EXACT REGULARIZATION OF CONVEX PROGRAMS

MICHAEL P. FRIEDLANDER* AND PAUL TSENG†

November 18, 2006

Abstract. The regularization of a convex program is *exact* if all solutions of the regularized problem are also solutions of the original problem for all values of the regularization parameter below some positive threshold. For a general convex program, we show that the regularization is exact if and only if a certain selection problem has a Lagrange multiplier. Moreover, the regularization parameter threshold is inversely related to the Lagrange multiplier. We use this result to generalize an exact regularization result of Ferris and Mangasarian (1991) involving a linearized selection problem. We also use it to derive necessary and sufficient conditions for exact penalization, similar to those obtained by Bertsekas (1975) and by Bertsekas, Nedić, and Ozdaglar (2003). When the regularization is not exact, we derive error bounds on the distance from the regularized solution to the original solution set. We also show that existence of a "weak sharp minimum" is in some sense close to being necessary for exact regularization. We illustrate the main result with numerical experiments on the ℓ_1 regularization of benchmark (degenerate) linear programs and semidefinite/second-order cone programs. The experiments demonstrate the usefulness of ℓ_1 regularization in finding sparse solutions.

Key words. convex program, conic program, linear program, regularization, exact penalization, Lagrange multiplier, degeneracy, sparse solutions, interior-point algorithms

1. Introduction.

A common approach to solving an ill-posed problem—one whose solution is not unique or is acutely sensitive to data perturbations—is to construct a related problem whose solution is well behaved and deviates only slightly from a solution of the original problem. This is known as regularization, and deviations from solutions of the original problem are generally accepted as a trade-off for obtaining solutions with other desirable properties. However, it would be more desirable if solutions of the regularized problem are also solutions of the original problem. We study necessary and sufficient conditions for this to hold, and their implications for general convex programs.

Consider the general convex program

$$\text{(P)} \qquad \underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad x \in \mathcal{C},$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, and $\mathcal{C} \subseteq \mathbb{R}^n$ is a nonempty closed convex set. In cases where (P) is ill-posed or lacks a smooth dual, a popular technique is to regularize the problem by adding a convex function to the objective. This yields the regularized problem

$$(\text{P}_\delta) \qquad \underset{x}{\text{minimize}} \quad f(x) + \delta\phi(x) \quad \text{subject to} \quad x \in \mathcal{C},$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ is a convex function and δ is a nonnegative regularization parameter. The regularization function ϕ may be nonlinear and/or nondifferentiable.

*Department of Computer Science, University of British Columbia, Vancouver V6T 1Z4, BC, Canada (mpf@cs.ubc.ca). Research supported by the National Science and Engineering Council of Canada.

†Department of Mathematics, University of Washington, Seattle, Washington 98195, USA (tseng@math.washington.edu). Research supported by National Science Foundation Grant DMS-0511283.

In general, solutions of the regularized problem (P_δ) need not be solutions of (P). (Here and throughout, "solution" is used in lieu of "optimal solution.") We say that the regularization is *exact* if the solutions of (P_δ) are also solutions of (P) for all values of δ below some positive threshold value $\bar{\delta}$. We choose the term *exact* to draw an analogy with exact penalization that is commonly used for solving constrained nonlinear programs. An exact penalty formulation of a problem can recover a solution of the original problem for all values of the penalty parameter beyond a threshold value. See, for example, [4, 5, 9, 21, 24, 31] and, for more recent discussions, [7, 15].

Exact regularization can be useful for various reasons. If a convex program does not have a unique solution, exact regularization may be used to select solutions with desirable properties. In particular, Tikhonov regularization [45], which corresponds to $\phi(x) = \|x\|_2^2$, can be used to select a least two-norm solution. Specialized algorithms for computing least two-norm solutions of linear programs (LPs) have been proposed by [25, 26, 27, 30, 33, 48], among others. Saunders [42] and Altman and Gondzio [1] use Tikhonov regularization as a tool for influencing the conditioning of the underlying linear systems that arise in the implementation of large-scale interior-point algorithms for LPs. Bertsekas [4, Proposition 4] and Mangasarian [30] use Tikhonov regularization to form a smooth convex approximation of the dual LP.

More recently, there has been much interest in ℓ_1 regularization, which corresponds to $\phi(x) = \|x\|_1$. Recent work related to signal processing has focused on using LPs to obtain sparse solutions (i.e., solutions with few nonzero components) of underdetermined systems of linear equations $Ax = b$ (with the possible additional condition $x \geq 0$); for examples, see [13, 12, 14, 18]. In machine learning and statistics, ℓ_1 regularization of linear least-squares problems (sometimes called lasso regression) plays a prominent role as an alternative to Tikhonov regularization; for examples, see [19, 44]. Further extensions to regression and maximum likelihood estimation are studied in [2, 41], among others.

There have been some studies of exact regularization for the case of differentiable ϕ, mainly for LP [4, 30, 34], but to our knowledge there has been only one study, by Ferris and Mangasarian [20], for the case of nondifferentiable ϕ. However, their analysis is mainly for the case of strongly convex ϕ and thus is not applicable to regularization functions such as the one-norm. In this paper, we study exact regularization of the convex program (P) by (P_δ) for a general convex ϕ.

Central to our analysis is a related convex program that selects solutions of (P) of least ϕ-value:

$$(\mathrm{P}^\phi) \qquad \underset{x}{\text{minimize}} \quad \phi(x) \quad \text{subject to} \quad x \in \mathcal{C}, \quad f(x) \leq p^*,$$

where p^* denotes the optimal value of (P). We assume a nonempty solution set of (P), which we denote by $\mathcal{S}$, so that p^* is finite and (P^ϕ) is feasible. Clearly, any solution of (P^ϕ) is also a solution of (P). The converse, however, does not generally hold. In §2 we prove our main result: the regularization (P_δ) is exact if and only if the selection problem (P^ϕ) has a Lagrange multiplier μ^*. Moreover, the solution set of (P_δ) coincides with the solution set of (P^ϕ) for all $\delta < 1/\mu^*$; see Theorem 2.1 and Corollary 2.2.

A particular case of special interest is conic programs, which correspond to

$$f(x) = c^T x \qquad \text{and} \qquad \mathcal{C} = \{x \in \mathcal{K} \mid Ax = b\}, \tag{1.1}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and $\mathcal{K} \subseteq \mathbb{R}^n$ is a nonempty closed convex cone. In the further case where $\mathcal{K}$ is polyhedral, (P$^\phi$) always has a Lagrange multiplier. Thus we extend a result obtained by Mangasarian and Meyer for LPs [34, Theorem 1]; their (weaker) result additionally assumes differentiability (but not convexity) of ϕ on $\mathcal{S}$, and proves that existence of a Lagrange multiplier for (P$^\phi$) implies existence of a common solution x^* of (P) and (P$_\delta$) for all positive δ below some threshold. In general, however, (P$^\phi$) need not have a Lagrange multiplier even if $\mathcal{C}$ has nonempty interior. This is because the additional constraint $f(x) = c^Tx \le p^*$ may exclude points in the interior of $\mathcal{C}$. We discuss this further in §2.

1.1. Applications. We present four applications of our main result. The first three show how to extend existing results in convex optimization. The fourth shows how exact regularization can be used in practice.

Linearized selection (§3). In the case where f is differentiable, $\mathcal{C}$ is polyhedral, and ϕ is strongly convex, Ferris and Mangasarian [20, Theorem 9] show that the regularization (P$_\delta$) is exact if and only if the solution of (P$^\phi$) is unchanged when f is replaced by its linearization at any $\bar{x} \in \mathcal{S}$. We generalize this result by relaxing the strong convexity assumption on ϕ; see Theorem 3.2.

Exact penalization (§4). We show a close connection between exact regularization and exact penalization by applying our main results to obtain necessary and sufficient conditions for exact penalization of convex programs. The resulting conditions are similar to those obtained by Bertsekas [4, Proposition 1], Mangasarian [31, Theorem 2.1], and Bertsekas, Nedić, and Ozdaglar [7, §7.3]; see Theorem 4.2.

Error bounds (§5). We show that in the case where f is continuously differentiable, $\mathcal{C}$ is polyhedral, and $\mathcal{S}$ is bounded, a necessary condition for exact regularization with any ϕ is that f have a "weak sharp minimum" [10, 11] over $\mathcal{C}$. In the case where the regularization is not exact, we derive error bounds on the distance from each solution of the regularized problem (P$_\delta$) to $\mathcal{S}$ in terms of δ and the growth rate of f on $\mathcal{C}$ away from $\mathcal{S}$.

Sparse solutions (§6). As an illustration of our main result, we apply exact ℓ_1 regularization to select sparse solutions of conic programs. In §6.1 we report numerical results on a set of benchmark LPs from the Netlib [36] test set and on a set of randomly generated LPs with prescribed dual degeneracy (i.e., nonunique primal solutions). Analogous results are reported in §6.2 for a set of benchmark semidefinite programs (SDPs) and second-order cone programs (SOCPs) from the DIMACS test set [37]. The numerical results highlight the effectiveness of this approach for inducing sparsity in the solutions obtained via an interior-point algorithm.

1.2. Assumptions. The following assumptions hold implicitly throughout.

ASSUMPTION 1.1 (Feasibility and finiteness). *The feasible set $\mathcal{C}$ is nonempty and the solution set $\mathcal{S}$ of* (P) *is nonempty.*

ASSUMPTION 1.2 (Bounded level sets). *The level set $\{x \in \mathcal{S} \mid \phi(x) \le \beta\}$ is bounded for each $\beta \in \mathbb{R}$, and* $\inf_{x\in\mathcal{C}} \phi(x) > -\infty$. (For example, this assumption holds when ϕ is coercive.)

Assumption 1.1 implies that the optimal value p^* of (P) is finite. Assumptions 1.1 and 1.2 together ensure that the solution set of (P$^\phi$), denoted by $\mathcal{S}^\phi$, is nonempty and compact, and that the solution set of (P$_\delta$), denoted by $\mathcal{S}_\delta$, is nonempty and compact for all $\delta > 0$. The latter is true because, for any $\delta > 0$ and $\beta \in \mathbb{R}$, any point x in the level set $\{x \in \mathcal{C} \mid f(x) + \delta\phi(x) \le \beta\}$ satisfies $f(x) \ge p^*$ and $\phi(x) \ge \inf_{x'\in\mathcal{C}} \phi(x')$, so that $\phi(x) \le (\beta - p^*)/\delta$ and $f(x) \le \beta - \delta \inf_{x'\in\mathcal{C}} \phi(x')$. Assumptions 1.1 and 1.2 then imply that ϕ, f, and $\mathcal{C}$ have no nonzero recession direction in common, so the above level set must be bounded [40, Theorem 8.7].

Our results can be extended accordingly if the above assumptions are relaxed to the assumption that $\mathcal{S}^\phi \neq \emptyset$ and $\mathcal{S}_\delta \neq \emptyset$ for all $\delta > 0$ below some positive threshold.

2. Main results. Ferris and Mangasarian [20, Theorem 7] prove that if the objective function f is linear, then

$$\bigcap_{0<\delta<\bar{\delta}} \mathcal{S}_\delta \subseteq \mathcal{S}^\phi \tag{2.1}$$

for any $\bar{\delta} > 0$. However, an additional constraint qualification on $\mathcal{C}$ is needed to ensure that the set on the left-hand side of (2.1) is nonempty (see [20, Theorem 8]). The following example shows that the set can be empty:

$$\underset{x}{\text{minimize}} \quad x_3 \quad \text{subject to} \quad x \in \mathcal{K}, \tag{2.2}$$

where $\mathcal{K} = \{(x_1, x_2, x_3) \mid x_1^2 \le x_2 x_3,\ x_2 \ge 0,\ x_3 \ge 0\}$, i.e., $\mathcal{K}$ defines the cone of 2×2 symmetric positive semidefinite matrices. Clearly $\mathcal{K}$ has a nonempty interior, and the solutions have the form $x_1^* = x_3^* = 0$, $x_2^* \ge 0$, with $p^* = 0$. Suppose that the convex regularization function ϕ is

$$\phi(x) = |x_1 - 1| + |x_2 - 1| + |x_3|. \tag{2.3}$$

(Note that ϕ is coercive, but not strictly convex.) Then (P^ϕ) has the singleton solution set $\mathcal{S}^\phi = \{(0,1,0)\}$. However, for any $\delta > 0$, (P_δ) has the unique solution

$$x_1 = \frac{1}{2(1+\delta^{-1})}, \quad x_2 = 1, \quad x_3 = \frac{1}{4(1+\delta^{-1})^2},$$

which converges to the solution of (P^ϕ) as $\delta \to 0$, but is never equal to it. Therefore $\mathcal{S}_\delta$ differs from $\mathcal{S}^\phi$ for all $\delta > 0$ sufficiently small.

