Department of Computer Science Tech. Rep. TR-2005-31, December 2005
University of British Columbia

EXACT REGULARIZATION OF LINEAR PROGRAMS

MICHAEL P. FRIEDLANDER*

Abstract. We show that linear programs (LPs) admit regularizations that either contract the original (primal) solution set or leave it unchanged. Any regularization function that is convex and has compact level sets is allowed—differentiability is not required. This is an extension of the result first described by Mangasarian and Meyer (*SIAM J. Control Optim.*, 17(6), pp. 745–752, 1979). We show that there always exist positive values of the regularization parameter such that a solution of the regularized problem simultaneously minimizes the original LP and minimizes the regularization function over the original solution set. We illustrate the main result using the nondifferentiable ℓ_1 regularization function on a set of degenerate LPs. Numerical results demonstrate how such an approach yields sparse solutions from the application of an interior-point method.

Key words. linear programming, regularization, degeneracy, sparse solutions, interior-point methods

1. Introduction.

Consider the feasible linear program (LP)

$$\text{(P)} \qquad \begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b, \quad x \geq 0, \end{array}$$

with bounded optimal value, where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. We show that (P) always admits perturbations of its objective function such that solutions of the perturbed problem, henceforth referred to as the *regularized LP*, continue to be solutions of (P). In fact, a solution of the regularized LP simultaneously solves (P) and minimizes the regularization function over the optimal solution set of (P). We prove this result for any regularization function that is convex and has compact level sets—differentiability is not required. This may be regarded as an extension of Mangasarian and Meyer's result [MM79, Theorem 1], which additionally requires differentiability of the regularization function.

Consider any regularization function $\phi : \mathbb{R}^n \to \mathbb{R}$ that is compact and has compact level sets. The regularized LP is given by

$$\text{(P}_\delta\text{)} \qquad \begin{array}{ll} \underset{x}{\text{minimize}} & c^T x + \delta\phi(x) \\ \text{subject to} & Ax = b, \quad x \geq 0, \end{array}$$

where δ is a constant nonnegative regularization parameter. The regularization function ϕ may be nonlinear and/or nondifferentiable, so that (P$_\delta$) is not necessarily an LP (although it is convex). In §3 we prove that solutions of (P$_\delta$) are also solutions of (P) for all values of δ below some positive threshold value. Moreover, this set of solutions minimizes ϕ over the optimal solution set of (P). Interestingly, this positive threshold

*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.
December 26, 2005

value always exists. Our proof is constructive, and we show how this threshold value may be computed.

We call such regularizations *exact*, and choose this term to draw an analogy with exact penalty functions that are commonly used in algorithms for nonlinear optimization. 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, [HM79, Ber82, Fle85], and [CGT00] for a more recent discussion.) As we have described, an analogous property holds for LPs: a solution of the regularized problem is a solution of the original problem for all values of the regularization parameter less than some threshold value.

Regularization is a technique used commonly for solving ill-posed problems whose solutions may not be unique or may be acutely sensitive to the problem data. Regularization amounts to changing the problem statement in order to compute (usually approximate) solutions that are "well-behaved." For example, approximate solutions with moderate norms may be preferred over very-large-norm solutions; or solutions to nearby and stable problems may be preferred. (See [Han98] for a comprehensive study of regularization for linear systems, including least-squares problems.)

As with other problem classes, exact regularization of LPs may be useful for various reasons. If an LP does not have a unique optimal solution, exact regularization may be used to select solutions with desirable properties. In particular, Tikhonov [TA77] regularization can be used to select a minimum two-norm solution. Specialized algorithms for computing minimum two-norm solutions of LPs have been proposed by [Man84, Luc87a, Luc87b, ZL02, KQQ03, Man04], among others. Saunders and Tomlin [ST96] and Altman and Gondzio [AG99] consider Tikhonov regularization as a tool for influencing the conditioning of the underlying linear systems that arise in the implementation of large-scale interior-point methods.

If we were to assume differentiability of ϕ, then our main result may be obtained by using the KKT conditions of (P) and (P_δ). This is the approach used by Mangasarian and Meyer [MM79]. We can discard the differentiability assumption, however, by instead appealing to Lagrange duality theory. The advantage of this approach is evident for the need in many applications for nondifferentiable regularization functions.

