July 2004 **Submitted for publication**

Enhanced Fritz John Conditions for Convex Programming[1]

by

Dimitri P. Bertsekas, Asuman E. Ozdaglar,[2] and Paul Tseng[3]

Abstract

We consider convex constrained optimization problems, and we enhance the classical Fritz John optimality conditions to assert the existence of multipliers with special sensitivity properties. In particular, we prove the existence of Fritz John multipliers that are informative in the sense that they identify constraints whose relaxation, at rates proportional to the multipliers, strictly improves the primal optimal value. Moreover, we show that if the set of geometric multipliers is nonempty, then the minimum-norm vector of this set is informative, and defines the optimal rate of cost improvement per unit constraint violation. Our assumptions are very general, and allow for the presence of duality gap and the non-existence of optimal solutions. In particular, for the case where there is a duality gap, we establish enhanced Fritz John conditions involving the dual optimal value and dual optimal solutions.

[1] Research supported by NSF Grant ECS-0218328.
[2] D. Bertsekas and A. Ozdaglar are with the Dept. EECS, M.I.T., Cambridge, MA 02139.
[3] P. Tseng is with the Dept. Mathematics, University of Washington, Seattle, WA 98195

1. INTRODUCTION

We consider the convex constrained optimization problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in X, \quad g(x) = \big(g_1(x), \ldots, g_r(x)\big)' \le 0, \end{aligned} \tag{P}$$

where X is a nonempty convex subset of $\Re^n$, and $f : X \to \Re$ and $g_j : X \to \Re$ are convex functions. Here and throughout the paper, we denote by $\Re$ the real line, by $\Re^n$ the space of n-dimensional real column vectors with the standard Euclidean norm, $\|\cdot\|$, and we denote by $'$ the transpose of a vector. We say that a function $f : X \to \Re$ is convex (respectively, closed) if its epigraph $\big\{(x, w) \mid x \in X,\, f(x) \le w\big\}$ is convex (respectively, closed). For some of our results, we will assume that $f, g_1, \ldots, g_r$ are also closed. We note that our analysis readily extends to the case where there are affine equality constraints by replacing each affine equality constraint with two affine inequality constraints.

We refer to problem (P) as the *primal problem* and we consider the *dual problem*

$$\begin{aligned} &\text{maximize} && q(\mu) \\ &\text{subject to} && \mu \ge 0, \end{aligned} \tag{D}$$

where q is the *dual function*:

$$q(\mu) = \inf_{x \in X} \big\{ f(x) + \mu' g(x) \big\}, \qquad \mu \in \Re^r.$$

We denote by f^* and q^* the optimal values of (P) and (D), respectively:

$$f^* = \inf_{x \in X,\, g(x) \le 0} f(x), \qquad q^* = \sup_{\mu \ge 0} q(\mu).$$

We write $f^* < \infty$ or $q^* > -\infty$ to indicate that (P) or (D), respectively, has at least one feasible solution. The weak duality theorem states that $q^* \le f^*$. If $q^* = f^*$, we say that there is *no duality gap*.

An important result, attributed to Fritz John [Joh48], is that there exist a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)'$ satisfying the following conditions:

(i) $\mu_0^* f^* = \inf_{x \in X} \big\{ \mu_0^* f(x) + \mu^{*\prime} g(x) \big\}$.

(ii) $\mu_j^* \ge 0$ for all $j = 0, 1, \ldots, r$.

(iii) $\mu_0^*, \mu_1^*, \ldots, \mu_r^*$ are not all equal to 0.

We call such a pair (μ_0^*, μ^*) a *FJ-multiplier*.

If the coefficient μ_0^* of a FJ-multiplier is nonzero, by normalization one can obtain a FJ-multiplier of the form $(1, \mu^*)$, and we have

$$\mu^* \geq 0, \qquad f^* = q(\mu^*). \tag{1.1}$$

A vector μ^* thus obtained is called a *geometric multiplier*. It is well known and readily seen from the weak duality theorem that μ^* is a geometric multiplier if and only if there is no duality gap and μ^* is an optimal solution of the dual problem. It is further known that the set of geometric multipliers is closed and coincides with the negative of the subdifferential of the *perturbation function*

$$p(u) = \inf_{x \in X,\, g(x) \leq u} f(x)$$

at $u = 0$, provided that p is convex, proper, and $p(0)$ is finite [Roc70, Theorem 29.1], [BNO03, Prop. 6.5.8]. If in addition the origin is in the relative interior of $\mathrm{dom}(p)$ (a constraint qualification that guarantees that the set of geometric multipliers is nonempty), and μ^* is the geometric multiplier of minimum norm, then either $\mu^* = 0$, in which case $0 \in \partial p(0)$ and $u = 0$ is a global minimum of p, or else $\mu^* \neq 0$, in which case μ^* is a direction of steepest descent for p at $u = 0$. More specifically, if $\mu^* \neq 0$, the directional derivative $p'(0; d)$ of p at 0 in the direction d satisfies

$$\inf_{\|d\|=1} p'(0; d) = p'\big(0; \mu^*/\|\mu^*\|\big) = -\|\mu^*\| < 0. \tag{1.2}$$

based on the calculation

$$\inf_{\|d\|=1} p'(0; d) = \inf_{\|d\|=1} \sup_{v \in \partial p(0)} d'v = \sup_{v \in \partial p(0)} \inf_{\|d\|=1} d'v = \sup_{v \in \partial p(0)} \{-\|v\|\} = -\|\mu^*\|.$$

Thus, the minimum-norm geometric multiplier provides useful sensitivity information, namely, relaxing the inequality constraints at rates equal to the components of $\mu^*/\|\mu^*\|$ yields a decrease of the optimal value at the optimal rate, which is equal to $\|\mu^*\|$.

On the other hand, if the origin is not a relative interior point of $\mathrm{dom}(p)$ [the effective domain $\{u \mid p(u) < \infty\}$ of p], there may be no direction of steepest descent, because the directional derivative function $p'(0; \cdot)$ is discontinuous, and the infimum over $\|d\| = 1$ in Eq. (1.2) may not be attained. This can happen even if there is no duality gap and there exists a geometric multiplier. As an example, consider the following two-dimensional problem:

$$\begin{aligned} &\text{minimize} \quad -x_2 \\ &\text{subject to} \quad x \in X = \{x \mid x_2^2 \leq x_1\}, \quad g_1(x) = x_1 \leq 0, \quad g_2(x) = x_2 \leq 0. \end{aligned}$$

It can be verified that

$$\mathrm{dom}(p) = \{u \mid u_2^2 \leq u_1\} + \{u \mid u \geq 0\},$$

and

$$p(u) = \begin{cases} -u_2 & \text{if } u_2^2 \le u_1, \\ -\sqrt{u_1} & \text{if } u_1 \le u_2^2,\ u_1 \ge 0,\ u_2 \ge 0, \\ \infty & \text{otherwise,} \end{cases}$$

while

$$q(\mu) = \begin{cases} -\frac{(\mu_2 - 1)^2}{4\mu_1} & \text{if } \mu_1 > 0, \\ 0 & \text{if } \mu_1 = 0,\ \mu_2 = 1, \\ -\infty & \text{otherwise.} \end{cases}$$

We have $f^* = q^* = 0$, and the set of geometric multipliers is

$$\{\mu \ge 0 \mid \mu_2 = 1\}.$$

However, the geometric multiplier of minimum norm, $\mu^* = (0,1)$, is not a direction of steepest descent, since starting at $u = 0$ and going along the direction $(0,1)$, $p(u)$ is equal to 0, so

$$p'(0; \mu^*) = 0.$$

In fact p has no direction of steepest descent at $u = 0$, because $p'(0; \cdot)$ is not continuous. To see this, note that directions of descent $d = (d_1, d_2)$ are those for which $d_1 > 0$ and $d_2 > 0$, and that along any such direction, we have

$$p'(0; d) = -d_2.$$

It follows that

$$\inf_{\|d\|=1} p'(0; d) = -1 = -\|\mu^*\|,$$

but there is no direction of descent that attains the infimum above. On the other hand, there are sequences $\{u^k\} \subset \text{dom}(p)$ and $\{x^k\} \subset X$ of infeasible points [in fact, the sequences $u^k = x^k = (1/k^2,\ 1/k)$] such that

$$\lim_{k\to\infty} \frac{p(0) - p(u^k)}{\|{u^k}^+\|} = \lim_{k\to\infty} \frac{f^* - f(x^k)}{\|g^+(x^k)\|} = \|\mu^*\| = 1,$$

where we denote

$$u_j^+ = \max\{0, u_j\},\ u^+ = (u_1^+, \ldots, u_r^+)', \quad g_j^+(x) = \max\bigl\{0, g_j(x)\bigr\},\ g^+(x) = \bigl(g_1^+(x), \ldots, g_r^+(x)\bigr)'.$$

Thus, the minimum norm of geometric multipliers can still be interpreted as the optimal rate of improvement of the cost per unit constraint violation. However, this rate of improvement cannot be obtained by approaching 0 along a straight line, but only by approaching it along a curve.

In this paper, we derive more powerful versions of the Fritz John conditions, which provide sensitivity information like the one discussed above. In particular, in addition to conditions (i)-(iii) above, we obtain an additional necessary condition [e.g., condition (CV) of Prop. 2.1 in the next section] that narrows down the set of candidates for optimality. Furthermore, our conditions also apply in the exceptional case where the set of geometric multipliers is empty. In this case, we will show that a certain degenerate FJ-multiplier, i.e., one of the form $(0, \mu^*)$ with $\mu^* \neq 0$ and $0 = \inf_{x \in X} \mu^{*\prime} g(x)$, provides sensitivity information analogous to that provided by the minimum-norm geometric multiplier. In particular, there exists a FJ-multiplier $(0, \mu^*)$ such that by relaxing the inequality constraints at rates proportional to the components of $\mu^*/\|\mu^*\|$, we can strictly improve the primal optimal value. Furthermore, $\|\mu^*\|$ is the optimal rate of improvement per unit constraint violation. In the case where there is a duality gap, we also prove dual versions of these results, involving the dual optimal value, and dual FJ-multipliers. To our knowledge, except for a preliminary version of our work that appeared in the book [BNO03], these are the first results that provide enhanced, sensitivity-related Fritz John conditions for convex programming, and also derive the optimal sensitivity rate under very general assumptions, i.e., without any constraint qualification and even in the presence of a duality gap.

This paper is organized as follows. In Section 2, we present enhanced Fritz John conditions for convex problems that have optimal solutions. In Section 3, we present analogous results for convex problems that have dual optimal solutions. In particular, we show that the dual optimal solution of minimum norm provides useful sensitivity information, even in the presence of a duality gap. In Section 4, we present Fritz John conditions for problems that may not have optimal solutions, we introduce the notion of pseudonormality, and we discuss its connections to classical constraint qualifications. We also prove dual versions of these conditions involving the dual optimal value.

2. ENHANCED FRITZ JOHN CONDITIONS

The existence of FJ-multipliers is often used as the starting point for the analysis of the existence of geometric multipliers. Unfortunately, these conditions in their classical form are not sufficient to deduce the existence of geometric multipliers under some of the standard constraint qualifications, such as when $X = \Re^n$ and the constraint functions g_j are affine. Recently, the classical Fritz John conditions have been enhanced through the addition of an extra necessary condition, and their effectiveness has been significantly improved (see Hestenes [Hes75] for the case $X = \Re^n$, Bertsekas [Ber99], Prop. 3.3.11, for the case where X is a closed convex set, and Bertsekas and Ozdaglar [BeO02] for the case where X is a closed set). All of these results assume that an optimal solution exists, and that the cost and the constraint functions are smooth (but possibly nonconvex). In this section, we retain the assumption of existence of an optimal solution, and instead of smoothness we assume the following.

Assumption 2.1: (Closedness) The functions f and $g_1, \ldots, g_r$ are closed.

We note that f and $g_1, \ldots, g_r$ are closed if and only if they are lower semicontinuous on X, i.e., for each $\bar{x} \in X$, we have

$$f(\bar{x}) \le \liminf_{x \in X,\ x \to \bar{x}} f(x), \qquad g_j(\bar{x}) \le \liminf_{x \in X,\ x \to \bar{x}} g_j(x), \quad j = 1, \ldots, r,$$

(see e.g., [BNO03], Prop. 1.2.2). Under the preceding assumption, we prove the following version of the enhanced Fritz John conditions. Because we assume that f and $g_1, \ldots, g_r$ are convex over X rather than over $\Re^n$, the lines of proof from the preceding references (based on the use of gradients or subgradients) break down. We use a different line of proof, which is based instead on minimax arguments. The proof also uses the following lemma.

Lemma 2.1: Consider the convex problem (P) and assume that $-\infty < q^*$. If μ^* is a dual optimal solution, then

$$\frac{q^* - f(x)}{\|g^+(x)\|} \le \|\mu^*\|, \qquad \text{for all } x \in X \text{ that are infeasible.}$$

Proof: For any $x \in X$ that is infeasible, we have from the definition of the dual function that

$$q^* = q(\mu^*) \le f(x) + {\mu^*}' g(x) \le f(x) + {\mu^*}' g^+(x) \le f(x) + \|\mu^*\| \|g^+(x)\|.$$

Q.E.D.

Note that the preceding lemma shows that the minimum distance to the set of dual optimal solutions is an upper bound for the cost improvement/constraint violation ratio $\big(q^* - f(x)\big)/\|g^+(x)\|$. The next proposition shows that, under certain assumptions including the absence of a duality gap, this upper bound is sharp, and is asymptotically attained by an appropriate sequence $\{x^k\} \subset X$. The same fact will also be shown in Section 3, but under considerably more general assumptions (see Prop. 3.3).

Proposition 2.1: Consider the convex problem (P) under Assumption 2.1 (Closedness), and assume that x^* is an optimal solution. Then there exists a FJ-multiplier (μ_0^*, μ^*) satisfying the following condition (CV). Moreover, if $\mu_0^* \neq 0$, then μ^*/μ_0^* must be the geometric multiplier of minimum norm.