Note that the left-hand side of (2.1) can be empty even when ϕ is strongly convex and infinitely differentiable. As an example, consider the strongly convex quadratic regularization function

$$\phi(x) = |x_1 - 1|^2 + |x_2 - 1|^2 + |x_3|^2.$$

As with (2.3), it can be shown in this case that $\mathcal{S}_\delta$ differs from $\mathcal{S}^\phi = \{(0,1,0)\}$ for all $\delta > 0$ sufficiently small. In particular, $(\delta/2, 1, \delta^2/4)$ is feasible for (P_δ), and its objective function value is strictly less than that of $(0,1,0)$. Thus the latter cannot be a solution of (P_δ) for any $\delta > 0$.

In general, one can show that as $\delta \to 0$, each cluster point of solutions of (P_δ) belongs to $\mathcal{S}^\phi$. Moreover, there is no duality gap between (P^ϕ) and its dual because $\mathcal{S}^\phi$ is compact (see [40, Theorem 30.4(i)]). However, the supremum in the dual problem might not be attained, in which case there would be no Lagrange multiplier for (P^ϕ)—and hence no exact regularization property. Thus additional constraint qualifications are needed when f is not affine or $\mathcal{C}$ is not polyhedral.

The following theorem and corollary are our main results. They show that the regularization (P_δ) is exact if and only if the selection problem (P^ϕ) has a Lagrange multiplier μ^*. Moreover, $\mathcal{S}_\delta = \mathcal{S}^\phi$ for all $\delta < 1/\mu^*$. Parts of our proof bear similarity to the arguments used by Mangasarian and Meyer [34, Theorem 1], who consider the two cases $\mu^* = 0$ and $\mu^* > 0$ separately in proving the "if" direction. However, instead of working with the KKT conditions for (P) and (P_δ), we work with saddle-point conditions.

THEOREM 2.1.
(a) *For any* $\delta > 0$, $\mathcal{S} \cap \mathcal{S}_\delta \subseteq \mathcal{S}^\phi$.
(b) *If there exists a Lagrange multiplier* μ^* *for* (P^ϕ)*, then* $\mathcal{S} \cap \mathcal{S}_\delta = \mathcal{S}^\phi$ *for all* $\delta \in (0, 1/\mu^*]$.
(c) *If there exists* $\bar{\delta} > 0$ *such that* $\mathcal{S} \cap \mathcal{S}_{\bar{\delta}} \neq \emptyset$*, then* $1/\bar{\delta}$ *is a Lagrange multiplier for* (P^ϕ)*, and* $\mathcal{S} \cap \mathcal{S}_\delta = \mathcal{S}^\phi$ *for all* $\delta \in (0, \bar{\delta}]$.
(d) *If there exists* $\bar{\delta} > 0$ *such that* $\mathcal{S} \cap \mathcal{S}_{\bar{\delta}} \neq \emptyset$*, then* $\mathcal{S}_\delta \subseteq \mathcal{S}$ *for all* $\delta \in (0, \bar{\delta})$.

Proof.
Part (a). Consider any $x^* \in \mathcal{S} \cap \mathcal{S}_\delta$. Then, because $x^* \in \mathcal{S}_\delta$,

$$f(x^*) + \delta\phi(x^*) \le f(x) + \delta\phi(x) \quad \text{for all} \quad x \in \mathcal{C}.$$

Also, $x^* \in \mathcal{S}$, so $f(x) = f(x^*) = p^*$ for all $x \in \mathcal{S}$. This implies that

$$\phi(x^*) \le \phi(x) \quad \text{for all} \quad x \in \mathcal{S}.$$

Thus $x^* \in \mathcal{S}^\phi$, and it follows that $\mathcal{S} \cap \mathcal{S}_\delta \subseteq \mathcal{S}^\phi$.

Part (b). Assume that there exists a Lagrange multiplier μ^* for (P^ϕ). We consider the two cases $\mu^* = 0$ and $\mu^* > 0$ in turn.

First, suppose that $\mu^* = 0$. Then, for any solution x^* of (P^ϕ),

$$x^* \in \operatorname*{arg\,min}_{x \in \mathcal{C}} \ \phi(x),$$

or, equivalently,

$$\phi(x^*) \le \phi(x) \quad \text{for all} \quad x \in \mathcal{C}. \tag{2.4}$$

Also, x^* is feasible for (P^ϕ), so $x^* \in \mathcal{S}$. Thus

$$f(x^*) \le f(x) \quad \text{for all} \quad x \in \mathcal{C}.$$

Multiplying the inequality in (2.4) by $\delta \ge 0$ and adding it to the above inequality yields

$$f(x^*) + \delta\phi(x^*) \le f(x) + \delta\phi(x) \quad \text{for all} \quad x \in \mathcal{C}.$$

Thus $x^* \in \mathcal{S}_\delta$ for all $\delta \in [0, \infty)$.

Second, suppose that $\mu^* > 0$. Then, for any solution x^* of (P^ϕ),

$$x^* \in \operatorname*{arg\,min}_{x \in \mathcal{C}} \ \phi(x) + \mu^*(f(x) - p^*),$$

or, equivalently,

$$x^* \in \operatorname*{arg\,min}_{x \in \mathcal{C}} \ f(x) + \frac{1}{\mu^*}\phi(x).$$

Thus

$$f(x^*) + \frac{1}{\mu^*}\phi(x^*) \le f(x) + \frac{1}{\mu^*}\phi(x) \quad \text{for all} \quad x \in \mathcal{C}.$$

Also, x^* is feasible for (P^ϕ), so that $x^* \in \mathcal{S}$. Therefore

$$f(x^*) \le f(x) \quad \text{for all} \quad x \in \mathcal{C}.$$

Then, for any $\lambda \in [0,1]$, multiplying the above two inequalities by λ and $1-\lambda$, respectively, and summing them yields

$$f(x^*) + \frac{\lambda}{\mu^*}\phi(x^*) \le f(x) + \frac{\lambda}{\mu^*}\phi(x) \quad \text{for all} \quad x \in \mathcal{C}.$$

Thus $x^* \in \mathcal{S}_\delta$ for all $\delta \in [0, 1/\mu^*]$.

The above arguments show that $\mathcal{S}^\phi \subseteq \mathcal{S}_\delta$ for all $\delta \in [0, 1/\mu^*]$, and therefore $\mathcal{S}^\phi \subseteq \mathcal{S} \cap \mathcal{S}_\delta$ for all $\delta \in (0, 1/\mu^*]$. By part (a) of the theorem, we must have $\mathcal{S}^\phi = \mathcal{S} \cap \mathcal{S}_\delta$ as desired.

Part (c). Assume that there exists $\bar\delta > 0$ such that $\mathcal{S} \cap \mathcal{S}_{\bar\delta} \ne \emptyset$. Then for any $x^* \in \mathcal{S} \cap \mathcal{S}_{\bar\delta}$, we have $x^* \in \mathcal{S}_{\bar\delta}$, and thus

$$x^* \in \operatorname*{arg\,min}_{x \in \mathcal{C}} \; f(x) + \bar\delta\phi(x),$$

or, equivalently,

$$x^* \in \operatorname*{arg\,min}_{x \in \mathcal{C}} \; \phi(x) + \frac{1}{\bar\delta}(f(x) - p^*).$$

By part (a), $x^* \in \mathcal{S}^\phi$. This implies that any $x \in \mathcal{S}^\phi$ attains the minimum because $\phi(x) = \phi(x^*)$ and $f(x) = p^*$. Therefore $1/\bar\delta$ is a Lagrange multiplier for (P^ϕ). By part (b), $\mathcal{S} \cap \mathcal{S}_\delta = \mathcal{S}^\phi$ for all $\delta \in (0, \bar\delta]$.

Part (d). To simplify notation, define $f_\delta(x) = f(x) + \delta\phi(x)$. Assume that there exists a $\bar\delta > 0$ such that $\mathcal{S} \cap \mathcal{S}_{\bar\delta} \ne \emptyset$. Fix any $x^* \in \mathcal{S} \cap \mathcal{S}_{\bar\delta}$. For any $\delta \in (0, \bar\delta)$ and any $x \in \mathcal{C} \setminus \mathcal{S}$, we have

$$f_{\bar\delta}(x^*) \le f_{\bar\delta}(x) \quad \text{and} \quad f(x^*) < f(x).$$

Because $0 < \delta/\bar\delta < 1$, this implies that

$$f_\delta(x^*) = \frac{\delta}{\bar\delta} f_{\bar\delta}(x^*) + \left(1 - \frac{\delta}{\bar\delta}\right) f(x^*) < \frac{\delta}{\bar\delta} f_{\bar\delta}(x) + \left(1 - \frac{\delta}{\bar\delta}\right) f(x) = f_\delta(x).$$

Because $x^* \in \mathcal{C}$, this shows that $x \in \mathcal{C} \setminus \mathcal{S}$ cannot be a solution of (P_δ), and so $\mathcal{S}_\delta \subseteq \mathcal{S}$, as desired. □

Theorem 2.1 shows that existence of a Lagrange multiplier μ^* for (P^ϕ) is necessary and sufficient for exact regularization of (P) by (P_δ) for all $0 < \delta < 1/\mu^*$. Coerciveness of ϕ on $\mathcal{S}$ is needed only to ensure that $\mathcal{S}^\phi$ is nonempty. If $\delta = 1/\mu^*$, then $\mathcal{S}_\delta$ need not be a subset of $\mathcal{S}$. For example, suppose that

$$n = 1, \quad \mathcal{C} = [0, \infty), \quad f(x) = x, \quad \text{and} \quad \phi(x) = |x - 1|.$$

Then $\mu^* = 1$ is the only Lagrange multiplier for (P^ϕ), but $\mathcal{S}_1 = [0,1] \not\subseteq \mathcal{S} = \{0\}$. If $\mathcal{S}_\delta$ is a singleton for $\delta \in (0, 1/\mu^*]$, such as when ϕ is strictly convex, then Theorem 2.1(b) and $\mathcal{S}^\phi \ne \emptyset$ imply that $\mathcal{S}_\delta \subseteq \mathcal{S}$.

The following corollary readily follows from Theorem 2.1(b)–(c) and $\mathcal{S}^\phi \neq \emptyset$.

Corollary 2.2.
(a) *If there exists a Lagrange multiplier* μ^* *for* (P^ϕ)*, then* $\mathcal{S}_\delta = \mathcal{S}^\phi$ *for all* $\delta \in (0, 1/\mu^*)$.
(b) *If there exists* $\bar\delta > 0$ *such that* $\mathcal{S}_{\bar\delta} = \mathcal{S}^\phi$*, then* $1/\bar\delta$ *is a Lagrange multiplier for* (P^ϕ)*, and* $\mathcal{S}_\delta = \mathcal{S}^\phi$ *for all* $\delta \in (0, \bar\delta]$.

2.1. Conic programs. Conic programs (CPs) correspond to (P) with f and $\mathcal{C}$ given by (1.1). They include several important problem classes. LPs correspond to $\mathcal{K} = \mathbb{R}^n_+$ (the nonnegative orthant); SOCPs correspond to

$$\mathcal{K} = \mathcal{K}^{\mathrm{soc}}_{n_1} \times \cdots \times \mathcal{K}^{\mathrm{soc}}_{n_K} \qquad \text{with} \qquad \mathcal{K}^{\mathrm{soc}}_{n} := \left\{ x \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n-1} x_i^2 \leq x_n^2,\ x_n \geq 0 \right\}$$

(a product of second-order cones); SDPs correspond to $\mathcal{K} = \mathbb{S}^n_+$ (the cone of symmetric positive semidefinite $n \times n$ real matrices). CPs are discussed in detail in [3, 8, 35, 38], among others.

It is well known that when $\mathcal{K}$ is polyhedral, the selection problem (P^ϕ), with f and $\mathcal{C}$ given by (1.1), must have a Lagrange multiplier [40, Theorem 28.2]. In this important case, Corollary 2.2 immediately yields the following exact-regularization result for polyhedral CPs.

Corollary 2.3. *Suppose that* f *and* $\mathcal{C}$ *have the form given by* (1.1) *and that* $\mathcal{K}$ *is polyhedral. Then there exists a positive* $\bar\delta$ *such that* $\mathcal{S}_\delta = \mathcal{S}^\phi$ *for all* $\delta \in (0, \bar\delta)$.

Corollary 2.3 extends [34, Theorem 1], which additionally assumes differentiability (though not convexity) of ϕ on $\mathcal{S}$ and proves a weaker result that there exists a common solution $x^* \in \mathcal{S} \cap \mathcal{S}_\delta$ for all positive δ below some threshold. If $\mathcal{S}$ is furthermore bounded, then an "excision lemma" of Robinson [39, Lemma 3.5] can be applied to show that $\mathcal{S}_\delta \subseteq \mathcal{S}$ for all positive δ below some threshold. This result is still weaker than Corollary 2.3 however.