A vital example is ℓ_1 regularization, which has lately received significant attention in applications closely related to linear programming. Recent work related to signal processing has focused on using LPs to obtain sparse solutions (i.e., solutions with few nonzero elements) of underdetermined systems of linear equations $Ax = b$ (with the possible additional condition $x \geq 0$); for examples, see [CDS01, CRT04, CRT05, DT05]. 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 [Tib96, EHJT04].

To illustrate our main result, we consider the nondifferentiable ℓ_1 regularization function as a technique for selecting sparse solutions of LPs that do not have unique primal solutions (i.e., LPs that are dual degenerate). In §4 we give numerical results of ℓ_1 regularization on a set of standard LPs from the Netlib test set and on a set of randomly generated LPs with prescribed degeneracy. The numerical results highlight the effectiveness of this approach for encouraging sparse solutions obtained via an interior-point method. With the main result described in this paper, we can select a regularization parameter δ so that the solution of the regularized LP gives a minimum ℓ_1 solution over the original LP optimal solution set.

2. Preliminaries. The following assumptions hold implicitly throughout:

ASSUMPTION 2.1 (Feasibility and finiteness). *The feasible set is nonempty—i.e., there exists a point $\bar{x}$ such that $A\bar{x} = b$ and $\bar{x} \geq 0$. Moreover, the optimal value of (P) is bounded—i.e., $p^* > -\infty$, where p^* is the optimal value of (P).*

ASSUMPTION 2.2 (Convexity and compactness). *The regularization function ϕ is convex, and the level sets $\{x \mid \phi(x) \leq \beta\}$ are compact for each β.*

The first assumption guarantees that (P) and its dual have optimal solutions. The first and second assumptions together ensure that (P_δ) attains its optimal value and that strong duality holds. (See, for example, Propositions 5.2.1 and 5.2.2 of [Ber99] for these well-known results on linearly constrained convex programs and LPs, respectively.)

Our main result hinges on relating the optimal solutions of (P) and (P_δ) via the (convex) optimization problem

$$(\mathrm{P}_\phi) \qquad \begin{array}{ll} \underset{x}{\text{minimize}} & \phi(x) \\ \text{subject to} & Ax = b, \quad c^T x \leq p^*, \quad x \geq 0. \end{array}$$

By construction, any solution of (P_ϕ) is also a solution of (P). (The converse, however, does not generally hold.) We show that for all values of δ (including 0) below a certain *positive* threshold value, every solution of (P_δ) is also a solution of (P_ϕ), and hence also a solution of (P). The dual solutions of one problem are linear combinations of the dual solutions of the other problem; the linear combination depends on δ and on the Lagrange multiplier corresponding to the constraint $c^T x \leq p^*$. Note that Lagrange multipliers for (P_ϕ) always exist because its constraints are linear. (Again, see [Ber99, Proposition 5.2.1].)

We make the following definitions: The Lagrangian functions of (P_δ) and (P_ϕ) are respectively given by

$$\begin{aligned} L_\delta(x, y, z) &= c^T x + \delta\phi(x) - y^T(Ax - b) - z^T x, \\ L_\phi(x, y, z, \mu) &= \phi(x) - y^T(Ax - b) + \mu(c^T x - p^*) - z^T x, \end{aligned}$$

where $y \in \mathbb{R}^m$, $z \in \mathbb{R}^n$, and $\mu \in \mathbb{R}$ are independent variables. The associated Lagrange dual functions are given by

$$\begin{aligned} g_\delta(y, z) &= \inf_{x \in \mathbb{R}^n} L_\delta(x, y, z), \\ g_\phi(y, z, \mu) &= \inf_{x \in \mathbb{R}^n} L_\phi(x, y, z, \mu). \end{aligned}$$

Denote by $\mathcal{S}$ and $\mathcal{S}_\delta$ the set of (primal) solutions of (P) and (P_δ), respectively. Assumption 2.1 ensures that $\mathcal{S}$ and $\mathcal{S}_\delta$ are nonempty. Denote by p_δ^* the optimal value of (P_δ), and as defined above, p^* as the optimal value of (P).

