ACTIVE-SET METHODS FOR BASIS PURSUIT DENOISING

MICHAEL P. FRIEDLANDER* AND MICHAEL A. SAUNDERS†

Abstract. Many imaging and compressed sensing applications seek sparse solutions to large under-determined least-squares problems. The basis pursuit (BP) approach minimizes the 1-norm of the solution, and the BP denoising (BPDN) approach balances it against the least-squares fit. The duals of these problems are conventional linear and quadratic programs. We introduce a modified parameterization of the BPDN problem and explore the effectiveness of active-set methods for solving its dual. Our basic algorithm for the BP dual unifies several existing algorithms and is applicable to large-scale examples.

Key words. basis pursuit, basis pursuit denoising, active-set method, quadratic program, convex program, duality, regularization, sparse solutions, one-norm

AMS subject classifications. 49M29, 65K05, 90C25, 90C06

1. Introduction.

Consider the linear system $Ax + r = b$, where A is an m-by-n matrix and b is a m-vector. In many statistical and signal processing applications the aim is to obtain a solution (x, r) such that the residual vector r is small in norm and the vector x is sparse. Typically, $m \ll n$ and the problem is ill-posed. To obtain well defined solutions for any m and n, we study the parameterized problem

$$\text{BP}_{\delta\lambda}: \quad \underset{x,\,y}{\text{minimize}} \quad \|x\|_1 + \tfrac{1}{2}\delta\|x\|_2^2 + \tfrac{1}{2}\lambda\|y\|_2^2 \quad \text{subject to} \quad Ax + \lambda y = b$$

and its dual

$$\text{BPdual}_{\delta\lambda}: \quad \underset{x,\,y}{\text{maximize}} \quad b^T y - \tfrac{1}{2}\delta\|x\|_2^2 - \tfrac{1}{2}\lambda\|y\|_2^2 \quad \text{subject to} \quad -e \le -\delta x + A^T y \le e,$$

where $\delta \ge 0$, $\lambda \ge 0$, and e is a vector of ones. The problems are duals of each other in the sense that the Karush-Kuhn-Tucker (KKT) conditions for optimality for each problem are satisfied by the same vector pair (x, y). (The KKT conditions require the constraints in each problem to be satisfied and the objective values to be equal.) Typically δ will be a small regularization parameter (say $\delta = 10^{-6}$ or 10^{-8}), while λ may be small or large. If $\|A\| \approx \|b\| \approx 1$, we expect $\|x\| \approx \|y\| \approx 1$. Thus the problem variables (x, y) are well scaled in these formulations.

When $\delta = \lambda = 0$, $\text{BP}_{\delta\lambda}$ is the basis pursuit (BP) problem of Chen et al. [3,4]. This insists on a zero residual $r = \lambda y$ and often yields a sparse solution x. In some cases, it yields the sparsest solution possible (Candès, Romberg, and Tao [1], Donoho [5]).

When $\delta = 0$ and $\lambda > 0$, $\text{BP}_{\delta\lambda}$ is equivalent to the basis pursuit denoising (BPDN) problem in [3,4]. It allows a nonzero residual, but the sparsity of x remains of prime importance. Our chosen form of the problems reduce naturally to the BP problem and its dual when $\lambda = 0$.

*Department of Computer Science, University of British Columbia, Vancouver V6T 1Z4, BC, Canada (mpf@cs.ubc.ca). Partially supported by NSERC Collaborative Research and Development Grant 334810-05.

†Systems Optimization Laboratory, Department of Management Science and Engineering, Stanford University, Stanford, CA 94305-4026, USA (saunders@stanford.edu). Partially supported by Office of Naval Research grant N00014-08-1-0191. Draft of April 13, 2019

When $\delta > 0$ and $\lambda > 0$, the objective of problem BPdual$_{\delta\lambda}$ is minimized at the point $x = 0$, $y = b/\lambda$, which satisfies the problem's constraints if $\lambda \geq \|A^T b\|_\infty$. We can show that $x = 0$ and $y = b/\lambda$ is the unique solution of both problems for all $\lambda \geq \|A^T b\|_\infty$. Also, both problems have unique optimal solutions (x, y) for any $\delta > 0$ and $\lambda > 0$. In this sense, δ and λ are regularization parameters.

We present active-set algorithms, suitable for large problems, that can solve both BP$_{\delta\lambda}$ and BPdual$_{\delta\lambda}$. The flexibility of our algorithms provides a base from which more involved algorithms can be easily implemented. Some examples are

- Homotopy [10], which solves BP$_{0\lambda}$ for all values of λ;
- Lars [6], which greedily approximates the solution of BP$_{0\lambda}$ for all λ;
- Reweighted one-norm minimization [2], to approximate zero-norm solutions;
- Sequential compressed sensing [8], in which rows are added to A and b.