(CV) If $\mu^* \neq 0$, then there exists a sequence $\{x^k\} \subset X$ of infeasible points that converges to x^* and satisfies

$$f(x^k) \to f^*, \qquad g^+(x^k) \to 0, \tag{2.1}$$

$$\frac{f^* - f(x^k)}{\|g^+(x^k)\|} \to \begin{cases} \|\mu^*\|/\mu_0^* & \text{if } \mu_0^* \neq 0, \\ \infty & \text{if } \mu_0^* = 0, \end{cases} \tag{2.2}$$

$$\frac{g^+(x^k)}{\|g^+(x^k)\|} \to \frac{\mu^*}{\|\mu^*\|}. \tag{2.3}$$

Proof: For positive integers k and m, we consider the saddle function

$$L_{k,m}(x, \xi) = f(x) + \frac{1}{k^3}\|x - x^*\|^2 + \xi' g(x) - \frac{1}{2m}\|\xi\|^2.$$

We note that, for fixed $\xi \geq 0$, $L_{k,m}(x, \xi)$, viewed as a function from X to $\Re$, is closed and convex, because of the Closedness Assumption. Furthermore, for a fixed x, $L_{k,m}(x, \xi)$ is negative definite quadratic in ξ. For each k, we consider the set

$$X^k = X \cap \big\{x \mid \|x - x^*\| \leq k\big\}.$$

Since f and g_j are closed and convex when restricted to X, they are closed, convex, and coercive when restricted to X^k. Thus, we can use the Saddle Point Theorem (e.g., [BNO03, Prop. 2.6.9]) to assert that $L_{k,m}$ has a saddle point over $x \in X^k$ and $\xi \geq 0$. This saddle point is denoted by $(x^{k,m}, \xi^{k,m})$.

The infimum of $L_{k,m}(x, \xi^{k,m})$ over $x \in X^k$ is attained at $x^{k,m}$, implying that

$$\begin{aligned} &f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m} - x^*\|^2 + \xi^{k,m\prime} g(x^{k,m}) \\ &\quad = \inf_{x \in X^k} \left\{ f(x) + \frac{1}{k^3}\|x - x^*\|^2 + \xi^{k,m\prime} g(x) \right\} \\ &\quad \leq \inf_{x \in X^k,\, g(x) \leq 0} \left\{ f(x) + \frac{1}{k^3}\|x - x^*\|^2 + \xi^{k,m\prime} g(x) \right\} \\ &\quad \leq \inf_{x \in X^k,\, g(x) \leq 0} \left\{ f(x) + \frac{1}{k^3}\|x - x^*\|^2 \right\} \\ &\quad = f(x^*). \end{aligned} \tag{2.4}$$

Hence, we have

$$\begin{aligned} L_{k,m}(x^{k,m},\xi^{k,m}) &= f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m}-x^*\|^2 + \xi^{k,m\prime} g(x^{k,m}) - \frac{1}{2m}\|\xi^{k,m}\|^2 \\ &\le f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m}-x^*\|^2 + \xi^{k,m\prime} g(x^{k,m}) \\ &\le f(x^*). \end{aligned} \tag{2.5}$$

Since $L_{k,m}(x^{k,m},\xi)$ is quadratic in ξ, the supremum of $L_{k,m}(x^{k,m},\xi)$ over $\xi \ge 0$ is attained at

$$\xi^{k,m} = mg^+(x^{k,m}). \tag{2.6}$$

This implies that

$$\begin{aligned} L_{k,m}(x^{k,m},\xi^{k,m}) &= f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m}-x^*\|^2 + \frac{m}{2}\|g^+(x^{k,m})\|^2 \\ &\ge f(x^{k,m}) + \frac{1}{k^3}\|x^{k,m}-x^*\|^2 \\ &\ge f(x^{k,m}). \end{aligned} \tag{2.7}$$

From Eqs. (2.5) and (2.7), we see that the sequence $\{x^{k,m}\}$, with k fixed, belongs to the set $\{x \in X^k \mid f(x) \le f(x^*)\}$, which is compact. Hence, $\{x^{k,m}\}$ has a cluster point (as $m \to \infty$), denoted by $\overline{x}^k$, which belongs to $\{x \in X^k \mid f(x) \le f(x^*)\}$. By passing to a subsequence if necessary, we can assume without loss of generality that $\{x^{k,m}\}$ converges to $\overline{x}^k$ as $m \to \infty$. For each k, the sequence $\{f(x^{k,m})\}$ is bounded from below by $\inf_{x\in X^k} f(x)$, which is finite by Weierstrass' Theorem since f is closed and coercive when restricted to X^k. Also, for each k, $L_{k,m}(x^{k,m},\xi^{k,m})$ is bounded from above by $f(x^*)$ [cf. Eq. (2.5)], so the equality in Eq. (2.7) implies that

$$\limsup_{m\to\infty} g_j(x^{k,m}) \le 0, \qquad \forall\ j = 1,\ldots,r.$$

Therefore, by using the lower semicontinuity of g_j, we obtain $g(\overline{x}^k) \le 0$, implying that $\overline{x}^k$ is a feasible solution of problem (P), so that $f(\overline{x}^k) \ge f(x^*)$. Using Eqs. (2.5) and (2.7) together with the lower semicontinuity of f, we also have

$$f(\overline{x}^k) \le \liminf_{m\to\infty} f(x^{k,m}) \le \limsup_{m\to\infty} f(x^{k,m}) \le f(x^*),$$

thereby showing that for each k,

$$\lim_{m\to\infty} f(x^{k,m}) = f(x^*).$$

Together with Eqs. (2.5) and (2.7), this also implies that for each k,

$$\lim_{m\to\infty} x^{k,m} = x^*.$$

Combining the preceding relations with Eqs. (2.5) and (2.7), for each k, we obtain

$$\lim_{m\to\infty}\big(f(x^{k,m}) - f(x^*) + \xi^{k,m\prime} g(x^{k,m})\big) = 0. \tag{2.8}$$

Denote

$$\delta^{k,m} = \sqrt{1 + \|\xi^{k,m}\|^2}, \qquad \mu_0^{k,m} = \frac{1}{\delta^{k,m}}, \qquad \mu^{k,m} = \frac{\xi^{k,m}}{\delta^{k,m}}. \tag{2.9}$$

Since $\delta^{k,m}$ is bounded from below by 1, by dividing Eq. (2.8) by $\delta^{k,m}$, we obtain

$$\lim_{m\to\infty}\Big(\mu_0^{k,m} f(x^{k,m}) - \mu_0^{k,m} f(x^*) + \mu^{k,m\prime} g(x^{k,m})\Big) = 0.$$

By the preceding relations, for each k we can find a sufficiently large integer m_k such that

$$\left|\mu_0^{k,m_k} f(x^{k,m_k}) - \mu_0^{k,m_k} f(x^*) + \mu^{k,m_k\prime} g(x^{k,m_k})\right| \le \frac{1}{k}, \tag{2.10}$$

and

$$\|x^{k,m_k} - x^*\| \le \frac{1}{k}, \quad |f(x^{k,m_k}) - f(x^*)| \le \frac{1}{k}, \quad \|g^+(x^{k,m_k})\| \le \frac{1}{k}. \tag{2.11}$$

Dividing both sides of the first relation in Eq. (2.4) by δ^{k,m_k}, we obtain

$$\begin{aligned}\mu_0^{k,m_k} f(x^{k,m_k}) &+ \frac{1}{k^3\delta^{k,m_k}}\|x^{k,m_k} - x^*\|^2 + \mu^{k,m_k\prime} g(x^{k,m_k})\\ &\le \mu_0^{k,m_k} f(x) + \mu^{k,m_k\prime} g(x) + \frac{1}{k\delta^{k,m_k}}, \qquad \forall\, x \in X^k,\end{aligned}$$

where we also use the fact that $\|x - x^*\| \le k$ for all $x \in X^k$ (see the definition of X^k). Since the sequence $\big\{(\mu_0^{k,m_k}, \mu^{k,m_k})\big\}$ is bounded, it has a cluster point, denoted by (μ_0^*, μ^*), which satisfies conditions (ii), (iii) in the definition of a FJ-multiplier. For any $x \in X$, we have $x \in X^k$ for all k sufficiently large. Without loss of generality, we will assume that the entire sequence $\big\{(\mu_0^{k,m_k}, \mu^{k,m_k})\big\}$ converges to (μ_0^*, μ^*). Taking the limit as $k \to \infty$, and using Eq. (2.10), we obtain

$$\mu_0^* f(x^*) \le \mu_0^* f(x) + \mu^{*\prime} g(x), \qquad \forall\, x \in X.$$

Since $\mu^* \ge 0$, this implies that

$$\begin{aligned}\mu_0^* f(x^*) &\le \inf_{x\in X}\big\{\mu_0^* f(x) + \mu^{*\prime} g(x)\big\}\\ &\le \inf_{x\in X,\, g(x)\le 0}\big\{\mu_0^* f(x) + \mu^{*\prime} g(x)\big\}\\ &\le \inf_{x\in X,\, g(x)\le 0}\mu_0^* f(x)\\ &= \mu_0^* f(x^*).\end{aligned}$$

Thus we have

$$\mu_0^* f(x^*) = \inf_{x\in X}\big\{\mu_0^* f(x) + \mu^{*\prime} g(x)\big\},$$

so that (μ_0^*, μ^*) also satisfies condition (i) in the definition of a FJ-multiplier.

If $\mu^* = 0$, then $\mu_0^* \neq 0$, (CV) is automatically satisfied, and $\mu^*/\mu_0^* = 0$ has minimum norm. Moreover, condition (i) yields $f^* = \inf_{x \in X} f(x)$, so that (CV) [in particular, Eq. (2.2)] is satisfied by only $\mu^* = 0$.

Assume now that $\mu^* \neq 0$, so that the index set $J = \{j \neq 0 \mid \mu_j^* > 0\}$ is nonempty. Then, for sufficiently large k, we have $\xi_j^{k,m_k} > 0$ and hence $g_j(x^{k,m_k}) > 0$ for all $j \in J$. Thus, for each k, we can choose the index m_k to further satisfy $x^{k,m_k} \neq x^*$, in addition to Eqs. (2.10), (2.11). Using Eqs. (2.6), (2.9) and the fact that $\mu^{k,m_k} \to \mu^*$, we obtain

$$\frac{g^+(x^{k,m_k})}{\|g^+(x^{k,m_k})\|} = \frac{\mu^{k,m_k}}{\|\mu^{k,m_k}\|} \to \frac{\mu^*}{\|\mu^*\|}.$$

Using also Eq. (2.5) and $f(x^*) = f^*$, we have that

$$\frac{f^* - f(x^{k,m_k})}{\|g^+(x^{k,m_k})\|} \geq \frac{\xi^{k,m_k\prime} g(x^{k,m_k})}{\|g^+(x^{k,m_k})\|} = \|\xi^{k,m_k}\| = \frac{\|\mu^{k,m_k}\|}{\mu_0^{k,m_k}}. \tag{2.12}$$

If $\mu_0^* = 0$, then $\mu_0^{k,m_k} \to 0$, so Eq. (2.12) together with $\|\mu^{k,m_k}\| \to \|\mu^*\| > 0$ yields

$$\frac{f^* - f(x^{k,m_k})}{\|g^+(x^{k,m_k})\|} \to \infty.$$

If $\mu_0^* \neq 0$, then Eq. (2.12) together with $\mu_0^{k,m_k} \to \mu_0^*$ and $\|\mu^{k,m_k}\| \to \|\mu^*\|$ yields

$$\liminf_{k \to \infty} \frac{f^* - f(x^{k,m_k})}{\|g^+(x^{k,m_k})\|} \geq \frac{\|\mu^*\|}{\mu_0^*}.$$

Since μ^*/μ_0^* is a geometric multiplier and $f^* = q^*$, Lemma 2.1 implies that in fact μ^*/μ_0^* is of minimum norm and the inequality holds with equality. From Eq. (2.11), we have $f(x^{k,m_k}) \to f(x^*)$, $g^+(x^{k,m_k}) \to 0$, and $x^{k,m_k} \to x^*$. Hence, the sequence $\{x^{k,m_k}\}$ also satisfies conditions (2.1)-(2.3) of the proposition, concluding the proof. **Q.E.D.**

Note that Eq. (2.3) implies that, for all k sufficiently large,

$$g_j(x^k) > 0, \quad \forall\, j \in J, \qquad g_j^+(x^k) = o\left(\min_{j \in J} g_j^+(x^k)\right), \quad \forall\, j \notin J,$$

where $J = \{j \neq 0 \mid \mu_j^* > 0\}$. Thus, the (CV) condition (Complementarity Violation) in Prop. 2.1 refines that used in [BNO03, Sec. 5.7] by also estimating the rate of cost improvement. As an illustration of Prop. 2.1, consider the 2-dimensional example of Duffin:

$$\begin{aligned} &\text{minimize} \quad x_2 \\ &\text{subject to} \quad x = (x_1, x_2)' \in \Re^2, \quad \|x\| - x_1 \leq 0. \end{aligned}$$

Here $f^* = 0$ and $x^* = (x_1^*, 0)$ is an optimal solution for any $x_1^* \geq 0$. Also, $q(\mu) = -\infty$ for all $\mu \geq 0$, so $q^* = -\infty$ and there is duality gap. It can be seen that $\mu_0^* = 0$, $\mu^* = 1$ form a FJ-multiplier and, together with $x^k = (x_1^*, -1/k)'$, satisfy condition (CV).

The proof of Prop. 2.1 can be explained in terms of the construction shown in Fig. 2.1. Consider the function $L_{k,m}$ introduced in the proof,

$$L_{k,m}(x, \xi) = f(x) + \frac{1}{k^3}\|x - x^*\|^2 + \xi' g(x) - \frac{1}{2m}\|\xi\|^2.$$

Note that the term $(1/k^3)\|x - x^*\|^2$ ensures that x^* is a strict local minimum of the function $f(x) + (1/k^3)\|x - x^*\|^2$. To simplify the following discussion, let us assume that f is strictly convex, so that this term can be omitted from the definition of $L_{k,m}$. This assumption is satisfied by the above example if its cost function is changed to e^x, for which $f^* = 1$ and $q^* = 0$.

For any nonnegative vector $u \in \Re^r$, let $p^k(u)$ denote the optimal value of the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g(x) \leq u, \\ &\qquad\qquad\quad x \in X^k = X \cap \left\{x \;\middle|\; \|x - x^*\| \leq k\right\}. \end{aligned} \tag{2.13}$$

For each k and m, the saddle point of the function $L_{k,m}(x, \xi)$, denoted by $(x^{k,m}, \xi^{k,m})$, can be characterized in terms of $p^k(u)$ as follows.