3. Main result.

The following theorem is a precise statement of our main result.

THEOREM 3.1. *There exists a constant $\bar{\delta} > 0$ such that $\mathcal{S}_\delta \subseteq \mathcal{S}$ for all $\delta \in [0, \bar{\delta}]$. Moreover, each element in $\mathcal{S}_\delta$ minimizes ϕ over $\mathcal{S}$.*

Proof. We first establish a few key properties of the dual optimal values of (P) and (P_ϕ). Assumption 2.1 implies that there exist optimal dual variables y^* and $z^* \geq 0$ of (P) that satisfy (among other conditions)

$$A^T y^* + z^* = c \qquad \text{and} \qquad b^T y^* = p^*. \tag{3.1}$$

Assumptions 2.1 and 2.2 imply that there exist optimal dual variables y_ϕ^*, $z_\phi^* \geq 0$, and $\mu_\phi^* \geq 0$, of (P_ϕ) such that

$$g_\phi(y_\phi^*, z_\phi^*, \mu_\phi^*) = b^T y_\phi^* - \mu_\phi^* p^* + \inf_{x\in\mathbb{R}^n} \phi(x) - (A^T y_\phi^* + z_\phi^* - \mu_\phi^* c)^T x \tag{3.2}$$

is maximal. Hence, the infimum on the right-hand side of (3.2) attains its minimum at some point x_ϕ^*. (Recall that ϕ has compact level sets over $\mathbb{R}^n$.) By strong duality, x_ϕ^* is optimal for (P_ϕ), and

$$\phi(x_\phi^*) = g_\phi(y_\phi^*, z_\phi^*). \tag{3.3}$$

We next consider the regularized problem (P_δ). Its feasible set coincides with (P), so the assumptions ensure that there exist optimal dual variables y_δ^* and $z_\delta^* \geq 0$. We show that these optimal dual variables are in fact linear combinations of the optimal dual variables of (P) and (P_ϕ). The linear combination appears in two forms, depending on whether μ_ϕ^* is zero or positive. We consider these two cases in turn.

Case 1: $\mu_\phi^* = 0$. We claim that the optimal dual variables of (P_δ) are given by

$$y_\delta^* := y^* + \delta y_\phi^*, \qquad z_\delta^* := z^* + \delta z_\phi^*. \tag{3.4}$$

Clearly $z_\delta^* \geq 0$, as required. Evaluate $g_\delta(y, z)$ with the pair (y_δ^*, z_δ^*), and use (3.4) to obtain

$$\begin{aligned} g_\delta(y_\delta^*, z_\delta^*) &\equiv \inf_{x\in\mathbb{R}^n} L_\delta(x, y_\delta^*, z_\delta^*) \\ &= b^T y_\delta^* + \inf_{x\in\mathbb{R}^n} \delta\phi(x) - (A^T y_\delta^* + z_\delta^* - c)^T x \\ &= b^T y^* + \delta b^T y_\phi^* + \inf_{x\in\mathbb{R}^n} \delta\phi(x) - \{A^T y^* + z^* - c + \delta(A^T y_\phi^* + z_\phi^*)\}^T x \\ &= p^* + \delta\big(b^T y_\phi^* + \inf_{x\in\mathbb{R}^n} \phi(x) - (A^T y_\phi^* + z_\phi^*)^T x\big), \end{aligned} \tag{3.5}$$

in which we use (3.1) to arrive at the last expression. Moreover, (3.2) and (3.3) (with $\mu_\phi^* = 0$) imply that the infimum on the right-hand-side of (3.5) achieves its minimum at the point x_ϕ^*. Substitute (3.2) and (3.3) into (3.5) to obtain

$$g_\delta(y_\delta^*, z_\delta^*) = p^* + \delta\phi(x_\phi^*). \tag{3.6}$$

Note that x_ϕ^* is feasible for (P_ϕ) and for (P_δ), so that

$$p^* \geq c^T x_\phi^* \qquad \text{and} \qquad c^T x_\phi^* + \delta\phi(x_\phi^*) \geq p_\delta^*. \tag{3.7}$$

It then follows from (3.6) and (3.7) that $g_\delta(y_\delta^*, z_\delta^*) \geq p_\delta^*$. By weak duality, however, $g_\delta(y_\delta^*, z_\delta^*) \leq p_\delta^*$, so that in fact $g_\delta(y_\delta^*, z_\delta^*) = p_\delta^*$.

