Cahier du GERAD G-2015-00

A REGULARIZED INTERIOR-POINT METHOD FOR CONSTRAINED LINEAR LEAST SQUARES

MOHSEN DEHGHANI AND DOMINIQUE ORBAN

Abstract. We propose an infeasible interior-point algorithm for constrained linear least-squares problems based on the primal-dual regularization of convex programs of Friedlander and Orban (2012). At each iteration, the sparse LDL$^{\mathrm{T}}$ factorization of a symmetric quasi-definite matrix is computed. This matrix is shown to be uniformly bounded and nonsingular. We establish conditions under which a solution of the original problem is recovered. Regularization allows us to dispense with the assumption that the active gradients are linearly independent. Although the implementation described here is factorization based, it paves the way to a matrix-free implementation in which a regularized unconstrained linear least-squares problem is solved at each iteration. We report computational experience and illustrate the potential advantages of our approach.

1. Introduction

We are concerned with the constrained linear least-squares problem in standard form:

eq:lsq
$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ c^T x + \tfrac{1}{2}\|Ax-d\|^2 \quad \text{subject to}\ Bx = b,\ x \geq 0, \tag{1.1}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{p\times n}$, $d \in \mathbb{R}^p$, $B \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and inequalities are understood componentwise. It is typically assumed that $p > n$ and $m < n$ but the approach proposed in this paper allows us to do away with these restrictions. If $A = 0$, (1.1) reduces to the linear programming problem in standard form. In all cases, (1.1) is a convex quadratic program. Numerous applications give rise to (1.1), including signal decomposition by basis pursuit (Chen et al., 1998), ℓ_1-regularized linear least squares (Kim et al., 2007), compressed sensing (Donoho, 2006), and machine learning (Koh et al., 2007).

An interior-point method applied directly to (1.1) might suffer several difficulties. Firstly, the matrix A^TA, which may be rather dense and ill conditioned, will appear explicitly in the Newton step computation. Secondly, numerical instabilities will arise if the constraint matrix B does not have full row rank. We remove the first difficulty in two different ways that lead to two slightly different implementations. The second difficulty disappears if we consider the following regularization of (1.1)

Date: January 20, 2015.

2010 *Mathematics Subject Classification.* 90C06, 90C20, 90C30, 90C51, 90C53, 90C55, 65F10, 65F50.

Key words and phrases. linear least squares, linear constraints, symmetric quasi-definite system, interior-point method, primal-dual regularization, proximal point, augmented Lagrangian.

Research partially supported by NSERC Discovery Grant 299010-04.

proposed by Friedlander and Orban (2012):

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n, w\in\mathbb{R}^m}{\text{minimize}} & c^T x + \frac{1}{2}\|Ax - d\|^2 + \frac{1}{2}\rho\|x - x_k\|^2 + \frac{1}{2}\delta\|w + y_k\|^2 \\ \text{subject to} & Bx + \delta w = b,\ x \geq 0, \end{array}$$

where $\rho > 0$ and $\delta > 0$ are regularization parameters, x_k and y_k are the current approximations of the optimal primal variables and Lagrange multipliers, respectively, and w are auxiliary variables playing the role of a constraint residual.

In this paper, we specialize the interior-point framework of Friedlander and Orban (2012) and apply it to (1.2) with ultimately constant regularization parameters. At each iteration, a step is computed by solving a large and sparse symmetric quasi-definite linear system (Vanderbei, 1995). Contrary to most interior-point implementations, partial block elimination is not applied to this system to reduce it to the so-called *augmented system* form or to the *normal equations*. Instead, a similarity transformation is applied that guarantees that the system remains uniformly bounded and nonsingular throughout the iterations and in the limit, provided strict complementarity is satisfied at a solution. We establish global convergence under weak assumptions. In particular, no assumption on the rank of B or A. A distinctive feature of the regularization (1.2) is that it allows to be recovered a solution of (1.1) in many situations, not a solution of a perturbed problem. In addition, (1.2) is never solved to optimality for fixed values of ρ, δ, x_k and y_k. Instead, it is used to compute a single Newton step before attention turns to the next regularized subproblem.

In (1.2), the *primal* regularization term $\frac{1}{2}\rho\|x - x_k\|^2$ serves the dual purpose of regularizing A whenever it is rank deficient and simplifying the implementation of the interior-point method in the presence of free variables. The *dual* regularization term $\frac{1}{2}\delta\|w + y_k\|^2$ regularizes B whenever it is rank–deficient.

The implementation proposed below relies on a sparse LDL^T factorization of the symmetric quasi-definite matrix. This factorization may be obtained at lower cost than the symmetric indefinite factorization, such as that of Duff (2004), and typically yields sparser factors. Its stability on symmetric quasi-definite systems has been analyzed by Gill et al. (1996).

Many applications only provide A and B in the form of linear operators instead of explicit matrices. Iterative methods specialized to symmetric quasi-definite systems have been proposed recently by Arioli and Orban (2012). Our algorithm paves the way to a matrix-free implementation using such iterative methods. This yields an elegant framework in which an unconstrained regularized linear least-squares problem must be solved at each iteration.

Our analysis and implementation differ from those of Friedlander and Orban (2012) in several respects. First, the linear systems used in the definition of the Newton steps are larger, sparser and tailored to the special structure of (1.1). If strict complementarity holds at a solution, they also have uniformly bounded condition number. Second, our approach illustrates how to apply the primal-dual regularization of Friedlander and Orban (2012) selectively, leaving some variables and some constraints untouched. This has the benefit of exploiting the structure of the problem at hand.

1.1. **Related Research.** Most of the literature concentrates on the box-constrained variant of (1.1), i.e., with no linear equality constraints. Portugal et al. (1994) compare infeasible active-set-type methods with interior-point algorithms and conclude that it is beneficial to use the normal equations formulation of the step-finding subproblem in the interior-point method. Bellavia et al. (2006) propose an inexact Newton scheme under the assumption that A has full rank but specifically target problems failing to satisfy strict complementarity at the solution. Bellavia et al. (2008) regularize the linear systems arising in the method of Bellavia et al. (2006) with the aim of decreasing their condition number. The resulting linear systems are subsequently partially eliminated, preconditioned and solved with an iterative method. Bellavia et al. (2011) propose and compare an interior-point method and a method derived from that of Barzilai and Borwein (1988).

Active-set methods for the box-constrained problems have also been actively researched. Among others, Bierlaire et al. (1991) describe three methods for reasonably well-conditioned problems, two of which are inspired by trust-region methods. Friedlander (2007) implements a dual-metric method in which search directions are linear combinations of the Newton direction and a scaled gradient direction.

Finally, background and pointers to more literature may be found in the seminal books of Lawson and Hanson (1995) and Björck (1996).

1.2. **Notation.** The notation X and Z is used to denote the diagonal matrices $\mathrm{diag}(x)$ and $\mathrm{diag}(z)$. The vector e denotes the vector of all ones of appropriate dimension. The notation $\|\cdot\|$ denotes the Euclidian norm throughout. The i-th component of a vector x is denoted $[x]_i$ while the value of x at the k-th iteration of a process is denoted x_k. For a given positive definite matrix M, the M-norm is defined as $\|x\|_M^2 = x^T M x$. The notation $\mathrm{blkdiag}(A_1, \ldots, A_k)$ denotes a block-diagonal matrix having the blocks A_1 through A_k consecutively on the diagonal. Whenever a block A_j is an identity block, its size is dictated by the context. For two related sequences $\{\alpha_k\}$ and $\{\beta_k\}$ of positive numbers, we write $\alpha_k = O(\beta_k)$ if there exists a constant $C > 0$ such that $\alpha_k \leq C\beta_k$ for all sufficiently large k. We write $\alpha_k = \Theta(\beta_k)$ if $\alpha_k = O(\beta_k)$ and $\beta_k = O(\alpha_k)$.

