the system

$$c_j + (A^T)_j \lambda \begin{cases} = 0 & j \in M \\ \geq 0 & j \in J \backslash M \end{cases}, \quad c_{J^c} + (A^T)_{J^c} \lambda \geq 0, \quad \lambda_I \geq 0, \; \lambda_{I^c} = 0, \tag{6}$$

has a solution (namely, any cluster point of λ^ϵ as $\epsilon \to 0$), it follows from a lemma of Hoffman [Hof52] that there exists a solution μ^ϵ satisfying

$$\|\lambda^\epsilon - \mu^\epsilon\| \leq \kappa \; \epsilon \|\nabla f_j(x_j^\epsilon)_{j \in J}\|, \tag{7}$$

where $\kappa > 0$ is a constant depending on A only.

Let Λ denote the set of $\lambda \in \Re^m$ satisfying (6) and $\|\lambda\| \leq \sup_{\epsilon \in \Upsilon} \|\mu^\epsilon\|$. We claim that there exists a scalar $\rho > 0$ such that, for every $\lambda \in \Lambda$ there exists a $K \subset J^c$ such that

$$c_j + (A^T)_j \lambda > \rho \; \forall j \in K \quad \text{and} \quad c_L + (A^T)_L \mu = 0 \text{ for some } \mu \in \Lambda, \tag{8}$$

where $L := J^c \backslash K$. If not, then for every sequence of scalars $\rho^k > 0$, $k = 1, 2, ...$, tending to zero, there would exist a $\nu^k \in \Lambda$ such that, for every $K \subset J^c$ we have

$$c_j + (A^T)_j \nu^k \leq \rho^k \;\text{ for some } j \in K \quad \text{or} \quad c_L + (A^T)_L \mu \neq 0 \; \forall \mu \in \Lambda,$$

where $L := J^c \backslash K$. Since Λ is bounded and closed, then ν^k, $k = 1, 2, ...$, has a cluster point $\nu \in \Lambda$ such that, for every $K \subset J^c$ we have

$$c_j + (A^T)_j \nu = 0 \;\text{ for some } j \in K \quad \text{or} \quad c_L + (A^T)_L \mu \neq 0 \; \forall \mu \in \Lambda,$$

where $L := J^c \backslash K$. However, this cannot be true since the above relations fail to hold for $K = \{j \in J^c : c_j + (A^T)_j \nu > 0\}$ and $\mu = \nu$.

For each $\epsilon \in \Upsilon$, we have $\mu^\epsilon \in \Lambda$ and hence there exists a $K \subset J^c$ such that (8) holds with $\lambda = \mu^\epsilon$ and $L := J^c \backslash K$. Since the number of such subset K is finite and independent of ϵ, by passing into a subsequence if necessary, we can assume it is the same K and L for all $\epsilon \in \Upsilon$. Since, by (7), $\lambda^\epsilon - \mu^\epsilon \to 0$ as $\epsilon \to 0$, we have $c_j + (A^T)_j \lambda^\epsilon \geq \rho$ for all $j \in K$ and sufficiently small $\epsilon \in \Upsilon$, in which case (4) and the nondecreasing property of ∇f_j^{-1} would imply

$$x_j^\epsilon = \nabla f_j^{-1} \left(-(c_j + (A^T)_j \lambda^\epsilon)/\epsilon \right) \leq \nabla f_j^{-1} \left(-\rho/\epsilon \right) \quad \forall j \in K. \tag{9}$$

Since $x^\epsilon \to x^*$, we have from (5) that $x^*_{J\backslash M} = 0$ and $I \subset I^*$, where $I^* := \{i \in \{1, ..., m\} : A_i x^* = b_i\}$. Also, $x^*_{J^c} = 0$. For convenience, let $\Delta^\epsilon := b^\epsilon - b$. We claim that there exists a constant $\tau_1 > 0$ such that

$$\|x_L^\epsilon\| \leq \tau_1 (\|\Delta^\epsilon_{I^*}\| + \|x^\epsilon_K\|). \tag{10}$$

If not, then $(\|\Delta^\epsilon_{I^*}\| + \|x^\epsilon_K\|)/\|x^\epsilon_L\| \to 0$ as $\epsilon \to 0$ along some subsequence of Υ. Since $x^*_{J^c} = 0$ and $A_I x^* = b$ and $A_I x^\epsilon = b^\epsilon_I$ so that $A_{IJ}(x^\epsilon_J - x^*_J) + A_{IL} x^\epsilon_L = \Delta^\epsilon_I - A_{IK} x^\epsilon_K$, dividing both sides by $\|x^\epsilon_L\|$ would yield in the limit (note that $I \subset I^*$)

$$A_{IJ}(x^\epsilon_J - x^*_J)/\|x^\epsilon_L\| + A_{IL} x^\epsilon_L/\|x^\epsilon_L\| = (\Delta^\epsilon_I - A_{IK} x^\epsilon_K)/\|x^\epsilon_L\| \to 0.$$

