MATHEMATICS OF OPERATIONS RESEARCH
Vol. 29, No. 3, August 2004, pp. 479–491
ISSN 0364-765X | EISSN 1526-5471 | 04 | 2903 | 0479
DOI 10.1287/moor.1040.0103

Some Properties of the Augmented Lagrangian in Cone Constrained Optimization

Alexander Shapiro
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205, ashapiro@isye.gatech.edu

Jie Sun
Department of Decision Sciences and Singapore-MIT Alliance, National University of Singapore, Republic of Singapore, jsun@nus.edu.sg

A large class of optimization problems can be modeled as minimization of an objective function subject to constraints given in a form of set inclusions. In this paper, we discuss augmented Lagrangian duality for such optimization problems. We formulate the augmented Lagrangian dual problems and study conditions ensuring existence of the corresponding augmented Lagrange multipliers. We also discuss sensitivity of optimal solutions to small perturbations of augmented Lagrange multipliers.

Key words: cone constraints; augmented Lagrangian; conjugate duality; duality gap; Lagrange multipliers; sensitivity analysis
MSC2000 subject classification: Primary: 90C30
OR/MS subject classification: Primary: Programming/nonlinear
History: Received September 16, 2003; revised February 1, 2004.

1. Introduction. The augmented Lagrangian approach to finite-dimensional problems with a finite number of equality constraints was introduced in Hestenes (1969) and Powell (1969). This was extended to finite-dimensional inequality constrained problems by Buys (1972). Theoretical properties of the augmented Lagrangian duality method, in a finite-dimensional setting with a finite number of constraints, were thoroughly investigated in Rockafellar (1974a).

In this paper, we consider optimization problems defined in the form

$$\min_{x\in\mathscr{X}} f(x) \quad \text{subject to} \quad G(x)\in K, \tag{1.1}$$

where $\mathscr{X}$ is a vector space, K is a nonempty convex subset of a vector space $\mathscr{Y}$, $f\colon \mathscr{X}\to\overline{\mathbb{R}}$ is an extended real-valued function, and $G\colon \mathscr{X}\to\mathscr{Y}$. We assume that $\mathscr{Y}$ is a Hilbert space equipped with a scalar product, denoted $\langle\cdot,\cdot\rangle$, and that the set K is closed in the strong (norm) topology of $\mathscr{Y}$. A large class of optimization problems can be formulated in the form (1.1). For example, in case $\mathscr{Y}$ is the linear space of $p\times p$ symmetric matrices and $K\subset\mathscr{Y}$ is the set (cone) of positive semidefinite matrices, problem (1.1) becomes a (nonlinear) semidefinite programming problem.

In the next section, we introduce the augmented Lagrangian dual of problem (1.1) and study its basic properties. The developments of that section follow basic ideas outlined in Rockafellar (1974a). In §3, we study the existence of augmented Lagrange multipliers. Some results of that section, and the following §4, seem to be new even in the finite-dimensional setting. In particular, second-order necessary and sufficient conditions for existence of an augmented Lagrange multiplier are given in Theorem 3.4. In §4, we discuss sensitivity of minimizers of the augmented Lagrangian to small perturbations of the augmented Lagrange multipliers. The analysis of §§3 and 4 is based on a perturbation theory of optimization problems. We use Bonnans and Shapiro (2000) as a reference book for that theory.

We use the following notation and terminology throughout this paper. For a mapping $G\colon \mathscr{X}\to\mathscr{Y}$, we denote by $DG(x)$ its derivative at $x\in\mathscr{X}$. If $\mathscr{X}$ and $\mathscr{Y}$ are finite dimensional, we can write $DG(x)h = [\nabla G(x)]^T h$, where $\nabla G(x)$ is the Jacobian matrix of $G(\cdot)$ at x. For a set $S \subset \mathcal{Y}$, we denote by $\text{int}(S)$ its interior and by

$$\sigma(y, S) := \sup_{z \in S} \langle y, z \rangle \tag{1.2}$$

its support function. The metric projection $P_K(y)$ of $y \in \mathcal{Y}$ onto the set K is defined as point $z \in K$ closest to y. That is, $P_K(y) \in K$ and $\text{dist}(y, K) = \|y - P_K(y)\|$. Because $\mathcal{Y}$ is a Hilbert space and K is closed and convex, $P_K(y)$ exists and is uniquely defined. By $T_K(y)$ and $N_K(y)$ we denote the tangent and normal cones, respectively, to the set K at $y \in K$. By definition the sets $T_K(y)$ and $N_K(y)$ are empty if $y \notin K$. The set

$$T_K^2(y, z) := \left\{ w\colon \text{dist}(y + tz + \tfrac{1}{2}t^2 w, K) = o(t^2),\ \ t \geq 0 \right\}$$

is called the second-order tangent set to K at y in direction z. The set $T_K^2(y, z)$ can be nonempty only if $y \in K$ and $z \in T_K(y)$. If K is a cone, then

$$K^* := \{ y \in \mathcal{Y}\colon \langle y, z \rangle \leq 0,\ \ \forall z \in K \} \tag{1.3}$$

defines the (negative) dual of K. For a function $v\colon \mathcal{Y} \to \overline{\mathbb{R}}$, we denote by $v^*(\cdot)$ its conjugate,

$$v^*(y^*) := \sup_{y \in \mathcal{Y}} \{ \langle y^*, y \rangle - v(y) \}. \tag{1.4}$$

2. Augmented duality. Consider the optimization problem (1.1), to which we refer as problem (P). The natural way of introducing the augmented Lagrangian dual for this problem is by the following construction (cf., Rockafellar and Wets 1998, Chapter 11, Section K*). With problem (P) is associated the parameterized problem, denoted (P_y):

$$\min_{x \in \mathcal{X}} f(x) \quad \text{subject to} \quad G(x) + y \in K. \tag{2.1}$$

Clearly, for $y = 0$, problem (P_0) coincides with problem (P). We denote by $v(y)$ the optimal value of problem (P_y), that is $v(y) := \text{val}(P_y)$. Consider the function

$$v_\tau(y) := v(y) + \tau \|y\|^2. \tag{2.2}$$

We say that $\lambda \in \mathcal{Y}$ is an *augmented Lagrange multiplier* of (P) if $\text{val}(P)$ is finite and there exists $\tau \geq 0$ such that

$$v_\tau(y) \geq v_\tau(0) + \langle \lambda, y \rangle \quad \forall y \in \mathcal{Y} \tag{2.3}$$

(cf., Rockafellar and Wets 1998, Example 11.62). The above condition (2.3) means that λ is a subgradient, at $y = 0$, of the function $v_\tau(\cdot)$. The set of all λ satisfying (2.3) is called the subdifferential of $v_\tau(y)$, at $y = 0$, and denoted $\partial v_\tau(0)$. Note that $\partial v_\tau(0)$ is defined only if $v_\tau(0) = \text{val}(P)$ is finite.

We denote by $\mathcal{A}$ the set of all augmented Lagrange multipliers. Because $v_\tau(0) = v(0)$, it immediately follows from the definition that if $\tau \leq \tau'$, then $\partial v_\tau(0) \subset \partial v_{\tau'}(0)$, and hence

$$\mathcal{A} = \bigcup_{\tau \in \mathbb{R}_+} \partial v_\tau(0). \tag{2.4}$$

It follows that the set $\mathcal{A}$ is convex. Note also that if $v(\cdot)$ is convex, then $\partial v_\tau(0)$ coincides with $\partial v(0)$ for any $\tau \geq 0$, and hence in that case $\mathcal{A} = \partial v(0)$.