The maximization of $L_{k,m}(x, \xi)$ over $\xi \geq 0$ for any fixed $x \in X^k$ yields

$$\xi = m g^+(x), \tag{2.14}$$

so that we have

$$\begin{aligned} L_{k,m}(x^{k,m}, \xi^{k,m}) &= \inf_{x \in X^k} \sup_{\xi \geq 0} \left\{ f(x) + \xi' g(x) - \frac{1}{2m}\|\xi\|^2 \right\} \\ &= \inf_{x \in X^k} \left\{ f(x) + \frac{m}{2}\|g^+(x)\|^2 \right\}. \end{aligned}$$

This minimization can also be written as

$$\begin{aligned} L_{k,m}(x^{k,m}, \xi^{k,m}) &= \inf_{x \in X^k} \inf_{u \in \Re^r,\ g(x) \leq u} \left\{ f(x) + \frac{m}{2}\|u^+\|^2 \right\} \\ &= \inf_{u \in \Re^r} \inf_{x \in X^k,\ g(x) \leq u} \left\{ f(x) + \frac{m}{2}\|u^+\|^2 \right\} \\ &= \inf_{u \in \Re^r} \left\{ p^k(u) + \frac{m}{2}\|u^+\|^2 \right\}. \end{aligned} \tag{2.15}$$

The vector $u^{k,m} = g(x^{k,m})$ attains the infimum in the preceding relation. This minimization can be visualized geometrically as in Fig. 2.1. The point of contact of the graphs of the functions $p^k(u)$

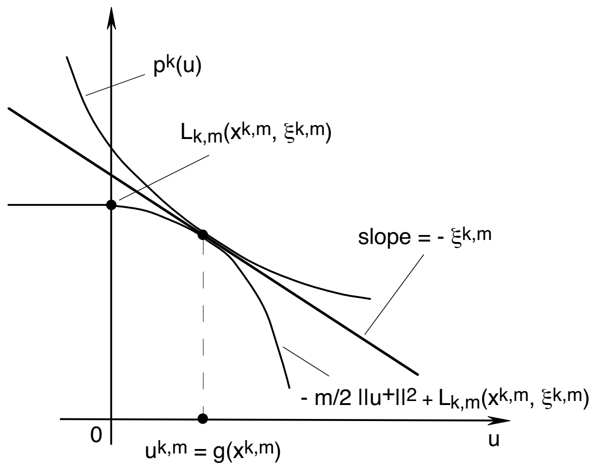

Figure 2.1. Illustration of the saddle point of the function $L_{k,m}(x,\xi)$ over $x \in X^k$ and $\xi \geq 0$ in terms of the function $p^k(u)$, which is the optimal value of problem (2.13) as a function of u.

and $L_{k,m}(x^{k,m},\xi^{k,m}) - (m/2)\|u^+\|^2$ corresponds to the vector $u^{k,m}$ that attains the infimum in Eq. (2.15). A similar compactification-regularization technique is used in [Roc93, Sec. 9].

We can also interpret $\xi^{k,m}$ in terms of the function p^k. In particular, the infimum of $L_{k,m}(x,\xi^{k,m})$ over $x \in X^k$ is attained at $x^{k,m}$, implying that

$$\begin{aligned} f(x^{k,m}) + \xi^{k,m\prime} g(x^{k,m}) &= \inf_{x \in X^k} \bigl\{ f(x) + \xi^{k,m\prime} g(x) \bigr\} \\ &= \inf_{u \in \Re^r} \bigl\{ p^k(u) + \xi^{k,m\prime} u \bigr\}. \end{aligned}$$

Replacing $g(x^{k,m})$ by $u^{k,m}$ in the preceding relation, and using the fact that $x^{k,m}$ is feasible for problem (2.13) with $u = u^{k,m}$, we obtain

$$p^k(u^{k,m}) \leq f(x^{k,m}) = \inf_{u \in \Re^r} \bigl\{ p^k(u) + \xi^{k,m\prime}(u - u^{k,m}) \bigr\}.$$

Thus, we see that

$$p^k(u^{k,m}) \leq p^k(u) + \xi^{k,m\prime}(u - u^{k,m}), \qquad \forall\ u \in \Re^r,$$

which, by the definition of the subgradient of a convex function, implies that

$$-\xi^{k,m} \in \partial p^k(u^{k,m})$$

(cf. Fig. 2.1). It can be seen from this interpretation that the limit of $L_{k,m}(x^{k,m},\xi^{k,m})$ as $m \to \infty$ is equal to $p^k(0)$, which is equal to $f(x^*)$ for each k. The limit of the normalized sequence

$$\left\{ \frac{(1, \xi^{k,m_k})}{\sqrt{1 + \|\xi^{k,m_k}\|^2}} \right\}$$

as $k \to \infty$ yields the FJ-multiplier (μ_0^*, μ^*), and the sequence $\{x^{k,m_k}\}$ is used to construct the sequence that satisfies condition (CV) of the proposition.

3. MINIMUM-NORM DUAL OPTIMAL SOLUTIONS

In the preceding section we focussed on the case where a primal optimal solution exists and we showed that the geometric multiplier of minimum norm is informative. Notice that a geometric multiplier is automatically a dual optimal solution. When there is duality gap, there exists no geometric multiplier, even if there is a dual optimal solution. In this section we focus on the case where a dual optimal solution exists and we will see that, analogously, the dual optimal solution of minimum norm is informative. In particular, it satisfies a condition analogous to condition (CV), with primal optimal value f^* replaced by q^*. Consistent with our analysis in Section 2, we call such a dual optimal solution *informative* [BNO03, Section 6.6.2], since it indicates the constraints to relax and the rate of relaxation in order to obtain a primal cost reduction by an amount that is strictly greater than the size of the duality gap $f^* - q^*$.

We begin with the following proposition, which is a classical result and requires no additional assumptions on (P). It will be used to prove Lemma 3.1. We provide its proof for completeness.

Proposition 3.1: (Fritz John Conditions) Consider the convex problem (P), and assume that $f^* < \infty$. Then there exists a FJ-multiplier (μ_0^*, μ^*).

Proof: If $f^* = -\infty$, then $\mu_0^* = 1$ and $\mu^* = 0$ form a FJ-multiplier. We may thus assume that f^* is finite. Consider the subset of $\Re^{r+1}$ given by

$$M = \big\{(u_1, \ldots, u_r, w) \mid \text{there exists } x \in X \text{ such that}$$
$$g_j(x) \le u_j,\ j = 1, \ldots, r,\ f(x) \le w\big\}$$

(cf. Fig. 3.1). We first show that M is convex. To this end, we consider vectors $(u, w) \in M$ and $(\tilde{u}, \tilde{w}) \in M$, and we show that their convex combinations lie in M. The definition of M implies that for some $x \in X$ and $\tilde{x} \in X$, we have

$$f(x) \le w, \qquad g_j(x) \le u_j, \qquad j = 1, \ldots, r,$$

$$f(\tilde{x}) \le \tilde{w}, \qquad g_j(\tilde{x}) \le \tilde{u}_j, \qquad j = 1, \ldots, r.$$

For any $\alpha \in [0, 1]$, we multiply these relations with α and $1 - \alpha$, respectively, and add them. By using the convexity of f and g_j, we obtain

$$f\big(\alpha x + (1-\alpha)\tilde{x}\big) \le \alpha f(x) + (1-\alpha) f(\tilde{x}) \le \alpha w + (1-\alpha)\tilde{w},$$

$$g_j\big(\alpha x + (1-\alpha)\tilde{x}\big) \le \alpha g_j(x) + (1-\alpha) g_j(\tilde{x}) \le \alpha u_j + (1-\alpha)\tilde{u}_j, \quad j = 1, \ldots, r.$$

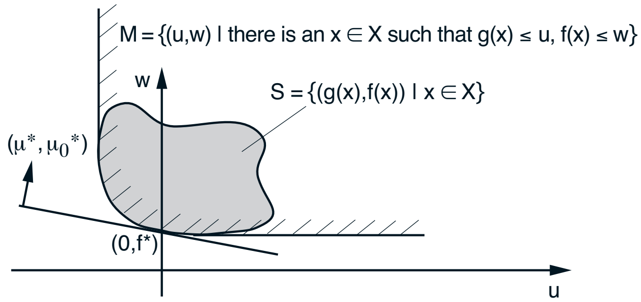

Figure 3.1. Illustration of the set

$$S = \Big\{ \big(g(x), f(x)\big) \mid x \in X \Big\}$$

and the set

$$M = \Big\{ (u, w) \mid \text{ there exists } x \in X \text{ such that } \; g(x) \le u, \; f(x) \le w \Big\}$$

used in the proof of Prop. 3.1. The idea of the proof is to show that M is convex and that $(0, f^*)$ is not an interior point of M. A hyperplane passing through $(0, f^*)$ and supporting M is used to define the FJ-multipliers.

In view of the convexity of X, we have $\alpha x + (1 - \alpha)\tilde{x} \in X$, so these inequalities imply that the convex combination of (u, w) and $(\tilde{u}, \tilde{w})$, i.e., $\big(\alpha u + (1-\alpha)\tilde{u}, \alpha w + (1-\alpha)\tilde{w}\big)$, belongs to M. This proves the convexity of M.

We next observe that $(0, f^*)$ is not an interior point of M; otherwise, for some $\epsilon > 0$, the point $(0, f^* - \epsilon)$ would belong to M, contradicting the definition of f^* as the optimal value. Therefore, there exists a hyperplane passing through $(0, f^*)$ and containing M in one of its closed halfspaces, i.e., there exists a vector $(\mu^*, \mu_0^*) \neq (0, 0)$ such that

$$\mu_0^* f^* \le \mu_0^* w + \mu^{*\prime} u, \qquad \forall \; (u, w) \in M. \tag{3.1}$$

This relation implies that

$$\mu_0^* \ge 0, \qquad \mu_j^* \ge 0, \qquad \forall \; j = 1, \ldots, r,$$

since for each $(u, w) \in M$, we have that $(u, w + \gamma) \in M$ and $(u_1, \ldots, u_j + \gamma, \ldots, u_r, w) \in M$ for all $\gamma > 0$ and j.

Finally, since for all $x \in X$, we have $\big(g(x), f(x)\big) \in M$, Eq. (3.1) implies that

$$\mu_0^* f^* \le \mu_0^* f(x) + \mu^{*\prime} g(x), \qquad \forall \; x \in X.$$

Taking the infimum over all $x \in X$, it follows that

$$\begin{aligned}\mu_0^* f^* &\le \inf_{x \in X} \{\mu_0^* f(x) + \mu^{*\prime} g(x)\} \\ &\le \inf_{x \in X,\ g(x) \le 0} \{\mu_0^* f(x) + \mu^{*\prime} g(x)\} \\ &\le \inf_{x \in X,\ g(x) \le 0} \mu_0^* f(x) \\ &= \mu_0^* f^*.\end{aligned}$$

Hence, equality holds throughout above, proving the desired result. **Q.E.D.**

If the scalar μ_0^* in the preceding proposition can be proved to be positive, then μ^*/μ_0^* is a geometric multiplier for problem (P). This can be used to show the existence of a geometric multiplier in the case where the Slater condition holds, i.e., there exists a vector $\overline{x} \in X$ such that $g(\overline{x}) < 0$. Indeed, in this case the scalar μ_0^* cannot be 0, since if it were, then according to the proposition, we would have

$$0 = \inf_{x \in X} \mu^{*\prime} g(x)$$

for some vector $\mu^* \ge 0$ with $\mu^* \ne 0$, while for this vector, we would also have $\mu^{*\prime} g(\overline{x}) < 0$, a contradiction.

Using Prop. 3.1, we have the following lemma which will be used to prove the next proposition, as well as Prop. 4.3 in the next section.

Lemma 3.1: Consider the convex problem (P), and assume that $f^* < \infty$. For each $\delta > 0$, let

$$f^\delta = \inf_{\substack{x \in X \\ g_j(x) \le \delta,\, j=1,\ldots,r}} f(x). \tag{3.2}$$

Then the dual optimal value q^* satisfies $f^\delta \le q^*$ for all $\delta > 0$ and

$$q^* = \lim_{\delta \downarrow 0} f^\delta.$$

Proof: We first note that either $\lim_{\delta \downarrow 0} f^\delta$ exists and is finite, or else $\lim_{\delta \downarrow 0} f^\delta = -\infty$, since f^δ is monotonically nondecreasing as $\delta \downarrow 0$, and $f^\delta \le f^*$ for all $\delta > 0$. Since $f^* < \infty$, there exists some $\overline{x} \in X$ such that $g(\overline{x}) \le 0$. Thus, for each $\delta > 0$ such that $f^\delta > -\infty$, the Slater condition is satisfied for problem (3.2), and by Prop. 3.1 and the subsequent discussion, there exists a $\mu^\delta \ge 0$ satisfying

$$f^\delta = \inf_{x \in X} \left\{ f(x) + \mu^{\delta\prime} g(x) - \delta \sum_{j=1}^{r} \mu_j^\delta \right\}$$

$$\begin{aligned} &\le \inf_{x \in X} \{ f(x) + \mu^{\delta\prime} g(x) \} \\ &= q(\mu^\delta) \\ &\le q^*. \end{aligned}$$

For each $\delta > 0$ such that $f^\delta = -\infty$, we also have $f^\delta \le q^*$, so that

$$f^\delta \le q^*, \qquad \forall\ \delta > 0.$$

By taking the limit as $\delta \downarrow 0$, we obtain

$$\lim_{\delta \downarrow 0} f^\delta \le q^*.$$

To show the reverse inequality, we consider two cases: (1) $f^\delta > -\infty$ for all $\delta > 0$ that are sufficiently small, and (2) $f^\delta = -\infty$ for all $\delta > 0$. In case (1), for each $\delta > 0$ with $f^\delta > -\infty$, choose $x^\delta \in X$ such that $g_j(x^\delta) \le \delta$ for all j and $f(x^\delta) \le f^\delta + \delta$. Then, for any $\mu \ge 0$,

$$q(\mu) = \inf_{x \in X} \{ f(x) + \mu' g(x) \} \le f(x^\delta) + \mu' g(x^\delta) \le f^\delta + \delta + \delta \sum_{j=1}^r \mu_j.$$

Taking the limit as $\delta \downarrow 0$, we obtain

$$q(\mu) \le \lim_{\delta \downarrow 0} f^\delta,$$

so that $q^* \le \lim_{\delta \downarrow 0} f^\delta$. In case (2), choose $x^\delta \in X$ such that $g_j(x^\delta) \le \delta$ for all j and $f(x^\delta) \le -1/\delta$. Then, similarly, for any $\mu \ge 0$, we have

$$q(\mu) \le f(x^\delta) + \mu' g(x^\delta) \le -\frac{1}{\delta} + \delta \sum_{j=1}^r \mu_j,$$

so by taking $\delta \downarrow 0$, we obtain $q(\mu) = -\infty$ for all $\mu \ge 0$, and hence also $q^* = -\infty = \lim_{\downarrow 0} f^\delta$. **Q.E.D.**

Using Lemmas 2.1 and 3.1, we prove below the main result of this section, which shows under very general assumptions that the minimum-norm dual optimal solution is informative.

Proposition 3.2: (Existence of Informative Dual Optimal Solution) Consider the convex problem (P) under Assumption 2.1 (Closedness), and assume that $f^* < \infty$ and $-\infty < q^*$. If there exists a dual optimal solution, then the dual optimal solution μ^* of minimum norm satisfies the following condition (dCV). Moreover, it is the only dual optimal solution that satisfies this condition.