Similarly, since $x^*_{J^c} = 0$ and $A_H x^* = b_H$ and $A_H x^\epsilon \leq b^\epsilon_H$, where we denote $H := I^* \backslash I$, so that $A_{HJ}(x^\epsilon_J - x^*_J) + A_{HL} x^\epsilon_L \leq \Delta^\epsilon_H - A_{HK} x^\epsilon_K$, dividing both sides by $\|x^\epsilon_L\|$ would yield in the limit

$$A_{HJ}(x^\epsilon_J - x^*_J)/\|x^\epsilon_L\| + A_{HL} x^\epsilon_L/\|x^\epsilon_L\| \leq (\Delta^\epsilon_H - A_{HK} x^\epsilon_K)/\|x^\epsilon_L\| \to 0.$$

Thus, $A_{IJ}\theta_J + A_{IL}\theta_L = 0$ and $A_{HJ}\theta_J + A_{HL}\theta_L \leq 0$ for some $\theta_J \in \Re^{|J|}$ and some nonzero $\theta_L \in [0, \infty)^{|L|}$. Moreover, $\theta_j \geq 0$ for all $j \in J$ with $x^*_j = 0$ and $\theta_{J\backslash M} = 0$. Then, for $\alpha > 0$ sufficiently small, the vector $x \in \Re^n$ given by $x_J := x^*_J + \alpha\theta_J, x_L := \alpha\theta_L$, $x_K := 0$ would satisfy $x \geq 0$, $x_{J\backslash M} = 0$, $A_I x = A_I x^* = b_I$, and $A_{I^c} x \leq b_{I^c}$. Then $x \in X$ and, together with any $\mu \in \Lambda$ satisfying $c_L + (A^T)_L \mu = 0$ (see (8)), satisfies the Kuhn-Tucker conditions for (P), so $x \in S$. Moreover, the directional derivative of f at x^* in the direction $x - x^*$ would be negative (since $x^*_j = 0$ and $\nabla f_j(0) = -\infty$ for all $j \in K \cup L = J^c$), contradicting the definition of x^*.

For $j \in J \backslash M$, we have from (5) and $x^\epsilon_j \to x^*_j$ that $x^\epsilon_j = x^*_j = 0$ for all $\epsilon \in \Upsilon$. Lastly, we estimate $x^\epsilon_M - x^*_M$. From (4) and μ^ϵ being a solution of (6) we have

$$\nabla f_j(x^\epsilon_j) + (A^T)_j (\lambda^\epsilon - \mu^\epsilon)/\epsilon = 0 \ \forall j \in M, \quad \lambda^\epsilon_{I^c} - \mu^\epsilon_{I^c} = 0$$

for all $\epsilon \in \Upsilon$. Since, by (7), $(\lambda^\epsilon - \mu^\epsilon)/\epsilon$ is bounded as $\epsilon \to 0$, then it has a cluster point $\pi \in \Re^m$ satisfying

$$\nabla f_j(x^*_j) + (A^T)_j \pi = 0 \ \forall j \in M, \quad \pi_{I^c} = 0.$$

Multiplying the first set of equations by ϵ and adding them to $c_j + (A^T)_j \mu^\epsilon = 0$, $j \in M$ (since μ^ϵ is a solution of (6)), gives

$$\epsilon \nabla f_j(x_j^*) + c_j + (A^T)_j(\mu^\epsilon + \epsilon\pi) = 0 \ \forall j \in M.$$

Subtracting this from the first set of equations in (4) corresponding to $j \in M$ gives

$$\epsilon \left(\nabla f_j(x_j^\epsilon) - \nabla f_j(x_j^*) \right) + (A^T)_j(\lambda^\epsilon - \mu^\epsilon - \epsilon\pi) = 0 \ \forall j \in M.$$

Multiplying both sides by $(x_j^\epsilon - x_j^*)$ and summing over all $j \in M$ gives the single equation

$$\epsilon (x_M^\epsilon - x_M^*)^T \left(\nabla f_j(x_j^\epsilon) - \nabla f_j(x_j^*) \right)_{j\in M} = (A_M x_M^\epsilon - A_M x_M^*)^T (\mu^\epsilon - \lambda^\epsilon + \epsilon\pi). \quad (11)$$