The function

$$\mathcal{L}(x, \lambda, \tau) := \inf_{y \in K - G(x)} \{ f(x) - \langle \lambda, y \rangle + \tau \|y\|^2 \} \tag{2.5}$$

is called the *augmented Lagrangian* of problem (P). We have that

$$\inf_{y\in\mathcal{Y}}\big\{v_\tau(y)-\langle\lambda, y\rangle\big\} = \inf_{y\in\mathcal{Y}}\inf_{x\in\mathcal{X}}\big\{f(x)-\langle\lambda, y\rangle+\tau\|y\|^2\colon G(x)+y\in K\big\}$$
$$= \inf_{x\in\mathcal{X}}\inf_{y\in K-G(x)}\big\{f(x)-\langle\lambda, y\rangle+\tau\|y\|^2\big\},$$

and hence (cf., Rockafellar and Wets 1998)

$$\inf_{y\in\mathcal{Y}}\{v_\tau(y)-\langle\lambda, y\rangle\} = \inf_{x\in\mathcal{X}}\mathcal{L}(x,\lambda,\tau). \tag{2.6}$$

It is straightforward to verify that for $\tau>0$ the augmented Lagrangian can be written in the form (cf., Rockafellar and Wets 1998):

$$\mathcal{L}(x,\lambda,\tau) = f(x)+\tau[\mathrm{dist}(G(x)+(2\tau)^{-1}\lambda, K)]^2-(4\tau)^{-1}\|\lambda\|^2, \tag{2.7}$$

while $\mathcal{L}(x,\lambda,0)=L(x,\lambda)-\sigma(\lambda,K)$ for $\tau=0$. Here,

$$L(x,\lambda):=f(x)+\langle\lambda, G(x)\rangle \tag{2.8}$$

is the (standard) Lagrangian of problem (P). It follows from (2.5) and (2.6), respectively, that the functions $\mathcal{L}(x,\cdot,\cdot)$ and

$$g(\lambda,\tau):=\inf_{x\in\mathcal{X}}\mathcal{L}(x,\lambda,\tau)$$

are concave and upper semicontinuous on $\mathcal{Y}\times\mathbb{R}$.

Consider the metric projection operator $P_K(\cdot)$ onto the set K. If K is a convex *cone*, then $\mathrm{dist}(y,K)=\|P_{K^*}(y)\|$. Consequently, in that case for $\tau>0$, the augmented Lagrangian can be written as follows:

$$\mathcal{L}(x,\lambda,\tau) = f(x)+\frac{1}{4\tau}(\|P_{K^*}(\lambda+2\tau G(x))\|^2-\|\lambda\|^2), \tag{2.9}$$

while for $\tau=0$, it is given by $\mathcal{L}(x,\lambda,0)=L(x,\lambda)$ if $\lambda\in K^*$ and $\mathcal{L}(x,\lambda,0)=-\infty$ if $\lambda\notin K^*$.

In the finite-dimensional setting the following result is well known (cf., Rockafellar 1993, p. 212), we give its proof for the sake of completeness.

Lemma 2.1. *For any $\tau\geq 0$ we have that*

$$\sup_{\lambda\in\mathcal{Y}}\mathcal{L}(x,\lambda,\tau) = \begin{cases} f(x) & \text{if } G(x)\in K,\\ +\infty & \text{otherwise,}\end{cases} \tag{2.10}$$

and

$$\mathrm{val}(P) = \inf_{x\in\mathcal{X}}\sup_{\lambda\in\mathcal{Y}}\mathcal{L}(x,\lambda,\tau). \tag{2.11}$$

Proof. Suppose first that $\tau>0$. Then,

$$[\mathrm{dist}(G(x)+(2\tau)^{-1}\lambda, K)]^2-\|(2\tau)^{-1}\lambda\|^2 = \mathrm{dist}(\mu,S)^2-\|\mu\|^2, \tag{2.12}$$

where $\mu:=(2\tau)^{-1}\lambda$ and $S:=K-G(x)$. If $0\in S$, i.e., $G(x)\in K$, then the right-hand side of (2.12) is less than or equal to zero, and hence in that case the supremum of the right-hand side of (2.12), over $\mu\in\mathcal{Y}$, is zero. On the other hand, if $0\notin S$, then for $\mu:=-aP_S(0)$, with $a>0$, we have $\mathrm{dist}(\mu,S)=(1+a)\|P_S(0)\|$. By letting $a\to+\infty$, we obtain that the supremum of the right-hand side of (2.12) is $+\infty$, and hence (2.10) follows. For $\tau=0$, the proof of (2.10) is similar.

Equation (2.11) follows immediately from (2.10). □

For $\tau \geq 0$, consider the following dual, denoted (D_τ), of problem (P):

$$\max_{\lambda \in \mathcal{Y}} \left\{ g(\lambda, \tau) = \inf_{x \in \mathcal{X}} \mathcal{L}(x, \lambda, \tau) \right\}. \tag{2.13}$$

Let us make the following observations. Equation (2.6) can be written in the form

$$g(\lambda, \tau) = -v_\tau^*(\lambda), \tag{2.14}$$

where $v_\tau^*(\cdot)$ is the conjugate of the function $v_\tau(\cdot)$. Therefore, we obtain that

$$\text{val}(D_\tau) = v_\tau^{**}(0). \tag{2.15}$$

Recall that by the Fenchel-Moreau Theorem (see, e.g., Rockafellar 1974b), we have that

$$v_\tau^{**} = \text{cl}(\text{conv}\, v_\tau). \tag{2.16}$$

The results presented in the following theorem are quite standard in the conjugate duality theory (cf., Rockafellar 1974a, 1993, p. 213).

THEOREM 2.1. *For any $\tau \geq 0$, the following holds*: (i) $\text{val}(P) \geq \text{val}(D_\tau)$ *and*

$$\text{val}(D_\tau) = \text{cl}(\text{conv}\, v_\tau)(0), \tag{2.17}$$

(ii) $\text{val}(P) = \text{val}(D_\tau)$ *if and only if* $\text{cl}(\text{conv}\, v_\tau)(0) = v_\tau(0)$, (iii) *if $\partial v_\tau(0)$ is nonempty, then* $\text{val}(P) = \text{val}(D_\tau)$ *and the set of optimal solutions of (D_τ) coincides with* $\partial v_\tau(0)$, (iv) *if* $\text{val}(P) = \text{val}(D_\tau)$ *and is finite, then the (possibly empty) set of optimal solutions of (D_τ) coincides with $\partial v_\tau(0)$, and* (v) $\text{val}(P) = \text{val}(D_\tau)$ *and $\bar{x}$ and $\bar{\lambda}$ are optimal solutions of (P) and (D_τ), respectively, if and only if $(\bar{x}, \bar{\lambda})$ is a saddle point of $\mathcal{L}(\cdot, \cdot, \tau)$, i.e.,*

$$\mathcal{L}(\bar{x}, \lambda, \tau) \leq \mathcal{L}(\bar{x}, \bar{\lambda}, \tau) \leq \mathcal{L}(x, \bar{\lambda}, \tau) \quad \forall\, (x, \lambda) \in \mathcal{X} \times \mathcal{Y}. \tag{2.18}$$

PROOF. The inequality $\text{val}(P) \geq \text{val}(D_\tau)$ follows from the min-max representations (2.11) and (2.13), and (2.17) is a consequence of (2.15) and (2.16). Property (ii) is a consequence of (2.17). Properties (iii) and (iv) follow by general duality theory (e.g., Bonnans and Shapiro 2000, Theorem 2.142). Property (v) follows from (2.11) and (2.13). □

It is also clear that the optimal value in the left-hand side of (2.6), and hence $g(\lambda, \tau)$ and $\text{val}(D_\tau)$, are monotonically nondecreasing with increase of τ. It follows that if $\text{val}(P) = \text{val}(D_\tau)$ for some $\tau \geq 0$, then $\text{val}(P) = \text{val}(D_{\tau'})$ for any $\tau' \geq \tau$. As a consequence of the above results we obtain the following.