2.2. Relaxing the assumptions on the regularization function. The assumption that ϕ is coercive on $\mathcal{S}$ and is bounded from below on $\mathcal{C}$ (Assumption 1.2) ensures that the selection problem (P^ϕ) and the regularized problem (P_δ) have solutions. This assumption is preserved under the introduction of slack variables for linear inequality constraints. For example, if $\mathcal{C} = \{x \in \mathcal{K} \mid Ax \leq b\}$ for some closed convex set $\mathcal{K}$, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$, then

$$\tilde\phi(x, s) = \phi(x) \quad \text{with} \quad \tilde{\mathcal{C}} = \{(x, s) \in \mathcal{K} \times [0, \infty)^m \mid Ax + s = b\}$$

also satisfies Assumption 1.2. Here $\tilde\phi(x, s)$ depends only on x. Can Assumption 1.2 be relaxed?

Suppose that $\phi(x)$ depends only on a subset of coordinates x_J and is coercive with respect to x_J, where $x_J = (x_j)_{j\in J}$ and $J \subseteq \{1, \ldots, n\}$. Using the assumption that (P) has a feasible point x^*, it is readily seen that (P_δ) has a solution with respect to x_J for each $\delta > 0$, i.e., the minimization in (P_δ) is attained at some x_J. For an LP (f linear and $\mathcal{C}$ polyhedral) it can be shown that (P_δ) has a solution with respect to all coordinates of x. However, in general this need not be true, even for an SOCP. An example is

$$n = 3, \quad f(x) = -x_2 + x_3, \quad \mathcal{C} = \{x \mid \sqrt{x_1^2 + x_2^2} \le x_3\}, \quad \text{and} \quad \phi(x) = |x_1 - 1|.$$

Here, $p^* = 0$ (since $\sqrt{x_1^2 + x_2^2} - x_2 \ge 0$ always) and solutions are of the form $(0, \xi, \xi)$ for all $\xi \ge 0$. For any $\delta > 0$, (P_δ) has optimal value of zero (achieved by setting $x_1 = 1$, $x_3 = \sqrt{1 + x_2^2}$, and taking $x_2 \to \infty$) but has no solution. In general, if we define

$$\widehat{f}(x_J) := \min_{(x_j)_{j \notin J} | x \in \mathcal{C}} f(x),$$

then it can be shown, using convex analysis results [40], that $\widehat{f}$ is convex and lower semicontinuous—i.e., the epigraph of $\widehat{f}$ is convex and closed. Then (P_δ) is equivalent to

$$\underset{x_J}{\text{minimize}} \;\; \widehat{f}(x_J) + \delta\phi(x_J),$$

with ϕ viewed as a function of x_J. Thus, we can in some sense reduce this case to the one we currently consider. Note that $\widehat{f}$ may not be real-valued, but this does not pose a problem with the proof of Theorem 2.1.

3. Linearized selection. Ferris and Mangasarian [20] develop a related exact-regularization result for the special case where f is differentiable, $\mathcal{C}$ is polyhedral, and ϕ is strongly convex. They show that (P_δ) is an exact regularization if the solution set of the selection problem (P^ϕ) is unchanged when f is replaced by its linearization at any $\bar{x} \in \mathcal{S}$. In this section we show how Theorem 2.1 and Corollary 2.2 can be applied to generalize this result. We begin with a technical lemma, closely related to some results given by Mangasarian [32].

LEMMA 3.1. *Suppose that f is differentiable on $\mathbb{R}^n$ and is constant on the line segment joining two points x^* and $\bar{x}$ in $\mathbb{R}^n$. Then*

$$\nabla f(x^*)^T (x - x^*) = \nabla f(\bar{x})^T (x - \bar{x}) \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{3.1}$$

Moreover, ∇f is constant on the line segment.

Proof. Because f is convex differentiable and is constant on the line segment joining x^* and $\bar{x}$, $\nabla f(x^*)^T (\bar{x} - x^*) = 0$. Because f is convex,

$$f(y) - f(\bar{x}) \ge f(y) - f(x^*) \ge \nabla f(x^*)^T (y - x^*) \quad \text{for all} \quad y \in \mathbb{R}^n.$$

Fix any $x \in \mathbb{R}^n$. Taking $y = \bar{x} + \alpha(x - \bar{x})$ with $\alpha > 0$ yields

$$f(\bar{x} + \alpha(x - \bar{x})) - f(\bar{x}) \ge \nabla f(x^*)^T (\bar{x} + \alpha(x - \bar{x}) - x^*) = \alpha \nabla f(x^*)^T (x - \bar{x}).$$

Dividing both sides by α and then taking $\alpha \to 0$ yields in the limit

$$\nabla f(\bar{x})^T (x - \bar{x}) \ge \nabla f(x^*)^T (x - \bar{x}) = \nabla f(x^*)^T (x - x^*). \tag{3.2}$$

Switching $\bar{x}$ and x^* in the above argument yields an inequality in the opposite direction. Thus (3.1) holds, as desired.

By taking $x = \alpha(\nabla f(x^*) - \nabla f(\bar{x}))$ in (3.2) (where the inequality is now replaced by equality) and letting $\alpha \to \infty$, we obtain that $\|\nabla f(x^*) - \nabla f(\bar{x})\|_2^2 = 0$ and hence that $\nabla f(x^*) = \nabla f(\bar{x})$. This shows that ∇f is constant on the line segment. ◻

Suppose that f is differentiable at every $\bar{x} \in \mathcal{S}$, and consider a variation of the selection problem (P^ϕ) in which the constraint is linearized about $\bar{x}$:

$$(\mathrm{P}^{\phi,\bar{x}}) \qquad \underset{x}{\text{minimize}} \quad \phi(x) \quad \text{subject to} \quad x \in \mathcal{C}, \quad \nabla f(\bar{x})^T(x - \bar{x}) \le 0.$$

Lemma 3.1 shows that the feasible set of ($\mathrm{P}^{\phi,\bar{x}}$) is the same for all $\bar{x} \in \mathcal{S}$. Since f is convex, the feasible set of ($\mathrm{P}^{\phi,\bar{x}}$) contains $\mathcal{S}$, which is the feasible set of (P^ϕ). Let $\mathcal{S}^{\phi,\bar{x}}$ denote the solution set of ($\mathrm{P}^{\phi,\bar{x}}$). In general $\mathcal{S}^\phi \ne \mathcal{S}^{\phi,\bar{x}}$. In the case where ϕ is strongly convex and $\mathcal{C}$ is polyhedral, Ferris and Mangasarian [20, Theorem 9] show that exact regularization (i.e., $\mathcal{S}^\phi = \mathcal{S}_\delta$ for all $\delta > 0$ sufficiently small) holds if and only if $\mathcal{S}^\phi = \mathcal{S}^{\phi,\bar{x}}$. By using Theorem 2.1, Corollary 2.2, and Lemma 3.1, we can generalize this result by relaxing the assumption that ϕ is strongly convex.

THEOREM 3.2. *Suppose that f is differentiable on $\mathcal{C}$.*
(a) *If there exists a $\bar{\delta} > 0$ such that $\mathcal{S}_{\bar{\delta}} = \mathcal{S}^\phi$, then*

$$\mathcal{S}^\phi \subseteq \mathcal{S}^{\phi,\bar{x}} \quad \textit{for all} \quad \bar{x} \in \mathcal{S}. \tag{3.3}$$

(b) *If $\mathcal{C}$ is polyhedral and* (3.3) *holds, then there exists a $\bar{\delta} > 0$ such that $\mathcal{S}_\delta = \mathcal{S}^\phi$ for all $\delta \in (0, \bar{\delta})$.*

Proof.
Part (a). Suppose that there exists a $\bar{\delta} > 0$ such that $\mathcal{S}_{\bar{\delta}} = \mathcal{S}^\phi$. Then by Corollary 2.2, $\mu^* := 1/\bar{\delta}$ is a Lagrange multiplier for (P^ϕ), and for any $x^* \in \mathcal{S}^\phi$,

$$x^* \in \underset{x\in\mathcal{C}}{\arg\min} \ \phi(x) + \mu^* f(x). \tag{3.4}$$

Because ϕ and f are real-valued and convex, x^* and μ^* satisfy the optimality condition

$$0 \in \partial\phi(x^*) + \mu^*\nabla f(x^*) + N_{\mathcal{C}}(x^*).$$

Then x^* satisfies the KKT condition for the linearized selection problem

$$\underset{x\in\mathcal{C}}{\text{minimize}} \quad \phi(x) \quad \text{subject to} \quad \nabla f(x^*)^T(x - x^*) \le 0, \tag{3.5}$$

and is therefore a solution of this problem. By Lemma 3.1, the feasible set of this problem remains unchanged if we replace $\nabla f(x^*)^T(x - x^*) \le 0$ with $\nabla f(\bar{x})^T(x - \bar{x}) \le 0$ for any $\bar{x} \in \mathcal{S}$. Thus $x^* \in \mathcal{S}^{\phi,\bar{x}}$. The choice of x^* was arbitrary, and so $\mathcal{S}^\phi \subseteq \mathcal{S}^{\phi,\bar{x}}$.

Part (b). Suppose that $\mathcal{C}$ is polyhedral and (3.3) holds. By Lemma 3.1, the solution set of ($\mathrm{P}^{\phi,\bar{x}}$) remains unchanged if we replace $\nabla f(\bar{x})^T(x - \bar{x}) \le 0$ by $\nabla f(x^*)^T(x - x^*) \le 0$ for any $x^* \in \mathcal{S}^\phi$. The resulting problem (3.5) is linearly constrained and therefore has a Lagrange multiplier $\bar{\mu} \in \mathbb{R}$. Moreover, $\bar{\mu}$ is independent of x^*. By Corollary 2.2(a), the problem

$$\underset{x\in\mathcal{C}}{\text{minimize}} \ \phi(x) + \mu^*\nabla f(x^*)^T x$$

has the same solution set as (3.5) for all $\mu^* > \bar{\mu}$. The necessary and sufficient optimality condition for this convex program is

$$0 \in \partial\phi(x) + \mu^*\nabla f(x^*) + N_{\mathcal{C}}(x).$$

Because (3.3) holds, x^* satisfies this optimality condition. Thus (3.4) holds for all $\mu^* > \bar{\mu}$ or, equivalently, $x^* \in \mathcal{S}_\delta$ for all $\delta \in (0, 1/\bar{\mu})$. Because $\bar{\mu}$ is independent of x^*, this shows that $\mathcal{S}^\phi \subseteq \mathcal{S}_\delta$ for all $\delta \in (0, 1/\bar{\mu})$. And because $\emptyset \neq \mathcal{S}^\phi \subseteq \mathcal{S}$, it follows that $\mathcal{S} \cap \mathcal{S}_\delta \neq \emptyset$ for all $\delta \in (0, 1/\bar{\mu})$. By Theorem 2.1(a) and (d), $\mathcal{S}_\delta \subseteq \mathcal{S}^\phi$ for all $\delta \in (0, 1/\bar{\mu})$. Therefore $\mathcal{S}_\delta = \mathcal{S}^\phi$ for all $\delta \in (0, 1/\bar{\mu})$. ☐

In the case where ϕ is strongly convex, $\mathcal{S}^\phi$ and $\mathcal{S}^{\phi,\bar{x}}$ are both singletons, so (3.3) is equivalent to $\mathcal{S}^\phi = \mathcal{S}^{\phi,\bar{x}}$ for all $\bar{x} \in \mathcal{S}$. Thus, when $\mathcal{C}$ is also polyhedral, Theorem 3.2 reduces to [20, Theorem 9]. Note that in Theorem 3.2(b) the polyhedrality of $\mathcal{C}$ is needed only to ensure the existence of a Lagrange multiplier for (3.5), and can be relaxed by assuming an appropriate constraint qualification. In particular, if $\mathcal{C}$ is given by inequality constraints, then it suffices that ($\mathrm{P}^{\phi,\bar{x}}$) has a feasible point that strictly satisfies all nonlinear constraints [40, Theorem 28.2].

Naturally, (3.3) holds if f is linear. Thus Theorem 3.2(b) is false if we drop the polyhedrality assumption on $\mathcal{C}$, as we can find examples of convex coercive ϕ, linear f, and closed convex (but not polyhedral) $\mathcal{C}$ for which exact regularization fails; see example (2.2).

4. Exact penalization. In this section we show a close connection between exact regularization and exact penalization by applying Corollary 2.2 to obtain necessary and sufficient conditions for exact penalization of convex programs. Consider the convex program

$$\underset{x}{\text{minimize}} \quad \phi(x) \quad \text{subject to} \quad x \in \mathcal{C}, \quad g(x) := \big(g_i(x)\big)_{i=1}^m \leq 0, \tag{4.1}$$

where $\phi, g_1, \ldots, g_m$ are real-valued convex functions defined on $\mathbb{R}^n$, and $\mathcal{C} \subseteq \mathbb{R}^n$ is a nonempty closed convex set. The penalized form of (4.1) is

$$\underset{x}{\text{minimize}} \quad \phi(x) + \sigma P(g(x)) \quad \text{subject to} \quad x \in \mathcal{C}, \tag{4.2}$$

where σ is a positive penalty parameter and $P : \mathbb{R}^m \to [0, \infty)$ is a convex function having the property that $P(u) = 0$ if and only if $u \leq 0$; see [7, §7.3]. A well-known example of such a penalty function is

$$P(u) = \|\max\{0, u\}\|_p, \tag{4.3}$$

where $\|\cdot\|_p$ is the p-norm ($1 \leq p \leq \infty$) [22, §14.3].