(dCV) If $\mu^* \neq 0$, then there exists a sequence $\{x^k\} \subset X$ of infeasible points that satisfies

$$f(x^k) \to q^*, \qquad g^+(x^k) \to 0, \tag{3.3}$$

$$\frac{q^* - f(x^k)}{\|g^+(x^k)\|} \to \|\mu^*\|, \tag{3.4}$$

$$\frac{g^+(x^k)}{\|g^+(x^k)\|} \to \frac{\mu^*}{\|\mu^*\|}. \tag{3.5}$$

Proof: Let μ^* be the dual optimal solution of minimum norm. Assume that $\mu^* \neq 0$. For $k = 1, 2, \ldots$, consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \qquad g_j(x) \leq \dfrac{1}{k^4},\ k = 1, \ldots, r. \end{array}$$

By Lemma 3.1, for each k, the optimal value of this problem is less than or equal to q^*. Since q^* is finite (in view of the assumptions $-\infty < q^*$ and $f^* < \infty$, and the weak duality relation $q^* \leq f^*$), we may select for each k, a vector $\tilde{x}^k \in X$ that satisfies

$$f(\tilde{x}^k) \leq q^* + \frac{1}{k^2}, \qquad g_j(\tilde{x}^k) \leq \frac{1}{k^4},\ j = 1, \ldots, r.$$

Consider also the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_j(x) \leq \dfrac{1}{k^4},\ j = 1, \ldots, r, \\ & x \in \tilde{X}^k = X \cap \left\{ x \;\middle|\; \|x\| \leq k \left(\max\limits_{1 \leq i \leq k} \|\tilde{x}^i\| + 1 \right) \right\}. \end{array}$$

By the Closedness Assumption, f and g_j are closed and convex when restricted to X, so they are closed, convex, and coercive when restricted to $\tilde{X}^k$. Thus, the problem has an optimal solution, which we denote by $\overline{x}^k$. Note that since $\tilde{x}^k$ belongs to the feasible solution set of this problem, we have

$$f(\overline{x}^k) \leq f(\tilde{x}^k) \leq q^* + \frac{1}{k^2}. \tag{3.6}$$

For each k, we consider the saddle function

$$L_k(x,\mu) = f(x) + \mu' g(x) - \frac{\|\mu\|^2}{2k},$$

and the set

$$X^k = \tilde{X}^k \cap \big\{x \mid g_j(x) \le k,\ j = 1,\ldots,r\big\}.$$

We note that $L_k(x,\mu)$, for fixed $\mu \ge 0$, is closed, convex, and coercive in x, when restricted to X^k, and negative definite quadratic in μ for fixed x. Hence, using the Saddle Point Theorem (e.g., [BNO03, Prop. 2.6.9]), we can assert that L_k has a saddle point over $x \in X^k$ and $\mu \ge 0$, denoted by (x^k,μ^k).

Since L_k is quadratic in μ, the supremum of $L_k(x^k,\mu)$ over $\mu \ge 0$ is attained at

$$\mu^k = kg^+(x^k). \tag{3.7}$$

Similarly, the infimum in $\inf_{x\in X^k} L_k(x,\mu^k)$ is attained at x^k, implying that

$$\begin{aligned} f(x^k) + \mu^{k\prime} g(x^k) &= \inf_{x\in X^k} \big\{f(x) + \mu^{k\prime} g(x)\big\} \\ &= \inf_{x\in X^k} \big\{f(x) + kg^+(x^k)' g(x)\big\} \\ &\le \inf_{x\in X^k,\ g_j(x)\le \frac{1}{k^4},\, j=1,\ldots,r} \left\{f(x) + k\sum_{j=1}^r g_j^+(x^k)' g_j(x)\right\} \\ &\le \inf_{x\in X^k,\ g_j(x)\le \frac{1}{k^4},\, j=1,\ldots,r} \left\{f(x) + \frac{r}{k^2}\right\} \\ &= f(\overline{x}^k) + \frac{r}{k^2} \\ &\le q^* + \frac{r+1}{k^2}, \end{aligned} \tag{3.8}$$

where the second inequality holds in view of the fact $x^k \in X^k$, implying that $g_j^+(x^k) \le k$, $j = 1,...,r$, and the third inequality follows from Eq. (3.6).

We also have

$$\begin{aligned} L_k(x^k,\mu^k) &= \sup_{\mu\ge 0} \inf_{x\in X^k} L_k(x,\mu) \\ &\ge \sup_{\mu\ge 0} \inf_{x\in X} L_k(x,\mu) \\ &= \sup_{\mu\ge 0} \left\{\inf_{x\in X} \big\{f(x) + \mu' g(x)\big\} - \frac{\|\mu\|^2}{2k}\right\} \\ &= \sup_{\mu\ge 0} \left\{q(\mu) - \frac{\|\mu\|^2}{2k}\right\} \\ &\ge q(\mu^*) - \frac{\|\mu^*\|^2}{2k} \\ &= q^* - \frac{\|\mu^*\|^2}{2k}, \end{aligned} \tag{3.9}$$

where we recall that μ^* is the dual optimal solution with the minimum norm.

Combining Eqs. (3.9) and (3.8), we obtain

$$\begin{aligned} q^* - \frac{1}{2k}\|\mu^*\|^2 &\le L_k(x^k, \mu^k) \\ &= f(x^k) + \mu^{k\prime} g(x^k) - \frac{1}{2k}\|\mu^k\|^2 \\ &\le q^* + \frac{r+1}{k^2} - \frac{1}{2k}\|\mu^k\|^2. \end{aligned} \tag{3.10}$$

This relation shows that $\|\mu^k\|^2 \le \|\mu^*\|^2 + 2(r+1)/k$, so the sequence $\{\mu^k\}$ is bounded. Let $\overline{\mu}$ be a cluster point of $\{\mu^k\}$. Without loss of generality, we assume that the entire sequence $\{\mu^k\}$ converges to $\overline{\mu}$. We also have from Eq. (3.10) that

$$\lim_{k\to\infty} \big\{ f(x^k) + \mu^{k\prime} g(x^k) \big\} = q^*.$$

Hence, taking the limit as $k \to \infty$ in Eq. (3.8) yields

$$q^* \le \inf_{x\in X} \big\{ f(x) + \overline{\mu}' g(x) \big\} = q(\overline{\mu}) \le q^*.$$

Hence $\overline{\mu}$ is a dual optimal solution, and since $\|\overline{\mu}\| \le \|\mu^*\|$ [which follows by taking the limit in Eq. (3.10)], by using the minimum norm property of μ^*, we conclude that any cluster point $\overline{\mu}$ of μ^k must be equal to μ^*. Thus $\mu^k \to \mu^*$, and using Eq. (3.10), we obtain

$$\lim_{k\to\infty} k\big(L_k(x^k, \mu^k) - q^*\big) = -\frac{1}{2}\|\mu^*\|^2. \tag{3.11}$$

Using Eq. (3.7), it follows that

$$L_k(x^k, \mu^k) = \sup_{\mu\ge 0} L_k(x^k, \mu) = f(x^k) + \frac{1}{2k}\|\mu^k\|^2,$$

which combined with Eq. (3.11) yields

$$\lim_{k\to\infty} k\big(f(x^k) - q^*\big) = -\|\mu^*\|^2,$$

implying that $f(x^k) < q^*$ for all sufficiently large k, since $\mu^* \ne 0$. Since, $\mu^k \to \mu^*$, Eq. (3.7) also implies that

$$\lim_{k\to\infty} k g^+(x^k) = \mu^*.$$

It follows that the sequence $\{x^k\}$ satisfies Eqs. (3.3), (3.4), and (3.5). Moreover, Lemma 2.1 shows that $\{x^k\}$ satisfies (3.4) only when μ^* is the dual optimal solution of minimum norm. This completes the proof. **Q.E.D.**

Our final result of this section shows that Assumption 2.1 in Prop. 3.2 can in fact be relaxed. We denote by $\overline{f}$ the closure of f, i.e., the function whose epigraph is the closure of f. Similarly, for each j, we denote by $\overline{g}_j$ the closure of g_j. A key fact we use is that *replacing f and g_j by their closures, does not affect the closure of the primal function, and hence also the dual function.* This is based on the following lemma on the closedness of functions generated by partial minimization.

Lemma 3.2: Consider a function $F : \Re^{n+r} \mapsto (-\infty, \infty]$ and the function $p : \Re^n \mapsto [-\infty, \infty]$ defined by

$$p(u) = \inf_{x \in \Re^n} F(x, u).$$

Then the following hold:

(a)

$$P\big(\mathrm{epi}(F)\big) \subset \mathrm{epi}(p) \subset \mathrm{cl}\Big(P\big(\mathrm{epi}(F)\big)\Big), \tag{3.12}$$

$$P\Big(\mathrm{cl}\big(\mathrm{epi}(F)\big)\Big) \subset \mathrm{cl}\big(\mathrm{epi}(p)\big), \tag{3.13}$$

where $P(\cdot)$ denotes projection on the space of (u, w), i.e., $P(x, u, w) = (u, w)$.

(b) If $\overline{F}$ is the closure of F and $\overline{p}$ is defined by

$$\overline{p}(u) = \inf_{x \in \Re^n} \overline{F}(x, u),$$

then the closures of p and $\overline{p}$ coincide.

Proof: (a) The left-hand side of Eq. (3.12) follows from the definition

$$\mathrm{epi}(p) = \left\{ (u, w) \;\middle|\; \inf_{x \in \Re^n} F(x, u) \le w \right\}.$$

To show the right-hand side of Eq. (3.12), note that for any $(u, w) \in \mathrm{epi}(p)$ and every integer $k \ge 1$, there exists an x^k such that $(x^k, u, w + 1/k) \in \mathrm{epi}(F)$, so that $(u, w + 1/k) \in P\big(\mathrm{epi}(F)\big)$ and $(u, w) \in \mathrm{cl}\Big(P\big(\mathrm{epi}(F)\big)\Big)$.

To show Eq. (3.13), let $(\overline{u}, \overline{w})$ belong to $P\Big(\mathrm{cl}\big(\mathrm{epi}(F)\big)\Big)$. Then there exists $\overline{x}$ such that $(\overline{x}, \overline{u}, \overline{w}) \in \mathrm{cl}\big(\mathrm{epi}(F)\big)$, and hence there is a sequence $(x^k, u^k, w^k) \in \mathrm{epi}(F)$ such that $x^k \to \overline{x}$, $u^k \to \overline{u}$, and $w^k \to \overline{w}$. Thus we have $p(u^k) \le F(x^k, u^k) \le w^k$, implying that $(u^k, w^k) \in \mathrm{epi}(p)$ for all k. It follows that $(\overline{u}, \overline{w}) \in \mathrm{cl}\big(\mathrm{epi}(p)\big)$.

(b) By taking closure in Eq. (3.12), we see that

$$\mathrm{cl}\big(\mathrm{epi}(p)\big) = \mathrm{cl}\Big(P\big(\mathrm{epi}(F)\big)\Big), \tag{3.14}$$

and by replacing F with $\overline{F}$, we also have

$$\mathrm{cl}\big(\mathrm{epi}(\overline{p})\big) = \mathrm{cl}\Big(P\big(\mathrm{epi}(\overline{F})\big)\Big). \tag{3.15}$$

On the other hand, by taking closure in Eq. (3.13), we have

$$\mathrm{cl}\Big(P\big(\mathrm{epi}(\overline{F})\big)\Big) \subset \mathrm{cl}\Big(P\big(\mathrm{epi}(F)\big)\Big),$$

which implies that

$$\mathrm{cl}\Big(P\big(\mathrm{epi}(\overline{F})\big)\Big) = \mathrm{cl}\Big(P\big(\mathrm{epi}(F)\big)\Big). \tag{3.16}$$

By combining Eqs. (3.14)-(3.16), we see that

$$\mathrm{cl}\big(\mathrm{epi}(p)\big) = \mathrm{cl}\big(\mathrm{epi}(\overline{p})\big).$$

Q.E.D.

Using Lemmas 2.1 and 3.2, we now prove the last main result of this section.

Proposition 3.3: (Relaxing Closedness Assumption in Prop. 3.2) Consider the convex problem (P), and assume that $f^* < \infty$, $-\infty < q^*$, and $\mathrm{dom}(\overline{f}) = \mathrm{dom}(\overline{g}_j)$, $j = 1, ..., r$. If μ^* is the dual optimal solution of minimum norm, then it satisfies condition (dCV) of Prop. 3.2. Moreover, it is the only dual optimal solution that satisfies this condition.

Proof: We apply Lemma 3.2 to the primal function $p(u)$, which is defined by partial minimization over $x \in \Re^n$ of the extended real-valued function

$$F(x,u) = \begin{cases} f(x) & \text{if } x \in X,\ g(x) \le u, \\ \infty & \text{otherwise.} \end{cases}$$

Note that the closure of F is

$$\overline{F}(x,u) = \begin{cases} \overline{f}(x) & \text{if } x \in \overline{X},\ \overline{g}(x) \le u, \\ \infty & \text{otherwise,} \end{cases}$$

where $\overline{g} = (\overline{g}_1, \ldots, \overline{g}_r)'$ and $\overline{X} = \mathrm{dom}(\overline{f}) = \mathrm{dom}(\overline{g}_j)$, $j = 1, ..., r$.[4] Thus, by Lemma 3.2, replacing X, f and g with $\overline{X}$, $\overline{f}$ and $\overline{g}$ does not change the closure of the primal function, and therefore does not change the dual function.

[4] Why? By definition of the closure of F, $\overline{F}(x,u) = \liminf_{(x^k,u^k)\to(x,u)} F(x^k,u^k)$. Suppose $\overline{F}(x,u) < \infty$. Then there exist $x^k \in X$ and u^k such that $(x^k,u^k) \to (x,u)$, $f(x^k) \to \overline{F}(x,u)$ and $g(x^k) \le u^k$ for all $k = 1, 2, ...$ Passing to the limit yields $\overline{f}(x) \le \overline{F}(x,u)$ and $\overline{g}(x) \le u$. Conversely, suppose $\overline{f}(x) < \infty$ and $\overline{g}(x) \le u$. Fix any $\overline{x} \in \mathrm{ri}(X)$, and let $x^\epsilon = (1-\epsilon)x + \epsilon\overline{x}$, $u^\epsilon = (1-\epsilon)u + \epsilon\overline{u}$, where $\overline{u} = g(\overline{x})$. Then $x^\epsilon \in \mathrm{ri}(X) = \mathrm{ri}(\overline{X})$ and $g(x^\epsilon) \le u^\epsilon$ for $\epsilon \in (0,1)$. Since $\overline{f}$ coincide with f on $\mathrm{ri}(\overline{X})$ and $\overline{f}$ is continuous along any line segment in $\overline{X}$, this implies

$$\lim_{\epsilon\to 0} f(x^\epsilon) = \lim_{\epsilon\to 0} \overline{f}(x^\epsilon) = \overline{f}(x).$$

Thus $\lim_{\epsilon\to 0} F(x^\epsilon,u^\epsilon) = \overline{f}(x)$, implying $\overline{F}(x,u) \le \overline{f}(x)$.