THEOREM 2.2. *If for some $\bar{\tau} \geq 0$ the set $\partial v_{\bar{\tau}}(0)$ is nonempty (i.e., there exists an augmented Lagrange multiplier), then for any $\tau \geq \bar{\tau}$ there is no duality gap between problems (P) and (D_τ), and for any $\bar{\lambda} \in \partial v_\tau(0)$ the (possibly empty) set of optimal solutions of (P) is contained in the set* $\arg\min_{x \in \mathcal{X}} \mathcal{L}(x, \bar{\lambda}, \tau)$.

The following condition was introduced in Rockafellar (1974a) and called there "the quadratic growth condition." Because we use the term "quadratic growth" for a different meaning, we refer to it as condition (R).

CONDITION (R). There exist constants a and b such that

$$v(y) \geq a - b\|y\|^2 \quad \forall\, y \in \mathcal{Y}. \tag{2.19}$$

If, for example, $f(x)$ is bounded from below on $\mathcal{X}$ by a constant c, then (2.19) holds with $a := c$ and $b := 0$.

LEMMA 2.2. *Suppose that* $\liminf_{y\to 0} v(y) < +\infty$ *and condition* (R) *holds. Then, for any* $\lambda \in \mathcal{Y}$,

$$\lim_{\tau\to+\infty} g(\lambda, \tau) = \liminf_{y\to 0} v(y). \tag{2.20}$$

PROOF. By (2.6) we have

$$g(\lambda, \tau) = \inf_{y\in\mathcal{Y}} \{v(y) + \tau\|y\|^2 - \langle\lambda, y\rangle\}. \tag{2.21}$$

It follows that for any τ, $g(\lambda, \tau)$ is less than or equal to the right-hand side of (2.20), and hence

$$\lim_{\tau\to+\infty} g(\lambda, \tau) \le \liminf_{y\to 0} v(y).$$

Now it follows from condition (R) that, for any $y \in \mathcal{Y}$,

$$v_\tau(y) - \langle\lambda, y\rangle \ge a - \|\lambda\|\,\|y\| + (\tau - b)\|y\|^2, \tag{2.22}$$

and hence $\liminf_{y\to 0} v(y) > -\infty$. It follows that $\liminf_{y\to 0} v(y)$ is finite. We have then that for any $r > 0$, there exists τ^* such that for $\tau \ge \tau^*$ the infimum of the right-hand side of (2.22), over all y satisfying $\|y\| > r$, is bigger than $\liminf_{y\to 0} v(y)$. Consequently, by (2.22) we obtain that for any $r > 0$,

$$\sup_{\tau\in\mathbb{R}_+} \inf_{\|y\|>r} \{v_\tau(y) - \langle\lambda, y\rangle\} \ge \liminf_{y\to 0} v(y) = \liminf_{y\to 0} [v_\tau(y) - \langle\lambda, y\rangle].$$

Together with (2.21) this implies that

$$\lim_{\tau\to+\infty} g(\lambda, \tau) \ge \liminf_{y\to 0} v(y),$$

and hence (2.20) follows. □

Consider the following problem, denoted $(\hat{D})$:

$$\max_{(\lambda,\tau)\in\mathcal{Y}\times\mathbb{R}_+} g(\lambda, \tau). \tag{2.23}$$

Clearly $\mathrm{val}(\hat{D}) = \sup_{\tau\in\mathbb{R}_+} \mathrm{val}(D_\tau)$. Because for any $\tau \ge 0$ we have that $\mathrm{val}(D_\tau) \le \mathrm{val}(P)$, it follows that $\mathrm{val}(\hat{D}) \le \mathrm{val}(P)$.

THEOREM 2.3. *Suppose that* $\liminf_{y\to 0} v(y) < +\infty$ *and condition* (R) *holds. Then,*

$$\mathrm{val}(\hat{D}) = \liminf_{y\to 0} v(y) \le \mathrm{val}(P). \tag{2.24}$$

PROOF. By Lemma 2.2 we have that

$$\mathrm{val}(\hat{D}) \ge \liminf_{y\to 0} v(y). \tag{2.25}$$

Because of (2.17) we have that for any $\tau \ge 0$,

$$\mathrm{val}(D_\tau) \le \liminf_{y\to 0} v_\tau(y) = \liminf_{y\to 0} v(y),$$

and hence the opposite of inequality (2.25) holds. This proves the first equality statement in (2.24). The inequality statement of (2.24) follows from $\mathrm{val}(\hat{D}) \le \mathrm{val}(P)$. □

It is said that there is *no duality gap* between problems $(\hat{D})$ and (P) if

$$\mathrm{val}(\hat{D}) = \mathrm{val}(P). \tag{2.26}$$

It follows from (2.24) that the "no duality gap" condition (2.26) holds iff the optimal value function $v(y)$ is lower semicontinuous at $y = 0$. There exist various conditions ensuring lower semicontinuity of $v(y)$ at $y = 0$. One such condition is that the space $\mathscr{X}$ is a topological vector space, the function $f(\cdot)$ is lower semicontinuous, the mapping $G(\cdot)$ is continuous, and the so-called inf-compactness condition holds (e.g., Bonnans and Shapiro 2000, Proposition 4.4).

THEOREM 2.4. *Suppose that condition* (R) *is satisfied,* $\text{val}(P)$ *is finite, and the optimal value function* $v(y)$ *is lower semicontinuous at* $y = 0$. *Then, the following holds*: (i) *there is no duality gap between* $(\hat{D})$ *and* (P), (ii) *if* $(\bar{\lambda}, \bar{\tau})$ *is an optimal solution of* $(\hat{D})$, *then* $\text{val}(D_{\bar{\tau}}) = \text{val}(P)$ *and* $\bar{\lambda} \in \partial v_{\bar{\tau}}(0)$, *and* (iii) *if* $\bar{\lambda} \in \partial v_{\bar{\tau}}(0)$, *then* $(\bar{\lambda}, \bar{\tau})$ *is an optimal solution of* $(\hat{D})$.