The conjugate and polar functions of P [40, §§12, 15] are defined, respectively, by

$$P^*(w) := \sup_u \; w^T u - P(u) \qquad \text{and} \qquad P^\circ(w) := \sup_{u \not\leq 0} \frac{w^T u}{P(u)}.$$

Note that $P^\circ(\alpha w) = \alpha P^\circ(w)$ for all $\alpha \geq 0$. For P given by (4.3), $P^\circ(w)$ equals the q-norm of w whenever $w \geq 0$, where $1/p + 1/q = 1$. The following lemma gives key properties of these functions that are implicit in the analysis of [7, §7.3].

Lemma 4.1. *Suppose that $P : \mathbb{R}^m \to [0, \infty)$ is a convex function and $P(u) = 0$ if and only if $u = 0$. Then*
(a) *$P(u) \leq P(v)$ whenever $u \leq v$; and*

(b)

$$P^*(w) \begin{cases} = \infty & \text{if } w \not\geq 0; \\ > 0 & \text{if } w \geq 0 \text{ and } P^\circ(w) > 1; \\ = 0 & \text{if } w \geq 0 \text{ and } P^\circ(w) \leq 1. \end{cases}$$

Proof.
Part (a). Fix any $u, v \in \mathbb{R}^m$ with $u < v$, and define

$$\pi(\alpha) := P(u + \alpha(v - u)) \quad \text{for all} \quad \alpha \in \mathbb{R}.$$

We have $u + \alpha(v - u) < 0$ for all $\alpha < 0$ sufficiently negative, in which case $\pi(\alpha) = 0$. Because π is convex, this implies that π is nondecreasing and hence $\pi(0) \leq \pi(1)$—i.e., $P(u) \leq P(v)$. Thus $P(u) \leq P(v)$ whenever $u < v$. Because P is continuous on $\mathbb{R}^m$ [40, Theorem 10.1], this yields $P(u) \leq P(v)$ whenever $u \leq v$.

Part (b). Fix any $w \in \mathbb{R}^m$. If $w_i < 0$ for some $i \in \{1, \ldots, m\}$, then by letting $u_i \to -\infty$ and setting all other components of u to zero, we obtain $w^T u - P(u) = w_i u_i \to \infty$ and thus $P^*(w) = \infty$. If $w \geq 0$ and $P^\circ(w) > 1$, then $w^T u > P(u)$ for some $u \not\leq 0$ and thus $P^*(w) \geq w^T u - P(u) > 0$. If $w \geq 0$ and $P^\circ(w) \leq 1$, then $w^T u \leq 0 = P(u)$ for all $u \leq 0$, and $w^T u \leq P(u)$ for all $u \not\leq 0$, so that $w^T u \leq P(u)$ for all $u \in \mathbb{R}^m$ (with equality holding when $u = 0$). Therefore $P^*(w) = 0$. □

THEOREM 4.2. *Suppose that* (4.1) *has a nonempty compact solution set. If there exist Lagrange multipliers* y^* *for* (4.1)*, then the penalized problem* (4.2) *has the same solution set as* (4.1) *for all* $\sigma > P^\circ(y^*)$*. Conversely, if* (4.1) *and* (4.2) *have the same solution set for some* $\sigma = \mu^* > 0$*, then* (4.1) *and* (4.2) *have the same solution set for all* $\sigma \geq \mu^*$*, and there exists a Lagrange multiplier vector* y^* *for* (4.1) *with* $\mu^* \geq P^\circ(y^*)$*.*

Proof. Set $f(x) = P\big(g(x)\big)$ for all $x \in \mathbb{R}^n$. By the convexity of $g_1, \ldots, g_m$, P, and Lemma 4.1(a), f is a convex function and thus (4.2) is a convex program. Moreover, any feasible point x^* of (4.1) is a solution of (P) with optimal value $p^* = 0$. Accordingly, we identify (4.2) with (P_δ) (where ϕ is the regularization function and $\delta = 1/\sigma$ is the regularization parameter), and we identify the problem

$$\underset{x}{\text{minimize}} \quad \phi(x) \quad \text{subject to} \quad x \in \mathcal{C}, \quad P\big(g(x)\big) \leq 0 \tag{4.4}$$

with (P^ϕ). Assumptions 1.1 and 1.2 are satisfied because (4.1) has a nonempty compact solution set.

A primal-dual solution pair (x^*, y^*) of (4.1) satisfies the KKT conditions

$$0 \in \partial\phi(x) + \sum_{i=1}^m y_i \partial g_i(x) + N_{\mathcal{C}}(x), \quad y \geq 0, \quad g(x) \leq 0, \quad y^T g(x) = 0. \tag{4.5}$$

By [40, Theorem 23.5], the subdifferential of P at u has the expression $\partial P(u) = \{w \mid w^T u = P(u) + P^*(w)\}$. If $u \leq 0$, then $P(u) = 0$ and, by Lemma 4.1(b), $w^T u = P^*(w)$ only if $w \geq 0$ and $P^\circ(w) \leq 1$. This implies that

$$\partial P(u) = \{w \mid w \geq 0,\ P^\circ(w) \leq 1,\ w^T u = 0\} \quad \text{for all} \quad u \leq 0.$$

We can then express the KKT conditions for (4.4) as

$$0 \in \partial\phi(x) + \mu \sum_{i=1}^{m} w_i \partial g_i(x) + N_{\mathcal{C}}(x), \quad \left\{ \begin{array}{r} w \geq 0 \\ P^\circ(w) \leq 1 \\ \mu \geq 0 \end{array} \right\}, \quad g(x) \leq 0, \quad w^T g(x) = 0. \tag{4.6}$$

Comparing (4.5) and (4.6) and using the positive homogeneous property of P°, we see that they are equivalent in the sense that (x^*, y^*) satisfies (4.5) if and only if (x^*, μ^*) satisfies (4.6), where

$$\mu^* w^* = y^* \quad \text{and} \quad \mu^* = P^\circ(y^*),$$

for some $w^* \geq 0$ with $P^\circ(w^*) \leq 1$. Note that μ^* is a Lagrange multiplier for (4.4). Therefore, by Corollary 2.2(a), (4.2) and (4.4) have the same solution set for all $\sigma > \mu^* = P^\circ(y^*)$.

Conversely, suppose that (4.2) and (4.4) have the same solution set for $\sigma = \mu^* > 0$. Then (P_δ) and (P^ϕ) have the same solution set for $\delta = 1/\mu^*$. By Corollary 2.2(b), μ^* is a Lagrange multiplier for (P^ϕ), and (P_δ) and (P^ϕ) have the same solution set for all $\delta \in (0, 1/\mu^*]$. Therefore, (4.1) and (4.2) have the same solution set for all $\sigma \geq \mu^*$. Moreover, for any $x^* \in \mathcal{S}^\phi$ there exists a vector w^* such that (x^*, μ^*, w^*) satisfies (4.6), and so $y^* := \mu^* w^*$ is a Lagrange multiplier vector for (4.1) that satisfies $P^\circ(y^*) = \mu^* P^\circ(w^*) \leq \mu^*$. □

We can consider a minimum P°-value Lagrange multiplier vector y^* and, similarly, a minimum exact penalty parameter σ. Theorem 4.2 asserts that these two quantities are equal—that is,

$$\left\{ \begin{array}{ll} \inf & P^\circ(y^*) \\ \text{such that} & y^* \in \mathbb{R}^m \text{ is a Lagrange} \\ & \text{multiplier for (4.1)} \end{array} \right\} = \left\{ \begin{array}{ll} \inf & \sigma \\ \text{such that} & \text{(4.2) has the same} \\ & \text{solution set as (4.1)} \end{array} \right\}.$$

Theorem 4.2 shows that the existence of Lagrange multipliers y^* with $P^\circ(y^*) < \infty$ is necessary and sufficient for exact penalization. There has been much study of sufficient conditions for exact penalization; see, e.g., [4], [5, Proposition 4.1], and [9]. The results in [4, Propositions 1 and 2] assume the existence of Lagrange multipliers y^* and, for the case of separable P (i.e., $P(u) = \sum_i P_i(u_i)$), prove necessary and sufficient conditions on P and y^* for exact penalization. For separable P, the condition $P^\circ(y^*) \leq \sigma$ reduces to

$$y_i^* \leq \sigma \lim_{u_i \downarrow 0} \frac{P_i(u_i)}{u_i}, \quad i = 1, \ldots, m, \tag{4.7}$$

as derived in [4, Proposition 1]. A similar result was obtained in [31, Theorem 2.1] for the further special case of $P_i(u_i) = \max\{0, u_i\}$. Thus Theorem 4.2 may be viewed as a generalization of these results. (For the standard quadratic penalty $P_i(u_i) = \max\{0, u_i\}^2$, the right-hand side of (4.7) is zero, so (4.7) holds only if $y_i^* = 0$, i.e., the constraint $g_i(x) \leq 0$ is redundant.)

The results in [9, Corollary 2.5.1 and Theorem 5.3] assume either the linear-independence or Slater constraint qualifications in order to ensure existence of Lagrange multipliers. Theorem 4.2 is partly motivated by and very similar to the necessary and sufficient conditions obtained in [7, Proposition 7.3.1]. The connection with exact regularization, however, appears to be new.

Although our results for exact regularization can be used to deduce results for exact penalization, the reverse direction does not appear possible. In particular, applying exact penalization to the selection problem (P^ϕ) yields a penalized problem very different from (P_δ).

5. Error bounds and weak sharp minimum. Even when exact regularization cannot be achieved, we can still estimate the distance from $\mathcal{S}_\delta$ to $\mathcal{S}$ in terms of δ and the growth rate of f away from $\mathcal{S}$. We study this type of error bound in this section.

THEOREM 5.1.
(a) *For any* $\bar\delta > 0$, $\cup_{0<\delta\le\bar\delta}\mathcal{S}_\delta$ *is bounded.*
(b) *Suppose that there exist* $\tau > 0, \gamma \ge 1$ *such that*

$$f(x) - p^* \ge \tau \operatorname{dist}(x,\mathcal{S})^\gamma \quad \textit{for all} \quad x \in \mathcal{C}, \tag{5.1}$$

where $\operatorname{dist}(x,\mathcal{S}) = \min_{x^*\in\mathcal{S}} \|x - x^*\|_2$. *Then, for any* $\bar\delta > 0$ *there exists* $\tau' > 0$ *such that*

$$\operatorname{dist}(x_\delta,\mathcal{S})^{\gamma-1} \le \tau'\delta \quad \textit{for all} \quad x_\delta \in \mathcal{S}_\delta,\ \delta \in (0,\bar\delta].$$

Proof.
Part (a). Fix any $x^* \in \mathcal{S}$ and any $\bar\delta > 0$. For any $\delta \in (0,\bar\delta]$ and $x_\delta \in \mathcal{S}_\delta$,

$$f(x^*) + \delta\phi(x^*) \ge f(x_\delta) + \delta\phi(x_\delta) \ge f(x^*) + \delta\phi(x_\delta),$$

and thus $\phi(x^*) \ge \phi(x_\delta)$. Using $\phi(x_\delta) \ge \inf_{x\in\mathcal{C}} \phi(x)$, we have, similarly, that

$$f(x_\delta) \le f(x^*) + \delta\Big(\phi(x^*) - \inf_{x\in\mathcal{C}} \phi(x)\Big) \le f(x^*) + \bar\delta\Big(\phi(x^*) - \inf_{x\in\mathcal{C}} \phi(x)\Big).$$

This shows that $\cup_{0<\delta\le\bar\delta}\mathcal{S}_\delta \subseteq \{x \in \mathcal{C} \mid \phi(x) \le \beta,\ f(x) \le \beta\}$ for some $\beta \in \mathbb{R}$. Since ϕ, f, and $\mathcal{C}$ have no nonzero recession direction in common (see Assumptions 1.1 and 1.2), the second set is bounded and therefore so is the first set.