2. Background and Preliminaries

As a convex quadratic program, the dual of (1.1) may be written as the constrained linear least-squares problem

eq.lsq-dual

$$\begin{aligned} \underset{x,y,z}{\text{maximize}} \quad & b^T y - (A^T d)^T x - \tfrac{1}{2}\|Ax - d\|^2 \\ \text{subject to} \quad & B^T y + z - A^T A x = c - A^T d, \ z \geq 0, \end{aligned} \tag{2.1}$$

where $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ are the vectors of Lagrange multipliers associated with the equality constraints and bounds of (1.1), respectively.

Friedlander and Orban (2012) justify the regularization (1.2) of (1.1) as an application of the proximal method of multipliers of Rockafellar (1976). It consists of the addition of a proximal-point term $\frac{1}{2}\rho\|x - x_k\|^2$, which we will refer to as a *primal regularization* term, and of augmented Lagrangian terms consisting of the objective term $\frac{1}{2}\delta\|w + y_k\|^2$ and the constraint residual term δw, which we will collectively refer to as a *dual regularization* term. The regularized problem (1.2) is

still a convex quadratic program and its dual may be written as the regularized constrained linear least-squares problem

eq:lsq-reg-dual
$$\begin{aligned} \underset{x,s,y,z}{\text{maximize}} \quad & b^T y - (A^T d)^T x - \tfrac{1}{2}\|Ax - d\|^2 - \tfrac{1}{2}\delta\|y - y_k\|^2 - \tfrac{1}{2}\rho\|s + x_k\|^2 \\ \text{subject to} \quad & B^T y + z - A^T A x - \rho s = c - A^T d, \; z \geq 0, \end{aligned} \tag{2.2}$$

where $s = x - x_k$ are auxiliary variables playing the same role as w in (1.2). The strength of this regularization approach is that (2.2) coincides with the primal-dual regularization of (2.1).

For convenience, we let $u := (x, s, w, y, z)$ and we define the function

$$F_k(u; \rho, \delta, \tau) := \begin{bmatrix} c - A^T d - B^T y - z + A^T A x + \rho s \\ \rho x - \rho(s + x_k) \\ \delta y - \delta(w + y_k) \\ Bx + \delta w - b \\ Xz - \tau e \end{bmatrix},$$

where $\tau \geq 0$ is a parameter. Using this notation, the common necessary and sufficient optimality conditions of (1.2) and (2.2) may be written compactly as

$$F_k(u; \rho, \delta, 0) = 0, \qquad (x, z) \geq 0.$$

Note also that additionally setting $\rho = \delta = 0$ recovers the optimality conditions of the original primal-dual pair (1.1) and (2.1).

An interior-point method applied to the primal-dual pair (1.2) and (2.2) iteratively seeks approximate solutions to the nonlinear system

$$F_k(u; \rho, \delta, \tau_k) = 0, \quad (x, z) > 0,$$

for a sequence of positive parameters $\{\tau_k\} \downarrow 0$. For each fixed value of τ_k, a Newton step Δu is computed from the current approximation u_k as the solution of the linear system

$$\nabla_u F_k(u_k; \rho, \delta, \tau_k) \Delta u = -F_k(u_k; \rho, \delta, \tau_k),$$

where the Jacobian is given by

$$\nabla_u F_k(u_k; \rho, \delta, \tau_k) = \begin{bmatrix} A^T A & \rho I & & -B^T & -I \\ \rho I & -\rho I & & & \\ & & -\delta I & \delta I & \\ B & & \delta I & & \\ Z & & & & X \end{bmatrix}.$$

Of particular concern is that the matrix $A^T A$ is (nearly) dense if A has a (nearly) dense row. One way to circumvent this difficulty is to introduce $\xi := d - A\Delta x - Ax$. An equivalent way to write the previous system is then

$$\begin{bmatrix} & A^T & \rho I & -B^T & -I \\ A & I & & & \\ & & \vdots & & \end{bmatrix} \begin{bmatrix} \Delta x \\ \xi \\ \vdots \end{bmatrix} = \begin{bmatrix} -c + B^T y + z - \rho s \\ d - Ax \\ \vdots \end{bmatrix},$$

where the ellipses indicate that the rest of the system is unchanged. This system is larger but sparser and does away with the matrix-matrix product $A^T A$.

A more natural way to give rise to the previous sparse Jacobian is to systematically transform every problem of the form (1.1) to the form

eq:lsq-r
$$\begin{aligned} \underset{x,r}{\text{minimize}} \quad & c^T x + \tfrac{1}{2}\|r\|^2 \\ \text{subject to} \quad & Bx = b, \quad Ax + r = d, \quad x \geq 0, \end{aligned} \tag{2.3}$$

whose dual may be written

eq:lsq-r-dual
$$\begin{aligned} \underset{x,r,y,z}{\text{maximize}} \quad & b^T y - (A^T d)^T x - \tfrac{1}{2}\|r\|^2 \\ \text{subject to} \quad & B^T y + z + A^T r = c, \quad Ax + r = d, \quad z \geq 0. \end{aligned} \tag{2.4}$$

Note that the quadratic term in r is the same in the objective function of both the primal and dual problem. An important difference from standard quadratic programming is the appearance of the linear term in x in the dual objective. We return to the importance of this term in §5. It is interesting to note that the constraints $Ax + r = d$, defining the residual r, appear in both the primal and the dual problem. This simple fact turns out to guide our choice of regularization.

We now consider the following regularization of (2.3):

eq:lsq-r-reg
$$\begin{aligned} \underset{x,r,w}{\text{minimize}} \quad & c^T x + \tfrac{1}{2}\|r\|^2 + \tfrac{1}{2}\rho\|x - x_k\|^2 + \tfrac{1}{2}\delta\|w + y_k\|^2 \\ \text{subject to} \quad & Bx + \delta w = b, \quad Ax + r = d, \quad x \geq 0. \end{aligned} \tag{2.5}$$

This regularization differs from (1.2) in two respects. First, no primal regularization term is added for the variables r because the objective function of (2.3) is already strictly convex in r. Second, the constraints $Ax + r = d$ are not regularized because they already have full row rank. Since the variables r do not appear elsewhere in the constraints, the equality constraints of (2.5) have full row rank. No harm would be done in regularizing $Ax + r = d$ although it would result in a larger system. The dual of (2.5) may be stated as

eq:lsq-r-dual-reg
$$\begin{aligned} \underset{x,r,y,s}{\text{maximize}} \quad & b^T y - (A^T d)^T x - \tfrac{1}{2}\|r\|^2 - \tfrac{1}{2}\delta\|y - y_k\|^2 - \tfrac{1}{2}\rho\|s + x_k\|^2 \\ \text{subject to} \quad & A^T r + B^T y + z - \rho s = c, \quad Ax + r = d, \quad z \geq 0, \end{aligned} \tag{2.6}$$

where we introduced the auxiliary variables $s = x - x_k$. Because the dual (2.6) also features the full-rank constraints $Ax + r = d$, it offers an additional elegant justification for omitting the primal regularization term for the variables r in (2.5). Indeed, regularizing those constraints in (2.6) precisely amounts to adding the primal regularization term in question. It is now evident that upon setting $\rho = \delta = 0$, (2.5) and (2.6) coincide with (2.3) and (2.4), respectively. In the rest of this paper, we concentrate on the formulation (2.5)–(2.6).

Proceeding as before, we let $v := (x, r, s, w, y, z)$ and define

eq:def-psi
$$\Psi_k(v; \rho, \delta, \tau) := \begin{bmatrix} c + \rho s - B^T y - A^T r - z \\ \rho x - \rho(s + x_k) \\ \delta y - \delta(w + y_k) \\ Bx + \delta w - b \\ Ax + r - d \\ Xz - \tau e \end{bmatrix}. \tag{2.7}$$

Note that the definition of Ψ does not involve the Lagrange multipliers associated with the constraints $Ax + r = d$. Because those can be readily eliminated and

are always equal to r. Once again, the optimality conditions of (2.5)–(2.6) can be succinctly stated as $\Psi_k(v;\rho,\delta,0)=0$ and $(x,z)\geq 0$ while those of (2.3)–(2.4) can be expressed as $\Psi_k(v;0,0,0)=0$ and $(x,z)\geq 0$.