This proves that strong duality holds, and that the dual optimum is attained, at the pair (y_δ^*, z_δ^*) defined by (3.4). Therefore, (3.5) and (3.6) together imply that $\inf_{x\in\mathbb{R}^n} L_\delta(x, y_\delta^*, z_\delta^*)$ attains its minimum at x_ϕ^*, so that in fact x_ϕ^* is a solution of (P_δ). Moreover, x_ϕ^* is feasible for (P), so (3.7) implies that it is also a solution of (P). We therefore conclude that, when $\mu_\phi^* = 0$, then x_ϕ^* (and in fact, every solution of (P_ϕ)) is an optimal solution of (P_δ) and also of (P), for all values of $\delta \geq 0$. In this case, $\bar{\delta} = +\infty$.

Case 2: $\mu_\phi^* > 0$. We claim that the optimal dual variables of (P_δ) are given by

$$y_\delta^* := (1-\lambda)y^* + \frac{\lambda}{\mu_\phi^*} y_\phi^*, \qquad z_\delta^* := (1-\lambda)z^* + \frac{\lambda}{\mu_\phi^*} z_\phi^*, \tag{3.8}$$

for any $\lambda \in [0,1]$. Clearly $z_\delta^* \geq 0$, as required. Evaluate $g_\delta(y,z)$ with this pair, and consider values of $\delta \equiv \lambda/\mu_\phi^*$, to get

$$\begin{aligned} g_\delta(y_\delta^*, z_\delta^*) &\equiv \inf_{x\in\mathbb{R}^n} L_\delta(x, y_\delta^*, z_\delta^*) \\ &= b^T y_\delta^* + \inf_{x\in\mathbb{R}^n} \delta\phi(x) - (A^T y_\delta^* + z_\delta^* - c)^T x \\ &= (1-\lambda) b^T y^* + \frac{\lambda}{\mu_\phi^*} b^T y_\phi^* \\ &\qquad + \inf_{x\in\mathbb{R}^n} \delta\phi(x) - \left((1-\lambda)(A^T y^* + z^*) - c + \frac{\lambda}{\mu_\phi^*}(A^T y_\phi^* + z_\phi^*) \right)^T x \\ &= (1-\lambda)p^* + \frac{\lambda}{\mu_\phi^*} b^T y_\phi^* + \inf_{x\in\mathbb{R}^n} \frac{\lambda}{\mu_\phi^*}\phi(x) - \left(\frac{\lambda}{\mu_\phi^*}(A^T y_\phi^* + z_\phi^*) - \lambda c \right)^T x \\ &= p^* + \frac{\lambda}{\mu_\phi^*} \left(b^T y_\phi^* - \mu_\phi^* p^* + \inf_{x\in\mathbb{R}^n} \phi(x) - (A^T y_\phi^* + z_\phi^* - \mu_\phi^* c)^T x \right), \end{aligned} \tag{3.9}$$

where we use (3.1) to arrive at the second-to-last inequality. As we noted for (3.5), the last infimum of the right-hand side of (3.9) achieves its minimum at the point x_ϕ^* (but this time with $\mu_\phi^* > 0$). Substitute (3.2) and (3.3) into (3.9) to obtain

$$g_\delta(y_\delta^*, z_\delta^*) = p^* + \frac{\lambda}{\mu_\phi^*}\phi(x_\phi^*). \tag{3.10}$$

With the same arguments as used for Case 1, we can conclude that, when $\mu_\phi^* > 0$, then x_ϕ^* (and in fact, every solution of (P_ϕ)) is an optimal solution of (P_δ) and also of (P), for all values of

$$\delta \in [0, \bar{\delta}] \quad \text{with} \quad \bar{\delta} := \frac{1}{\mu_\phi^*} > 0. \tag{3.11}$$

The last statement of the theorem follows immediately from the fact that x_ϕ^* solves both (P) and (P_ϕ). □

Note that the range of values admitted by (3.11) is consistent with the extreme case $\mu_\phi^* \searrow 0$. In that situation, all nonnegative values of the penalty parameter are allowed by Theorem 3.1, as is predicted when $\mu_\phi^* = 0$ (Case 1 of the theorem).