Part (b). For any $\delta > 0$ and $x_\delta \in \mathcal{S}_\delta$, let $x^*_\delta \in \mathcal{S}$ satisfy $\|x_\delta - x^*_\delta\|_2 = \operatorname{dist}(x_\delta,\mathcal{S})$. Then

$$\begin{aligned} f(x^*_\delta) + \delta\phi(x^*_\delta) &\ge f(x_\delta) + \delta\phi(x_\delta) \\ &\ge f(x^*_\delta) + \tau\|x_\delta - x^*_\delta\|_2^\gamma + \delta\phi(x_\delta), \end{aligned}$$

which implies that

$$\tau\|x_\delta - x^*_\delta\|_2^\gamma \le \delta\big(\phi(x^*_\delta) - \phi(x_\delta)\big).$$

Because ϕ is convex and real-valued,

$$\phi(x_\delta) \ge \phi(x^*_\delta) + \eta_\delta^T(x_\delta - x^*_\delta) \ge \phi(x^*_\delta) - \|\eta_\delta\|_2\|x_\delta - x^*_\delta\|_2,$$

for some $\eta_\delta \in \partial\phi(x^*_\delta)$. Combining the above two inequalities yields

$$\tau\|x_\delta - x^*_\delta\|_2^{\gamma-1} \le \delta\|\eta_\delta\|_2.$$

By part (a), x_δ lies in a bounded set for all $\delta > 0$, so x^*_δ lies in a bounded subset of $\mathcal{S}$ for all $\delta > 0$. Then η_δ lies in a bounded set [40, Theorem 24.7], so that $\|\eta_\delta\|_2$ is uniformly bounded. This proves the desired bound. ∎

Error bounds of the form (5.1) have been much studied, especially in the cases of linear growth ($\gamma = 1$) and quadratic growth ($\gamma = 2$); see [6, 10, 11, 28, 29, 46] and references therein. In general, it is known that (5.1) holds for some $\tau > 0$ and $\gamma \geq 1$ whenever f is analytic and $\mathcal{C}$ is bounded [28, Theorem 2.1].

Theorem 5.1 does not make much use of the convexity of f and ϕ, and it readily extends to nonconvex f and ϕ. In the case of $\gamma = 1$ in (5.1) (i.e., f has a "weak sharp minimum" over $\mathcal{C}$), Theorem 5.1(b) implies that $\operatorname{dist}(x_\delta, \mathcal{S}) = 0$ for all $x_\delta \in \mathcal{S}_\delta$—i.e., $\mathcal{S}_\delta \subseteq \mathcal{S}$, whenever $\delta < 1/\tau'$. In this case, then, $\mathcal{S}_\delta = \mathcal{S}^\phi$ whenever $\delta < 1/\tau'$ and $\mathcal{S}_\delta \neq \emptyset$. This gives another exact-regularization result.

The following result shows that it is nearly necessary for f to have a weak sharp minimum over $\mathcal{C}$ in order for there to be exact regularization by any strongly convex quadratic regularization function.

THEOREM 5.2. *Suppose that f is continuously differentiable on $\mathbb{R}^n$ and $\mathcal{S}$ is bounded. If there does not exist $\tau > 0$ such that (5.1) holds with $\gamma = 1$, then either*

(i) there exists a strongly convex quadratic function of the form $\phi(x) = \|x - \widehat{x}\|_2^2$ ($\widehat{x} \in \mathbb{R}^n$) and a scalar $\bar{\delta} > 0$ for which $\mathcal{S}_\delta \neq \mathcal{S}^\phi$ for all $\delta \in (0, \bar{\delta}]$;

or

(ii) for every sequence $x^k \in \mathcal{C} \setminus \mathcal{S}$, $k = 1, 2, \ldots$, satisfying

$$\frac{f(x^k) - p^*}{\operatorname{dist}(x^k, \mathcal{S})} \to 0, \tag{5.2}$$

and every cluster point (x^*, v^*) of $\{(s^k, \frac{x^k - s^k}{\|x^k - s^k\|_2})\}$, we have $x^* + \alpha v^* \notin \mathcal{C}$ for all $\alpha > 0$, where $s^k \in \mathcal{S}$ satisfies $\|x^k - s^k\|_2 = \operatorname{dist}(x^k, \mathcal{S})$.

If case (ii) occurs, then $\mathcal{C}$ is not polyhedral, and for any $\bar{x} \in \mathcal{S}$,

$$\mathcal{S} = \operatorname*{arg\,min}_{x \in \mathcal{C}} \ \nabla f(\bar{x})^T x. \tag{5.3}$$

Proof. Suppose that there does not exist $\tau > 0$ such that (5.1) holds with $\gamma = 1$. Then there exists a sequence $x^k \in \mathcal{C} \setminus \mathcal{S}$, $k = 1, 2, \ldots$, that satisfies (5.2). Let $s^k \in \mathcal{S}$ satisfy $\|x^k - s^k\|_2 = \operatorname{dist}(x^k, \mathcal{S})$. Let $v^k = (x^k - s^k)/\|x^k - s^k\|_2$, so that $\|v^k\|_2 = 1$. Because $\mathcal{S}$ is bounded, $\{s^k\}$ is bounded. By passing to a subsequence if necessary, we can assume that $(s^k, v^k) \to$ some (x^*, v^*). Because s^k is the nearest point projection of x^k onto $\mathcal{S}$, we have $v^k \in N_{\mathcal{S}}(s^k)$, i.e., $(x - s^k)^T v^k \leq 0$ for all $x \in \mathcal{S}$. Taking the limit yields $v^* \in N_{\mathcal{S}}(x^*)$, i.e., $(x - x^*)^T v^* \leq 0$ for all $x \in \mathcal{S}$.

Note that $\{x^k\}$ need not converge to x^* or even be bounded. Now, consider the auxiliary sequence

$$y^k = s^k + \epsilon^k (x^k - s^k) \quad \text{with} \quad \epsilon^k = \frac{1}{\max\{k, \|x^k - s^k\|_2\}},$$

$k = 1, 2, \ldots$. Then $\epsilon^k \in (0, 1]$, $y^k \in \mathcal{C} \setminus \mathcal{S}$, $(y^k - s^k)/\|y^k - s^k\|_2 = v^k$ for all k, and $y^k - s^k \to 0$ (so $y^k \to x^*$). Also, the convexity of f implies $f(y^k) \leq (1 - \epsilon^k) f(s^k) + \epsilon^k f(x^k)$ which, together with $\|y^k - s^k\|_2 = \epsilon^k \|x^k - s^k\|_2$ and $f(s^k) = p^*$, implies

$$0 \leq \frac{f(y^k) - f(s^k)}{\|y^k - s^k\|_2} \leq \frac{\epsilon^k f(x^k) - \epsilon^k f(s^k)}{\|y^k - s^k\|_2} = \frac{f(x^k) - p^*}{\operatorname{dist}(x^k, \mathcal{S})} \to 0. \tag{5.4}$$

Because $f(y^k) - f(s^k) = \nabla f(s^k)^T(y^k - s^k) + o(\|y^k - s^k\|_2)$ and f is continuously differentiable, (5.4) and $y^k - x^k \to 0$ yield, in the limit,

$$\nabla f(x^*)^T v^* = 0. \tag{5.5}$$

Let $f_\delta(x) = f(x) + \delta\phi(x)$, with

$$\phi(x) = \|x - (x^* + v^*)\|_2^2.$$

Because $v^* \in N_{\mathcal{S}}(x^*)$, we have $\mathcal{S}^\phi = \{x^*\}$.

Suppose that there exists $\alpha > 0$ such that $x^* + \alpha v^* \in \mathcal{C}$. Then, for any $\beta \in (0, \alpha]$,

$$\begin{aligned} f_\delta(x^* + \beta v^*) &= f(x^* + \beta v^*) + \|\beta v^* - v^*\|_2^2 \\ &= f(x^*) + \beta \nabla f(x^*)^T v^* + o(\beta) + \delta(\beta - 1)^2\|v^*\|_2^2 \\ &= f(x^*) + o(\beta) + \delta(1 - 2\beta + \beta^2) \\ &= f_\delta(x^*) + o(\beta) - \delta\beta(2 - \beta), \end{aligned}$$

where the third equality uses (5.5) and $\|v^*\|_2 = 1$. Thus $x^* + \beta v^* \in \mathcal{C}$ and $f_\delta(x^* + \beta v^*) < f_\delta(x^*)$ for all $\beta > 0$ sufficiently small, implying $\mathcal{S}_\delta \neq \mathcal{S}^\phi$. Therefore, if case (ii) does not occur, then case (i) must occur.

Suppose that case (ii) occurs. First, we claim that, for any $\bar{x} \in \mathcal{S}$,

$$\nabla f(\bar{x})^T(x - \bar{x}) > 0 \quad \text{for all} \quad x \in \mathcal{C} \setminus \mathcal{S}.$$[1]

Fix any $\bar{x} \in \mathcal{S}$. Because $\nabla f(\bar{x})^T(x - \bar{x}) = 0$ for all $x \in \mathcal{S}$, this yields (5.3). Next, we claim that $\mathcal{C}$ cannot be polyhedral. If $\mathcal{C}$ were polyhedral, then the minimization in (5.3) would be an LP, for which weak sharp minimum holds. Then there would exist $\tau > 0$ such that

$$\nabla f(\bar{x})^T(x - \bar{x}) \geq \tau \operatorname{dist}(x, \mathcal{S}) \quad \text{for all} \quad x \in \mathcal{C}.$$

Because f is convex and thus $f(x) - p^* = f(x) - f(\bar{x}) \geq \nabla f(\bar{x})^T(x - \bar{x})$ for all $x \in \mathcal{C}$, this would imply that (5.1) holds with $\gamma = 1$, contradicting our assumption. ☐

An example of case (ii) occurring in Theorem 5.2 is

$$n = 2, \quad f(x) = x_2, \quad \text{and} \quad \mathcal{C} = \{x \in \mathbb{R}^2 \mid x_1^2 \leq x_2\}.$$

Here $\mathcal{S} = \{(0,0)\}$, $p^* = 0$, and

$$\frac{f(x) - p^*}{\operatorname{dist}(x, \mathcal{S})} = \frac{x_2}{\|x\|_2} = \frac{1}{\sqrt{(x_1/x_2)^2 + 1}} \quad \text{for all} \quad x \in \mathcal{C} \setminus \mathcal{S}.$$

The right-hand side goes to 0 if and only if $x_1/x_2 \to \infty$, in which case $x/\|x\|_2 \to (\pm 1, 0)$, and $\alpha(\pm 1, 0) \notin \mathcal{C}$ for all $\alpha > 0$. Interestingly, we can still find $\phi(x) =$

[1]If this were false, then there would exist $\bar{x} \in \mathcal{S}$ and $x \in \mathcal{C} \setminus \mathcal{S}$ such that $\nabla f(\bar{x})^T(x - \bar{x}) = 0$. ($\nabla f(\bar{x})^T(x - \bar{x}) < 0$ cannot occur because $\bar{x} \in \mathcal{S}$.) Let $s \in \mathcal{S}$ satisfy $\|x - s\|_2 = \operatorname{dist}(x, \mathcal{S})$. By Lemma 3.1, $\nabla f(s)^T(x - s) = 0$. Then for $x^k = s + (x - s)/k$, we would have $x^k \in \mathcal{C} \setminus \mathcal{S}$, $f(x^k) - f(s) = o(1/k)$, and $\operatorname{dist}(x^k, \mathcal{S}) = \|x - s\|_2/k$, so x^k satisfies (5.2) and $s^k = s$ for $k = 1, 2, \ldots$. Because $(s^k, \frac{x^k - s^k}{\|x^k - s^k\|_2}) \to (s, \frac{x - s}{\|x - s\|_2})$ and $s + \alpha\frac{x - s}{\|x - s\|_2} \in \mathcal{C}$ for all $\alpha \in (0, \|x - s\|_2]$, this would contradict case (ii) occurring.

$\|x - \widehat{x}\|_2^2$ for which $\mathcal{S}_\delta \neq \mathcal{S}^\phi$ for all $\delta > 0$ sufficiently small. For example, take $\phi(x) = (x_1 - 1)^2 + (x_2 - 1)^2$. Then (P$_\delta$) becomes

$$\underset{x}{\text{minimize}} \quad x_2 + \delta(x_1 - 1)^2 + \delta(x_2 - 1)^2 \quad \text{subject to} \quad x_1^2 \le x_2.$$

It is straightforward to check that $(0, 0)$ does not satisfy the necessary optimality conditions for (P$_\delta$) for all $\delta > 0$. This raises the question of whether case (ii) is subsumed by case (i) when $\mathcal{C}$ is not polyhedral. In §8, we give an example showing that the answer is "no".

6. Sparse solutions. In this section we illustrate a practical application of Corollary 2.2. Our aim is to find sparse solutions of linear and conic programs that may not have unique solutions. To this end, we let $\phi(x) = \|x\|_1$, which clearly satisfies the required Assumption 1.2. (In general, however, some components of x may be more significant or be at different scales, in which case we may not wish to regularize all components or regularize them equally.)

Regularization based on the one-norm has been used in many applications, with the goal of obtaining sparse or even *sparsest* solutions of underdetermined systems of linear equations and least-squares problems. Some recent examples include [14, 16, 17, 18].

The AMPL model and data files and the MATLAB scripts used to generate all of the numerical results presented in the following subsections can be obtained at `http://www.cs.ubc.ca/~mpf/exactreg/`.