In the next section, we outline the main features of a long-step interior-point method applied to (2.5)–(2.6).

3. Interior-Point Method

This section describes the linear systems to be solved at each iteration of an interior-point method applied to (2.5)–(2.6) and the neighborhood of the central path used to guide the iterates to a solution of (1.1) and (2.1). We end the section by stating our algorithm formally.

3.1. Linear Systems.

As in the previous section, the Newton correction Δv for (2.7) from the current approximation v_k with barrier parameter τ_k solves the system $\nabla_v \Psi_k(v_k;\rho,\delta,\tau_k)\Delta v = -\Psi_k(v_k;\rho,\delta,\tau_k)$. After eliminating Δs and Δw, and slightly rearranging, we have

eq:newton-unsym

$$\begin{bmatrix} -\rho I & A^T & B^T & I \\ A & I & & \\ B & & \delta I & \\ Z_k & & & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta r \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} c - B^T y_k - A^T r_k - z_k \\ d - Ax_k - r_k \\ b - Bx_k \\ \tau_k e - X_k z_k \end{bmatrix}. \tag{3.1}$$

The remaining directions may be recovered via

eq:other-dirs

$$\Delta w = \Delta y - w_k, \qquad \Delta s = \Delta x - s_k. \tag{3.2}$$

Note that upon setting $\rho=\delta=0$, we recover the Newton equations used to compute a step from the k-th iterate of an interior-point method applied to (2.3)–(2.4).

Rather than using (3.1) directly, our implementation, described in §5, makes use of the following symmetrization, obtained via the similarity transformation defined by the diagonal matrix $\operatorname{blkdiag}(I, I, I, Z_k^{-\frac{1}{2}})$:

eq:newton-sym

$$\begin{bmatrix} -\rho I & A^T & B^T & Z_k^{\frac{1}{2}} \\ A & I & & \\ B & & \delta I & \\ Z_k^{\frac{1}{2}} & & & X_k \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta r \\ \Delta y \\ Z_k^{-\frac{1}{2}} \Delta z \end{bmatrix} = \begin{bmatrix} c - B^T y_k - A^T \lambda_k - z_k \\ d - Ax_k - r_k \\ b - Bx_k \\ \tau_k Z_k^{-\frac{1}{2}} e - X_k Z_k^{\frac{1}{2}} e \end{bmatrix}. \tag{3.3}$$

The above symmetric system differs from that traditionally used in interior-point methods, which results from an additional step of block Gaussian elimination about the $(4,4)$ block. Our motivation for using (3.3) stems from recent results of Greif et al. (2012), who establish that as long as ρ and δ remain bounded away from zero and strict complementarity holds at the limiting solution, the above matrix remains uniformly bounded and uniformly nonsingular. Moreover, its condition number remains sufficiently small along the iterations that a reasonable number of significant digits in the solution may be expected (Greif et al., 2012, §4). By contrast, the matrix of the traditional system is increasingly ill-conditioned as $\tau_k \downarrow 0$ and typically diverges, even if strict complementarity holds (Greif et al., 2012, §3).

Note that the matrix of (3.3) is symmetric and quasi definite (Vanderbei, 1995). It is therefore strongly factorizable, i.e., any symmetric permutation of it possesses a LDL^T factorization with L unit lower triangular and D diagonal indefinite. The computation of this factorization is typically cheaper than that of a sparse symmetric

indefinite factorization because pivoting need only be concerned with sparsity (Gill et al., 1996).

There is an elegant interpretation of (3.3) that is particularly fitting in the present least-square framework. For simplicity, let us rewrite (3.3) as

$$\begin{bmatrix} \rho I & C^T \\ C & -D \end{bmatrix} \begin{bmatrix} \Delta x \\ t \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix},$$

where we defined

$$C^T := \begin{bmatrix} A^T & B^T & Z_k^{\frac{1}{2}} \end{bmatrix}, \quad D = \text{blkdiag}(I, \delta I, X_k), \quad t = (\Delta r, \Delta y, Z_k^{-\frac{1}{2}} \Delta z)$$

and the right-hand side is defined accordingly. This last system may be solved in two stages. Firstly, let $\bar{t} := -D^{-1} g$. Since D is diagonal, computing $\bar{t}$ is trivial. The system may now equivalently be written

$$\begin{bmatrix} \rho I & C^T \\ C & -D \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta t \end{bmatrix} = \begin{bmatrix} \bar{f} \\ 0 \end{bmatrix},$$

where $t = \bar{t} + \Delta t$ and $\bar{f} := f - C^T \bar{t}$. This shifted system represents the necessary and sufficient optimality conditions of the unconstrained regularized linear least-squares problem

eq:lsq-dir (3.4) $$\underset{\Delta t}{\text{minimize}} \ \tfrac{1}{2} \|C^T \Delta t - \bar{f}\|^2_{M^{-1}} + \tfrac{1}{2} \|\Delta t\|^2_D,$$

where $M := \rho I$.

Any method requiring the solution of a symmetric quasi-definite system at each iteration may be interpreted as solving a regularized linear least-squares problem of the form (3.4) at each iteration. In the present context of solving (1.1), this interpretation is particularly fitting. It also forms the basis for the iterative methods developed by Arioli and Orban (2012) and hence paves the way to a matrix-free interior-point method for (1.1). In §5, however, we solve (3.3) using a sparse LDL^T factorization.

3.2. **Neighborhood of the Central Path.** The *central path* is the set of exact roots $v(\tau)$ of $\Psi_k(v; \rho, \delta, \tau)$ for $\tau > 0$. For fixed ρ and δ, as τ approaches zero, it can be shown that $v(\tau)$ approaches a solution of (2.3)–(2.4) (Fiacco and McCormick, 1990). Note that the last block equation of $\Psi_k(v; \rho, \delta, \tau) = 0$ implies that $[x(\tau)]_i [z(\tau)]_i = \tau$ for all $i = 1, \ldots, n$.

Typical interior-point methods compute estimates v_k for some sequence $\{\tau_k\} \downarrow 0$ that are close, in some sense, to $v(\tau_k)$. This concept of proximity is formalized by a neighborhood of the central path. A usual choice is $\tau_k := \sigma_k \mu_k$, where $\sigma_k \in (0, 1)$ is a *centering* parameter and $\mu_k := x_k^T z_k / n$ measures the average centrality. If v_k lies exactly on the central path, we have $[x_k]_i [z_k]_i = \mu_k$ for all $i = 1, \ldots, n$. The quantity μ_k is also directly proportional to the duality gap between (2.3) and (2.4) if v_k is primal-dual feasible.

A difference between our approach and traditional interior-point methods is that during the course of the iterations, the regularization parameters ρ and δ may be updated. At the k-th iteration, the current iterate is v_k and the regularization parameters have values ρ_k and δ_k. The neighborhood $\mathcal{N}_k$ is defined by an appropriate

eq:neighborhood *subset* of the following conditions:

eq:centrality
$$\bar{\gamma}_C x^T z/n \geq [x]_i[z]_i \geq \gamma_C x^T z/n, \tag{3.5a}$$

eq:pFeas
$$x^T z \geq \gamma_P \|Bx + \delta_k w - b\|, \tag{3.5b}$$

eq:lsqFeas
$$x^T z \geq \gamma_R \|Ax + r - d\|, \tag{3.5c}$$

eq:dFeas
$$x^T z \geq \gamma_D \|c + \rho_k s - B^T y - A^T r - z\|, \tag{3.5d}$$

eq:sx
$$x^T z \geq \gamma_S \|\rho_k x - \rho_k(s + x_k)\|, \tag{3.5e}$$

eq:yw
$$x^T z \geq \gamma_W \|\delta_k y - \delta_k(w + y_k)\|, \tag{3.5f}$$

where $0 < \gamma_C < 1 < \bar{\gamma}_C$ and $(\gamma_P, \gamma_R, \gamma_D, \gamma_S, \gamma_W) > 0$ are given constants. Our interior-point scheme computes a steplength $\alpha_k \in (0, 1]$ as well as updated regularization parameters ρ_{k+1} and δ_{k+1} so that the next iterate $v_{k+1} = v_k(\alpha_k) := v_k + \alpha_k \Delta v \in \mathcal{N}_{k+1}$.