Assume $\mu^* \neq 0$. By Prop. 3.2, there exists a sequence $\{x^k\} \subset \overline{X}$ of infeasible points that satisfies

$$\frac{q^* - \overline{f}(x^k)}{\|\overline{g}^+(x^k)\|} \to \|\mu^*\|, \qquad \frac{\overline{g}^+(x^k)}{\|\overline{g}^+(x^k)\|} \to \frac{\mu^*}{\|\mu^*\|}, \qquad \|\overline{g}^+(x^k)\| \to 0.$$

We will now perturb the sequence $\{x^k\}$ so that it lies in $\mathrm{ri}(X)$, while it still satisfies the preceding relations. Indeed, fix any $\overline{x} \in \mathrm{ri}(X)$. For each k, we can choose a sufficiently small $\epsilon \in (0,1)$ such that $\overline{f}(\epsilon\overline{x} + (1-\epsilon)x^k)$ and $\|\overline{g}^+(\epsilon\overline{x} + (1-\epsilon)x^k)\|$ are arbitrarily close to $\overline{f}(x^k)$ and $\|\overline{g}^+(x^k)\|$, respectively. This is possible because $\overline{f}$, $\overline{g}_1, ..., \overline{g}_r$ are closed and hence continuous along the line segment that connects x^k and $\overline{x}$. Thus, for each k, we can choose $\epsilon_k \in (0,1)$ so that the corresponding vector $\overline{x}^k = \epsilon_k\overline{x} + (1-\epsilon_k)x^k$ satisfies

$$\left| \frac{q^* - \overline{f}(x^k)}{\|\overline{g}^+(x^k)\|} - \frac{q^* - \overline{f}(\overline{x}^k)}{\|\overline{g}^+(\overline{x}^k)\|} \right| \leq \frac{1}{k}, \qquad \left| \frac{\overline{g}^+(x^k)}{\|\overline{g}^+(x^k)\|} - \frac{\overline{g}^+(\overline{x}^k)}{\|\overline{g}^+(\overline{x}^k)\|} \right| \leq \frac{1}{k}, \qquad \|\overline{g}^+(\overline{x}^k)\| \to 0.$$

Since $\overline{x}$ lies in $\mathrm{ri}(X) = \mathrm{ri}(\overline{X})$, every point in the open line segment that connects x^k and $\overline{x}$, including $\overline{x}^k$, lies in $\mathrm{ri}(X)$, so that $\overline{f}(\overline{x}^k) = f(\overline{x}^k)$ and $\overline{g}(\overline{x}^k) = g(\overline{x}^k)$. We thus obtain a sequence $\{\overline{x}^k\}$ in the relative interior of X satisfying

$$\frac{q^* - f(\overline{x}^k)}{\|g^+(\overline{x}^k)\|} \to \|\mu^*\|, \qquad \frac{g^+(\overline{x}^k)}{\|g^+(\overline{x}^k)\|} \to \frac{\mu^*}{\|\mu^*\|}, \qquad \|g^+(\overline{x}^k)\| \to 0.$$

The first and the third relations imply $f(\overline{x}^k) \to q^*$. Thus μ^* satisfies condition (dCV) of Prop. 3.2. By Lemma 2.1, μ^* is the only dual optimal solution that satisfies this condition. **Q.E.D.**

Fritz John Conditions and Constraint Qualifications

We close this section by discussing the connection of the Fritz John conditions with classical constraint qualifications that guarantee the existence of a geometric multiplier (and hence also the existence of a dual optimal solution, which makes the analysis of the present section applicable). As mentioned earlier in this section, the classical Fritz John conditions of Prop. 3.1 can be used to assert the existence of a geometric multiplier when the Slater condition holds. However, Prop. 3.1 is insufficient to show that a geometric multiplier exists in the case of affine constraints. The following proposition strengthens the Fritz John conditions for this case, so that they suffice for the proof of the corresponding existence result.

Proposition 3.4: (Fritz John Conditions for Affine Constraints) Consider the convex problem (P), and assume that the functions $g_1, \ldots, g_r$ are affine, and $f^* < \infty$. Then there exists a FJ-multiplier (μ_0^*, μ^*) satisfying the following condition:

(CV') If $\mu^* \neq 0$, then there exists a vector $\tilde{x} \in X$ satisfying

$$f(\tilde{x}) < f^*, \qquad {\mu^*}' g(\tilde{x}) > 0.$$

Proof: If $\inf_{x \in X} f(x) = f^*$, then $\mu_0^* = 1$ and $\mu^* = 0$ form a FJ-multiplier, and condition (CV') is automatically satisfied. We will thus assume that $\inf_{x \in X} f(x) < f^*$, which also implies that f^* is finite.

Let the affine constraint function be represented as

$$g(x) = Ax - b,$$

for some real matrix A and vector b. Consider the nonempty convex sets

$$C_1 = \big\{(x, w) \mid \text{there is a vector } x \in X \text{ such that } f(x) < w\big\},$$

$$C_2 = \big\{(x, f^*) \mid Ax - b \leq 0\big\}.$$

Note that C_1 and C_2 are disjoint. The reason is that if $(x, f^*) \in C_1 \cap C_2$, then we must have $x \in X$, $Ax - b \leq 0$, and $f(x) < f^*$, contradicting the fact that f^* is the optimal value of the problem.

Since C_2 is polyhedral, by the Polyhedral Proper Separation Theorem (see [Roc70], Th. 20.2, or Prop. 3.5.1 of [BNO03]), there exists a hyperplane that separates C_1 and C_2 and does not contain C_1, i.e., there exists a vector (ξ, μ_0^*) such that

$$\mu_0^* f^* + \xi' z \leq \mu_0^* w + \xi' x, \qquad \forall\, x \in X, w, z \text{ with } f(x) < w,\ Az - b \leq 0, \tag{3.17}$$

$$\inf_{(x,w) \in C_1} \{\mu_0^* w + \xi' x\} < \sup_{(x,w) \in C_1} \{\mu_0^* w + \xi' x\}.$$

These relations imply that

$$\mu_0^* f^* + \sup_{Az - b \leq 0} \xi' z \leq \inf_{(x,w) \in C_1} \{\mu_0^* w + \xi' x\} < \sup_{(x,w) \in C_1} \{\mu_0^* w + \xi' x\}, \tag{3.18}$$

and that $\mu_0^* \geq 0$ [since w can be taken arbitrarily large in Eq. (3.17)].

Consider the linear program in Eq. (3.18):

$$\text{maximize } \xi' z$$
$$\text{subject to } \; Az - b \leq 0.$$

By Eq. (3.18), this program is bounded and therefore it has an optimal solution, which we denote by z^*. The dual of this program is

$$\text{minimize } b'\mu$$
$$\text{subject to } \; \xi = A'\mu, \qquad \mu \geq 0.$$

By linear programming duality, it follows that this problem has a dual optimal solution $\mu^* \geq 0$ satisfying

$$\sup_{Az-b\leq 0} \xi' z = \xi' z^* = {\mu^*}' b, \qquad \xi = A'\mu^*. \tag{3.19}$$

Note that μ_0^* and μ^* satisfy the nonnegativity condition (ii). Furthermore, we cannot have both $\mu_0^* = 0$ and $\mu^* = 0$, since then by Eq. (3.19), we would also have $\xi = 0$, and Eq. (3.18) would be violated. Thus, μ_0^* and μ^* also satisfy condition (iii) in the definition of a FJ-multiplier.

From Eq. (3.18), we have

$$\mu_0^* f^* + \sup_{Az-b\leq 0} \xi' z \leq \mu_0^* w + \xi' x, \qquad \forall\, x \in X \text{ with } f(x) < w,$$

which together with Eq. (3.19), implies that

$$\mu_0^* f^* + {\mu^*}' b \leq \mu_0^* w + {\mu^*}' A x, \qquad \forall\, x \in X \text{ with } f(x) < w,$$

or

$$\mu_0^* f^* \leq \inf_{x\in X,\, f(x)<w} \big\{ \mu_0^* w + {\mu^*}'(Ax - b) \big\}. \tag{3.20}$$

Similarly, from Eqs. (3.18) and (3.19), we have

$$\mu_0^* f^* < \sup_{x\in X,\, f(x)<w} \big\{ \mu_0^* w + {\mu^*}'(Ax - b) \big\}. \tag{3.21}$$

Using Eq. (3.20), we obtain

$$\begin{aligned}
\mu_0^* f^* &\leq \inf_{x\in X} \big\{ \mu_0^* f(x) + {\mu^*}'(Ax - b) \big\} \\
&\leq \inf_{x\in X,\, Ax-b\leq 0} \big\{ \mu_0^* f(x) + {\mu^*}'(Ax - b) \big\} \\
&\leq \inf_{x\in X,\, Ax-b\leq 0} \mu_0^* f(x) \\
&= \mu_0^* f^*.
\end{aligned}$$

Hence, equality holds throughout above, which proves condition (i) in the definition of FJ-multiplier.

We will now show that the vector μ^* also satisfies condition (CV'). To this end, we consider separately the cases where $\mu_0^* > 0$ and $\mu_0^* = 0$.

If $\mu_0^* > 0$, let $\tilde{x} \in X$ be such that $f(\tilde{x}) < f^*$ [based on our earlier assumption that $\inf_{x\in X} f(x) < f^*$]. Then condition (i) yields

$$\mu_0^* f^* \leq \mu_0^* f(\tilde{x}) + \mu^{*\prime}(A\tilde{x} - b),$$

implying that $0 < \mu_0^*\big(f^* - f(\tilde{x})\big) \leq \mu^{*\prime}(A\tilde{x} - b)$, and showing condition (CV').

If $\mu_0^* = 0$, condition (i) together with Eq. (3.21) yields

$$0 = \inf_{x\in X} \mu^{*\prime}(Ax - b) < \sup_{x\in X} \mu^{*\prime}(Ax - b). \tag{3.22}$$

The above relation implies the existence of a vector $\hat{x} \in X$ such that $\mu^{*'}(A\hat{x} - b) > 0$. Let $\overline{x} \in X$ be such that $f(\overline{x}) < f^*$, and consider a vector of the form

$$\tilde{x} = \alpha\hat{x} + (1 - \alpha)\overline{x},$$

where $\alpha \in (0, 1)$. Note that $\tilde{x} \in X$ for all $\alpha \in (0, 1)$, since X is convex. From Eq. (3.22), we have $\mu^{*\prime}(A\overline{x} - b) \geq 0$, which combined with the inequality $\mu^{*'}(A\hat{x} - b) > 0$, implies that

$$\mu^{*\prime}(A\tilde{x} - b) = \alpha\mu^{*\prime}(A\hat{x} - b) + (1 - \alpha)\mu^{*\prime}(A\overline{x} - b) > 0, \qquad \forall\ \alpha \in (0, 1). \tag{3.23}$$

Furthermore, since f is convex, we have

$$f(\tilde{x}) \leq \alpha f(\hat{x}) + (1 - \alpha) f(\overline{x}) = f^* + \big(f(\overline{x}) - f^*\big) + \alpha\big(f(\hat{x}) - f(\overline{x})\big), \qquad \forall\ \alpha \in (0, 1).$$

Thus, for α small enough so that $\alpha\big(f(\hat{x}) - f(\overline{x})\big) < f^* - f(\overline{x})$, we have $f(\tilde{x}) < f^*$ as well as $\mu^{*\prime}(A\tilde{x} - b) > 0$ [cf. Eq. (3.23)]. **Q.E.D.**

We now introduce the following constraint qualification, which is analogous to one introduced for nonconvex problems by Bertsekas and Ozdaglar [BeO02].

Definition 3.1: The constraint set of the convex problem (P) is said to be *pseudonormal* if one cannot find a vector $\mu \geq 0$ and a vector $\tilde{x} \in X$ satisfying the following conditions:

(i) $0 = \inf_{x\in X} \mu' g(x)$.

(ii) $\mu' g(\tilde{x}) > 0$.

To provide a geometric interpretation of pseudonormality, let us introduce the set

$$G = \{g(x) \mid x \in X\}$$

and consider hyperplanes that support this set and pass through 0. As Fig. 3.2 illustrates, *pseudonormality means that there is no hyperplane with a normal $\mu \geq 0$ that properly separates the sets $\{0\}$ and G, and contains G in its positive halfspace.*

It is evident (see also Fig. 3.2) that pseudonormality holds under the Slater condition, i.e., if there exists an $\bar{x} \in X$ such that $g(\bar{x}) < 0$. Prop. 3.4 also shows that if $f^* < \infty$, the constraint functions $g_1, \ldots, g_r$ are affine, and the constraint set is pseudonormal, then there exists a geometric multiplier satisfying the special condition (CV') of Prop. 3.4. As illustrated also in Fig. 3.2, the constraint set is pseudonormal if X is an affine set and g_j, $j = 1, \ldots, r$, are affine functions. In conclusion, *if $f^* < \infty$, and either the Slater condition holds, or X and $g_1, \ldots, g_r$ are affine, then the constraint set is pseudonormal, and a geometric multiplier is guaranteed to exist.* Since in this case there is no duality gap, Prop. 3.3 guarantees the existence of a geometric multiplier (the one of minimum norm) that satisfies the corresponding (CV) condition and sensitivity properties.

Finally, consider the question of pseudonormality and existence of geometric multipliers in the case where X is the intersection of a polyhedral set and a convex set C, and there exists a feasible solution that belongs to the relative interior of C. Then, the constraint set need not be pseudonormal, as Fig. 3.2(a) illustrates. However, it is pseudonormal in the extended representation (i.e., when the affine inequalities that represent the polyhedral part are lumped with the remaining affine inequality constraints), and it follows that there exists a geometric multiplier in the extended representation. From this, it follows that there exists a geometric multiplier in the original representation as well (see Exercise 6.2 of [BNO03]).

4. FRITZ JOHN CONDITIONS WHEN THERE IS NO OPTIMAL SOLUTION

In the preceding sections, we studied sensitivity properties of the geometric multiplier or dual optimal solution of minimum norm in the case where there exists a primal optimal solution or a dual optimal solution. In this section, we allow the problem to have neither a primal nor a dual optimal solution, and we develop several analogous results.