4. Sparse solutions. In this section we illustrate a practical application of Theorem 3.1. The aim is to obtain sparse solutions of linear programs that do not have unique optimal solutions. For the following discussion we let $\phi(x) = \|x\|_1$, which clearly satisfies both standing assumptions.

Regularization based on the ℓ_1 norm has been used in many applications with the goal of obtaining (or approximating) the *sparsest* solution of underdetermined systems of linear equations and least-squares problems. Some recent examples include [CDS01, DE03, DT05, DET05].

For underdetermined systems of equations that arise in fields such as signal processing, [CDS01], [CRT04], and [DT05] advocate solving the problem

$$\underset{x}{\text{minimize}}\ \|x\|_1 \quad \text{subject to} \quad Ax = b \quad (\text{and possibly } x \geq 0) \tag{4.1}$$

in order to obtain a sparse solution. (Well-known techniques exist for recasting (4.1) as an LP.) The sparsest solution is in fact given by minimizing the so-called 0-norm $\|x\|_0$, which counts the number of nonzero elements in x. The combinatorial nature of such a problem, however, makes it computationally intractable for all but the most trivial cases. Interestingly, there exist conditions under which a solution of (4.1) is the *sparsest* feasible solution; see [CRT04] and [DT05].

Following this approach, we use Theorem 3.1 as a guide for obtaining least ℓ_1-norm solutions of a generic LP

$$\underset{x}{\text{minimize}}\ c^T x \quad \text{subject to} \quad Ax = b, \quad l \leq x \leq u, \tag{4.2}$$

by solving its regularized version:

$$\underset{x}{\text{minimize}}\ c^T x + \delta\|x\|_1 \quad \text{subject to} \quad Ax = b, \quad l \leq x \leq u. \tag{4.3}$$

(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 (4.2) (given by (4.3) for small-enough values of δ) is considerably sparser than the solution obtained by solving (4.2) directly. In each case, we solve the regularized and unregularized problems with the same interior-point solver. We emphasize that, with the appropriate choice of the regularization parameter, the solution of the regularized LP is *also* a solution of the original LP.

We consider two sets of test problems in our numerical experiments. The problems of the first set are constructed from random data using a degenerate LP generator described by [Gon03]. Those of the second are derived from the infeasible LPs in the Netlib collection (`http://www.netlib.org/lp/infeas/`). Both sets of test problems are further described in §§4.1–4.2.

For each test problem we follow the same procedure. We first reformulate the problem as an LP; this problem corresponds to (P). (There may, of course, be upper and lower bounds on x.) We solve the unregularized LP in order to obtain x^* (an unregularized solution), and thus the optimal value $p^* := c^T x^*$. Next, in order to obtain the threshold value $\bar{\delta}$ of the regularization parameter, we solve the LP that corresponds to (P_ϕ). (In the nontrivial case, this value is given by (3.11).) Finally, an exact regularized solution x_δ^* is obtained by solving (P_δ) with $\delta := \bar{\delta}/2$.

We use the interior-point algorithm implemented in CPLEX 9.1 to solve each subproblem. The default CPLEX options are used, except for `crossover = 0` and `comptol = 1e-10`. The first option forces CPLEX to use its barrier algorithm and to not "cross over" to an optimal basic solution. Under certain conditions, we expect the interior method to converge to the analytic center of the optimal face (see [Ye97, Theorems 2.16 and 2.17]). The second 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 this change aids in computing the sparsity of a computed solution, which we determine as

$$\|x\|_0 = \text{card}\{x_i \mid |x_i| > \sqrt{\epsilon}\}. \tag{4.4}$$

TABLE 4.1
Random LPs generated with various levels of increasing dimension of the optimal primal face.