3.3. **Algorithm.** Our algorithm is the same as (Friedlander and Orban, 2012, Algorithm 4.1) except for of the linear system used in Step 2, and is formalized as Algorithm 3.1.

alg:pd-reg **Algorithm 3.1** Primal-Dual Regularized Interior-Point Algorithm

Step 0: **[Initialize]** Choose minimum and maximum centering parameters $0 < \sigma_{\min} \leq \sigma_{\max} < 1$, a constant $\sigma_{max} < \beta < 1$, proximity parameters $0 < \gamma_C < 1 < \bar{\gamma}_C$ and $(\gamma_P, \gamma_R, \gamma_D, \gamma_S, \gamma_W) > 0$, initial regularization parameters $\rho_0 > 0$, $\delta_0 > 0$, and a stopping tolerance $\epsilon > 0$. Let the neighborhood of the central path be defined by (3.5a)–(3.5d). Choose initial primal $x_0 \in \mathbb{R}^n_{++}$, $r_0 \in \mathbb{R}^m$, $w_0 \in \mathbb{R}^m$ and dual guesses $s_0 \in \mathbb{R}^n$, $y_0 \in \mathbb{R}^m$, $z_0 \in \mathbb{R}^n_{++}$ so that $v_0 \in \mathcal{N}_0$. Set $\mu_0 := x_0^T z_0/n$ and $k = 0$.

Step 1: **[Test convergence]** If $x_k^T z_k \leq \epsilon$, declare convergence.

Step 2: **[Step computation]** Choose a centering parameter $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$. Compute the Newton step Δv_k, e.g., by solving (3.1) with $\tau_k := \sigma_k \mu_k$ and recovering the remaining components from (3.2).

Step 3: **[Linesearch]** Select $\delta_{k+1} \in (0, \delta_k]$ and $\rho_{k+1} \in (0, \rho_k]$ and compute α_k as the largest $\alpha \in (0, 1]$ such that

$$v_k(\alpha) \in \mathcal{N}_{k+1} \quad \text{and} \quad \mu_k(\alpha) \leq (1 - \alpha(1 - \beta))\mu_k,$$

where $\mu_k(\alpha) := x_k(\alpha)^T z_k(\alpha)/n$.

Step 4: **[Update iterates]** Set $v_{k+1} := v_k(\alpha_k)$, $\mu_{k+1} := \mu_k(\alpha_k)$. Increment k by 1 and go to Step 1.

Friedlander and Orban (2012) present two variants of Algorithm 3.1. The first variant keeps ρ and δ fixed throughout the iterations and only considers (3.5a)–(3.5d) in the definition of $\mathcal{N}_k$. In the second variant, ρ_k and δ_k are allowed to decrease at most linearly and $\mathcal{N}_k$ is defined by (3.5a), (3.5e) and (3.5f). Both variants have similar convergence properties. In our implementation, described in §5, we initially decrease ρ_k and δ_k to speed up convergence, but eventually keep them fixed at a level that guarantees numerical stability of the factorization of the matrix of (3.3). For this reason, in the next section, we only cover the convergence properties of the variant with fixed regularization parameters.

4. Convergence Analysis

Our convergence analysis rests upon a variation on (Armand and Benoist, 2011, Theorem 1) stating that the inverse of the matrix of (3.1) remains uniformly bounded as long as $\{x_k\}$ and $\{z_k\}$ remain bounded away from zero. The reason for this unconventional last assumption is that the convergence proof proceeds by contradiction on the fact that $\{\mu_k\}$ converges to zero. We begin by stating the result on the boundedness of the inverse matrix.

thm:inv-bounded

Theorem 4.1. *Let $\{M_k\}$ be a sequence of $n \times n$ real symmetric matrices, $\{A_k\}$ be a sequence of $p \times n$ real matrices, $\{B_k\}$ be a sequence of $m \times n$ real matrices, and $\{\delta_k\}$ be a sequence of positive numbers. Let $\{x_k\}$ and $\{z_k\}$ be two sequences of $\mathbb{R}^n$ with positive components. Define for all $k \in \mathbb{N}$,*

$$J_k := \begin{bmatrix} M_k & A_k^T & B_k^T & -I \\ A_k & -I & & \\ B_k & & -\delta_k I & \\ Z_k & & & X_k \end{bmatrix}.$$

Assume the following properties are satisfied:

item:bounded (1) *The sequences $\{M_k\}$, $\{A_k\}$ and $\{B_k\}$ are bounded.*

item:delta (2) *The sequence $\{\delta_k\}$ is bounded away from zero.*

item:mu (3) *There exists $\eta > 0$ such that for all $k \in \mathbb{N}$ and all $i \in \{1, \ldots, n\}$,*

eq:contradiction
$$[x_k]_i [z_k]_i \geq \eta. \tag{4.1}$$

item:convex (4) *There exists $\lambda > 0$ such that for all $k \in \mathbb{N}$ and all $d \in \mathbb{R}^n$,*

$$d^T H_k d \geq \lambda \|d\|^2,$$

where $H_k := M_k + X_k^{-1} Z_k + \delta_k^{-1} B_k^T B_k + A_k^T A_k$.

Then the sequence $\{J_k^{-1}\}$ is well defined and bounded.

Proof. It suffices to apply (Armand and Benoist, 2011, Theorem 1) to the matrix $\mathrm{blkdiag}(I, \sqrt{\delta_k} I, I, I) J_k \, \mathrm{blkdiag}(I, \sqrt{\delta_k} I, I, I)$. □

Our interest in Theorem 4.1 is to set $M_k := \rho I$, $A_k := -A$ and $B_k := -B$ for all k. This guarantees that Assumption (1) is satisfied. Assumption (2) is satisfied because Algorithm 3.1 works with fixed regularization parameters. Assumption (4) is also satisifed because of our definition of M_k. Assumption (3) will be the main contradiction assumption. Note also that J_k of Theorem 4.1 is the matrix of (3.1) multiplied by the diagonal matrix $\mathrm{blkdiag}(-I, -I, -I, I)$. Therefore, their inverses are simultaneously uniformly bounded.

Our first result gives conditions under which the right-hand side of (3.1) is bounded.

lem:rhs-bounded

Lemma 4.2. *Let $\{v_k\}$ be the sequence generated by Algorithm 3.1. Assume $\{(s_k, w_k)\}$ remains bounded. Then the right-hand side of (3.1) is uniformly bounded.*

Proof. Using our assumption that $\{w_k\}$ is bounded, we have

$$\|b - Bx_k\| \leq \|b - Bx_k - \delta w_k\| + \delta\|w_k\| \leq \gamma_P^{-1} n\mu_k + \delta \sup_k \|w_k\|,$$

where we used the definition of $\mu_k = x_k^T z_k/n$ and (3.5b). Since Algorithm 3.1 ensures that $\{\mu_k\}$ is decreasing, the above establishes boundedness of $\{b - Bx_k\}$. The boundedness of $\{c - A^T r_k - B^T y_k - z_k\}$ follows similarly from the boundedness of $\{s_k\}$ and (3.5d). Boundedness of $\{Ax_k + r_k - d\}$ follows directly from (3.5c). Finally, boundedness of $\{\sigma_k\mu_k - X_k z_k\}$ follows from (3.5a) and the boundedness of $\{\sigma_k\}$ and $\{\mu_k\}$. □

Using Theorem 4.1 and Lemma 4.2 we obtain uniform boundedness of the direction Δv under the contradiction assumption.

lem:dir-bounded

Lemma 4.3. *Let $\{v_k\}$ be the sequence generated by Algorithm 3.1. Assume $\{(s_k, w_k)\}$ remains bounded and assume that there exists $\eta > 0$ such that (4.1) is satisfied. Then the direction Δv is uniformly bounded.*