PROOF. The no duality gap property (i) was already discussed above. Now if $(\bar{\lambda}, \bar{\tau})$ is an optimal solution of $(\hat{D})$, then, because $\text{val}(D_\tau)$ is nondecreasing as a function τ, we have that $\text{val}(D_{\bar{\tau}}) = \text{val}(\hat{D})$. It follows that $\text{val}(D_{\bar{\tau}}) = \text{val}(P)$ and $\bar{\lambda} \in \partial v_{\bar{\tau}}(0)$. Conversely, if this holds, then again by the monotonicity we obtain that $\text{val}(D_{\bar{\tau}}) = \text{val}(P)$ and $(\bar{\lambda}, \bar{\tau})$ is an optimal solution of $(\hat{D})$. □

3. Existence of augmented Lagrange multipliers.

In this section, we discuss the existence of augmented Lagrange multipliers. We assume that condition (R) holds and $\text{val}(P)$ is finite. The following lemma shows that in this case verification of condition (2.3) can be reduced to a local analysis. Denote $B_r := \{y: \|y\| \le r\}$.

LEMMA 3.1. *Suppose that condition* (R) *holds and* $\text{val}(P)$ *is finite. Then,* λ *is an augmented Lagrange multiplier if and only if for any* $\varepsilon > 0$, *there exists* $\tau \ge 0$ *such that*

$$v_\tau(y) \ge v_\tau(0) + \langle \lambda, y \rangle \quad \forall\, y \in B_\varepsilon. \tag{3.1}$$

PROOF. Necessity of condition (3.1) follows directly from the definition. Let us show sufficiency. It follows from condition (R) that

$$v_\tau(y) - v_\tau(0) - \langle \lambda, y \rangle \ge a - v(0) - \|\lambda\| \|y\| + (\tau - b)\|y\|^2. \tag{3.2}$$

This implies that, for $\tau > b$, the left-hand side of (3.2) is nonnegative for all y such that $\|y\| \ge r(\tau)$, where

$$r(\tau) := \frac{\|\lambda\|}{2(\tau - b)} + \sqrt{\frac{\|\lambda\|^2}{4(\tau - b)^2} + \frac{v(0) - a}{\tau - b}}. \tag{3.3}$$

Therefore, λ is an augmented Lagrange multiplier if, for $\tau > b$, the inequality in (2.3) holds for all $y \in B_r$, where $r = r(\tau)$ is defined in (3.3). Clearly, for a fixed λ, $r(\tau) \to 0$ as $\tau \to +\infty$. Now if condition (3.1) holds for some $\tau \ge 0$ and $\varepsilon > 0$, then it holds for any bigger value of τ and the same ε. By taking τ large enough so that $\tau > b$ and $\varepsilon > r(\tau)$, we obtain that (3.1) implies condition (2.3) for all $y \in \mathscr{Y}$, and hence λ is an augmented Lagrange multiplier. □

We assume now that the space $\mathscr{X}$ is a Banach space, the function $f: \mathscr{X} \to \mathbb{R}$ is real valued, and $f(\cdot)$ and $G(\cdot)$ are continuously differentiable. Let x_0 be an optimal solution of problem (P). Existence of such an optimal solution implies, of course, that $\text{val}(P)$ is finite. We denote by $\Lambda(x_0)$ the set of Lagrange multipliers satisfying the first-order optimality conditions at the point x_0:

$$D_x L(x_0, \lambda) = 0, \quad \lambda \in N_K(G(x_0)). \tag{3.4}$$

Recall that if Robinson's constraint qualification holds at x_0, then the set $\Lambda(x_0)$ is nonempty and bounded.

Consider the function $\delta(\cdot) := \text{dist}(\cdot, K)^2$. We have that

$$\delta(y) = \inf_{z\in K} \|y - z\|^2, \tag{3.5}$$

and hence the function $\delta(\cdot)$ is convex and differentiable with

$$D\delta(y) = 2(y - P_K(y)), \tag{3.6}$$

where (3.6) follows, for example, by the Danskin Theorem (e.g., Bonnans and Shapiro 2000). Moreover, $P_K(\cdot)$ is Lipschitz continuous (modulus one), and hence $D\delta(\cdot)$ is Lipschitz continuous. By the chain rule of differentiation we obtain that

$$D_\lambda \mathscr{L}(x, \lambda, \tau) = G(x) - P_K(G(x) + (2\tau)^{-1}\lambda). \tag{3.7}$$

Let $\bar\lambda$ be an augmented Lagrange multiplier, i.e., $\bar\lambda \in \partial v_\tau(0)$ for some $\tau \geq 0$. Recall that existence of the augmented Lagrange multiplier $\bar\lambda$ implies that $\text{val}(P) = \text{val}(D_\tau)$. Suppose also that problem (P) has an optimal solution x_0. Then, $(x_0, \bar\lambda)$ is a saddle point of $\mathscr{L}(\cdot, \cdot, \tau)$, and hence $D_\lambda \mathscr{L}(x_0, \bar\lambda, \tau) = 0$ and $D_x \mathscr{L}(x_0, \bar\lambda, \tau) = 0$. By (3.7) the first of these equations means that

$$G(x_0) = P_K(G(x_0) + (2\tau)^{-1}\bar\lambda). \tag{3.8}$$

We also have that $y - P_K(y) \in N_K(P_K(y))$ for any $y \in \mathscr{Y}$. By applying this to $y := G(x_0) + (2\tau)^{-1}\bar\lambda$ and using (3.8) we obtain that $\bar\lambda \in N_K(G(x_0))$. By using (3.6) and (3.8) it is also straightforward to verify that

$$D_x \mathscr{L}(x_0, \bar\lambda, \tau) = D_x L(x_0, \bar\lambda), \tag{3.9}$$

and hence $D_x L(x_0, \bar\lambda) = 0$. It follows that $\bar\lambda \in \Lambda(x_0)$. We obtain the following result.

Proposition 3.1. *If x_0 is an optimal solution of problem (P), then $\mathscr{A} \subset \Lambda(x_0)$.*

It may be interesting to remark that it follows from the above proposition that if S^* is the set of optimal solutions of (P), then $\mathscr{A} \subset \bigcap_{x\in S^*} \Lambda(x)$. In particular, if there are two points in S^* with disjoint sets of Lagrange multipliers, then problem (P) does not possess augmented Lagrange multipliers.

Let us now consider the lower directional derivative of $v(\cdot)$ at $y = 0$, defined as

$$v'_-(0, d) := \liminf_{t\downarrow 0} \frac{v(td) - v(0)}{t}. \tag{3.10}$$

The upper directional derivative $v'_+(0, d)$ is defined similarly by taking "lim sup" instead of "lim inf" in (3.10). It is said that $v(y)$ is directionally differentiable at $y = 0$ if $v'_-(0, d) = v'_+(0, d)$ for all $d \in \mathscr{Y}$. In that case the directional derivative $v'(0, d)$ is equal to $v'_-(0, d)$ and $v'_+(0, d)$.