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	9.1e+01	100	100	1.5e−04
random-20	5.6e−13	6.6e−13	2.9e+02	2.0e+02	278	100	2.2e−02
random-40	3.8e−12	3.7e−12	4.9e+02	2.9e+02	459	100	2.9e−02
random-60	3.9e−14	9.2e−11	6.7e+02	3.6e+02	637	101	3.3e−02
random-80	9.1e−12	8.4e−13	8.9e+02	4.6e+02	816	100	2.1e−01
random-100	1.8e−16	3.2e−12	1.0e+03	5.4e+02	997	102	1.1e−01

Here, $\epsilon \approx 2.2 \cdot 10^{-16}$ is the relative machine precision. The value $\sqrt{\epsilon}$ is slightly larger than the specified convergence tolerance.

The AMPL model and data files used to generate all of the numerical results presented in §§4.1–4.2 can be obtained at `http://www.cs.ubc.ca/~mpf/exactreg/`.

4.1. Random LPs. Six dual-degenerate LPs were constructed using Gonzaga's MATLAB generator [Gon03]. 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

$$0 \le D_p \le n - m - 1 \qquad \text{and} \qquad 0 \le D_d \le m - 1. \tag{4.5}$$

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 (4.5)). The problems are respectively labeled *random-0*, *random-20*, *random-40*, and so on.

Table 4.1 summarizes the results. We confirm that in each case the optimal values of the unregularized and regularized problems are nearly identical (at least to within the specified tolerance). Except for the "control" problem *random-0*, the exact regularized solution x_δ^* has a strictly smaller ℓ_1 norm, and is considerably sparser, than the unregularized solution x^*.

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

$$\underset{x}{\text{minimize}} \ \|Ax - b\|_1 \qquad \text{subject to} \qquad l \le x \le u, \tag{Pi}$$

and its regularized counterpart

$$\underset{x}{\text{minimize}} \ \|Ax - b\|_1 + \delta\|x\|_1 \qquad \text{subject to} \qquad l \le x \le u. \tag{Pi$_\delta$}$$

The unregularized problem (Pi) mimics the plausible situation where we wish to fit a set of infeasible equations in the ℓ_1-norm sense. But because the ℓ_1 norm is not strictly convex, a solution of (Pi) may not be unique. The regularized problem (Pi$_\delta$) might be used in practice to further restrict that solution set.

The following infeasible Netlib LPs were eliminated because CPLEX returned some error during the solution of (Pi) or (Pi$_\delta$): *lpi-bgindy*, *lpi-cplex2*, *lpi-gran*, *lpi-klein1*, *lpi-klein2*, *lpi-klein3*, *lpi-qual*, *lpi-refinery*, *lpi-vol1*.

Table 4.2 summarizes the results. The exact regularized solution x_δ^* has a smaller or equal ℓ_1 norm than the unregularized solution x^* in all cases. In 60% of the

TABLE 4.2
Least ℓ_1 feasible solutions of the infeasible Netlib LPs.

LP	$c^T x^*$	$c^T x_\delta^*$	$\|x^*\|_1$	$\|x_\delta^*\|_1$	$\|x^*\|_0$	$\|x_\delta^*\|_0$	$\bar{\delta}$
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

cases, the regularized solution is sparser than the unregularized solution. In 25% of the cases, the solutions have the same amount of sparsity. In 15% of the cases (*lpi-galenet*, *lpi-itest6*, and *lpi-woodinfe*), the regularized solutions are actually less sparse, even though the ℓ_1 norm is lower.

5. Discussion. We emphasize that our main result is constructive, and offers a procedure for determining the threshold value of the regularization parameter. However, we do not know how the threshold value might be computed directly from (P) without also solving (P_ϕ). It seems that admissible values of δ can in general only be determined by first solving (P) to obtain p^*, and subsequently solving (P_ϕ) to obtain μ_ϕ^*, and thus $\bar{\delta}$ (c.f. (3.11)).

Suppose that a correct value of δ has been guessed, and an exact regularized solution x obtained. The dual variables (y, z) obtained from that regularized problem are necessarily perturbed (c.f. expressions (3.4) and (3.8)). Therefore, it is not possible to test the computed triple (x, y, z) against the optimality conditions of the original LP in order to verify that x is indeed an exact solution.