The Fritz John conditions of Props. 3.1 and 3.4 are weaker than Prop. 2.1 in that they do not include conditions analogous to condition (CV). Unfortunately, such a condition does not

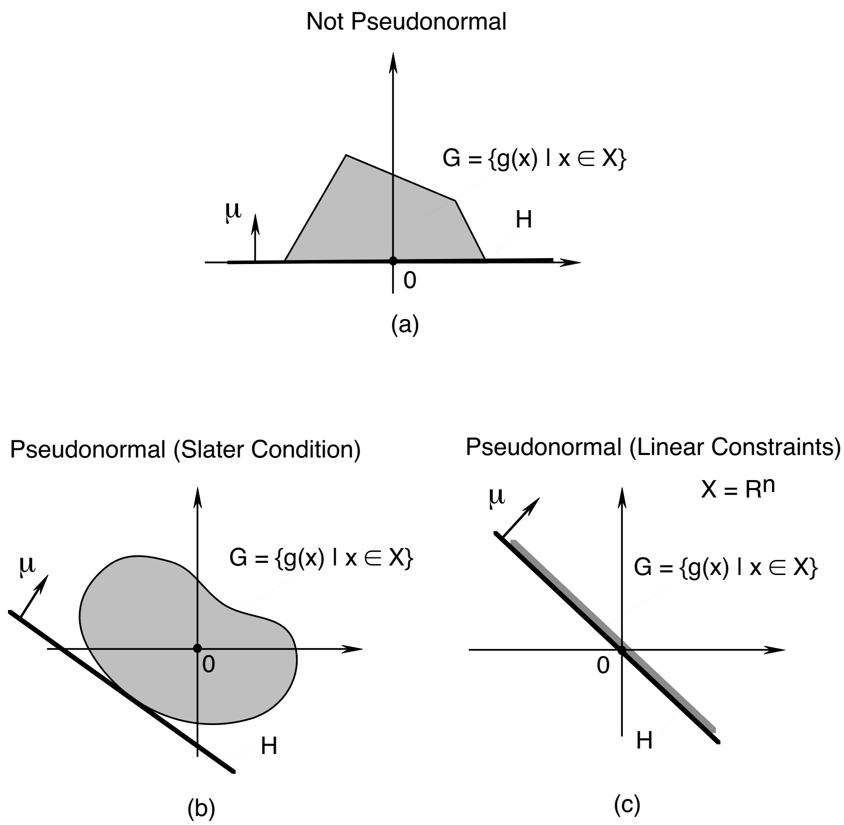

Figure 3.2. Geometric multiplier interpretation of pseudonormality. Consider the set

$$G = \left\{ g(x) \mid x \in X \right\}$$

and hyperplanes that support this set. For feasibility, G should intersect the nonpositive orthant $\{z \mid z \le 0\}$. The first condition $[0 = \inf_{x\in X} \mu' g(x)]$ in the definition of pseudonormality means that there is a hyperplane H with normal $\mu \ge 0$, which passes through 0, supports G, and contains G in its positive halfspace [note that, as illustrated in figure (b), this cannot happen if G intersects the interior of the nonpositive orthant; cf. the Slater criterion]. The second condition means that H does not fully contain G [cf. figures (a) and (c)]. If the Slater criterion holds, the first condition cannot be satisfied. If the linearity criterion holds, the set G is an affine set and the second condition cannot be satisfied (this depends critically on X being an affine set rather than X being a general polyhedron).

hold in the absence of additional assumptions, as can be seen from the following example.

Example 4.1

Consider the one-dimensional problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g(x) = x \le 0,\ x \in X = \{x \mid x \ge 0\}, \end{aligned}$$

where

$$f(x) = \begin{cases} -1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x < 0. \end{cases}$$

Then f is convex over X, and the assumptions of Props. 3.1 and 3.4 are satisfied. Indeed, each FJ-multiplier must have the form $\mu_0^* = 0$ and $\mu^* > 0$ (cf. Fig. 4.1). However, here we have $f^* = 0$, and for all x with $g(x) > 0$, we have $x > 0$ and $f(x) = -1$. Thus, there is no sequence $\{x^k\} \subset X$ satisfying (2.1)-(2.3).

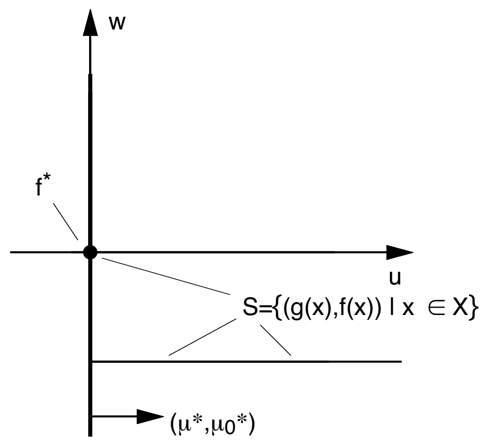

Figure 4.1. Illustration of the set $S = \Big\{\big(g(x), f(x)\big) \mid x \in X\Big\}$ in Example 4.1. Even though $\mu^* > 0$, there is no sequence $\{x^k\} \subset X$ such that $g(x^k) > 0$ for all k, and $f(x^k) \to f^*$.

The following proposition imposes the stronger Closedness Assumption in order to derive an enhanced set of Fritz John conditions analogous to those in Prop. 2.1. The proof uses ideas that are similar to the ones of the proof of Prop. 2.1, but is more complicated because an optimal solution of (P) may not exist. In particular, we approximate X by a sequence of expanding bounded convex subsets and we work with an optimal solution of the corresponding problem.

Proposition 4.1: (Enhanced Fritz John Conditions) Consider the convex problem (P) under Assumption 2.1 (Closedness), and assume that $f^* < \infty$. Then there exists a FJ-multiplier (μ_0^*, μ^*) satisfying the following condition (CV). Moreover, if $\mu_0^* \neq 0$, then μ^*/μ_0^* must be the geometric multiplier of minimum norm.

(CV) If $\mu^* \neq 0$, then there exists a sequence $\{x^k\} \subset X$ of infeasible points that satisfies Eqs. (2.1), (2.2), (2.3).

Proof: If $f(x) \geq f^*$ for all $x \in X$, then $\mu_0^* = 1$ and $\mu^* = 0$ form a FJ-multiplier, and condition (CV) is satisfied. Moreover, (CV) (in particular, Eq. (2.2)) is satisfied by only $\mu^* = 0$. We will thus assume that there exists some $\overline{x} \in X$ such that $f(\overline{x}) < f^*$. In this case, f^* is finite. Consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X^k, \qquad g(x) \leq 0, \end{aligned} \tag{4.1}$$

where

$$X^k = X \cap \left\{ x \;\middle|\; \|x\| \leq \beta k \right\}, \qquad k = 1, 2, \ldots,$$

and β is a scalar that is large enough so that for all k, the constraint set $\{x \in X^k \mid g(x) \leq 0\}$ is nonempty. Since f and g_j are closed and convex when restricted to X, they are closed, convex, and coercive when restricted to X^k. Hence, problem (4.1) has an optimal solution, which we denote by $\overline{x}^k$. Since this is a more constrained problem than the original, we have $f^* \leq f(\overline{x}^k)$ and $f(\overline{x}^k) \downarrow f^*$ as $k \to \infty$. Let

$$\gamma^k = f(\overline{x}^k) - f^*.$$

Note that if $\gamma^k = 0$ for some k, then $\overline{x}^k$ is an optimal solution for problem (P), and the result follows from Prop. 2.1 on enhanced Fritz John conditions for convex problems with an optimal solution. Therefore, we assume that $\gamma^k > 0$ for all k.

For positive integers k and positive scalars m, we consider the saddle function

$$L_{k,m}(x, \xi) = f(x) + \frac{(\gamma^k)^2}{4k^2} \|x - \overline{x}^k\|^2 + \xi' g(x) - \frac{\|\xi\|^2}{2m}.$$

We note that $L_{k,m}(x, \xi)$, viewed as a function from X^k to $\Re$, for fixed $\xi \geq 0$, is closed, convex, and coercive, in view of the Closedness Assumption. Furthermore, $L_{k,m}(x, \xi)$ is negative definite quadratic in ξ, for fixed x. Hence, we can use the Saddle Point Theorem (e.g., [BNO03, Prop. 2.6.9]) to assert that $L_{k,m}$ has a saddle point over $x \in X^k$ and $\xi \geq 0$, which we denote by $(x^{k,m}, \xi^{k,m})$.

We now derive several properties of the saddle points $(x^{k,m}, \xi^{k,m})$, which set the stage for the main argument. The first of these properties is

$$f(x^{k,m}) \leq L_{k,m}(x^{k,m}, \xi^{k,m}) \leq f(\overline{x}^k),$$

which is shown in the next paragraph.

The infimum of $L_{k,m}(x, \xi^{k,m})$ over $x \in X^k$ is attained at $x^{k,m}$, implying that

$$\begin{aligned} f(x^{k,m}) &+ \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x^{k,m}) \\ &= \inf_{x \in X^k} \left\{ f(x) + \frac{(\gamma^k)^2}{4k^2}\|x - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x) \right\} \\ &\le \inf_{x \in X^k,\, g(x) \le 0} \left\{ f(x) + \frac{(\gamma^k)^2}{4k^2}\|x - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x) \right\} \\ &\le \inf_{x \in X^k,\, g(x) \le 0} \left\{ f(x) + \frac{(\gamma^k)^2}{4k^2}\|x - \overline{x}^k\|^2 \right\} \\ &= f(\overline{x}^k). \end{aligned} \tag{4.2}$$

Hence, we have

$$\begin{aligned} L_{k,m}(x^{k,m}, \xi^{k,m}) &= f(x^{k,m}) + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x^{k,m}) - \frac{1}{2m}\|\xi^{k,m}\|^2 \\ &\le f(x^{k,m}) + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2 + \xi^{k,m\prime} g(x^{k,m}) \\ &\le f(\overline{x}^k). \end{aligned} \tag{4.3}$$

Since $L_{k,m}$ is quadratic in ξ, the supremum of $L_{k,m}(x^{k,m}, \xi)$ over $\xi \ge 0$ is attained at

$$\xi^{k,m} = m g^+(x^{k,m}). \tag{4.4}$$

This implies that

$$\begin{aligned} L_{k,m}(x^{k,m}, \xi^{k,m}) &= f(x^{k,m}) + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m} - \overline{x}^k\|^2 + \frac{m}{2}\|g^+(x^{k,m})\|^2 \\ &\ge f(x^{k,m}). \end{aligned} \tag{4.5}$$

We next show another property of the saddle points $(x^{k,m}, \xi^{k,m})$, namely that for each k, we have

$$\lim_{m \to \infty} f(x^{k,m}) = f(\overline{x}^k) = f^* + \gamma^k. \tag{4.6}$$

For a fixed k and any sequence of integers m that tends to ∞, consider the corresponding sequence $\{x^{k,m}\}$. From Eqs. (4.3) and (4.5), we see that $\{x^{k,m}\}$ belongs to the set $\big\{x \in X^k \mid f(x) \le f(\overline{x}^k)\big\}$, which is compact, since f is closed. Hence, $\{x^{k,m}\}$ has a cluster point, denoted by $\hat{x}^k$, which belongs to $\big\{x \in X^k \mid f(x) \le f(\overline{x}^k)\big\}$. By passing to a subsequence if necessary, we can assume without loss of generality that $\{x^{k,m}\}$ converges to $\hat{x}^k$. We claim that $\hat{x}^k$ is feasible for problem (4.1), i.e., $\hat{x}^k \in X^k$ and $g(\hat{x}^k) \le 0$. Indeed, the sequence $\big\{f(x^{k,m})\big\}$ is bounded from below by $\inf_{x \in X^k} f(x)$, which is finite by Weierstrass' Theorem since f is closed and coercive when restricted to X^k. Also, for each k, $L_{k,m}(x^{k,m}, \xi^{k,m})$ is bounded from above by $f(\overline{x}^k)$ [cf. Eq. (4.3)], so Eq. (4.5) implies that

$$\limsup_{m \to \infty} g_j(x^{k,m}) \le 0, \qquad \forall\ j = 1, \ldots, r.$$

Therefore, by using the closedness of g_j, we obtain $g(\hat{x}^k) \leq 0$, implying that $\hat{x}^k$ is a feasible solution of problem (4.1). Thus, $f(\hat{x}^k) \geq f(\overline{x}^k)$. Using Eqs. (4.3) and (4.5) together with the closedness of f, we also have

$$f(\hat{x}^k) \leq \liminf_{m\to\infty} f(x^{k,m}) \leq \limsup_{m\to\infty} f(x^{k,m}) \leq f(\overline{x}^k),$$

thereby showing Eq. (4.6).

The next step in the proof is given in the following lemma:

Lemma 4.1: For all sufficiently large k, and for all scalars $m \leq 1/\sqrt{\gamma^k}$, we have

$$f(x^{k,m}) \leq f^* - \frac{\gamma^k}{2}. \tag{4.7}$$

Furthermore, there exists a scalar $m_k \geq 1/\sqrt{\gamma^k}$ such that

$$f(x^{k,m_k}) = f^* - \frac{\gamma^k}{2}. \tag{4.8}$$

Proof: Let $\gamma = f^* - f(\overline{x})$, where $\overline{x}$ was defined earlier as the vector in X such that $f(\overline{x}) < f^*$. For sufficiently large k, we have $\overline{x} \in X^k$ and $\gamma^k < \gamma$. Consider the vector

$$z^k = \left(1 - \frac{2\gamma^k}{\gamma^k + \gamma}\right)\overline{x}^k + \frac{2\gamma^k}{\gamma^k + \gamma}\overline{x},$$

which belongs to X^k for sufficiently large k [by the convexity of X^k and the fact that $2\gamma^k/(\gamma^k + \gamma) < 1$]. By the convexity of f, we have

$$\begin{aligned} f(z^k) &\leq \left(1 - \frac{2\gamma^k}{\gamma^k + \gamma}\right) f(\overline{x}^k) + \frac{2\gamma^k}{\gamma^k + \gamma} f(\overline{x}) \\ &= \left(1 - \frac{2\gamma^k}{\gamma^k + \gamma}\right)(f^* + \gamma^k) + \frac{2\gamma^k}{\gamma^k + \gamma}(f^* - \gamma) \\ &= f^* - \gamma^k. \end{aligned} \tag{4.9}$$

Similarly, by the convexity of g_j, we have

$$g_j(z^k) \leq \left(1 - \frac{2\gamma^k}{\gamma^k + \gamma}\right) g_j(\overline{x}^k) + \frac{2\gamma^k}{\gamma^k + \gamma} g_j(\overline{x}) \leq \frac{2\gamma^k}{\gamma^k + \gamma} g_j(\overline{x}). \tag{4.10}$$

Using Eq. (4.5), we obtain

$$\begin{aligned} f(x^{k,m}) &\leq L_{k,m}(x^{k,m}, \xi^{k,m}) \\ &= \inf_{x \in X^k} \sup_{\xi \geq 0} L_{k,m}(x, \xi) \\ &= \inf_{x \in X^k} \left\{ f(x) + \frac{(\gamma^k)^2}{4k^2}\|x - \bar{x}^k\|^2 + \frac{m}{2}\|g^+(x)\|^2 \right\} \\ &\leq f(x) + (\beta\gamma^k)^2 + \frac{m}{2}\|g^+(x)\|^2, \qquad \forall\ x \in X^k, \end{aligned}$$

where in the last inequality we also use the definition of X^k so that $\|x - \bar{x}^k\| \leq 2\beta k$ for all $x \in X^k$. Substituting $x = z^k$ in the preceding relation, and using Eqs. (4.9) and (4.10), we see that for large k,

$$f(x^{k,m}) \leq f^* - \gamma^k + (\beta\gamma^k)^2 + \frac{2m(\gamma^k)^2}{(\gamma^k + \gamma)^2}\|g^+(\overline{x})\|^2.$$

Since $\gamma^k \to 0$, this implies that for sufficiently large k and for all scalars $m \leq 1/\sqrt{\gamma^k}$, we have

$$f(x^{k,m}) \leq f^* - \frac{\gamma^k}{2},$$

i.e., Eq. (4.7) holds.