6.1. Sparse solutions of linear programs. For underdetermined systems of linear equations $Ax = b$ that arise in fields such as signal processing, the studies in [13], [14], and [18] advocate solving

$$\underset{x}{\text{minimize}} \quad \|x\|_1 \quad \text{subject to} \quad Ax = b \quad (\text{and possibly } x \ge 0), \tag{6.1}$$

in order to obtain a sparse solution. This problem can be recast as an LP and be solved efficiently. The sparsest solution is given by minimizing the so-called zero-norm, $\|x\|_0$, which counts the number of nonzero components in x. However, the combinatorial nature of this minimization makes it computationally intractable for all but the simplest instances. Interestingly, there exist reasonable conditions under which a solution of (6.1) is a sparsest solution; see [13, 18].

Following this approach, we use Corollary 2.2 as a guide for obtaining least one-norm solutions of a generic LP,

$$\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad Ax = b, \quad l \le x \le u, \tag{6.2}$$

by solving its regularized version,

$$\underset{x}{\text{minimize}} \quad c^T x + \delta\|x\|_1 \quad \text{subject to} \quad Ax = b, \quad l \le x \le u. \tag{6.3}$$

The vectors l and u are lower and upper bounds on x. In many of the numerical tests given below, the exact ℓ_1 regularized solution of (6.2) (given by (6.3) for small-enough values of δ) is considerably sparser than the solution obtained by solving (6.2) directly. In each instance, we solve the regularized and unregularized problems with the same interior-point solver. We emphasize that, with an appropriate choice of the regularization parameter δ, the solution of the regularized LP is *also* a solution of the original LP.

We use two sets of test instances in our numerical experiments. The instances of the first set are randomly generated using a degenerate LP generator described in [23]. Those of the second set are derived from the infeasible LPs in the Netlib collection (`http://www.netlib.org/lp/infeas/`). Both sets of test instances are further described in §§6.1.1–6.1.2.

We follow the same procedure for each test instance. First, we solve the LP (6.2) to obtain an unregularized solution x^* and the optimal value $p^* := c^T x^*$. Next, we solve (P^ϕ), reformulated as an LP, to obtain a Lagrange multiplier μ^* and the threshold value $\bar{\delta} = 1/\mu^*$. Finally, we solve (6.3) with $\delta := \bar{\delta}/2$, reformulated as an LP, to obtain a regularized solution x^*_δ.

We use the log-barrier interior-point algorithm implemented in CPLEX 9.1 to solve each LP. The default CPLEX options are used, except for `crossover = 0` and `comptol = 1e-10`. Setting `crossover = 0` forces CPLEX to use the interior-point algorithm only, and to not "cross over" to find a vertex solution. In general, we expect the interior-point algorithm to find the analytic center of the solution set (see [47, Theorems 2.16 and 2.17]), which tends to be less sparse than vertex solutions. The `comptol` option tightens CPLEX's convergence tolerance from its default of `1e-8` to its smallest allowable setting. We do not advocate such a tight tolerance in practice, but the higher accuracy aids in computing the sparsity of a computed solution, which we determine as

$$\|x\|_0 = \mathrm{card}\{x_i \mid |x_i| > \epsilon\}, \tag{6.4}$$

where $\epsilon = 10^{-8}$ is larger than the specified convergence tolerance.

6.1.1. Randomly generated LPs. Six dual-degenerate LPs were constructed using Gonzaga's Matlab generator [23]. This Matlab program accepts as inputs the problem size and the dimensions of the optimal primal and dual faces, D_p and D_d, respectively. Gonzaga shows that these quantities must satisfy

$$0 \le D_p \le n - m - 1 \qquad \text{and} \qquad 0 \le D_d \le m - 1. \tag{6.5}$$

The six LPs are constructed with parameters $n = 1000$, $m = 100$, $D_d = 0$, and various levels of D_p set as 0%, 20%, 40%, 60%, 80%, and 100% of the maximum of 899 (given by (6.5)). The instances are respectively labeled `random-0`, `random-20`, `random-40`, and so on.

Table 6.1 summarizes the results. We confirm that in each instance the optimal values of the unregularized and regularized problems are nearly identical (at least to within the specified tolerance), so each regularized solution is exact. Except for the "control" instance `random-0`, the regularized solution x^*_δ has a strictly lower one-norm, and is considerably sparser than the unregularized solution x^*.

6.1.2. Infeasible LPs. The second set of test instances is derived from a subset of the infeasible Netlib LPs. For each infeasible LP, we discard the original objective, and instead form the problem

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_1 \quad \text{subject to} \quad l \le x \le u, \tag{$\mathrm{P}^{\mathrm{inf}}$}$$

and its regularized counterpart

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_1 + \delta\|x\|_1 \quad \text{subject to} \quad l \le x \le u. \tag{$\mathrm{P}^{\mathrm{inf}}_\delta$}$$

Table 6.1

Randomly generated LPs with increasing dimension of the optimal primal face. The arrows indicate differences between values in neighboring columns: $\rightarrow$ indicates that the value to the right is the same; $\searrow$ indicates that the value to the right is lower; $\nearrow$ indicates that the value to the right is larger.

LP	$c^T x^*$	$c^T x_\delta^*$	$\\|x^*\\|_1$		$\\|x_\delta^*\\|_1$	$\\|x^*\\|_0$		$\\|x_\delta^*\\|_0$	$\bar\delta$
random-0	2.5e−13	1.0e−13	9.1e+01	$\rightarrow$	9.1e+01	100	$\rightarrow$	100	1.5e−04
random-20	5.6e−13	6.6e−13	2.9e+02	$\searrow$	2.0e+02	278	$\searrow$	100	2.2e−02
random-40	3.8e−12	3.7e−12	4.9e+02	$\searrow$	2.9e+02	459	$\searrow$	100	2.9e−02
random-60	3.9e−14	9.2e−11	6.7e+02	$\searrow$	3.6e+02	637	$\searrow$	101	3.3e−02
random-80	9.1e−12	8.4e−13	8.9e+02	$\searrow$	4.6e+02	816	$\searrow$	100	2.1e−01
random-100	1.8e−16	3.2e−12	1.0e+03	$\searrow$	5.4e+02	997	$\searrow$	102	1.1e−01

The unregularized problem ($\mathrm{P}^{\mathrm{inf}}$) models the plausible situation where we wish to fit a set of infeasible equations in the least one-norm sense. But because the one-norm is not strictly convex or the equations are underdetermined, a solution of ($\mathrm{P}^{\mathrm{inf}}$) may not be unique, and the regularized problem ($\mathrm{P}_\delta^{\mathrm{inf}}$) is used to further select a sparse solution.

The following infeasible Netlib LPs were omitted because CPLEX returned an error message during the solution of ($\mathrm{P}^{\mathrm{inf}}$) or ($\mathrm{P}_\delta^{\mathrm{inf}}$): `lpi-bgindy`, `lpi-cplex2`, `lpi-gran`, `lpi-klein1`, `lpi-klein2`, `lpi-klein3`, `lpi-qual`, `lpi-refinery`, and `lpi-vol1`.

Table 6.2 summarizes the results. We can see that the regularized solution x_δ^* is exact (i.e., $c^T x_\delta^* = c^T x^*$) and has a one-norm lower than or equal to that of the unregularized solution x^* in all instances. In twelve of the twenty instances, x_δ^* is sparser than x^*. In five of the instances, they have the same sparsity. In three of the instances (`lpi-galenet`, `lpi-itest6`, and `lpi-woodinfe`), x_δ^* is actually less sparse, even though its one-norm is lower.

6.2. Sparse solutions of semidefinite/second-order cone programs. In §6.1 we used Corollary 2.2 to find sparse solutions of LPs. In this section, we report our numerical experience in finding sparse solutions of SDPs and SOCPs that may not have unique solutions. These are conic programs (P) with f and $\mathcal{C}$ given by (1.1), and $\mathcal{K}$ being the Cartesian product of real space, orthant, second-order cones, and semidefinite cones.

The regularized problem (P_δ) can be put in the conic form

$$\begin{array}{ll} \underset{x,u,v}{\text{minimize}} & c^T x + \delta e^T(u+v) \\ \text{subject to} & Ax = b, \quad x - u + v = 0, \\ & (x,u,v) \in \mathcal{K} \times [0,\infty)^{2n}, \end{array} \tag{6.6}$$

where e is the vector of ones. The selection problem (P^ϕ) can also be put in conic form:

$$\begin{array}{ll} \underset{x,u,v,s}{\text{minimize}} & e^T(u+v) \\ \text{subject to} & Ax = b, \quad x - u + v = 0, \quad c^T x + s = p^*, \\ & (x,u,v,s) \in \mathcal{K} \times [0,\infty)^{2n+1}. \end{array} \tag{6.7}$$

As in §6.1, we first solve (P) to obtain x^* and the optimal value $p^* := c^T x^*$. Then (6.7) is solved to obtain Lagrange multiplier μ^* and the corresponding threshold value

Table 6.2
Least one-norm residual solutions of the infeasible Netlib LPs.

LP	c^Tx^*	$c^Tx^*_\delta$	$\|x^*\|_1$		$\|x^*_\delta\|_1$	$\|x^*\|_0$		$\|x^*_\delta\|_0$	$\bar{\delta}$
lpi-bgdbg1	3.6e+02	3.6e+02	1.6e+04	↘	1.3e+04	518	↘	437	3.3e−03
lpi-bgetam	5.4e+01	5.4e+01	6.0e+03	↘	5.3e+03	633	↘	441	3.4e−04
lpi-bgprtr	1.9e+01	1.9e+01	4.7e+03	↘	3.0e+03	25	↘	20	3.7e−01
lpi-box1	1.0e+00	1.0e+00	5.2e+02	↘	2.6e+02	261	→	261	9.9e−01
lpi-ceria3d	2.5e−01	2.5e−01	8.8e+02	→	8.8e+02	1780	↘	1767	6.7e−04
lpi-chemcom	9.8e+03	9.8e+03	1.5e+05	↘	3.8e+04	711	↘	591	3.1e−01
lpi-cplex1	3.2e+06	3.2e+06	2.4e+09	↘	1.5e+09	3811	↘	3489	1.0e−02
lpi-ex72a	1.0e+00	1.0e+00	4.8e+02	↘	3.0e+02	215	→	215	1.6e−01
lpi-ex73a	1.0e+00	1.0e+00	4.6e+02	↘	3.0e+02	211	→	211	1.6e−01
lpi-forest6	8.0e+02	8.0e+02	4.0e+05	→	4.0e+05	54	→	54	1.2e−03
lpi-galenet	2.8e+01	2.8e+01	1.0e+02	↘	9.2e+01	10	↗	11	6.3e−01
lpi-gosh	4.0e−02	4.0e−02	1.5e+04	↘	7.1e+03	9580	↘	1075	3.9e−05
lpi-greenbea	5.2e+02	5.2e+02	1.4e+06	↘	5.6e+05	3658	↘	1609	1.1e−04
lpi-itest2	4.5e+00	4.5e+00	2.3e+01	→	2.3e+01	7	→	7	6.5e−01
lpi-itest6	2.0e+05	2.0e+05	4.8e+05	↘	4.6e+05	12	↗	14	4.8e−01
lpi-mondou2	1.7e+04	1.7e+04	3.2e+06	↘	2.7e+06	297	↘	244	9.5e−02
lpi-pang	2.4e−01	2.4e−01	1.4e+06	↘	8.2e+04	536	↘	336	1.4e−06
lpi-pilot4i	3.3e+01	3.3e+01	6.9e+05	↘	5.1e+04	773	↘	627	3.6e−06
lpi-reactor	2.0e+00	2.0e+00	1.5e+06	↘	1.1e+06	569	↘	357	4.1e−05
lpi-woodinfe	1.5e+01	1.5e+01	3.6e+03	↘	2.0e+03	60	↗	87	5.0e−01

$\bar{\delta} := 1/\mu^*$. Finally, we solve (6.6) with $\delta = \bar{\delta}/2$ to obtain x^*_δ. All three problems—(P), (6.6), and (6.7)—are solved using the MATLAB toolbox SeDuMi (version 1.05) [43], which is a C implementation of a log-barrier primal-dual interior-point algorithm for solving SDP/SOCP. The test instances are drawn from the DIMACS Implementation Challenge library [37], a collection of nontrivial medium-to-large SDP/SOCP arising from applications. We omit those instances for which either (P) is infeasible (e.g., `filtinf1`) or if one of (P), (6.6), (6.7) cannot be solved because of insufficient memory (e.g., `torusg3-8`). All runs were performed on a PowerPC G5 with 2GB of memory running MATLAB 7.3b.

Table 6.3 summarizes the results. For most of the instances, SeDuMi finds only an inaccurate solution (`info.numerr=1`) for at least one of (P), (6.6), (6.7). For most instances, however, SeDuMi also finds a value of μ^* that seems reasonable. In some instances (`nb_L2_bessel`, `nql30`, `nql80`, `qssp30`, `qssp60`, `qssp180`, `sch_100_100_scal`, `sch_200_100_scal`, `truss8`), the computed multiplier μ^* is quite large relative to the solution accuracy, and yet $c^Tx^*_\delta$ matches c^Tx^* in the first three significant digits; this suggests that the regularization is effectively exact. For `nb_L2`, `sch_100_50_scal`, and `sch_100_100_orig`, the discrepancies between $c^Tx^*_\delta$ and c^Tx^* may be attributed to a SeDuMi numerical failure or primal infeasibility in solving either (P) or (6.7) (thus yielding inaccurate μ^*), or (6.6). For `hinf12`, SeDuMi solved all three problems accurately, and μ^* looks reasonable, whereas for `hinf13`, SeDuMi solved all three problems inaccurately, but μ^* still looks reasonable. Yet $c^Tx^*_\delta$ is lower than c^Tx^* in both instances. We do not yet have an explanation for this.