Our approach is based on applying active-set methods to problems BP$_{\delta\lambda}$ and BPdual$_{\delta\lambda}$. Importantly, when $\delta = 0$ it is not necessary to reduce λ to zero in order to recover a solution of BP$_{00}$ and BPdual$_{00}$. As shown by Mangasarian and Meyer [9] and Friedlander and Tseng [7], there exists a positive parameter $\bar\lambda$ such that for all $\lambda \in (0, \bar\lambda)$ the solution y of BPdual$_{0\lambda}$ coincides with the unique least-norm solution of BPdual$_{00}$. This property is crucial in making our algorithm relevant for both BP and BPDN.

There are four components to this paper. The first two define active-set algorithms for solving BP$_{\delta\lambda}$ and BPdual$_{\delta\lambda}$. The third describes how to extend these algorithms to solve related problems. The fourth gives the results of a series of numerical experiments.

2. An active-set method for BP$_{\delta\lambda}$.

Since we expect many components of x to be zero, it is natural to partition the variables into two sets according to

$$AP = \begin{pmatrix} S & N \end{pmatrix}, \qquad x = P \begin{pmatrix} x_S \\ x_N \end{pmatrix}, \tag{1}$$

where P is a permutation. We assume that no component of x_S is zero, and we maintain $x_N = 0$ as P changes. For any such x, we can satisfy the BP$_{\delta\lambda}$ constraints $Ax + \lambda y = b$ by setting $y = (b - Sx_S)/\lambda$.

The objective function $\phi(x, y)$ and its gradient $g = \nabla\phi$ and Hessian $H = \nabla^2\phi$ are

$$\phi = \|x\|_1 + \tfrac{1}{2}\delta\|x\|_2^2 + \tfrac{1}{2}\lambda\|y\|_2^2,$$

$$g = \begin{pmatrix} \operatorname{sign}(x) + \delta x \\ \lambda y \end{pmatrix}, \qquad H = \begin{pmatrix} \delta I & \\ & \lambda I \end{pmatrix}.$$

To improve the values of (x, y), a search direction $p = (\Delta x, \Delta y)$ can be computed from the quadratic program

$$\min_p \quad g^T p + \tfrac{1}{2} p^T H p \quad \text{subject to} \quad \begin{pmatrix} A & \lambda I \end{pmatrix} p = 0.$$

With $\Delta x_N = 0$, this becomes

$$\begin{aligned} \min \quad & g_S^T \Delta x_S + \lambda y^T \Delta y + \tfrac{1}{2}\delta\|\Delta x_S\|^2 + \tfrac{1}{2}\lambda\|\Delta y\|^2 \\ \text{subject to} \quad & S\Delta x_S + \lambda \Delta y = 0, \end{aligned}$$

where $g_S = \operatorname{sign}(x_S) + \delta x_S$. The solution is given by

$$\begin{pmatrix} -\delta I & S^T \\ S & \lambda I \end{pmatrix} \begin{pmatrix} \Delta x_S \\ \Delta y \end{pmatrix} = \begin{pmatrix} g_S - S^T y \\ 0 \end{pmatrix}. \tag{2}$$

In some cases it may be effective to treat this "augmented system" directly. Since we expect S to have relatively few columns, it is reasonable to eliminate $\Delta y = -\frac{1}{\lambda} S \Delta x_S$ and solve the least-squares problem

$$\min \quad \left\| \begin{pmatrix} S \\ \beta I \end{pmatrix} \Delta x_S - \lambda \begin{pmatrix} y \\ -g_S/\beta \end{pmatrix} \right\|. \tag{3}$$

where $\beta = \sqrt{\delta\lambda}$. Alternatively we may eliminate $\Delta x_S = \frac{1}{\delta}(S^T \Delta y - (g_S - S^T y))$ and solve the damped least-squares problem

$$\min \quad \left\| \begin{pmatrix} S^T \\ \beta I \end{pmatrix} \Delta y - \begin{pmatrix} g_S - S^T y \\ 0 \end{pmatrix} \right\|. \tag{4}$$

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.

[2] E. J. Candés, M. B. Wakin, and S. P. Boyd, *Enhancing sparsity by reweighted L1 minimization*, tech. rep., California Institute of Technology, October 2007. Available at http://www.acm.caltech.edu/~emmanuel/papers/rwl1-oct2007.pdf.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1998), pp. 33–61.

[4] ———, *Atomic decomposition by basis pursuit*, SIAM Rev., 43 (2001), pp. 129–159.

[5] D. L. Donoho, *For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution*, Comm. Pure Appl. Math., 59 (2006), pp. 797–829.

[6] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Ann. Statist., 32 (2004), pp. 407–499.

[7] M. P. Friedlander and P. Tseng, *Exact regularization of convex programs*, SIAM J. Optim., 18 (2007), pp. 1326–1350.

[8] D. Malioutov, S. Sanghavi, and A. Willsky, *Compressed sensing with sequential observations*, in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, IEEE Signal Processing Society, April 2008.

[9] O. L. Mangasarian and R. R. Meyer, *Nonlinear perturbation of linear programs*, SIAM J. Control Optim., 17 (1979), pp. 745–752.

[10] M. R. Osborne, B. Presnell, and B. A. Turlach, *A new approach to variable selection in least squares problems*, IMA J. Numer. Anal., 20 (2000), pp. 389–403.