Theorem 3.1. *Let x_0 be an optimal solution of problem (P). Suppose that Robinson's constraint qualification holds at x_0, and $\mathscr{A} = \Lambda(x_0)$. Then, $v(\cdot)$ is directionally differentiable at $y = 0$ and $v'(0, d) = \sigma(d, \Lambda(x_0))$.*

Proof. We have that for any $\lambda \in \partial v_\tau(0)$,

$$v(td) - v(0) \geq -t^2\tau\|d\|^2 + t\langle\lambda, d\rangle, \tag{3.11}$$

and hence $v'_-(0, d) \geq \langle\lambda, d\rangle$. It follows that if the set $\mathscr{A}$ of augmented Lagrange multipliers is nonempty, then

$$v'_-(0, d) \geq \sigma(d, \mathscr{A}) \quad \forall\, d \in \mathscr{Y}. \tag{3.12}$$

Suppose that Robinson's constraint qualification holds at x_0. Then, the set $\Lambda(x_0)$ is nonempty and bounded (e.g., Bonnans and Shapiro 2000, Theorem 3.9) and

$$v'_+(0,d) \le \sigma(d, \Lambda(x_0)) \quad \forall\, d \in \mathcal{Y} \tag{3.13}$$

(see Bonnans and Shapiro 2000, Proposition 4.22). By (3.12) and (3.13) we have that for any $d \in \mathcal{Y}$,

$$\sigma(d, \mathcal{A}) \le v'_-(0,d) \le v'_+(0,d) \le \sigma(d, \Lambda(x_0)).$$

If, moreover, $\mathcal{A} = \Lambda(x_0)$, then $v'_-(0,d) = v'_+(0,d)$ and is equal to $\sigma(d, \Lambda(x_0))$. □

Let us observe at this point that even if (P) has unique optimal solution x_0 at which Robinson's constraint qualification holds, it still may happen that $v'_+(0,d)$ is strictly less than $\sigma(d, \Lambda(x_0))$ for some $d \in \mathcal{Y}$ (see Bonnans and Shapiro 2000, Proposition 4.108 and the following discussion). In such a case the inclusion $\mathcal{A} \subset \Lambda(x_0)$ is strict. A simple example of a finite-dimensional problem (3 decision variables and 3 constraints) with unique optimal solution, satisfying a weak second-order optimality condition, and having a nonempty and bounded set of Lagrange multipliers, and yet with empty set of augmented Lagrange multipliers, is given in (Bonnans et al. 2003, exercise 14.4).

Let x_0 be an optimal solution of problem (P) and suppose that Robinson's constraint qualification holds at x_0. If, further, $\mathcal{A} = \Lambda(x_0)$, then by Theorem 3.1 we have that $v'(0,\cdot) = \sigma(\cdot, \Lambda(x_0))$. This, in turn, implies that for any $d \in \mathcal{Y}$ and $\varepsilon > 0$, there exists a $(t\varepsilon)$-optimal $(t \ge 0)$ solution $\bar{x}(t)$ of (P_{td}) such that $\|\bar{x}(t) - x_0\| = O(t)$ (see Bonnans and Shapiro 2000, p. 282). Therefore, property $\mathcal{A} = \Lambda(x_0)$ is closely related to Lipschitz stability of optimal (nearly optimal) solutions of the parameterized problem (P_y).

Theorem 3.2. *Let x_0 be an optimal solution of problem (P). Suppose that:* (i) *condition* (R) *holds,* (ii) *$f(\cdot)$ and $G(\cdot)$ are $C^{1,1}$ (i.e., $f(\cdot)$ and $G(\cdot)$ are differentiable and their derivatives are locally Lipschitz continuous),* (iii) *Robinson's constraint qualification holds at x_0, and* (iv) *for all y in a neighborhood of $0 \in \mathcal{Y}$, problem (P_y) has an $\varepsilon(y)$-optimal solution $\bar{x}(y)$ such that $\varepsilon(y) = O(\|y\|^2)$ and*

$$\|\bar{x}(y) - x_0\| = O(\|y\|). \tag{3.14}$$

Then, $\mathcal{A} = \Lambda(x_0)$.

Proof. By Lemma 3.1 we need to verify the corresponding subgradient inequality only for all y near $0 \in \mathcal{Y}$. We have that

$$v(y) - v(0) = f(\bar{x}(y)) - f(x_0) + O(\|y\|^2). \tag{3.15}$$

Moreover, for any $\lambda \in \Lambda(x_0)$, we have that $\lambda \in N_K(G(x_0))$, and because $G(x_0) \in K$ and $G(\bar{x}(y)) + y \in K$, it follows that

$$\langle \lambda, G(\bar{x}(y)) + y - G(x_0) \rangle \le 0.$$

Consequently,

$$f(\bar{x}(y)) - f(x_0) \ge L(\bar{x}(y), \lambda) - L(x_0, \lambda) + \langle \lambda, y \rangle. \tag{3.16}$$

Because f and G are $C^{1,1}$, and by the first-order optimality conditions, $D_x L(x_0, \lambda) = 0$, and because of the assumption $\|\bar{x}(y) - x_0\| = O(\|y\|)$, we have that

$$|L(\bar{x}(y), \lambda) - L(x_0, \lambda)| = O(\|y\|^2). \tag{3.17}$$

It follows from (3.15)–(3.17) that for any $\lambda \in \Lambda(x_0)$,

$$v(y) - v(0) \ge \langle \lambda, y \rangle + O(\|y\|^2). \tag{3.18}$$

This implies (3.1) for some $\varepsilon > 0$ and $\tau \ge 0$ large enough, and hence $\lambda \in \mathcal{A}$. We obtain that $\Lambda(x_0) \subset \mathcal{A}$, which together with the opposite inclusion (see Proposition 3.1) imply that $\Lambda(x_0) = \mathcal{A}$. □

From assumptions (i)–(iv) of Theorem 3.2, the last one is the most delicate, of course. To verify this condition one can apply results from the theory of parametric optimization about Lipschitz stability of optimal (nearly optimal) solutions (see, e.g., Bonnans and Shapiro 2000, Chapter 4). It is known that Lipschitz stability of optimal solutions is closely related to the second-order properties of problem (P). We present a second analysis in the next section.

3.1. Second-order conditions. In this section, we discuss second-order optimality conditions ensuring the existence of augmented Lagrange multipliers. We assume throughout this section that the spaces $\mathscr{X}$ and $\mathscr{Y}$ are finite dimensional, and the function $f\colon \mathscr{X} \to \mathbb{R}$ and the mapping $G\colon \mathscr{X} \to \mathscr{Y}$ are twice continuously differentiable.

Let $x_0 \in \mathscr{X}$ be a stationary point of problem (P), i.e., the set $\Lambda(x_0)$ of Lagrange multipliers, satisfying the first-order necessary conditions (3.4), is nonempty. Let $\bar{\lambda} \in \Lambda(x_0)$ and consider the function $\ell(x) := \mathscr{L}(x, \bar{\lambda}, \bar{\tau})$ for some $\bar{\tau} \geq 0$. Recall that

$$\ell(x) = f(x) + \bar{\tau}\delta(G(x) + (2\bar{\tau})^{-1}\bar{\lambda}) - (4\bar{\tau})^{-1}\|\bar{\lambda}\|^2,$$

where $\delta(\cdot)$ is the squared distance function defined in (3.5), and that $D\ell(x_0) = D_x L(x_0, \bar{\lambda}) = 0$ by the first-order necessary conditions (compare with derivations of (3.7)–(3.9)). We say that the *quadratic growth* condition for the function $\ell(x)$, holds at x_0 if there exist a neighborhood $\mathcal{N} \subset \mathscr{X}$ of x_0 and constant $c > 0$ such that

$$\ell(x) \geq \ell(x_0) + c\|x - x_0\|^2 \quad \forall\, x \in \mathcal{N}. \tag{3.19}$$

Definition 3.1. It is said that the set K is second-order regular, at a point $\bar{y} \in K$, if for any $d \in T_K(\bar{y})$ and any sequence $y_k \in K$ of the form $y_k := \bar{y} + t_k d + \frac{1}{2}t_k^2 w_k$, where $t_k \downarrow 0$ and $t_k w_k \to 0$, the following condition holds:

$$\lim_{k\to\infty} \operatorname{dist}(w_k, T_K^2(\bar{y}, d)) = 0. \tag{3.20}$$

The above concept of second-order regularity (for general, not necessarily convex, sets) was developed in Bonnans et al. (1998, 1999; see also Bonnans and Shapiro 2000, §3.3.3) for a discussion of the concept of second-order regularity). The class of second-order regular sets is quite large; it contains polyhedral sets, cones of positive semidefinite matrices, etc.