Proof. Using Theorem 4.1 and Lemma 4.2, we obtain that $(\Delta x, \Delta r, \Delta y, \Delta z)$ is uniformly bounded. Finally, (3.2) and boundedness of $\{s_k\}$ and $\{w_k\}$ yield boundedness of $\{\Delta s\}$ and $\{\Delta w\}$. □

A careful inspection of (Friedlander and Orban, 2012, §5.2 and §5.4) reveals that Lemma 4.3 is all that is required to establish that the sequence $\{\mu_k\}$ converges to zero in Algorithm 3.1. More precisely, we have the following global convergence result.

thm:global-conv

Theorem 4.4 (Friedlander and Orban 2012, Theorem 5.10)**.** *Let $\{v_k\}$ be the sequence generated by Algorithm 3.1 with $\epsilon = 0$. Assume $\{(s_k, w_k)\}$ remains bounded. Then $\{\mu_k\} \to 0$.*

Proof. The proof is by contradiction. Assume that $\{\mu_k\}$ remains bounded away from zero. Because of (3.5a), there must exist $\eta > 0$ such that (4.1) is satisfied. We conclude from Lemma 4.3 that Δv is uniformly bounded. The rest of the proof proceeds as for (Friedlander and Orban, 2012, Theorem 5.4) to conclude that

$$0 < \eta \leq \mu_{k+1} \leq \gamma\mu_k \leq \cdots \leq \gamma^{k+1}\mu_0,$$

for some constant $\gamma \in (0, 1)$. Since the right-hand side of those inequalities converges to zero, we obtain the contradiction. □

The nature of the limit points of the sequence generated by Algorithm 3.1 is stated in the next two results. The first states that in general, a solution of a perturbed primal-dual pair is identified. This primal-dual pair coincides with the original pair (1.1) and (2.1) but has shifted linear terms and right-hand sides.

thm:perturbed-pairs

Theorem 4.5 (Friedlander and Orban 2012, Theorem 5.5)**.** *Let $\{v_k\}$ be the sequence generated by Algorithm 3.1 with $\epsilon = 0$. Assume $\{(s_k, w_k)\}$ remains bounded. If w_* and s_* denote particular limit points of $\{w_k\}$ and $\{s_k\}$ defined by subsequences indexed by the index set $\mathcal{K} \subseteq \mathbb{N}$, every limit point of $\{(x_k, r_k, z_k)\}_{\mathcal{K}}$ determines a primal-dual solution of the primal-dual pair*

$$\begin{aligned} \underset{x,r}{\text{minimize}} \quad & (c + \rho s_*)^T x + \tfrac{1}{2}\|r\|^2 \\ \text{subject to} \quad & Bx = b - \delta w_*, \quad Ax + r = d, \quad x \geq 0, \end{aligned}$$

and

$$\begin{aligned} \underset{x,r,y,z}{\text{maximize}} \quad & (b - \delta w_*)^T y - (A^T d)^T x - \tfrac{1}{2}\|r\|^2 \\ \text{subject to} \quad & B^T y + z + A^T r = c + \rho s_*, \quad Ax + r = d, \quad z \geq 0. \end{aligned}$$

It is now clear, in light of Theorem 4.5, that whenever $s_* = 0$ and $w_* = 0$, we recover a primal-dual solution of the original problem. That is the essence of the second result.

thm:limit-points

Theorem 4.6 (Friedlander and Orban 2012, Theorem 5.6)**.** *Let $\{v_k\}$ be the sequence generated by Algorithm 3.1 with $\epsilon = 0$. Assume $\{(s_k, w_k)\}$ remains bounded. Then*

1. *If $\{w_k\}_{\mathcal{K}} \to 0$ for some index set $\mathcal{K} \subseteq \mathbb{N}$, every limit point of $\{(x_k, r_k)\}_{\mathcal{K}}$ is feasible for (1.1).*
2. *If $\{s_k\}_{\mathcal{K}'} \to 0$ for some index set $\mathcal{K}' \subseteq \mathbb{N}$, every limit point of $\{(x_k, r_k, z_k)\}_{\mathcal{K}'}$ determines a feasible point for (2.1).*
3. *If $\{(s_k, w_k)\}_{\mathcal{K}''} \to 0$ for some index set $\mathcal{K}'' \subseteq \mathbb{N}$, every limit point of $\{(x_k, r_k, z_k)\}_{\mathcal{K}''}$ determines a primal-dual solution of (1.1)–(2.1).*

The typical case in practice is that $\liminf \|(s_k, w_k)\| = 0$ and the regularization approach recovers a solution of the original primal-dual pair (1.1)–(2.1). Cases where $\liminf \|s_k\| = 0$ but $\liminf \|w_k\| > 0$, or the other way around, suggest that either (1.1) or (2.1) is infeasible.

sec:implementation

5. Implementation and Numerical Results

5.1. **Implementation.** Our implementation is strongly based on that of Friedlander and Orban (2012) with the difference that a step Δv is computed using (3.3) instead of a partial reduction of this system. The implementation accepts problems with free variables in so-called *slack form*

eq:lsq-r-slack

$$\begin{aligned} \underset{x,r,t}{\text{minimize}} \quad & c^T x + \tfrac{1}{2}\|r\|^2 \\ \text{subject to} \quad & Bx + Ct = b, \quad Ax + r = d, \quad t \geq 0, \end{aligned} \tag{5.1}$$

whose dual may be written

$$\begin{aligned} \underset{x,r,y,z}{\text{maximize}} \quad & b^T y - (A^T d)^T x - \tfrac{1}{2}\|r\|^2 \\ \text{subject to} \quad & B^T y + A^T r = c, \quad C^T y + z = 0, \quad Ax + r = d, \quad z \geq 0. \end{aligned}$$

The problem is systematically turned to the form (2.3)–(2.4). The system used to compute a Newton step at iteration k now takes the form

eq:newton-sys-slack
$$\begin{bmatrix} -\rho I & & A^T & B^T & \\ & -\rho I & & C^T & Z^{\frac{1}{2}} \\ A & & I & & \\ B & C & & \delta I & \\ & Z^{\frac{1}{2}} & & & T \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta t \\ \Delta r \\ \Delta y \\ Z^{-\frac{1}{2}} \Delta z \end{bmatrix} = \begin{bmatrix} c - A^T r - B^T y \\ -C^T y - z \\ d - Ax - r \\ b - Bx - Ct \\ Z^{-\frac{1}{2}} (\sigma_k \mu_k e - T_k z_k) \end{bmatrix}. \tag{5.2}$$

This system is solved by sparse LDL^T factorization using MA57 from Duff (2004) as implemented in the HSL (2007) with the pivot tolerance to zero. Although in theory some components of T converge to zero, and therefore the limiting matrix is not symmetric and quasi definite, all components of T remain larger than, say, 10^{-8} in practice and we have not encountered numerical difficulties related to the factorization. In §6, we discuss alternatives that rule out such potential numerical difficulty. The important advantage of the matrix above is that its condition number remains uniformly bounded provided strict complementarity holds in the limit.