In practice, if it were prohibitively expensive to solve (P) and (P_ϕ) additionally, we might adopt an approach suggested by [Luc87b] and [Man04] for Tikhonov regularization. Lucidi and Mangasarian suggest successively solving the regularized LP 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 cases, the threshold values $\bar{\delta}$ shown in Tables 4.1 and 4.2 are comfortably large, and a value such as $\delta = 10^{-4}$ would cover 85% of the these cases.

Acknowledgments. I wish to thank my colleague Kevin Leyton-Brown for giving me access to his CPLEX installation for the numerical experiments in §4. Also, my

sincere thanks to Adam Oberman for many fruitful discussions and comments on earlier versions of this manuscript.

REFERENCES

[AG99] A. Altman and J. Gondzio, *Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization*, Optimization Methods and Software **11** (1999), 275–302.

[Ber82] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Academic Press, New York, 1982.

[Ber99] ———, *Nonlinear programming*, second ed., Athena Scientific, Belmont, MA, 1999.

[CDS01] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev. **43** (2001), no. 1, 129–159.

[CGT00] 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.

[CRT04] E. J. Candés, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, Tech. report, Applied and Computational Analysis, California Institute of Technology, 2004.

[CRT05] ———, *Stable signal recovery from incomplete and inaccurate measurements*, To appear in Comm. Pure Appl. Math., 2005.

[DE03] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization*, Proc. Nat. Acad. Sci. USA **100** (2003), no. 5, 2197–2202.

[DET05] D. L. Donoho, M. Elad, and V. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, To appear in IEEE Trans. Info. Theory, 2005.

[DT05] D. L. Donoho and J. Tanner, *Sparse nonnegative solution of underdetermined linear equations by linear programming*, Proc. Nat. Acad. Sci. USA **102** (2005), no. 27, 9446–9451.

[EHJT04] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, The Annals of Statistics **32** (2004), no. 2, 407–499.

[Fle85] R. Fletcher, *An ℓ_1 penalty method for nonlinear constraints*, Numerical Optimization 1984 (Philadelphia) (P. T. Boggs, R. H. Byrd, and R. B. Schnabel, eds.), Society of Industrial and Applied Mathematics, 1985, pp. 26–40.

[Gon03] C. C. Gonzaga, *Generation of degenerate linear programming problems*, Tech. report, Dept. of Mathematics, Federal University of Santa Catarina, Brazil, April 2003.

[Han98] P. C. Hansen, *Rank-deficient and discrete ill-posed problems*, Society of Industrial and Applied Mathematics, Philadelphia, 1998.

[HM79] S.-P. Han and O. L. Mangasarian, *Exact penalty functions in nonlinear programming*, Math. Prog. **17** (1979), 251–269.

[KQQ03] C. Kanzow, H. Qi, and L. Qi, *On the minimum norm solution of linear programs*, J. Optim. Theory Appl. **116** (2003), 333–345.

[Luc87a] S. Lucidi, *A finite algorithm for the least two-norm solution of a linear program*, Optimization **18** (1987), no. 6, 809–823.

[Luc87b] ———, *A new result in the theory and computation of the least-norm solution of a linear program*, J. Optim. Theory Appl. **55** (1987), no. 1, 103–117.

[Man84] O. L. Mangasarian, *Normal solutions of linear programs*, Math. Prog. Stud. **22** (1984), 206–216.

[Man04] ———, *A Newton method for linear programming*, J. Optim. Theory Appl. **121** (2004), 1–18.

[MM79] O. L. Mangasarian and R. R. Meyer, *Nonlinear perturbation of linear programs*, SIAM J. Control Optim. **17** (1979), no. 6, 745–752.

[ST96] M. A. Saunders and J. A. Tomlin, *Solving regularized linear programs using barrier methods and KKT systems*, Tech. Report SOL Report 96-4, Stanford University, Department of EES & OR, Stanford, CA 94305, December 1996, Also IRM Research Report RJ 10064.

[TA77] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*, V. H. Winston and Sons, Washington, D.C., 1977, Translated from Russian.

[Tib96] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, J. Royal. Statist. Soc B. **58** (1996), no. 1, 267–288.

[Ye97] Y. Ye, *Interior-point algorithms: Theory and analysis*, Wiley, Chichester, UK, 1997.

[ZL02] 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), no. 4, 893–912.