Suppose that the set K is second-order regular at $\bar{y} \in K$. Then, we have the following result (Bonnans and Shapiro 2000, Theorem 4.133): for any $y, d \in \mathscr{Y}$, such that $\bar{y} := P_K(y)$, it holds that

$$\lim_{\substack{t\downarrow 0\\ d'\to d}} \frac{\delta(y + td') - \delta(y) - tD\delta(y)d'}{\frac{1}{2}t^2} = \nu(d), \tag{3.21}$$

where $\nu(d)$ is the optimal value of the problem

$$\min_{z\in\mathscr{C}(y)} \{2\|d - z\|^2 - 2\sigma(y - \bar{y}, T_K^2(\bar{y}, z))\}, \tag{3.22}$$

and $\mathscr{C}(y) := \{z \in T_K(\bar{y})\colon \langle y - \bar{y}, z\rangle = 0\}$. We refer to the limit in the left-hand side of (3.21) as the *second-order Hadamard directional derivative* of $\delta(\cdot)$ at y in direction d, and denote it by $\delta''(y; d)$. Note that, when it exists, the second-order Hadamard directional derivative is continuous in d. Consequently, $\nu(\cdot)$ is a continuous function.

Because $\delta(\cdot)$ is locally Lipschitz continuous and $P_K(G(x_0) + (2\tau)^{-1}\bar{\lambda}) = G(x_0)$ for $\tau \geq 0$, it follows from (3.21) and (3.6) that for $h' \to h$ and $t \downarrow 0$,

$$\begin{aligned}
&\delta(G(x_0 + th') + (2\bar{\tau})^{-1}\bar{\lambda})\\
&\quad = \delta\big(G(x_0) + tDG(x_0)h' + \tfrac{1}{2}t^2 D^2G(x_0)(h, h) + (2\bar{\tau})^{-1}\bar{\lambda}\big) + o(t^2)
\end{aligned}$$

$$= \delta(G(x_0) + (2\bar{\tau})^{-1}\bar{\lambda}) + tD\delta(G(x_0) + (2\bar{\tau})^{-1}\bar{\lambda})\big(DG(x_0)h' + \tfrac{1}{2}tD^2G(x_0)(h,h)\big)$$
$$+ \tfrac{1}{2}t^2\delta''(G(x_0) + (2\bar{\tau})^{-1}\bar{\lambda}, DG(x_0)h) + o(t^2)$$
$$= (2\bar{\tau})^{-2}\|\bar{\lambda}\|^2 + t(\bar{\tau})^{-1}\big\langle\bar{\lambda}, DG(x_0)h' + \tfrac{1}{2}tD^2G(x_0)(h,h)\big\rangle + \tfrac{1}{2}t^2(\bar{\tau})^{-1}\vartheta_{\bar{\tau}}(h) + o(t^2),$$

where $\vartheta_{\bar{\tau}}(h)$ is the optimal value of the problem

$$\min_{z\in\mathscr{C}(x_0)}\{2\bar{\tau}\|DG(x_0)h - z\|^2 - \sigma(\bar{\lambda}, T_K^2(G(x_0), z)\}, \tag{3.23}$$

with

$$\mathscr{C}(x_0) := \{z \in T_K(G(x_0)) \colon \langle\bar{\lambda}, z\rangle = 0\}. \tag{3.24}$$

It follows from the above, and because $D\ell(x_0) = 0$, that $\ell(\cdot)$ is second-order Hadamard directionally differentiable at x_0, and

$$\ell''(x_0, h) = D_{xx}^2 L(x_0, \bar{\lambda})(h, h) + \vartheta_{\bar{\tau}}(h). \tag{3.25}$$

Note that this implies that $\ell''(x_0, \cdot)$, and hence $\vartheta_{\bar{\tau}}(\cdot)$, are continuous functions.

It is known (see, e.g., Bonnans and Shapiro 2000, §3.3.5), and is not difficult to see directly, that the following second-order optimality conditions follow: (i) If x_0 is a local minimizer of $\ell(\cdot)$, and hence $D\ell(x_0) = 0$, then

$$\ell''(x_0, h) \geq 0 \quad \forall\, h \in \mathscr{X}. \tag{3.26}$$

(ii) If $D\ell(x_0) = 0$, then the quadratic growth condition (3.19) holds iff

$$\ell''(x_0, h) > 0 \quad \forall\, h \in \mathscr{X}\backslash\{0\}. \tag{3.27}$$

Therefore, we obtain the following results.

THEOREM 3.3. *Let $\bar{\lambda} \in \Lambda(x_0)$ and suppose that the set K is second-order regular at $G(x_0)$. Then, the following holds*: (i) *if x_0 is a local minimizer of of $\mathscr{L}(\cdot, \bar{\lambda}, \bar{\tau})$, then necessarily*

$$D_{xx}^2 L(x_0, \bar{\lambda})(h, h) + \vartheta_{\bar{\tau}}(h) \geq 0 \quad \forall\, h \in \mathscr{X}. \tag{3.28}$$

(ii) *The quadratic growth condition* (3.19) *holds if and only if the following condition is satisfied*:

$$D_{xx}^2 L(x_0, \bar{\lambda})(h, h) + \vartheta_{\bar{\tau}}(h) > 0 \quad \forall\, h \in \mathscr{X}\backslash\{0\}, \tag{3.29}$$

where $\vartheta_{\bar{\tau}}(h)$ is the optimal value of problem (3.23).

Because K is convex, we have that the function

$$\phi(\cdot) := -\sigma(\bar{\lambda}, T_K^2(G(x_0), \cdot)) \tag{3.30}$$

is convex (Bonnans and Shapiro 2000, Proposition 3.48) and, because $\bar{\lambda} \in \Lambda(x_0)$, that

$$-\sigma(\bar{\lambda}, T_K^2(G(x_0), z)) \geq 0 \quad \forall\, z \in \mathscr{C}(x_0) \tag{3.31}$$

(e.g., Bonnans and Shapiro 2000). Note also that the function $\phi(\cdot)$, and hence the function $\vartheta_{\bar{\tau}}(\cdot)$, is second-order positively homogeneous, i.e., $\phi(th) = t^2\phi(h)$ for any $t > 0$ and h.

We can compare the second-order conditions (3.28) and (3.29) with the corresponding second-order optimality conditions for problem (P). Consider the critical cone

$$C(x_0) := \{h \in \mathscr{X} \colon DG(x_0)h \in T_K(G(x_0)), \langle\bar{\lambda}, DG(x_0)h\rangle = 0\} \tag{3.32}$$

of problem (P). We have that if $h \in C(x_0)$, then by taking $z := DG(x_0)h$ in (3.23), we obtain that

$$\vartheta_{\bar{\tau}}(h) \le -\sigma(\bar{\lambda}, T_K^2(G(x_0), DG(x_0)h)). \tag{3.33}$$

Therefore, condition (3.29) implies the following condition:

$$D_{xx}^2 L(x_0, \bar{\lambda})(h, h) - \sigma(\bar{\lambda}, T_K^2(G(x_0), DG(x_0)h)) > 0 \quad \forall\, h \in C(x_0)\backslash\{0\}. \tag{3.34}$$

Because $\bar{\lambda} \in \Lambda(x_0)$ and because of the second-order regularity of K, condition (3.34) is sufficient (but, in general, is not necessary) for the quadratic growth condition for problem (P) at the point x_0 (cf., Bonnans and Shapiro 2000, Theorem 3.86).