The algorithm implemented is a predictor-corrector variant of the long-step method described in Algorithm 3.1. It is worth briefly describing how we adapted the initial-point procedure of Mehrotra (1992). The initial values of (x, r, t) are chosen by solving the minimum-norm problem

$$\begin{aligned} \underset{x,r,t}{\text{minimize}} \quad & \tfrac{1}{2}\|r\|^2 + \tfrac{1}{2}\|t\|^2 \\ \text{subject to} \quad & Bx + Ct = b, \quad Ax + r = d, \end{aligned}$$

while (y, z) are initialized by solving

eq:init-yz
$$\begin{aligned} \underset{\xi,u,y,z}{\text{maximize}} \quad & -(A^T d)^T \xi - \tfrac{1}{2}\|u\|^2 - \tfrac{1}{2}\|z\|^2 \\ \text{subject to} \quad & B^T y + A^T u = c, \quad C^T y + z = 0, \quad A\xi + u = d. \end{aligned} \tag{5.3}$$

The inclusion of the linear term in x in the objective of (5.3) is important in order for the optimality conditions of both problems to share a common matrix. Indeed, the solution of both problems may be approximated via a linear system with two right-hand sides:

$$\begin{bmatrix} -\rho_0 I & & A^T & B^T & \\ & -\rho_0 I & & C^T & I \\ A & & I & & \\ B & C & & \delta_0 I & \\ & I & & & I \end{bmatrix} \begin{bmatrix} x & \xi \\ t & -z \\ r & u \\ \eta & y \\ \zeta & \omega \end{bmatrix} = \begin{bmatrix} 0 & c \\ 0 & 0 \\ d & d \\ b & 0 \\ 0 & 0 \end{bmatrix},$$

for some auxiliary variables η, ζ, ξ and ω. The auxiliary variables ζ and ω together with the last block equation of each system are introduced so the matrix above has the same sparsity pattern as that of (5.2). In addition, the regularization constants ρ_0 and δ_0, both set to 10^{-4}, are introduced so this matrix is symmetric and quasi definite and to guard against rank deficiency.

All other initializations and updates are as described by Friedlander and Orban (2012). In particular, ρ and δ are initialized to the value 1 and divided by 10 at each iteration but are not allowed to decrease below 10^{-8}. The method is implemented in the Python language as part of the NLPy library (Orban, 2012).

5.2. **Numerical Results.** We illustrate the performance of our approach on instances of the ℓ_1-regularized linear least-squares problem

eq:l1reg
$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|Ax-d\|_2^2+\lambda\|x\|_1, \tag{5.4}$$

which can be reformulated as the smooth bound-constrained linear least-squares problem

eq:l1reg-smooth
$$\underset{x\in\mathbb{R}^n,u\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|Ax-d\|_2^2+\lambda e^T u \quad \text{subject to } -u\le x\le u, \tag{5.5}$$

where $\lambda>0$ is a fixed regularization parameter and e is the vector of ones in $\mathbb{R}^n$. After introducing the residual, this last problem has the general form (5.1).

We contrast our results on sparse signal recovery problems generated randomly following the procedure given by Kim et al. (2007) against two comparable and closely related methods. The first is `l1_ls`, an inexact Newton interior-point method specifically designed for (5.4) also described by Kim et al. (2007). In `l1_ls`, a primal barrier method is applied to (5.5) and approximate Newton steps are computed using the preconditioned conjugate-gradient method. In particular, the Hessian contains a term of the form A^TA and this term is preconditioned by its diagonal. The remainder of the Hessian consists of diagonal terms that appear directly in the preconditioner. The second method is PDCO[1], a primal-dual interior-point method designed to solve convex problems of the form

$$\underset{x\in\mathbb{R}^n,r\in\mathbb{R}^m}{\text{minimize}} \ \phi(x)+\tfrac{1}{2}\|D_1x\|_2^2+\tfrac{1}{2}\|r\|_2^2 \quad \text{subject to } Ax+D_2r=b,\ l\le x\le u,$$

where $\phi:\mathbb{R}^n\to\mathbb{R}$ is smooth and convex, D_1 and D_2 are positive definite diagonal matrices, and l and u are vectors in $\mathbb{R}^n$ with possibly infinite components. PDCO has options to compute Newton steps using a direct or iterative method. In addition, we compare our results against those of a state-of-the-art primal-dual interior-point method for nonlinear optimization

6. Discussion

sec:discussion

6.1. **Linear Systems.** Gill et al. (1996) show that the LDL^T factorization of a symmetric and quasi–definite matrix becomes increasingly unstable if either diagonal block approaches singularity or if an off-diagonal block becomes large. Specifically, we have the result

thm:gss
Theorem 6.1 (Gill et al. 1996, Result 4.2)**.** *Let*

$$K:=\begin{bmatrix} H & V^T\\ V & -G\end{bmatrix}$$

be a symmetric quasi-definite matrix. The factorization $PKP^T=LDL^T$, where L is unit lower triangular and D is diagonal, is stable for every permutation matrix P if

$$\theta(K):=\left(\frac{\|V\|_2}{\max(\|G\|_2,\|H\|_2)}\right)^2\max(\kappa_2(G),\kappa_2(H)),$$

is not too large, where κ_2 denotes the spectral condition number.

[1] www.stanford.edu/group/SOL/software/pdco.html

Clearly, if G approaches singularity, $\theta(K)$ becomes large. This is not to say that there does not exist some permutation P for which the factorization is stable. However, this permutation, if it exists, may not yield particularly sparse factors.

One possibility in this case is to resort to the usual symmetric indefinite factorization with 1×1 and 2×2 pivots. The additional cost incurred may be acceptable because it should only be necessary in the last few iterations. Another possibility is to perform the usual block elimination on (3.3) and reduce it to a system with matrix

$$\begin{bmatrix} -(X_k^{-1}Z_k + \rho I) & A^T & B^T \\ A & I & \\ B & & \delta I \end{bmatrix}.$$

Zero elements on the diagonal no longer occur but unfortunately the above matrix no longer has a bounded condition number. Theorem 6.1 also suggests that the LDL^T factorization is still unstable. A third possibility is to perform an additional transform by multiplying the matrix of (3.3) on the left and right by $\text{blkdiag}(I, I, I, X^{-\frac{1}{2}})$ and scale the vector of unknowns and the right-hand side accordingly. This yields the matrix

$$\begin{bmatrix} -\rho I & A^T & B^T & X_k^{-\frac{1}{2}} Z_k^{\frac{1}{2}} \\ A & I & & \\ B & & \delta I & \\ X_k^{-\frac{1}{2}} Z_k^{\frac{1}{2}} & & & I \end{bmatrix}.$$

This time it is $\|V\|_2$ that becomes large in Theorem 6.1. Finally, eliminating the small diagonal elements of x in the vein of Gould (1986) again produces a limiting matrix that is not quasi definite. It appears difficult to maintain safe quasi definiteness in the limit if at least one bound constraint is active at a solution.

In our situation, however, we conjecture that there exists a permutation that produces a stable factorization of (3.3) for the following reason. Suppose the sequences $\{x_k\}$ and $\{z_k\}$ generated by Algorithm 3.1 converge to x_* and z_*, respectively. Define the index sets

$$\mathcal{A} := \{i \mid [x_*]_i = 0\} \qquad \mathcal{I} := \{i \mid [x_*]_i > 0\}.$$

By complementarity, we have $[z_*]_i = 0$ for all $i \in \mathcal{I}$. If we assume that strict complementarity holds at (x_*, z_*), then we also have $[z_*]_i > 0$ for all $i \in \mathcal{A}$. Therefore, for all sufficiently large indices k, (3.5a) implies that

$$\begin{aligned} [x_k]_i &= \Theta(\mu_k) \qquad [z_k]_i = \Theta(1) \quad (i \in \mathcal{A}) \\ [z_k]_i &= \Theta(\mu_k) \qquad [x_k]_i = \Theta(1) \quad (i \in \mathcal{I}). \end{aligned}$$

Consider the following example matrix representative of a problem with 3 variables and bound constraints only, two of which are active in the limit:

$$K = \begin{bmatrix} -1 & & & \sqrt{\mu} & & \\ & -1 & & & 1 & \\ & & -1 & & & 1 \\ \sqrt{\mu} & & & 1 & & \\ & 1 & & & \mu & \\ & & 1 & & & \mu \end{bmatrix}.$$

Without permutation, the LDL^T factorization of K yields

$$L = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ -\sqrt{\mu} & & & 1 & & \\ & -1 & & & 1 & \\ & & -1 & & & 1 \end{bmatrix} \quad D = \begin{bmatrix} -1 & & & & & \\ & -1 & & & & \\ & & -1 & & & \\ & & & 1+\mu & & \\ & & & & 1+\mu & \\ & & & & & 1+\mu \end{bmatrix}$$

Therefore, this factorization is stable in the sense that there exists a moderate constant $\gamma > 0$ such that

$$\| \, |L| \, |D| \, |L^T| \, \| \leq \gamma \|K\|,$$

where the absolute value of a matrix is the matrix of the absolute values of its elements—see Golub and Van Loan (1996). In fact, the relative error between the computed LDL^T and K is of the order of the machine precision for all μ larger than the machine epsilon. Note that μ never becomes that small in practice. If we exchange the first and second variables in K, the relative error for $\mu = 10^{-16}$ rises to about 70%. Though the above example is no proof, it leads us to speculate that as long as the small elements of X_k—i.e., those in $\mathcal{A}$—appear last on the diagonal, the factorization will be stable. This example is inspired by and related to similar examples by Vanderbei (1995) and Gill et al. (1996).