Since $\lambda_{I^c}^\epsilon = \mu_{I^c}^\epsilon = \pi_{I^c} = 0$ and $A_I x^* = b_I$ and $A_I x^\epsilon = b_I + \Delta_I^\epsilon$, then $(Ax^\epsilon - Ax^*)^T(\mu^\epsilon - \lambda^\epsilon + \epsilon\pi) = (\Delta_I^\epsilon)^T(\mu_I^\epsilon - \lambda_I^\epsilon + \epsilon\pi_I)$. This, together with $x_j^* = 0$ for $j \in \{1, ..., n\}\backslash M$ and $x_{J\backslash M}^\epsilon = 0$, implies

$$\begin{aligned} & (A_M x_M^\epsilon - A_M x_M^*)^T(\mu^\epsilon - \lambda^\epsilon + \epsilon\pi_I) \\ = \ & (Ax^\epsilon - Ax^*)^T(\mu^\epsilon - \lambda^\epsilon + \epsilon\pi) - (A_{J^c} x_{J^c}^\epsilon)^T(\mu^\epsilon - \lambda^\epsilon + \epsilon\pi) \\ = \ & (\Delta_I^\epsilon)^T(\mu_I^\epsilon - \lambda_I^\epsilon + \epsilon\pi_I) - (A_{J^c} x_{J^c}^\epsilon)^T(\mu^\epsilon - \lambda^\epsilon + \epsilon\pi). \quad (12) \end{aligned}$$

Using (7), (10), and (12) to bound the right-hand side of (11) (also noting that $\nabla f_j(x_j^\epsilon)_{j\in J}$ is bounded as $\epsilon \to 0$) yields

$$\epsilon \ (x_M^\epsilon - x_M^*)^T \left(\nabla f_j(x_j^\epsilon) - \nabla f_j(x_j^*) \right)_{j\in M} \leq \epsilon \ \tau_2 \left(\|\Delta_{I^*}^\epsilon\| + \|x_K^\epsilon\| \right),$$

for some constant $\tau_2 > 0$. The above inequality, together with (9) and (10) and $x_{J\backslash M}^\epsilon = x_{J\backslash M}^* = 0$, completes the proof ∎

We note that Proposition 2 still holds if we allow f_j to be non-differentiable at a finite number of points in $(0, \infty)$, and the proof of this entails only minor modifications. If in addition each f_j is locally uniformly convex of order $\beta \geq 1$ at x_j^* for all $j \in J$ (i.e., there exists a constant $\sigma > 0$ such that $\phi_j(x_j, x_j^*) \geq \sigma |x_j - x_j^*|^\beta$ for all x_j near x_j^*, $j \in J$), then Proposition 2 would imply the distance bound

$$\sigma \sum_{j\in J} |x_j^\epsilon - x_j^*|^\beta + \|x_{J^c}^\epsilon - x_{J^c}^*\| \leq \tau \left(\|b^\epsilon - b\| + \|\nabla f_j^{-1}(-\rho/\epsilon)_{j\in K}\| \right)$$

for all sufficiently small $\epsilon \in \Upsilon$. In the case where $f_j(x_j) = -\ln(x_j)$, we have $\beta = 2$ and $\nabla f_j^{-1}(-\rho/\epsilon) = \epsilon/\rho$ which tends to zero linearly with ϵ. In the case where $f_j(x_j) = x_j \ln(x_j)$, we have $\beta = 2$ and $\nabla f_j^{-1}(-\rho/\epsilon) = exp(-\rho/\epsilon + 1)$ which tends to zero exponentially with ϵ. In the case where each f_j is strictly convex and ∇f_j is defined and continuous on $[0, \infty)$ for all j (e.g., $f_j(x_j) = (x_j)^2/2$), we have $K = L = \emptyset$

and Proposition 2 implies that, if in addition $b^\epsilon = b$, then $x^\epsilon = x^*$ for all sufficiently small ϵ. This finite perturbation result is a special case of those shown in [MaR79] and, interestingly, it can also be deduced from Proposition 2. Notice that in the three cases considered above, the best approximation of x^* by x^ϵ, as $\epsilon \to 0$, is obtained in the third case and the worst approximation is obtained in the first case. For practical computation, however, the first case has been more favoured. Of course, we can use other choices of f_j, such as $f_j(x_j) = 1/x_j$ or $f_j(x_j) = -\sqrt{x_j}$, etc., and use Proposition 2 to obtain corresponding error bounds.