It also follows by (3.33) that condition (3.28) implies that

$$D_{xx}^2 L(x_0, \bar{\lambda})(h, h) - \sigma(\bar{\lambda}, T_K^2(G(x_0), DG(x_0)h)) \ge 0 \quad \forall\, h \in C(x_0). \tag{3.35}$$

Consequently, the above condition (3.35) is necessary for $\bar{\lambda}$ to be an augmented Lagrange multiplier.

THEOREM 3.4. *Let x_0 be an optimal solution of problem* (P) *and* $\bar{\lambda} \in \Lambda(x_0)$. *Suppose that the set K is second-order regular at $G(x_0)$. Then, the following holds*: (a) *If $\bar{\lambda}$ is an augmented Lagrange multiplier, then condition* (3.28) *holds for some $\bar{\tau} \ge 0$, and hence condition* (3.35) *follows.* (b) *Conversely, suppose* (*in addition*) *that*: (i) *condition* (R) *holds,* (ii) *the second-order condition* (3.34) *is satisfied,* (iii) *for all y in a neighborhood of* $0 \in \mathscr{Y}$ *problem (P_y) has an optimal solution $\bar{x}(y)$ converging to x_0 as $y \to 0$, and* (iv) *the function $\phi(\cdot)$, defined in* (3.30), *is lower semicontinuous on the set $\mathscr{C}(x_0)$. Then, $\bar{\lambda}$ is an augmented Lagrange multiplier.*

PROOF. Suppose that $\bar{\lambda}$ is an augmented Lagrange multiplier, i.e., $\bar{\lambda} \in \partial v_{\bar{\tau}}(0)$ for some $\bar{\tau} \ge 0$. Then, $(x_0, \bar{\lambda})$ is a saddle point of $\mathscr{L}(\cdot, \cdot, \bar{\tau})$ (see Theorem 2.1), and hence x_0 is a minimizer of $\mathscr{L}(\cdot, \bar{\lambda}, \bar{\tau})$. Consequently, the necessary condition (3.28) follows. Because of (3.33), condition (3.35) is implied by condition (3.28). This completes the proof of property (a).

To prove (b), we need to show that $(x_0, \bar{\lambda})$ is a saddle point of $\mathscr{L}(\cdot, \cdot, \tau)$ for some $\tau \ge 0$. We have that for any $\tau \ge 0$, the function $\mathscr{L}(x_0, \cdot, \tau)$ is concave and $D_\lambda \mathscr{L}(x_0, \bar{\lambda}, \tau) = 0$. It follows that $\mathscr{L}(x_0, \lambda, \tau) \le \mathscr{L}(x_0, \bar{\lambda}, \tau)$ for any $\lambda \in \mathscr{Y}$. Now by (2.6) and Lemma 2.2 we have, for some $\varepsilon > 0$ and $\tau \ge 0$ large enough,

$$\inf_{x \in \mathscr{X}} \mathscr{L}(x, \bar{\lambda}, \tau) = \inf_{y \in B_\varepsilon} \{f(\bar{x}(y)) + \tau \|y\|^2 - \langle \bar{\lambda}, y \rangle\}.$$

Consequently, it suffices to verify that $\mathscr{L}(x, \bar{\lambda}, \tau) \ge \mathscr{L}(x_0, \bar{\lambda}, \tau)$ only for all x in a neighborhood of x_0. Therefore, it only remains to show that condition (3.34) implies condition (3.29) for some $\bar{\tau} \ge 0$.

Note that because of the second-order regularity of the set K, we have here that the second-order tangent set $T_K^2(G(x_0), z)$ is nonempty for all $z \in \mathscr{C}(x_0)$. Together with (3.31) this implies that the function $\phi(z)$, defined in (3.30), is finite valued for all $z \in \mathscr{C}(x_0)$. We argue now by a contradiction. Suppose that condition (3.29) does not hold. This means that there exist sequences $\tau_k \to +\infty$ and $h_k \in \mathscr{X}\backslash\{0\}$ such that

$$D_{xx}^2 L(x_0, \bar{\lambda})(h_k, h_k) + \vartheta_{\tau_k}(h_k) \le 0. \tag{3.36}$$

Moreover, because the function in the left-hand side of (3.36) is second-order positively homogeneous, with respect to h, we can normalize h_k such that $\|h_k\| = 1$. Let z_k be the optimal solution of problem (3.23) associated with h_k (such a solution exists and is unique because the objective function in (3.23) is strongly convex and the feasible set is nonempty and convex), and hence

$$\vartheta_{\tau_k}(h_k) = 2\tau_k \|DG(x_0)h_k - z_k\|^2 + \phi(z_k).$$

Because $\tau_k \to +\infty$ and the first term in the left-hand side of (3.36) is bounded and $\phi(z_k)$ is nonnegative, it follows that

$$\lim_{k\to\infty} \|DG(x_0)h_k - z_k\| = 0. \tag{3.37}$$

Also, because $\|h_k\| = 1$, the sequence $\{z_k\}$ is bounded. Therefore, by passing to a subsequence if necessary, we can assume that $\{h_k\}$ and $\{z_k\}$ converge to points $\bar{h}$ and $\bar{z}$, respectively. Note that $\|\bar{h}\| = 1$, and hence $\bar{h} \neq 0$. Because $z_k \in \mathscr{C}(x_0)$ and the set $\mathscr{C}(x_0)$ is closed, it follows that $\bar{z} \in \mathscr{C}(x_0)$. Moreover, we have by (3.37) that $DG(x_0)\bar{h} = \bar{z}$, and hence $\bar{h} \in C(x_0)$. Because of the lower semicontinuity of $\phi(\cdot)$ and because $\vartheta_{\tau_k}(h_k) \geq \phi(z_k)$, it follows from (3.36) that

$$D^2_{xx}L(x_0, \bar{\lambda})(\bar{h}, \bar{h}) + \phi(\bar{z}) \leq 0. \tag{3.38}$$

It remains to note that because $\bar{z} = DG(x_0)\bar{h}$ and $\bar{h} \in C(x_0)\backslash\{0\}$, inequality (3.38) contradicts (3.34). □

Let us make the following remarks. Condition (iii) in Theorem 3.4 holds, for example, if the optimal solution x_0 is unique, Robinson's constraint qualification, at the point x_0, is satisfied and the so-called inf-compactness condition for problem (P) holds (see, e.g., Bonnans and Shapiro 2000, pp. 263–264).