6.2. **Extensions.** Many linear least-squares problems are already stated in regularized form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; c^T x + \tfrac{1}{2}\|Ax - d\|^2 + \tfrac{1}{2}\|Dx\|^2 \quad \text{subject to } Bx = b, \; x \geq 0,$$

where $D \in \mathbb{R}^{n \times n}$ is symmetric and positive semidefinite. Most often, D is a multiple of the identity but other situations are possible. In particular, we allow D to be singular for situations where the regularization applies to a subset of the variables.

We anticipate that our convergence theory carries over to more general convex terms in the objective function in the vein of PDCO—see `www.stanford.edu/group/SOL/software/pdco.html`

Finally, the generalization of our approach to nonlinear least-squares problems with linear constraints is the subject of current investigation.

6.3. **Numerical Aspects.** We consider problem (5.4) which can be reformulated as convex quadratic programs, and then solved by an interior-point method for example the regularization of (1.1) proposed by Friedlander and Orban (2012). Our method is implemented in the Python language as part of the NLPy library (Orban, 2012). We illustrate the performance of our approach against PDCO and $\ell_1 - \ell_s$ on sparse single recovery problems generated randomly following the procedure given by Kim et al. (2007). The parameters that we choose in (5.4) are different. For A we apply the Pseudorandom Number Generator PRNG with $m = 2^t$ and $n = 2^{t+2}$ and the other kind of choosing parameter A is the Discrete Cosine Transform DCT problem with $m = 2^t$ and $n = 2^{t+2}$. We take the parameter d as random vector. The regularization parameter was taken as $\lambda = 1e-4$ in (5.4). All large-scale problems, scaled form achieves significantly better performance in terms of number of iterations, cost, KKT residual, run time, and number of iteration in lsmr while

applying LSMR to the Newton system in LSQ and PDCO. We consider PCG iteration for $\ell_1 - \ell_s$.

TABLE 1. Comparison Solver **LSQ** for The Pseudorandom Number Generator **PRNG**) with $m = 2^t$ and $n = 2^{t+2}$

Size-$m-n$	Iter	Cost	KKTresidual	Time	Niter–lsmr
2–8	4	2.7e-06	6.9e-08	0.08	21
4–16	4	7.8e-06	1.9e-08	0.13	36
8–32	4	1.6e-05	1.3e-08	0.21	63
16–64	3	3.2e-05	5.4e-07	0.27	66
32–128	3	6.7e-05	1.7e-07	0.46	96
64–256	3	1.3e-04	5.5e-08	0.65	103
128–512	3	2.6e-04	1.7e-08	1.20	114
256–1024	3	5.2e-04	6.8e-09	2.85	118
512–2048	3	1.0e-03	2.0e-08	6.49	120
1024–4096	3	2.1e-03	1.1e-07	20.19	123
2048–8192	4	4.1e-03	1.5e-10	87.31	193
4096–16384	4	8.2e-03	4.2e-07	335.65	199

TABLE 2. Comparison Solver **LSQ** for The Discrete Cosine Transform **DCT** Problem with $m = 2^t$ and $n = 2^{t+2}$

Size-$m-n$	Iter	Cost	KKTresidual	Time	Niter–lsmr
2–8	5	4.2e-07	2.2e-07	0.07	8
4–16	5	1.4e-06	6.5e-07	0.07	8
8–32	5	3.4e-06	6.8e-07	0.08	8
16–64	5	8.0e-06	6.7e-07	0.09	8
32–128	5	1.6e-05	6.8e-07	0.11	8
64–256	5	3.3e-05	7.0e-07	0.16	8
128–512	5	6.7e-05	7.1e-07	0.24	8
256–1024	5	1.3e-04	7.1e-07	0.42	8
512–2048	5	2.6e-04	7.1e-07	0.83	8
1024–4096	5	5.2e-04	7.1e-07	1.55	8
2048–8192	5	1.0e-03	7.1e-07	2.89	8
4096–16384	5	2.0e-03	7.2e-07	5.93	8
8192–32768	5	4.1e-03	7.1e-07	11.87	8
16384–65536	5	8.2e-03	7.1e-07	24.00	8
32768–131072	5	1.6e-02	7.1e-07	48.46	8
65536–262144	5	3.3e-02	7.1e-07	98.36	8
131072–524288	5	6.5e-02	7.0e-07	199.75	8
262144–1048576	5	1.3e-01	7.3e-07	395.43	8

TABLE 3. Comparison Solver **l1–ls** for The Pseudorandom Number Generator **PRNG** with $m = 2^t$ and $n = 2^{t+2}$

Size-$m - n$	Iter	Cost	KKTresidual	Time	PCG–iter
2–8	24	6.38e-03	4.10e-09	0.33	48
4–16	27	1.20e-02	8.23e-09	0.27	103
8–32	31	1.50e-02	7.90e-09	0.22	240
16–64	35	1.64e-02	1.16e-08	0.53	486
32–128	32	2.64e-02	2.25e-08	0.88	1859
64–256	39	4.77e-02	3.69e-08	1.08	3092
128–512	85	8.98e-02	5.09e-08	2.45	5416
256–1024	220	1.63e-01	8.44e-08	15.38	16614
512–2048	626	3.27e-01	2.53e-07	153.21	82083
1024–4096	2024	7.07e-08	5.43e-14	4230.78	1542676
2048–8192	–	–	–	–	–
4096–16384	–	–	–	–	–

TABLE 4. Comparison Solver **l1–ls** for The Discrete Cosine Transform **DCT** Problem with $m = 2^t$ and $n = 2^{t+2}$

Size-$m - n$	Iter	Cost	KKTresidual	Time	PCG–iter
2–8	26	2.50e-03	1.70e-09	0.43	57
4–16	25	2.50e-03	2.45e-09	0.49	106
8–32	28	2.50e-03	1.85e-09	0.49	165
16–64	29	2.50e-03	2.14e-09	0.60	187
32–128	30	2.50e-03	1.66e-09	0.62	183
64–256	29	2.50e-03	2.34e-09	0.62	219
128–512	30	2.50e-03	2.39e-09	0.62	228
256–1024	30	2.50e-03	2.40e-09	0.69	225
512–2048	30	2.50e-03	1.61e-09	0.69	259
1024–4096	30	2.50e-03	2.47e-09	0.83	267
2048–8192	30	2.51e-03	1.46e-09	0.91	344
4096–16384	30	2.51e-03	2.20e-09	1.50	432
8192–32768	31	2.53e-03	1.46e-09	2.12	560
16384–65536	32	2.56e-03	2.20e-09	3.39	599
32768–131072	32	2.62e-03	1.95e-09	6.41	649
65536–262144	35	2.73e-03	1.95e-09	12.75	694
131072–524288	37	2.96e-03	1.95e-09	34.64	740
262144–1048576	38	3.43e-03	1.95e-09	87.72	861

TABLE 5. Comparison Solver **PDCO** for A Pseudorandom Number Generator **PRNG** Using **LSMR** with $m = 2^t$ and $n = 2^{t+2}$