Notice that the index subset K in Proposition 2 depends on ϵ. The following example shows that this dependence cannot be removed in general. Consider the perturbed problem:

$$\begin{array}{ll} \text{minimize} & x_1 + \epsilon f_1(x_1) + \epsilon f_2(x_2) \\ \text{subject to} & -x_1 + x_2 \leq 0, \quad x_1 \geq 0, \ x_2 \geq 0, \end{array} \tag{13}$$

where $f_1(x_1) = \frac{1}{p-1}(x_1)^{1-p}$ for $x_1 > 0$ and $f_2(x_2) = \frac{1}{q-1}(x_2)^{1-q}$ for $x_2 > 0$, with $p > 1, q > 1$ (so $\nabla f_1(x_1) = -(x_1)^{-p}$, $\nabla f_2(x_2) = -(x_2)^{-q}$). This problem satisfies the assumptions of Proposition 2 with $S^* = \{(0,0)\}$ and $J = \emptyset$. Moreover, direct calculation yields the optimal solution $x^\epsilon = (x_1^\epsilon, x_2^\epsilon)$ and Lagrange multiplier $\lambda^\epsilon \in [0,1]$ given by

$$x_1^\epsilon = (\epsilon/(1-\lambda^\epsilon))^{1/p} = x_2^\epsilon = (\epsilon/\lambda^\epsilon)^{1/q}, \qquad (1-\lambda^\epsilon)^q = \epsilon^{q-p}(\lambda^\epsilon)^p.$$

Thus, if $p < q$, then $1 - \lambda^\epsilon \to 0$ as $\epsilon \to 0$, so $x_1^\epsilon = x_2^\epsilon$ is in the order of $(\epsilon)^{1/q}$, with $K = \{2\}$. [Since $p < q$, x_1^ϵ cannot be below $\nabla f_1^{-1}(-\rho/\epsilon) = (\epsilon/\rho)^{1/p}$ for any positive constant ρ.] Symmetrically, if $p > q$, then $\lambda^\epsilon \to 0$ as $\epsilon \to 0$, so $x_1^\epsilon = x_2^\epsilon$ is in the order of $(\epsilon)^{1/p}$, with $K = \{1\}$. This shows that K is determined by the relative growth rate of ∇f_1 and ∇f_2 near 0. Now suppose we modify f_1 and f_2 so that these two growth rates alternate in dominance. In particular, let $\alpha_{k+1} = (\alpha_k)^2/2$ for $k = 1, 2, ...$, with $\alpha_1 = 1$, and choose strictly convex and continuously differentiable functions f_1 and f_2 on $(0, \infty)$ satisfying $\nabla f_1(\alpha_k) = -(\alpha_k)^{-1}$ and $\nabla f_2(\alpha_k) = -(\alpha_k)^{-2}$ for k odd and $\nabla f_1(\alpha_k) = -(\alpha_k)^{-2}$ and $\nabla f_2(\alpha_k) = -(\alpha_k)^{-1}$ for k even. [Notice that $\alpha_k \downarrow 0$ and $\nabla f_j(\alpha_k) \downarrow -\infty$ as $k \to \infty$ for $j = 1, 2$, so the function f_j can be constructed by interpolating ∇f_j using its value at $\alpha_1, \alpha_2, ...$ and then integrating.] With this choice of f_1 and f_2, the problem (13) still satisfies the assumptions of Proposition 2 with $S^* = \{(0,0)\}$ and $J = \emptyset$. Moreover, it can be seen that $x_1^\epsilon = x_2^\epsilon$ tends to zero continuously in ϵ, so for each k sufficiently large, we can find ϵ^k such that $x_1^{\epsilon^k} = x_2^{\epsilon^k} = \alpha_k$. Then, arguing as above, we obtain that $K = \{2\}$ for ϵ along the subsequence $\Upsilon_{\text{odd}} = \{\epsilon^k\}_{k \text{ odd}}$, and $K = \{1\}$ for ϵ along the subsequence $\Upsilon_{\text{even}} = \{\epsilon^k\}_{k \text{ even}}$. So, for ϵ along $\Upsilon = \Upsilon_{\text{odd}} \cup \Upsilon_{\text{even}}$, K would depend on ϵ. It is an open (and nontrivial) question whether K would still depend on ϵ if $f_1, ..., f_n$ are identical.

Acknowledgement. I thank the two referees for their reading of this paper and their comments.

References

[ACH97] Auslender, A., Cominetti, R., and Haddou, M., Asymptotic Analysis for Penalty and Barrier Methods in Convex and Linear Programming, *Mathematics of Operations Research*, Vol. 22, pp. 43–62, 1997.