If the set K is *polyhedral*, then it is second-order regular and the sigma term in (3.23) vanishes, i.e., $\phi(z) = 0$ for all $z \in \mathscr{C}(x_0)$. In that case, for $\bar{\tau}$ large enough, condition (3.29) is equivalent to the condition

$$D^2_{xx}L(x_0, \bar{\lambda})(h, h) > 0 \quad \forall\, h \in C(x_0)\backslash\{0\}. \tag{3.39}$$

If the set K is *cone reducible* (see Bonnans and Shapiro 2000, §3.4.4, for a discussion of the concept of cone reducibility), then K is second-order regular and the function (sigma term) $\phi(\cdot)$ is quadratic (and hence continuous) on $\mathscr{C}(x_0)$. A nontrivial example of a cone reducible set is (for any $p \in \mathbb{N}$) the cone $K := S^p_+$ of $p \times p$ positive semidefinite symmetric matrices. For $K := S^p_+$, the sigma term is quadratic and can be written explicitly.

It seems that in all interesting examples the function $\phi(\cdot)$ is lower semicontinuous (assumption (iv) of Theorem 3.4). In particular, this holds true if the set K is cone reducible. It is not known, however, whether $\phi(\cdot)$ is lower semicontinuous for a general convex set K.

4. Stability of augmented solutions. In this section, we discuss stability of minimizers of the augmented Lagrangian under small perturbations of the corresponding augmented Lagrange multipliers. We assume in this section that the spaces $\mathscr{X}$ and $\mathscr{Y}$ are *finite dimensional*, the function $f(x)$ is *lower semicontinuous*, and the mapping $G(x)$ is *continuous*.

Suppose that problem (P) has nonempty set S^* of optimal solutions. Let $\bar{\lambda}$ be an augmented Lagrange multiplier, i.e., $\bar{\lambda} \in \partial v_{\bar{\tau}}(0)$ for some $\bar{\tau} \geq 0$. It follows then that $S^* \subset \arg\min_{x\in\mathscr{X}} \mathscr{L}(x, \bar{\lambda}, \bar{\tau})$. For a given $\lambda \in \mathscr{Y}$, consider the problem

$$\min_{x\in\mathscr{X}} \mathscr{L}(x, \lambda, \bar{\tau}). \tag{4.1}$$

Now we study the question: What happens with the set of optimal solutions $S(\lambda) := \arg\min_{x\in\mathscr{X}} \mathscr{L}(x, \lambda, \bar{\tau})$ of the above problem under small perturbations of λ in a neighborhood of $\bar{\lambda}$?

It is said that the *inf-compactness* condition holds for problem (4.1), if there exist $\alpha \in \mathbb{R}$ and a bounded set $C \subset \mathscr{X}$ such that for every λ in a neighborhood of $\bar{\lambda}$ the level set $\{x \in \mathscr{X}: \mathscr{L}(x, \lambda, \bar{\tau}) \leq \alpha\}$ is nonempty and contained in C. The inf-compactness condition holds, for example, if the domain of f is bounded.

By the standard theory of sensitivity analysis we have that if the inf-compactness condition holds, then for all λ in a neighborhood of $\bar{\lambda}$, the set $S(\lambda)$ is nonempty and

$$\lim_{\lambda \to \bar{\lambda}} \left[\sup_{x \in S(\lambda)} \operatorname{dist}(x, S(\bar{\lambda})) \right] = 0. \tag{4.2}$$

Suppose that $S^* = \{x_0\}$, i.e., problem (P) has unique optimal solution x_0. We have then that $x_0 \in \arg\min_{x \in \mathscr{X}} \mathscr{L}(x, \bar{\lambda}, \bar{\tau})$. Suppose, further, that the quadratic growth condition (3.19) holds. It follows that if $\hat{x}(\lambda) \in S(\lambda) \cap \mathscr{N}$, then

$$\|\hat{x}(\lambda) - x_0\| \leq c^{-1} \kappa, \tag{4.3}$$

where $\kappa = \kappa(\lambda)$ is the Lipschitz constant, on the set $\mathscr{N}$, of the function

$$\psi_\lambda(\cdot) := \mathscr{L}(\cdot, \lambda, \bar{\tau}) - \mathscr{L}(\cdot, \bar{\lambda}, \bar{\tau})$$

(Bonnans and Shapiro 2000, Proposition 4.32). Now suppose that the mapping $G: \mathscr{X} \to \mathscr{Y}$ is continuously differentiable. We have then by (3.6) and the chain rule of differentiation that the function $\psi_\lambda(\cdot)$ is differentiable and

$$\nabla \psi_\lambda(x) = 2\bar{\tau} \nabla G(x)[(2\bar{\tau})^{-1}(\lambda - \bar{\lambda}) - P_K(G(x) + (2\bar{\tau})^{-1}\lambda) + P_K(G(x) + (2\bar{\tau})^{-1}\bar{\lambda})].$$

It follows that if $G(\cdot)$ is continuously differentiable, then the Lipschitz constant of $\psi_\lambda(\cdot)$, on a neighborhood of x_0, is of order $O(\|\lambda - \bar{\lambda}\|)$. We obtain the following result.

THEOREM 4.1. *Suppose that the inf-compactness condition holds. Then, for all* λ *in a neighborhood of* $\bar{\lambda}$, *the set* $S(\lambda)$ *is nonempty and* (4.2) *holds. Suppose, further, that* $S^* = \{x_0\}$, $G(\cdot)$ *is continuously differentiable,* $\bar{\lambda} \in \partial v_{\bar{\tau}}(0)$ *for some* $\bar{\tau} \geq 0$, *and the quadratic growth condition* (3.19) *is satisfied. Then, the following holds*:

$$\sup_{x \in S(\lambda) \cap \mathscr{N}} \|x - x_0\| = O(\|\lambda - \bar{\lambda}\|). \tag{4.4}$$

If, in addition to the assumptions of the above theorem, we assume that x_0 is the unique minimizer of $\mathscr{L}(\cdot, \bar{\lambda}, \bar{\tau})$, then the neighborhood $\mathscr{N}$ in formula (4.4) can be removed.

Acknowledgments. The authors are indebted to an anonymous referee for constructive comments that helped to improve the derivation of Theorem 3.4.

References

Bonnans, J. F., A. Shapiro. 2000. *Perturbation Analysis of Optimization Problems.* Springer, New York.

Bonnans, J. F., R. Cominetti, A. Shapiro. 1998. Sensitivity analysis of optimization problems under second order regular constraints. *Math. Oper. Res.* **23** 806–831.

Bonnans, J. F., R. Cominetti, A. Shapiro. 1999. Second order optimality conditions based on parabolic second order tangent sets. *SIAM J. Optim.* **9** 466–492.

Bonnans, J. F., J. C. Gilbert, C. Lemaréchal, C. A. Sagastizábal. 2003. *Numerical Optimization.* Springer, New York.

Buys, J. D. 1972. Dual algorithms for constrained optimization problems. Ph.D. thesis, University of Leiden, Leiden, The Netherlands.

Hestenes, M. R. 1969. Multiplier and gradient methods, *J. Optim. Theory Appl.* **4** 303–320.

Powell, M. J. D. 1969. A method for nonlinear constraints in minimization problems. R. Fletcher, ed. *Optimization.* Academic Press, London, U.K.

Rockafellar, R. T. 1974a. Augmented Lagrange multiplier functions and duality in convex programming. *SIAM J. Control* **12** 268–285.

Rockafellar, R. T. 1974b. *Conjugate Duality and Optimization*, *Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, PA.

Rockafellar, R. T. 1993. Lagrange multipliers and optimality, *SIAM Rev.* **35** 183–238.

Rockafellar, R. T., R. J-B. Wets. 1998. *Variational Analysis.* Springer-Verlag, New York, 519, 521, 524.