We next show that there exists a scalar $m_k \geq 1/\sqrt{\gamma^k}$ such that Eq. (4.8) holds. In the process, we show that, for fixed k, $L_{k,m}(x^{k,m}, \xi^{k,m})$ changes continuously with m, i.e, for all $\overline{m} > 0$, we have $L_{k,m}(x^{k,m}, \xi^{k,m}) \to L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}})$ as $m \to \overline{m}$. [By this we mean, for every sequence $\{m^t\}$ that converges to $\overline{m}$, the corresponding sequence $L_{k,m^t}(x^{k,m^t}, \xi^{k,m^t})$ converges to $L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}})$.] Denote

$$f^k(x) = f(x) + \frac{(\gamma^k)^2}{4k^2}\|x - \overline{x}^k\|^2.$$

From Eq. (4.5), we have

$$L_{k,m}(x^{k,m}, \xi^{k,m}) = \bar{f}(x^{k,m}) + \frac{m}{2}\|g^+(x^{k,m})\|^2 = \inf_{x \in X^k}\left\{\bar{f}(x) + \frac{m}{2}\|g^+(x)\|^2\right\},$$

so that for all $m \geq \overline{m}$, we obtain

$$\begin{aligned} L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}}) &= f^k(x^{k,\overline{m}}) + \frac{\overline{m}}{2}\|g^+(x^{k,\overline{m}})\|^2 \\ &\leq f^k(x^{k,m}) + \frac{\overline{m}}{2}\|g^+(x^{k,m})\|^2 \\ &\leq f^k(x^{k,m}) + \frac{m}{2}\|g^+(x^{k,m})\|^2 \\ &\leq f^k(x^{k,\overline{m}}) + \frac{m}{2}\|g^+(x^{k,\overline{m}})\|^2. \end{aligned}$$

It follows that $L_{k,m}(x^{k,m}, \xi^{k,m}) \to L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}})$ as $m \downarrow \overline{m}$. Similarly, we have for all $m \leq \overline{m}$,

$$\begin{aligned} f^k(x^{k,\overline{m}}) + \frac{m}{2}\|g^+(x^{k,\overline{m}})\|^2 &\leq f^k(x^{k,\overline{m}}) + \frac{\overline{m}}{2}\|g^+(x^{k,\overline{m}})\|^2 \\ &\leq f^k(x^{k,m}) + \frac{\overline{m}}{2}\|g^+(x^{k,m})\|^2 \\ &= f^k(x^{k,m}) + \frac{m}{2}\|g^+(x^{k,m})\|^2 + \frac{\overline{m} - m}{2}\|g^+(x^{k,m})\|^2 \\ &\leq f^k(x^{k,\overline{m}}) + \frac{m}{2}\|g^+(x^{k,\overline{m}})\|^2 + \frac{\overline{m} - m}{2}\|g^+(x^{k,m})\|^2. \end{aligned}$$

For each k, $f(x^{k,m})$ is bounded from below by $\inf_{x \in X^k} f(x)$, which is finite by Weierstrass' Theorem since f is closed and coercive when restricted to X^k. Since, by Eqs. (4.3) and (4.5),

$$f(x^{k,m}) + \frac{m}{2}\|g^+(x^{k,m})\|^2 \leq f(\overline{x}^k),$$

we see that $m\|g^+(x^{k,m})\|^2$ is bounded from above as $m \uparrow \overline{m} > 0$, so that $(\overline{m}-m)\|g^+(x^{k,m})\|^2 \to 0$. Therefore, we have from the preceding relation that $L_{k,m}(x^{k,m}, \xi^{k,m}) \to L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}})$ as $m \uparrow \overline{m}$, which shows that $L_{k,m}(x^{k,m}, \xi^{k,m})$ changes continuously with m.

Next, we show that, for fixed k, $x^{k,m} \to x^{k,\overline{m}}$ as $m \to \overline{m}$. Since, for each k, $x^{k,m}$ belongs to the compact set $\left\{x \in X^k \mid f(x) \le f(\overline{x}^k)\right\}$, it has a cluster point as $m \to \overline{m}$. Let $\hat{x}$ be a cluster point of $x^{k,m}$. Using the continuity of $L_{k,m}(x^{k,m}, \xi^{k,m})$ in m, and the closedness of f^k and g_j, we obtain

$$\begin{aligned} L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}}) &= \lim_{m\to\bar{m}} L_{k,m}(x^{k,m}, \xi^{k,m}) \\ &= \lim_{m\to\overline{m}} \left\{ f^k(x^{k,m}) + \frac{m}{2}\|g^+(x^{k,m})\|^2 \right\} \\ &\ge f^k(\hat{x}) + \frac{\overline{m}}{2}\|g^+(\hat{x})\|^2 \\ &\ge \inf_{x\in X^k} \left\{ f^k(x) + \frac{\overline{m}}{2}\|g^+(x)\|^2 \right\} \\ &= L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}}). \end{aligned}$$

This shows that $\hat{x}$ attains the infimum of $f^k(x) + \frac{\overline{m}}{2}\|g^+(x)\|^2$ over $x \in X^k$. Since this function is strictly convex, it has a unique optimal solution, showing that $\hat{x} = x^{k,\overline{m}}$.

Finally, we show that $f(x^{k,m}) \to f(x^{k,\overline{m}})$ as $m \to \overline{m}$. Since f is lower semicontinuous at $x^{k,\overline{m}}$, we have $f(x^{k,\overline{m}}) \le \liminf_{m\to\overline{m}} f(x^{k,m})$. Thus it suffices to show that $f(x^{k,\overline{m}}) \ge \limsup_{m\to\overline{m}} f(x^{k,m})$. Assume that $f(x^{k,\overline{m}}) < \limsup_{m\to\overline{m}} f(x^{k,m})$. Using the continuity of $L_{k,m}(x^{k,m}, \xi^{k,m})$ in m and the fact that $x^{k,m} \to x^{k,\overline{m}}$ as $m \to \overline{m}$, we have

$$\begin{aligned} f^k(x^{k,\overline{m}}) + \liminf_{m\to\overline{m}} \|g^+(x^{k,m})\|^2 &< \limsup_{m\to\overline{m}} L_{k,m}(x^{k,m}, \xi^{k,m}) \\ &= L_{k,\overline{m}}(x^{k,\overline{m}}, \xi^{k,\overline{m}}) \\ &= f^k(x^{k,\overline{m}}) + \|g^+(x^{k,\overline{m}})\|^2. \end{aligned}$$

This contradicts the lower semicontinuity of g_j, so that $f(x^{k,\overline{m}}) \ge \limsup_{m\to\overline{m}} f(x^{k,m})$. Thus $f(x^{k,m})$ is continuous in m.

From Eqs. (4.6), (4.7), and the continuity of $f(x^{k,m})$ in m, we see that there exists some scalar $m_k \ge 1/\sqrt{\gamma^k}$ such that Eq. (4.8) holds. **Q.E.D.**

We are now ready to construct FJ-multipliers with the desired properties. By combining Eqs. (4.8), (4.3), and (4.5) (for $m = m_k$), together with the facts that $f(\overline{x}^k) \to f^*$ and $\gamma^k \to 0$ as $k \to \infty$, we obtain

$$\lim_{k\to\infty} \left(f(x^{k,m_k}) - f^* + \frac{(\gamma^k)^2}{4k^2}\|x^{k,m_k} - \overline{x}^k\|^2 + \xi^{k,m_k'} g(x^{k,m_k}) \right) = 0. \tag{4.11}$$

Denote

$$\delta^k = \sqrt{1 + \|\xi^{k,m_k}\|^2}, \qquad \mu_0^k = \frac{1}{\delta^k}, \qquad \mu^k = \frac{\xi^{k,m_k}}{\delta^k}. \tag{4.12}$$

Since δ^k is bounded from below by 1, Eq. (4.11) yields

$$\lim_{k\to\infty}\left(\mu_0^k f(x^{k,m_k}) - \mu_0^k f^* + \frac{(\gamma^k)^2}{4k^2\delta^k}\|x^{k,m_k} - \overline{x}^k\|^2 + \mu^{k\prime} g(x^{k,m_k})\right) = 0. \tag{4.13}$$

Substituting $m = m_k$ in the first relation of Eq. (4.2) and dividing by δ^k, we obtain

$$\begin{aligned}\mu_0^k f(x^{k,m_k}) + \frac{(\gamma^k)^2}{4k^2\delta^k}\|x^{k,m_k} - \overline{x}^k\|^2 + \mu^{k\prime} g(x^{k,m_k})&\\ \leq \mu_0^k f(x) + \mu^{k\prime} g(x) + \frac{(\beta\gamma^k)^2}{\delta^k},& \qquad \forall\, x \in X^k,\end{aligned}$$

where we also use the fact that $\|x - \bar{x}^k\| \leq 2\beta k$ for all $x \in X^k$ (cf. the definition of X^k). Since the sequence $\{(\mu_0^k, \mu^k)\}$ is bounded, it has a cluster point, denoted by (μ_0^*, μ^*), which satisfies conditions (ii) and (iii) in the definition of FJ-multiplier. Without loss of generality, we will assume that the entire sequence $\{(\mu_0^k, \mu^k)\}$ converges to (μ_0^*, μ^*). For any $x \in X$, we have $x \in X^k$ for all k sufficiently large. Taking the limit as $k \to \infty$ in the preceding relation, and using Eq. (4.13) and $\gamma^k \to 0$, yields

$$\mu_0^* f^* \leq \mu_0^* f(x) + \mu^{*\prime} g(x), \qquad \forall\, x \in X,$$

which implies that

$$\begin{aligned}\mu_0^* f^* &\leq \inf_{x\in X}\{\mu_0^* f(x) + \mu^{*\prime} g(x)\}\\ &\leq \inf_{x\in X,\, g(x)\leq 0}\{\mu_0^* f(x) + \mu^{*\prime} g(x)\}\\ &\leq \inf_{x\in X,\, g(x)\leq 0} \mu_0^* f(x)\\ &= \mu_0^* f^*.\end{aligned}$$

Thus we have

$$\mu_0^* f^* = \inf_{x\in X}\{\mu_0^* f(x) + \mu^{*\prime} g(x)\},$$

so that μ_0^*, μ^* satisfy condition (i) in the definition of FJ-multiplier. Note that the existence of $\overline{x} \in X$ such that $f(\overline{x}) < f^*$, together with condition (i), imply that $\mu^* \neq 0$.

Finally, we establish condition (CV). Using Eqs. (4.4), (4.12) and the fact that $\mu^k \to \mu^*$, we obtain

$$\frac{g^+(x^{k,m_k})}{\|g^+(x^{k,m_k})\|} = \frac{\mu^{k,m_k}}{\|\mu^{k,m_k}\|} \to \frac{\mu^*}{\|\mu^*\|}.$$

We have from Eq. (4.8) and $\gamma^k \to 0$ that $f(x^{k,m_k}) \to f^*$. We also have from Eqs. (4.3), (4.5) with $m = m_k$, and (4.8) that

$$\frac{m_k}{2}\|g^+(x^{k,m_k})\|^2 \leq f(\bar{x}^k) - f(x^{k,m_k}) = \frac{3}{2}\gamma^k,$$

where the equality uses Eqs. (4.4), (4.12). Since $\gamma^k \to 0$ and $m_k \geq 1/\sqrt{\gamma^k} \to \infty$, this yields $g^+(x^{k,m_k}) \to 0$. Moreover, combining the above inequality with Eq. (4.8) yields

$$\frac{f^* - f(x^{k,m_k})}{\|g^+(x^{k,m_k})\|} = \frac{\gamma^k}{2\|g^+(x^{k,m_k})\|} \geq \frac{m_k\|g^+(x^{k,m_k})\|}{6} = \frac{\|\mu^{k,m_k}\|}{6\mu_0^{k,m_k}}. \tag{4.14}$$

If $\mu_0^* = 0$, then $\mu_0^{k,m_k} \to 0$, so Eq. (4.14) together with $\|\mu^{k,m_k}\| \to \|\mu^*\| > 0$ yields

$$\frac{f^* - f(x^{k,m_k})}{\|g^+(x^{k,m_k})\|} \to \infty.$$

It follows that the sequence $\{x^{k,m_k}\}$ satisfies condition (CV) of the proposition. If $\mu_0^* \neq 0$, then μ^*/μ_0^* is a geometric multiplier and $f^* = q^*$, so that μ^*/μ_0^* is also a dual optimal solution. Thus the set of dual optimal solutions is nonempty and coincides with the set of geometric multipliers. Then, the vector $(1, \bar{\mu})$, where $\bar{\mu}$ is the dual optimal solution of minimum norm, is a FJ-multiplier and, by Prop. 3.2 and the fact $f^* = q^*$, it satisfies condition (CV) and is the only dual optimal solution that satisfies this condition. This completes the proof. **Q.E.D.**

The FJ-multipliers of Props. 3.1, 3.4, 4.1 define a hyperplane with normal (μ^*, μ_0^*) that supports the set of constraint-cost pairs (i.e., the set M of Fig. 3.1) at $(0, f^*)$. On the other hand, it is possible to construct a hyperplane that supports the set M at the point $(0, q^*)$, where q^* is the dual optimal value, while asserting the existence of a sequence that satisfies a condition analogous to condition (CV) of Prop. 4.1. This is the subject of the next proposition. Its proof uses Lemmas 2.1 and 3.1.

In analogy with a FJ-multiplier, we consider a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)'$, satisfying the following conditions:

(i) $\mu_0^* q^* = \inf_{x \in X} \{\mu_0^* f(x) + \mu^{*\prime} g(x)\}$.

(ii) $\mu_j^* \geq 0$ for all $j = 0, 1, \ldots, r$.

(iii) $\mu_0^*, \mu_1^*, \ldots, \mu_r^*$ are not all equal to 0.