[Ber82] Bertsekas, D. P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, New York, 1982.

[CoD94] Cominetti, R. and Dussault, J. P., Stable Exponential-Penalty Algorithm with Superlinear Convergence, *Journal of Optimization Theory and Applications*, Vol. 83, pp. 285–309, 1994.

[CoS94] Cominetti, R. and San Martín, J., Asymptotic Analysis of the Exponential Penalty Trajectory in Linear Programming, *Mathematical Programming*, Vol. 67, pp. 169–187, 1994.

[Fan92] Fang, S. C., An Unconstrained Convex Programming View of Linear Programming, *Zeitschrift für Operations Research – Methods and Models of Operations Research*, Vol. 36, pp. 149–161, 1992.

[FRT97] Fang, S. C., Rajasekera, J. R., and Tsao, H.-S. J., *Entropy Optimization and Mathematical Programming*, Kluwer, Norwell, 1997.

[FaT93] Fang, S. C., and Tsao, H.-S. J., Linear Programming with Entropic Perturbation, *Zeitschrift für Operations Research – Methods and Models of Operations Research*, Vol. 37, pp. 171–186, 1993.

[FaT96] Fang, S. C., and Tsao, H.-S. J., On the Entropic Perturbation and Exponential Penalty Methods for Linear Programming, *Journal of Optimization Theory and Applications*, to appear.

[FeM91] Ferris, M. C. and Mangasarian, O. L., Finite Perturbation of Convex Programs, *Applied Mathematics and Optimization*, Vol. 23, pp. 263–273, 1991.

[FiM68] Fiacco, A. V. and McCormick, G. P., *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968; reprinted by SIAM, Philadelphia, 1990.

[Hof52] Hoffman, A. J., On Approximate Solutions of Systems of Linear Inequalities, *Journal of Research of the National Bureau of Standards*, 49 (1952), 263–265.

[Kar59] Karlin, S., *Mathematical Methods and Theory in Games, Programming, and Economics*, Vol. I, Addison-Wesley, Reading, 1959.

[MNP96] Madsen, K., Nielsen, H. B., and Pinar, M. Ç., A New Finite Continuation Algorithm for Linear Programming, *SIAM Journal on Optimization*, Vol. 6, pp. 600–616, 1996.

[Man81] Mangasarian, O. L., Iterative Solution of Linear Programs, *SIAM Journal on Numerical Analysis*, Vol. 18, pp. 606–614, 1981.

[Man84] Mangasarian, O. L., Normal Solution of Linear Programs, *Mathematical Programming Study*, Vol. 22, pp. 206-216, 1984.

[Man86] Mangasarian, O. L., Some Applications of Penalty Functions in Mathematical Programming, in *Optimization and Related Fields*, edited by R. Conti, E. De Giorgi, and F. Giannessi, pp. 307-329, Springer-Verlag, Heidelberg, 1986.

[MaR79] Mangasarian, O. L. and Meyer, R. R., Nonlinear Perturbation of Linear Programs, *SIAM Journal on Control and Optimization*, Vol. 17, pp. 745-752, 1979.

[McL80] McLinden, L., An Analogue of Moreau's Proximation Theorem, *Pacific Journal of Mathematics*, Vol. 88, pp. 101–161, 1980.

[Meg89] Megiddo, N., Pathways to the Optimal Set in Linear Programming, in *Progress in Mathematical Programming, Interior-Point and Related Methods*, edited by N. Megiddo, pp. 131–158, Springer-Verlag, 1989.

[Pin96] Pinar, M. Ç., Piecewise-Linear Pathways to the Optimal Solution Set in Linear Programming, *Journal of Optimization Theory and Applications*, to appear.

[RaF92] Rajasekera, J. R., and Fang, S. C., Deriving an Unconstrained Convex Program for Linear Programming, *Journal of Optimization Theory and Applications*, Vol. 75, pp. 603–612, 1992.

[Roc70] Rockafellar, R. T., *Convex Analysis*, Princeton University Press, Princeton, 1970.

[Tse95] Tseng, P., Linearly Constrained Convex Programming as Unconstrained Differentiable Concave Programming, *Journal of Optimization Theory and Applications*, Vol. 85, 489–494, 1995.

[WaW69] D.W. Walkup and R.J.-B. Wets, A Lipschitzian characterization of convex polyhedra, *Proceedings of the American Mathematical Society*, Vol. 23, pp. 167–173, 1969.

[Wri92] Wright, M. H., Interior Methods for Constrained Optimization, in *Acta Numerica 1992*, edited by A. Iserles, pp. 341–407, Cambridge University Press, New York, 1992.