The regularized solution x^*_δ has a one-norm lower than or equal to that of the unregularized solution x^* in all instances except `hinf12`, where $\|x^*_\delta\|_1$ is 1% higher (this small difference does not appear in Table 6.3). Solution sparsity is measured by

Table 6.3

Least one-norm solutions of the feasible DIMACS SDP/SOCPs. Three different types of SeDuMi failures are reported: [a]numerical error; [b]primal infeasibility detected in solving (6.7); [c]numerical error in solving (6.7). The "schedule" instances have been abbreviated from `sched_100_50_orig` *to* `sch_100_50_o`*, etc.*

SDP/SOCP	$c^T x^*$	$c^T x^*_\delta$	$\|x^*\|_1$		$\|x^*_\delta\|_1$	$\|x^*\|_0$		$\|x^*_\delta\|_0$	$\bar{\delta}$
nb	-5.07e−02	-5.07e−02	2.2e+0	$\searrow$	2.1e+0	142	$\searrow$	139	7.6e−3
nb_L1	-1.30e+01	-1.30e+01	3.1e+3	$\rightarrow$	3.1e+3	2407	$\searrow$	1613	1.2e−5
nb_L2	-1.63e+00	-1.63e+00	3.1e+1	$\rightarrow$	3.1e+1	847	$\rightarrow$	847	2.1e−5
nb_L2_bessel	-1.03e−01	-1.03e−01	1.0e+1	$\searrow$	9.7e+0	131	$\nearrow$	133	2.7e−6
copo14	-3.11e−12	-2.13e−10	4.6e+0	$\searrow$	2.0e+0	2128	$\searrow$	224	4.7e−1
copo23	-8.38e−12	-3.73e−09	6.6e+0	$\searrow$	2.0e+0	9430	$\searrow$	575	4.7e−1
filter48_socp	1.42e+00	1.42e+00	7.6e+2	$\rightarrow$	7.6e+2	3284	$\searrow$	3282	1.1e−6
minphase	5.98e+00	5.98e+00	1.6e+1	$\rightarrow$	1.6e+1	2304	$\rightarrow$	2304	5.8e−2
hinf12	-3.68e−02	-7.11e−02	1.0e+0	$\rightarrow$	1.0e+0	138	$\nearrow$	194	5.1e+0
hinf13	-4.53e+01	-4.51e+01	2.8e+4	$\searrow$	2.1e+4	322	$\searrow$	318	2.8e−4
nql30	-9.46e−01	-9.46e−01	5.8e+3	$\searrow$	2.8e+3	6301	$\rightarrow$	6301	1.0e−7
nql60	-9.35e−01	-9.35e−01	2.3e+4	$\searrow$	1.1e+4	25201	$\rightarrow$	25201	1.4e−6
nql180	-9.28e−01	-9.28e−01	2.1e+5	$\searrow$	1.0e+5	226776	$\searrow$	226767	6.3e−8
nql30old	-9.46e−01	-9.46e−01	5.5e+3	$\searrow$	1.0e+3	7502	$\searrow$	6244	3.2e−5
nql60old	-9.35e−01	-9.35e−01	2.2e+4	$\searrow$	4.0e+3	29515	$\searrow$	23854	2.0e−5
nql180old	[a]-9.31e−01	[a]-9.29e−01	1.9e+5	$\searrow$	6.8e+4	227097	$\searrow$	211744	1.4e−8
qssp30	-6.50e+00	-6.50e+00	4.5e+3	$\rightarrow$	4.5e+3	7383	$\rightarrow$	7383	4.1e−7
qssp60	-6.56e+00	-6.56e+00	1.8e+4	$\rightarrow$	1.8e+4	29163	$\rightarrow$	29163	1.2e−6
qssp180	-6.64e+00	-6.64e+00	1.6e+5	$\rightarrow$	1.6e+5	260283	$\rightarrow$	260283	[c]3.8e−7
sch_50_50_o	2.67e+04	2.67e+04	5.6e+4	$\rightarrow$	5.6e+4	1990	$\nearrow$	2697	8.7e−3
sch_50_50_s	7.85e+00	7.85e+00	1.1e+2	$\rightarrow$	1.1e+2	497	$\nearrow$	600	1.1e−5
sch_100_50_o	1.82e+05	1.82e+05	4.9e+5	$\rightarrow$	4.9e+5	3131	$\searrow$	3040	2.4e−4
sch_100_50_s	6.72e+01	[b]8.69e+01	6.0e+4	$\searrow$	1.3e+4	5827	$\nearrow$	7338	6.1e−3
sch_100_100_o	7.17e+05	[a]3.95e+02	1.8e+6	$\searrow$	8.4e+2	12726	$\nearrow$	18240	1.3e−0
sch_100_100_s	2.73e+01	2.73e+01	1.6e+5	$\rightarrow$	1.6e+5	17574	$\searrow$	16488	1.8e−8
sch_200_100_o	1.41e+05	1.41e+05	4.4e+5	$\rightarrow$	4.4e+5	24895	$\searrow$	16561	4.3e−4
sch_200_100_s	5.18e+01	5.18e+01	7.8e+4	$\rightarrow$	7.8e+4	37271	$\searrow$	37186	4.0e−8
truss5	1.33e+02	1.33e+02	2.1e+3	$\searrow$	1.5e+3	3301	$\rightarrow$	3301	1.6e−5
truss8	1.33e+02	1.33e+02	7.9e+3	$\searrow$	5.2e+3	11914	$\searrow$	11911	1.7e−7

the zero-norm defined in (6.4), where ϵ is based on the relative optimality gap

$$\epsilon = \frac{c^T x^*_\delta - b^T y^*_\delta}{1 + \|b\|\|y^*_\delta\| + \|c\|\|x^*_\delta\|}$$

of the computed solution of (6.6). For 52% of the instances, the regularized solution is sparser than the unregularized solution. For 28% of the instances, the solutions have the same sparsity. For the remaining six instances, the regularized solution is actually less sparse, even though its one-norm is lower (`nb_L2_bessel`, `sch_100_50_s`, `sch_100_100_o`) or the same (`hinf12`, `sch_50_50_o`, `sch_50_50_s`). SeDuMi implements an interior-point algorithm, so it is likely to find the analytic center of the solution set of (P).

The selection problem (6.7) is generally much harder to solve than (P) or (6.6). For example, on `nb_L2_bessel`, SeDuMi took 18, 99, and 16 iterations to solve (P), (6.7), and (6.6), respectively, and on `truss8` SeDuMi took, respectively, 24, 117, and

35 iterations. This seems to indicate that regularization is more efficient than solving the selection problem as a method for finding sparse solutions.

7. Discussion. We see from the numerical results in §6 that regularization can provide an effective way of selecting a solution with desirable properties, such as sparsity. However, finding the threshold value $\bar{\delta}$ for exact regularization entails first solving (P) to obtain p^*, and then solving (P^ϕ) to obtain μ^* and setting $\bar{\delta} = 1/\mu^*$; see Corollary 2.2. Can we find a $\delta < \bar{\delta}$ from (P) without also solving (P^ϕ)?

Consider the case of a CP, in which f and $\mathcal{C}$ have the form (1.1). Suppose that a value of $\delta < \bar{\delta}$ has been guessed (with $\bar{\delta}$ unknown), and a solution x^* of the regularized problem (P_δ) is obtained. By Corollary 2.2, x^* is also a solution of (P^ϕ). Suppose also that there exist Lagrange multipliers $y^* \in \mathbb{R}^m$ and $z^* \in \mathcal{K}^*$ for (P), where $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$ given by

$$\mathcal{K}^* := \{y \in \mathbb{R}^n \mid y^T x \geq 0 \text{ for all } x \in \mathcal{K}\}.$$

Then (y^*, z^*) satisfy, among other conditions,

$$A^T y^* + z^* = c \qquad \text{and} \qquad b^T y^* = p^*.$$

Suppose, furthermore, that there exist Lagrange multipliers $y^*_\phi \in \mathbb{R}^m$, $z^*_\phi \in \mathcal{K}^*$, and $\mu^* \geq 0$ for (P^ϕ) that satisfy, among other conditions,

$$0 \in \partial\phi(x^*) - (A^T y^*_\phi + z^*_\phi - \mu^* c).$$

Then, analogous to the proof of Theorem 2.1, we can construct Lagrange multipliers for (P_δ) as follows:

Case 1: $\mu^*_\phi = 0$. The Lagrange multipliers for (P_δ) are given by

$$y^*_\delta := y^* + \delta y^*_\phi \quad \text{and} \quad z^*_\delta := z^* + \delta z^*_\phi.$$

Case 2: $\mu^*_\phi > 0$. The Lagrange multipliers for (P_δ) are given by

$$y^*_\delta := (1-\lambda) y^* + \frac{\lambda}{\mu^*_\phi} y^*_\phi \quad \text{and} \quad z^*_\delta := (1-\lambda) z^* + \frac{\lambda}{\mu^*_\phi} z^*_\phi,$$

for any $\lambda \in [0,1]$. The Lagrange multipliers (y^*_δ, z^*_δ) obtained for the regularized problem are therefore necessarily perturbed. Therefore, it is not possible to test the computed triple $(x^*, y^*_\delta, z^*_\delta)$ against the optimality conditions for the original CP in order to verify that x^* is indeed an exact solution.

In practice, if it were prohibitively expensive to solve (P) and (P^ϕ), we might adopt an approach suggested by Lucidi [27] and Mangasarian [33] for Tikhonov regularization. They suggest solving the regularized problem successively with decreasing values $\delta_1 > \delta_2 > \cdots$. If successive regularized solutions do not change, then it is likely that a correct regularization parameter has been obtained. We note that in many instances, the threshold values $\bar{\delta}$ shown in Tables 6.1 and 6.2 are comfortably large, and a value such as $\delta = 10^{-4}$ would cover 85% of the these cases.

Acknowledgments. Sincere thanks to Dimitri Bertsekas for suggesting Lemma 4.1(a) and bringing to our attention the paper [4]. We also thank Kevin Leyton-Brown for generously giving us access to his CPLEX installation for the numerical experiments presented in §6.

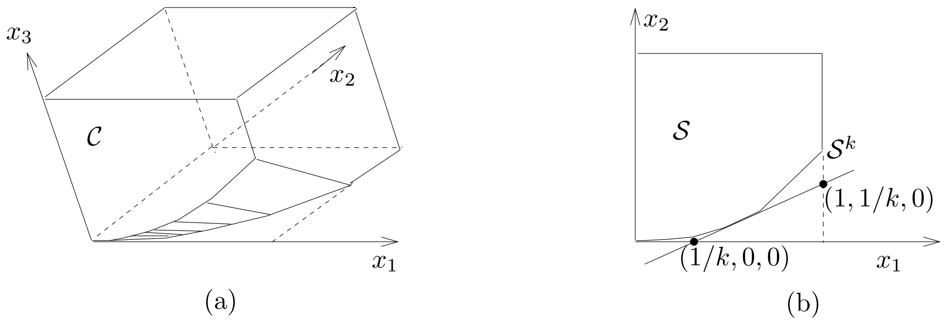

FIG. 8.1. *(a) The feasible set $\mathcal{C}$. (b) The solution set $\mathcal{S}$ and $\mathcal{S}^k$.*

8. Appendix. In this appendix, we give an example of f and $\mathcal{C}$ that satisfy the assumptions of Theorem 5.2 and for which weak sharp minimum fails to hold and yet exact regularization holds for $\phi(x) = \|x - \widehat{x}\|_2^2$ and any $\widehat{x} \in \Re^n$.