We call such a pair (μ_0^*, μ^*) a *dual FJ-multiplier.* If $\mu_0^* \neq 0$, then μ^*/μ_0^* is a dual optimal solution; otherwise $\mu_0^* = 0$ and $\mu^* \neq 0$.

Proposition 4.2: (Enhanced Dual Fritz John Conditions) Consider the convex problem (P) under Assumption 2.1 (Closedness), and assume that $f^* < \infty$ and $-\infty < q^*$. Then there exists a dual FJ-multiplier (μ_0^*, μ^*) satisfying the following condition (dCV). Moreover, if $\mu_0^* \neq 0$, then μ^*/μ_0^* must be the dual optimal solution of minimum norm.

(dCV) If $\mu^* \neq 0$, then there exists a sequence $\{x^k\} \subset X$ of infeasible points that satisfies

$$f(x^k) \to q^*, \qquad g^+(x^k) \to 0, \tag{4.15}$$

$$\frac{q^* - f(x^k)}{\|g^+(x^k)\|} \to \begin{cases} \|\mu^*\|/\mu_0^* & \text{if } \mu_0^* \neq 0, \\ \infty & \text{if } \mu_0^* = 0, \end{cases} \tag{4.16}$$

$$\frac{g^+(x^k)}{\|g^+(x^k)\|} \to \frac{\mu^*}{\|\mu^*\|}. \tag{4.17}$$

Proof: Since by assumption, we have $-\infty < q^*$ and $f^* < \infty$, it follows from the weak duality relation $q^* \leq f^*$ that both q^* and f^* are finite. For $k = 1, 2, \ldots$, consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X,\ g_j(x) \leq \frac{1}{k^4}, \quad j = 1, \ldots, r. \end{aligned}$$

By Lemma 3.1, for each k, the optimal value of this problem is less than or equal to q^*. Then, for each k, there exists a vector $\tilde{x}^k \in X$ that satisfies

$$f(\tilde{x}^k) \leq q^* + \frac{1}{k^2}, \qquad g_j(\tilde{x}^k) \leq \frac{1}{k^4}, \quad j = 1, \ldots, r.$$

Consider also the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g_j(x) \leq \frac{1}{k^2}, \quad j = 1, \ldots, r, \\ &\qquad\qquad\quad x \in \tilde{X}^k = X \cap \left\{ x \;\middle|\; \|x\| \leq k \left(\max_{1 \leq i \leq k} \|\tilde{x}^i\| + 1 \right) \right\}. \end{aligned} \tag{4.18}$$

Since f and g_j are closed and convex when restricted to X, they are closed, convex, and coercive when restricted to $\tilde{X}^k$. Hence, problem (4.18) has an optimal solution, which we denote by $\overline{x}^k$. Note that since $\tilde{x}^k$ belongs to the feasible solution set of this problem, we have

$$f(\overline{x}^k) \leq f(\tilde{x}^k) \leq q^* + \frac{1}{k^2}. \tag{4.19}$$

For each k, we consider the saddle function

$$L_k(x, \xi) = f(x) + \xi' g(x) - \frac{\|\xi\|^2}{2k},$$

and the set

$$X^k = \tilde{X}^k \cap \big\{x \mid g_j(x) \le k,\ j = 1, \ldots, r\big\}. \tag{4.20}$$

We note that $L_k(x, \xi)$, for fixed $\xi \ge 0$, is closed, convex, and coercive in x, when restricted to X^k, and negative definite quadratic in ξ for fixed x. Hence, using the Saddle Point Theorem (e.g., [BNO03], Prop. 2.6.9), we can assert that L_k has a saddle point over $x \in X^k$ and $\xi \ge 0$, denoted by (x^k, ξ^k).

Since L_k is quadratic in ξ, the supremum of $L_k(x^k, \xi)$ over $\xi \ge 0$ is attained at

$$\xi^k = k g^+(x^k). \tag{4.21}$$

Similarly, the infimum of $L_k(x, \xi^k)$ over $x \in X^k$ is attained at x^k, implying that

$$\begin{aligned} f(x^k) + \xi^{k\prime} g(x^k) &= \inf_{x \in X^k} \big\{f(x) + \xi^{k\prime} g(x)\big\} \\ &= \inf_{x \in X^k} \big\{f(x) + k g^+(x^k)' g(x)\big\} \\ &\le \inf_{x \in X^k,\ g_j(x) \le \frac{1}{k^4},\ j=1,\ldots,r} \left\{ f(x) + k \sum_{j=1}^r g_j^+(x^k)' g_j(x) \right\} \\ &\le \inf_{x \in X^k,\ g_j(x) \le \frac{1}{k^4},\ j=1,\ldots,r} \left\{ f(x) + \frac{r}{k^2} \right\} \\ &= f(\overline{x}^k) + \frac{r}{k^2} \\ &\le q^* + \frac{r+1}{k^2}, \end{aligned} \tag{4.22}$$

where the second inequality follows using the fact $g_j^+(x^k) \le k$, $j = 1, \ldots, r$ [cf. Eq. (4.20)], and the third inequality follows from Eq. (4.19).

Since q^* is finite, we may select a nonnegative sequence $\{\zeta^k\}$ such that

$$q(\zeta^k) \to q^*, \qquad \frac{\|\zeta^k\|^2}{2k} \to 0. \tag{4.23}$$

[For example, we can take ζ^k to be any maximizer of $q(\zeta)$ subject to $\zeta \ge 0$ and $\|\zeta\| \le k^{1/3}$.] Then, we have for all k,

$$\begin{aligned} L_k(x^k, \xi^k) &= \sup_{\xi \ge 0} \inf_{x \in X^k} L_k(x, \xi) \\ &\ge \sup_{\xi \ge 0} \inf_{x \in X} L_k(x, \xi) \\ &= \sup_{\xi \ge 0} \left\{ \inf_{x \in X} \big\{f(x) + \xi' g(x)\big\} - \frac{\|\xi\|^2}{2k} \right\} \\ &= \sup_{\xi \ge 0} \left\{ q(\xi) - \frac{\|\xi\|^2}{2k} \right\} \\ &\ge q(\zeta^k) - \frac{\|\zeta^k\|^2}{2k}. \end{aligned} \tag{4.24}$$

Combining Eqs. (4.24) and (4.22), we obtain

$$\begin{aligned} q(\zeta^k) - \frac{\|\zeta^k\|^2}{2k} &\le L_k(x^k, \xi^k) \\ &= f(x^k) + \xi^{k'} g(x^k) - \frac{\|\xi^k\|^2}{2k} \\ &\le f(x^k) + \xi^{k'} g(x^k) \\ &\le q^* + \frac{r+1}{k^2}. \end{aligned} \tag{4.25}$$

Taking the limit in the preceding relation, and using Eq. (4.23), we obtain

$$\lim_{k\to\infty} \big\{ f(x^k) - q^* + \xi^{k'} g(x^k) \big\} = 0. \tag{4.26}$$

Denote

$$\delta^k = \sqrt{1 + \|\xi^k\|^2}, \qquad \mu_0^k = \frac{1}{\delta^k}, \qquad \mu^k = \frac{\xi^k}{\delta^k}. \tag{4.27}$$

Since δ^k is bounded from below by 1, Eq. (4.26) yields

$$\lim_{k\to\infty} \big\{ \mu_0^k \big(f(x^k) - q^* \big) + \mu^{k'} g(x^k) \big\} = 0. \tag{4.28}$$

Dividing both sides of the first relation in Eq. (4.22) by δ^k, we get

$$\mu_0^k f(x^k) + \mu^{k'} g(x^k) \le \mu_0^k f(x) + \mu^{k'} g(x), \qquad \forall\ x \in X^k.$$

Since the sequence $\big\{(\mu_0^k, \mu^k)\big\}$ is bounded, it has a cluster point (μ_0^*, μ^*). This cluster point satisfies conditions (ii) and (iii) of the proposition. Without loss of generality, we assume that the entire sequence converges. For any $x \in X$, we have $x \in X^k$ for all k sufficiently large. Taking the limit as $k \to \infty$ in the preceding relation and using Eq. (4.28) yields

$$\mu_0^* q^* \le \mu_0^* f(x) + \mu^{*'} g_j(x), \qquad \forall\ x \in X.$$

We consider separately the two cases, $\mu_0^* > 0$ and $\mu_0^* = 0$, in the above relation to show that (μ_0^*, μ^*) satisfy condition (i) of the proposition. Indeed, if $\mu_0^* > 0$, by dividing with μ_0^*, we have

$$q^* \le \inf_{x\in X} \left\{ f(x) + \frac{\mu^{*'}}{\mu_0^*} g(x) \right\} = q\left(\frac{\mu^*}{\mu_0^*}\right) \le q^*.$$

Similarly, if $\mu_0^* = 0$, it can be seen that

$$0 = \inf_{x\in X} \mu^{*'} g(x)$$

[since $f^* < \infty$, so that there exists an $x \in X$ such that $g(x) \le 0$ and $\mu^{*\prime} g(x) \le 0$]. Hence, in both cases, we have

$$\mu_0^* q^* = \inf_{x\in X} \big\{ \mu_0^* f(x) + \mu^{*'} g(x) \big\},$$

thus showing condition (i) in the definition of dual FJ-multiplier.

If $\mu^* = 0$, then $\mu_0^* \neq 0$, (dCV) is automatically satisfied, and $\mu^*/\mu_0^* = 0$ has minimum norm. Assume now that $\mu^* \neq 0$. Using Eqs. (4.21), (4.27) and the fact that $\mu^k \to \mu^*$, we obtain

$$\frac{g^+(x^k)}{\|g^+(x^k)\|} = \frac{\mu^k}{\|\mu^k\|} \to \frac{\mu^*}{\|\mu^*\|}.$$

This proves (4.17). Also, we have from Eq. (4.25) that

$$k\big(f(x^k) - q^*\big) + \xi^{k\prime} k g(x^k) \leq \frac{r+1}{k}, \qquad \forall\ k = 1, 2, \ldots.$$

Using Eq. (4.21), this yields

$$k\big(f(x^k) - q^*\big) + \|\xi^k\|^2 \leq \frac{r+1}{k}.$$

Dividing both sides by $\|\xi^k\| = k\|g^+(x^k)\|$ and using Eq. (4.27) yields

$$\frac{q^* - f(x^k)}{\|g^+(x^k)\|} \geq \|\xi^k\| - \frac{r+1}{k\|\xi^k\|} = \frac{\|\mu^k\|}{\mu_0^k} - \frac{r+1}{k\|\mu^k\|/\mu_0^k}. \tag{4.29}$$

If $\mu_0^* = 0$, then $\mu_0^k \to 0$, so Eq. (4.29) together with $\|\mu^k\| \to \|\mu^*\| > 0$ yields

$$\frac{q^* - f(x^k)}{\|g^+(x^k)\|} \to \infty.$$

If $\mu_0^* \neq 0$, then Eq. (4.29) together with $\mu_0^k \to \mu_0^*$ and $\|\mu^k\| \to \|\mu^*\|$ yields

$$\liminf_{k\to\infty} \frac{q^* - f(x^k)}{\|g^+(x^k)\|} \geq \frac{\|\mu^*\|}{\mu_0^*}.$$

Since μ^*/μ_0^* is a dual optimal solution, Lemma 2.1 shows that in fact μ^*/μ_0^* is of minimum norm and the inequality holds with equality.

We finally show that $f(x^k) \to q^*$ and $g^+(x^k) \to 0$. By Eqs. (4.25) and (4.23), we have

$$\lim_{k\to\infty} \frac{\|\xi^k\|^2}{2k} = 0. \tag{4.30}$$

By Eq. (4.21), we have

$$\xi^{k\prime} g(x^k) = \frac{1}{k}\|\xi^k\|^2,$$

so using also Eqs. (4.25) and (4.23), we obtain

$$\lim_{k\to\infty} f(x^k) + \frac{\|\xi^k\|^2}{2k} = q^*,$$

which together with Eq. (4.30) shows that $f(x^k) \to q^*$. Moreover, Eqs. (4.30) and (4.21) imply that

$$\lim_{k\to\infty} k\|g^+(x^k)\|^2 = 0,$$

showing that $g^+(x^k) \to 0$. Therefore, the sequence $\{x^k\}$ satisfies condition (dCV) of the proposition, completing the proof. **Q.E.D.**

Note that the proof of Prop. 4.2 is similar to the proof of Prop. 2.1. The idea is to generate saddle points of the function

$$L_k(x,\xi) = f(x) + \xi' g(x) - \frac{\|\xi\|^2}{2k},$$

over $x \in X^k$ [cf. Eq. (4.20)] and $\xi \geq 0$. It can be shown that

$$L_k(x^k,\xi^k) = \inf_{u \in \Re^r} \left\{ p^k(u) + \frac{k}{2}\|u^+\|^2 \right\},$$

where $p^k(u)$ is the optimal value of the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g(x) \leq u, \ x \in X^k, \end{aligned} \tag{4.31}$$

(see the discussion following the proof of Prop. 2.1). For each k, the value $L_k(x^k,\xi^k)$ can be visualized geometrically as in Fig. 2.1. However, here the rate at which X^k approaches X is chosen high enough so that $L_k(x^k,\xi^k)$ converges to q^* as $k \to \infty$ [cf. Eq. (4.25)], and not to f^*, as in the proof of Props. 2.1 or 4.1.

5. REFERENCES

[BNO02] Bertsekas, D. P., Nedić, A., and Ozdaglar, A. E., 2003. Convex Analysis and Optimization, Athena Scientific, Belmont, MA.

[BeO02] Bertsekas, D. P., and Ozdaglar, A. E., 2002. “Pseudonormality and a Lagrange Multiplier Theory for Constrained Optimization,” J. Opt. Theory Appl., Vol. 114, 2002, pp. 287-343.

[Ber99] Bertsekas, D. P., 1999. Nonlinear Programming: 2nd Edition, Athena Scientific, Belmont, MA.

[Joh48] John, F., 1948. “Extremum Problems with Inequalities as Subsidiary Conditions,” in Studies and Essays: Courant Anniversary Volume, K. O. Friedrichs, Neugebauer, O. E., and Stoker, J. J., (Eds.), Wiley-Interscience, N. Y., pp. 187-204.

[Hes75] Hestenes, M. R., 1975. Optimization Theory: The Finite Dimensional Case, Wiley, N. Y.

[Roc70] Rockafellar, R. T., 1970. Convex Analysis, Princeton Univ. Press, Princeton, N. J.

[Roc93] Rockafellar, R. T., 1993. "Lagrange Multipliers and Optimality," SIAM Rev., Vol. 35, 1993, pp. 183-238.

[Sla50] Slater, M., 1950. "Lagrange Multipliers Revisited: A Contribution to Non-Linear Programming," Cowles Commission Discussion paper, Math. 403.