Size-$m-n$	Iter	Cost	KKTresidual	Time	Niter–lsmr
2–8	139	2.37e-05	6.93e-09	4.04	1366
4–16	140	4.55e-05	6.13e-08	4.80	1765
8–32	141	9.43e-05	9.16e-09	2.58	2114
16–64	147	2.06e-04	2.89e-09	6.91	2471
32–128	155	3.83e-04	2.73e-09	9.64	2415
64–256	162	7.23e-04	2.66e-09	11.34	2252
128–512	168	1.48e-03	2.86e-09	17.04	2072
256–1024	174	2.75e-03	2.88e-09	39.42	1834
512–2048	180	4.74e-03	2.70e-09	120.42	1716
1024–4096	184	9.56e-03	2.88e-09	475.15	1635
2048–8192	–	–	–	–	–
4096–16384	–	–	–	–	–

TABLE 6. Comparison Solver **PDCO** for The Discrete Cosine Transform **DCT** Problem Using **LSMR** with $m = 2^t$ and $n = 2^{t+2}$

Size-$m-n$	Iter	Cost	KKTresidual	Time	Niter–lsmr
2–8	144	1.29e-03	1.00e-06	4.14	282
4–16	144	1.32e-03	1.00e-06	3.48	281
8–32	144	1.39e-03	1.00e-06	3.86	281
16–64	144	1.52e-03	1.00e-06	4.18	281
32–128	144	1.77e-03	1.00e-06	4.37	281
64–256	144	2.29e-03	1.00e-06	4.43	281
128–512	144	3.35e-03	1.00e-06	5.29	281
256–1024	144	5.42e-03	1.00e-06	5.20	281
512–2048	144	9.51e-03	1.00e-06	5.52	281
1024–4096	144	1.77e-02	1.00e-06	6.29	281
2048–8192	144	3.40e-02	1.00e-06	7.10	281
4096–16384	144	6.66e-02	1.00e-06	11.51	281
8192–32768	146	1.25e-01	1.00e-06	21.04	285
16384–65536	147	2.61e-01	1.00e-06	46.84	287
32768–131072	148	5.11e-01	1.00e-06	91.84	288
65536–262144	149	9.91e-01	1.00e-06	179.74	291
131072–524288	150	1.98e+00	1.00e-06	344.74	293
262144–1048576	150	3.96e+00	1.00e-06	696.80	293

6.4. **Limitations of the Proposed Solution and Future Improvements.** Other future improvements include the solution of *constrained nonlinear least-squares* problems.

References

arioli-orban-2012 M. Arioli and D. Orban. Iterative methods for symmetric quasi-definite linear systems. Cahier du GERAD G-2012-xx, GERAD, Montréal, Québec, Canada, 2012. In preparation.

armand-benoist-2011 P. Armand and J. Benoist. Uniform boundedness of the inverse of a Jacobian matrix arising in regularized interior-point methods. *Mathematical Programming*, 2011. DOI: 10.1007/s10107-011-0498-3.

barzilai-borwein-1988 J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148, 1988. DOI: 10.1093/imanum/8.1.141.

bellavia-macconi-morini-2006 S. Bellavia, M. Macconi, and B. Morini. An interior point Newton-like method for non-negative least-squares problems with degenerate solution. *Numerical Linear Algebra with Applications*, 13:825–846, 2006. DOI: 10.1002/nla.502.

bellavia-gondzio-morini-2008 S. Bellavia, J. Gondzio, and B. Morini. Regularization and preconditioning of KKT systems arising in nonnegative least-squares problems. *Numerical Linear Algebra with Applications*, 16:39–61, 2008. DOI: 10.1002/nla.610.

bellavia-gondzio-morini-2011 S. Bellavia, J. Gondzio, and B. Morini. Computational experience with numerical methods for nonnegative least-squares problems. *Numerical Linear Algebra with Applications*, 18:363–385, 2011. DOI: 10.1002/nla.732.

bierlaire-toint-tuyttens-1991 M. Bierlaire, Ph. L. Toint, and D. Tuyttens. On iterative algorithms for linear least squares problems with bound constraints. *Linear Algebra and its Applications*, 143:111–143, 1991. DOI: 10.1016/0024-3795(91)90009-L.

bjorck-1996 Å. Björck. *Numerical Methods for Least Squares Problems*. Number OT51. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. DOI: 10.1137/1.9781611971484.

chen-donoho-saunders-1998 S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998. DOI: 10.1137/S1064827596304010.

donoho-2006 D. L. Donoho. For most large underdetermined systems of linear equations, the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006. DOI: 10.1002/cpa.20132.

duff-2004 I. S. Duff. MA57—a code for the solution of sparse symmetric definite and indefinite systems. *Transactions of the ACM on Mathematical Software*, 30(2):118–144, 2004. DOI: 10.1145/992200.992202.

FiaccoMcCormick A. V. Fiacco and G. P. McCormick. *Nonlinear programming: Sequential Unconstrained Minimization Techniques*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1990. DOI: 10.1137/1.9781611971316. Republication of original 1968 edition.

friedlander-2007 M. P. Friedlander. BCLS: Bound constrained least squares, 2007. URL `www.cs.ubc.ca/~mpf/bcls`. Accessed on 07/27/2012.

friedlander-orban-2012 M. P. Friedlander and D. Orban. A primal-dual regularized interior-point method for convex quadratic programs. *Mathematical Programming Computation*, 4(1): 71–107, 2012. DOI: 10.1007/s12532-012-0035-2.

gill-saunders-shinnerl-1996 P. E. Gill, M. A. Saunders, and J. R. Shinnerl. On the stability of Cholesky factorization for symmetric quasidefinite systems. *SIAM Journal on Matrix Analysis and Applications*, 17(1):35–46, 1996. DOI: 10.1137/S0895479893252623.

golub-vanloan-1996 G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins, Baltimore, third edition, 1996.

gould-1986 N. I. M. Gould. On the accurate determination of search directions for simple differentiable penalty functions. *IMA Journal of Numerical Analysis*, 6:357–372, 1986. DOI: 10.1093/imanum/6.3.357.

greif-moulding-orban-2012 C. Greif, E. Moulding, and D. Orban. Bounds on the eigenvalues of block matrices arising from interior-point methods. Cahier du GERAD G-2012-42, GERAD, Montréal, Québec, Canada, 2012.

HSL HSL. *The HSL Mathematical Software Library*. STFC Rutherford Appleton Laboratory, 2007. http://www.hsl.rl.ac.uk.

koh-lustig-boyd-gorinevsky-2007 S. J. Kim, K. Koh, M. Lusting, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale ℓ_1-regularized least squares. *IEEE Journal on Selected Topics in Signal Processing*, 1(4):606–617, 2007. DOI: 10.1109/JSTSP.2007.910971.

koh-kim-boyd-2007 K. Koh, S. J. Kim, and S Boyd. An interior-point method for large-scale ℓ_1-regularized logistic regression. *Journal of Machine Learning Research*, 8:1519–1555, 2007.

lawson-hanson-1995 C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Number 15 in Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995. DOI: 10.1137/1.9781611971217. Republication of original 1974 edition.

mehrotra-1992 S. Mehrotra. On the implementation of a primal-dual interior-point method. *SIAM Journal on Optimization*, 2(4):575–601, 1992. DOI: 10.1137/0802028.

orban-2012 D. Orban. NLPy—a large-scale optimization toolkit in Python. Technical Report Cahier du GERAD G-2012-xx, 2012. https://github.com/dpo/nlpy.

portugal-judice-vicente-1994 L. F. Portugal, J. J. Judice, and L. N. Vicente. Comparison of block pivoting and interior-point algorithms for linear least squares problems with nonnegative variables. *Mathematics of Computation*, 63:625–643, 1994. DOI: 10.1090/S0025-5718-1994-1250776-4.

rockafellar-1976 R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976. DOI: 10.1287/moor.1.2.97.

vanderbei-1995 R. J. Vanderbei. Symmetric quasi-definite matrices. *SIAM Journal on Optimization*, 5(1):100–113, 1995. DOI: 10.1137/0805005.

GERAD, Montréal, Canada
E-mail address: mohsen.dehghani@gerad.ca

GERAD and Mathematics and Industrial Engineering Department, École Polytechnique, Montréal, Canada
URL: www.gerad.ca/~orban
E-mail address: dominique.orban@gerad.ca