Consider the example

$$n = 3, \quad f(x) = x_3, \quad \text{and} \quad \mathcal{C} = [0,1]^3 \cap \left(\cap_{k=2}^{\infty} \mathcal{C}^k \right),$$

where $\mathcal{C}^k = \{x \in \mathbb{R}^3 \mid x_1 - (k-1)x_2 - k^2 x_3 \le 1/k\}$. Each $\mathcal{C}^k$ is a half-space in $\mathbb{R}^3$, so $\mathcal{C}$ is a closed convex set. Moreover, $\mathcal{C}$ is bounded and nonempty (since $0 \in \mathcal{C}$); see Figure 8.1(a). Clearly

$$p^* = 0 \quad \text{and} \quad \mathcal{S} = \{x \in \mathcal{C} \mid x_3 = 0\}. \tag{8.1}$$

First, we show that weak sharp minimum fails to hold, i.e., there does not exist $\tau > 0$ such that (5.1) holds with $\gamma = 1$. Let H^k be the hyperplane forming the boundary set of $\mathcal{C}^k$, i.e., $H^k = \{x \in \mathbb{R}^3 \mid x_1 - (k-1)x_2 - k^2 x_3 = 1/k\}$. Let x^k be the intersection point of H^k, H^{k+1} and the x_1x_3-plane. Direct calculation yields

$$x_1^k = \frac{1 - (1 + 1/k)^{-3}}{k(1 - (1 + 1/k)^{-2})}, \quad x_2^k = 0, \quad x_3^k = \frac{x_1^k - 1/k}{k^2}, \tag{8.2}$$

for $k = 2, 3, \ldots$. Since $\mathcal{C} \subset \mathcal{C}^k$, we have from (8.1) that $\mathcal{S} \subset \mathcal{S}^k$, where we let $\mathcal{S}^k = \{x \in \mathcal{C}^k \mid x_3 = 0\}$; see Figure 8.1(b). Thus

$$\operatorname{dist}(x^k, \mathcal{S}) \ge \operatorname{dist}(x^k, \mathcal{S}^k) \ge \operatorname{dist}((x_1^k, 0, 0), \mathcal{S}^k). \tag{8.3}$$

Since $\lim_{\alpha \to 1} \frac{1-\alpha^3}{1-\alpha^2} = \frac{3}{2}$, (8.2) implies that $kx_1^k \to 3/2$, i.e., $x_1^k = 1.5/k + o(1/k)$. The point in $\mathcal{S}^k$ nearest to $(x_1^k, 0, 0)$ lies on the line through $(1/k, 0, 0)$ and $(1, 1/k, 0)$ (with slope $1/(k-1)$ in the x_1x_2-plane), from which it follows that $\operatorname{dist}((x_1^k, 0, 0), \mathcal{S}^k) = 0.5/k^2 + o(1/k^2)$. Since $x_3^k = 1.5/k^3 + o(1/k^3)$ by (8.2), this together with (8.3) implies

$$\frac{x_3^k}{\operatorname{dist}(x^k, \mathcal{S}^k)} \le \frac{x_3^k}{\operatorname{dist}((x_1^k, 0, 0), \mathcal{S}^k)} = O(1/k) \to 0.$$

Moreover, for any $\ell \in \{2, 3, \dots\}$, we have from (8.2) and letting $\alpha = \ell/k$ that

$$\begin{aligned} x_1^k - (\ell-1)x_2^k - \ell^2 x_3^k - \frac{1}{\ell} &= \left(1-\alpha^2\right) x_1^k - \frac{1}{\ell}\left(1-\alpha^3\right) \\ &= (1-\alpha)\left((1+\alpha)x_1^k - \frac{1}{\ell}\left(1+\alpha+\alpha^2\right)\right) \\ &= \frac{(1-\alpha)}{k}\left((1+\alpha)\frac{1-(1+1/k)^{-3}}{1-(1+1/k)^{-2}} - \frac{1}{\alpha}\left(1+\alpha+\alpha^2\right)\right) \\ &= \frac{(1-\alpha)}{k}(1+\alpha)\left(\frac{(1+1/k)^{-2}}{1+(1+1/k)^{-1}} - \frac{1}{\alpha(1+\alpha)}\right) \\ &= \frac{(1-\alpha^2)}{k}\left(\frac{k^2}{(2k+1)(k+1)} - \frac{k^2}{\ell(k+\ell)}\right) \\ &= \frac{(1-\alpha^2)}{k}\frac{(2k+\ell+1)(\ell-k-1)}{(2k+1)(k+1)\ell(k+\ell)}, \end{aligned}$$

where the second equality uses $1-\alpha^2 = (1-\alpha)(1+\alpha)$, $1-\alpha^3 = (1-\alpha)(1+\alpha+\alpha^2)$; the fourth equality uses the same identities but with $(1+1/k)^{-1}$ in place of α. By considering the two cases $\ell \le k$ and $\ell \ge k+1$, it is readily seen that the above right-hand side is non-positive. This in turn shows that $x^k \in \mathcal{C}^\ell$ for $\ell = 2, 3, \dots,$ and hence $x^k \in \mathcal{C}$.

Second, fix any $\widehat{x} \in \mathbb{R}^3$ and let $\phi(x) = \|x - \widehat{x}\|_2^2$. Let $x^* = \arg\min_{x \in \mathcal{S}} \phi(x)$ and $f_\delta(x) = f(x) + \delta\phi(x)$. Suppose $x^* \neq 0$. Then $\mathcal{C}$ is polyhedral in a neighborhood $\mathcal{N}$ of x^*. Since $x_\delta = \arg\min_{x \in \mathcal{C}} f_\delta(x)$ converges to x^* as $\delta \to 0$, we have that $x_\delta \in \mathcal{C} \cap \mathcal{N}$ for all $\delta > 0$ below some positive threshold, in which case exact regularization holds (see Corollary 2.3). Suppose $x^* = 0$. Then

$$\widehat{x} = -\nabla\phi(x^*) \in N_{\mathcal{S}}(x^*) = (-\infty, 0]^2 \times \mathbb{R},$$

where the second equality follows from $[0,\infty)^2 \times \{0\}$ being the tangent cone of $\mathcal{S}$ at 0. Thus $\widehat{x}_2 \le 0, \widehat{x}_3 \le 0$ and we see from

$$\nabla f_\delta(x^*) = (0,0,1)^T - \delta\widehat{x}$$

that $\nabla f_\delta(x^*) \ge 0$ for all $\delta \in [0, \bar{\delta}]$, where $\bar{\delta} = \infty$ if $\widehat{x}_3 \le 0$ and $\bar{\delta} = 1/\widehat{x}_3$ if $\widehat{x}_3 > 0$. Because $\mathcal{C} \subset [0,\infty)^3$ this implies that, for $\delta \in [0, \bar{\delta}]$,

$$\nabla f_\delta(x^*)^T(x - x^*) = \nabla f_\delta(x^*)^T x \ge 0 \quad \text{for all} \quad x \in \mathcal{C}.$$

Because $x^* \in \mathcal{C}$ and f_δ is strictly convex for $\delta \in (0, \bar{\delta}]$, this implies that $x^* = \arg\min_{x \in \mathcal{C}} f_\delta(x)$ for all $\delta \in (0, \bar{\delta}]$. Hence exact regularization holds.

REFERENCES

[1] A. Altman and J. Gondzio, *Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization*, Optim. Methods Softw., 11 (1999), pp. 275–302.

[2] F. R. Bach, R. Thibaux, and M. I. Jordan, *Computing regularization paths for learning multiple kernels*, in Advances in Neural Information Processing Systems (NIPS) 17, L. Saul, Y. Weiss, and L. Bottou, eds., Morgan Kaufmann, San Mateo, CA, 2005.

[3] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, vol. 2 of MPS/SIAM Series on Optimization, Society of Industrial and Applied Mathematics, Philadelphia, 2001.

[4] D. P. Bertsekas, *Necessary and sufficient conditions for a penalty method to be exact*, Math. Program., 9 (1975), pp. 87–99.
[5] ———, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
[6] ———, *A note on error bounds for convex and nonconvex programs*, Comput. Optim. Appl., 12 (1999), pp. 41–51.
[7] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Belmont, MA, 2003.
[8] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, 2004.
[9] J. V. Burke, *An exact penalization viewpoint of constrained optimization*, SIAM J. Control Optim., 29 (1991), pp. 968–998.
[10] J. V. Burke and S. Deng, *Weak sharp minima revisited. II. Application to linear regularity and error bounds*, Math. Program., 104 (2005), pp. 235–261.
[11] J. V. Burke and M. C. Ferris, *Weak sharp minima in mathematical programming*, SIAM J. Control Optim., 31 (1993), pp. 1340–1359.
[12] E. J. Candés, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements.* To appear in Comm. Pure Appl. Math., 2005.
[13] ———, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Info. Theory, 52 (2006), pp. 489–509.
[14] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev., 43 (2001), pp. 129–159.
[15] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust-Region Methods*, MPS-SIAM Series on Optimization, Society of Industrial and Applied Mathematics, Philadelphia, 2000.
[16] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization*, Proc. Nat. Acad. Sci. USA, 100 (2003), pp. 2197–2202.
[17] D. L. Donoho, M. Elad, and V. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Info. Theory, 52 (2006), pp. 6–18.
[18] D. L. Donoho and J. Tanner, *Sparse nonnegative solution of underdetermined linear equations by linear programming*, Proc. Nat. Acad. Sci. USA, 102 (2005), pp. 9446–9451.
[19] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Ann. Statist., 32 (2004), pp. 407–499.
[20] M. C. Ferris and O. L. Mangasarian, *Finite perturbation of convex programs*, Appl. Math. Optim., 23 (1991), pp. 263–273.
[21] R. Fletcher, *An ℓ_1 penalty method for nonlinear constraints*, in Numerical Optimization 1984, P. T. Boggs, R. H. Byrd, and R. B. Schnabel, eds., Philadelphia, 1985, Society of Industrial and Applied Mathematics, pp. 26–40.
[22] ———, *Practical Methods of Optimization*, John Wiley and Sons, New York, second ed., 1987.
[23] C. C. Gonzaga, *Generation of degenerate linear programming problems*, tech. rep., Dept. of Mathematics, Federal University of Santa Catarina, Brazil, April 2003.
[24] S.-P. Han and O. L. Mangasarian, *Exact penalty functions in nonlinear programming*, Math. Program., 17 (1979), pp. 251–269.
[25] C. Kanzow, H. Qi, and L. Qi, *On the minimum norm solution of linear programs*, J. Optim. Theory Appl., 116 (2003), pp. 333–345.
[26] S. Lucidi, *A finite algorithm for the least two-norm solution of a linear program*, Optimization, 18 (1987), pp. 809–823.
[27] ———, *A new result in the theory and computation of the least-norm solution of a linear program*, J. Optim. Theory Appl., 55 (1987), pp. 103–117.
[28] Z.-Q. Luo and J.-S. Pang, *Error bounds for analytic systems and their applications*, Math. Program., 67 (1994), pp. 1–28.
[29] ———, eds., *Error bounds in mathematical programming,* Math. Program., vol. 88, 2000.
[30] O. L. Mangasarian, *Normal solutions of linear programs*, Math. Program. Study, 22 (1984), pp. 206–216.
[31] ———, *Sufficiency of exact penalty minimization*, SIAM J. Control Optim., 23 (1985), pp. 30–37.
[32] ———, *A simple characterization of solution sets of convex programs*, Oper. Res. Lett., 7 (1988), pp. 21–26.
[33] ———, *A Newton method for linear programming*, J. Optim. Theory Appl., 121 (2004), pp. 1–18.
[34] O. L. Mangasarian and R. R. Meyer, *Nonlinear perturbation of linear programs*, SIAM J. Control Optim., 17 (1979), pp. 745–752.
[35] Y. E. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Pro-*

gramming, vol. 14 of Studies in Applied Mathematics, Society of Industrial and Applied Mathematics, Philadelphia, 1994.
[36] *NETLIB linear programming library*, 2006.
[37] G. Pataki and S. Schmieta, *The DIMACS library of semidefinite-quadratic-linear programs*, Tech. Rep. Preliminary draft, Computational Optimization Research Center, Columbia University, New York, July 2002.
[38] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*, MPS/SIAM Series on Optimization, Society of Industrial and Applied Mathematics, Philadelphia, 2001.
[39] S. M. Robinson, *Local structure of feasible sets in nonlinear programming, part II: Nondegeneracy*, Math. Program. Study, 22 (1984), pp. 217–230.
[40] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
[41] S. Sardy and P. Tseng, *On the statistical analysis of smoothing by maximizing dirty markov random field posterior distributions*, J. Amer. Statist. Assoc., 99 (2004), pp. 191–204.
[42] M. A. Saunders, *Cholesky-based methods for sparse least squares: The benefits of regularization*, in Linear and Nonlinear Conjugate Gradient-Related Methods, L. Adams and J. L. Nazareth, eds., Society of Industrial and Applied Mathematics, Philadelphia, 1996, pp. 92–100.
[43] J. F. Sturm, *Using Sedumi 1.02, a Matlab toolbox for optimization over symmetric cones (updated for Version 1.05)*, tech. rep., Department of Econometrics, Tilburg University, Tilburg, The Netherlands, August 1998 – October 2001.
[44] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, J. Royal. Statist. Soc B., 58 (1996), pp. 267–288.
[45] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston and Sons, Washington, D.C., 1977. Translated from Russian.
[46] Z. Wu and J. J. Ye, *On error bounds for lower semicontinuous functions*, Math. Program., 92 (2002), pp. 301–314.
[47] Y. Ye, *Interior-Point Algorithms: Theory and Analysis*, Wiley, Chichester, UK, 1997.
[48] Y.-B. Zhao and D. Li, *Locating the least 2-norm solution of linear programs via a path-following method*, SIAM J. Optim., 12 (2002), pp. 893–912.