Solver for ℓ_1 plus ℓ_1 minimization

João F. C. Mota

September 27, 2014

We address

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 + \beta\|x-w\|_1 \\ \text{subject to} & Ax = b\,, \end{array} \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $w \in \mathbb{R}^n$, and $\beta > 0$ are given. The algorithm described here was designed to create the results in [1]. We solve (1) with the Alternating Direction Method of Multipliers (ADMM) [2, 3, 4], because each subproblem will have closed-form solutions, which will imply a fast algorithm. First, we recast the problem as

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(x) + g(y) \\ \text{subject to} & x = y\,, \end{array} \tag{2}$$

where $f(x) = \|x\|_1 + \beta\|x-w\|_1$ and $g(y) = \mathrm{i}_{\{x\,:\,b=Ax\}}(y)$, where $\mathrm{i}_S(x)$ is the indicator function of the set S, i.e.,

$$\mathrm{i}_S(x) = \begin{cases} 0 & ,\ \text{if } x \in S \\ +\infty & ,\ \text{if } x \notin S. \end{cases}$$

The augmented Lagrangian of (2) is

$$L_\rho(x, y; \lambda) = f(x) + g(y) + \lambda^\top(x-y) + \frac{\rho}{2}\|x-y\|^2\,,$$

and ADMM becomes

$$x^{k+1} = \underset{x}{\arg\min}\ f(x) + {\lambda^k}^\top x + \frac{\rho}{2}\|x - y^k\|^2 \tag{3}$$

$$y^{k+1} = \underset{y}{\arg\min}\ g(y) - {\lambda^k}^\top y + \frac{\rho}{2}\|x^{k+1} - y\|^2 \tag{4}$$

$$\lambda^{k+1} = \lambda^k + \rho(x^{k+1} - y^{k+1})\,.$$

It turns out that both (3) and (4) have closed-form solutions.

Problem in $\boldsymbol{x}$. Developing the square, problem (3) is equivalent to

$$x^{k+1} = \underset{x}{\arg\min}\ f(x) + (\lambda^k - \rho y^k)^\top x + \frac{\rho}{2}\|x\|^2\,. \tag{5}$$

Let $v := \lambda^k - \rho y^k$. Replacing the expression for function f in (5),

$$x^{k+1} = \underset{x}{\arg\min}\ \|x\|_1 + \beta\|x-w\|_1 + v^\top x + \frac{\rho}{2}\|x\|^2\,,$$

whose ith component is given by

$$x_i^{k+1} = \underset{x_i}{\arg\min}\ |x_i| + \beta|x_i - w_i| + v_i x_i + \frac{\rho}{2}x_i^2\,. \tag{6}$$

To find the solution of (6) in closed-form, we need to consider the following cases:

- $w_i > 0$:
  - $x_i < 0$: the optimality conditions in this case are

    $$0 = -1 - \beta + v_i + \rho x_i \quad \iff \quad x_i = \frac{1}{\rho}\Big(\beta + 1 - v_i\Big)\,,$$

    which hold when $v_i > \beta + 1$.
  - $0 < x_i < w_i$:

    $$0 = 1 - \beta + v_i + \rho x_i \quad \iff \quad x_i = \frac{1}{\rho}\Big(\beta - 1 - v_i\Big)\,,$$

    which hold when $-\rho w_i + \beta - 1 < v_i < \beta - 1$.
  - $x_i > w_i$:

    $$0 = 1 + \beta + v_i + \rho x_i \quad \iff \quad x_i = \frac{1}{\rho}\Big(-\beta - 1 - v_i\Big)\,,$$

    which holds when $v_i < -\rho w_i - \beta - 1$.

  We then have, for $w_i > 0$,

  $$x_i^\star = \begin{cases} \frac{1}{\rho}(-\beta - 1 - v_i) & ,\ v_i < -\rho w_i - \beta - 1 \\ w_i & ,\ -\rho w_i - \beta - 1 \leq v_i \leq -\rho w_i + \beta - 1 \\ \frac{1}{\rho}(\beta - 1 - v_i) & ,\ -\rho w_i + \beta - 1 < v_i < \beta - 1 \\ 0 & ,\ \beta - 1 \leq v_i \leq \beta + 1 \\ \frac{1}{\rho}(\beta + 1 - v_i) & ,\ v_i > \beta + 1\,. \end{cases}$$

- $w_i < 0$:
  - $x_i < w_i$: the optimality conditions are

    $$0 = -1 - \beta + v_i + \rho x_i \quad \iff \quad x_i = \frac{1}{\rho}\Big(\beta + 1 - v_i\Big)\,,$$

    which holds when $v_i > -\rho w_i + \beta + 1$.
  - $w_i < x_i < 0$:

    $$0 = -1 + \beta + v_i + \rho x_i \quad \iff \quad x_i = \frac{1}{\rho}\Big(-\beta + 1 - v_i\Big)\,,$$

    which holds when $-\beta + 1 < v_i < -\rho w_i - \beta + 1$.
  - $x_i > 0$:

    $$0 = 1 + \beta + v_i + \rho x_i \quad \iff \quad x_i = \frac{1}{\rho}\Big(-\beta - 1 - v_i\Big)\,,$$

    which holds when $v_i < -\beta - 1$.

  We then have, for $w_i < 0$,

  $$x_i^\star = \begin{cases} \frac{1}{\rho}(-\beta - 1 - v_i) & ,\ v_i < -\beta - 1 \\ 0 & ,\ -\beta - 1 \leq v_i \leq -\beta + 1 \\ \frac{1}{\rho}(-\beta + 1 - v_i) & ,\ -\beta + 1 < v_i < -\rho w_i - \beta + 1 \\ w_i & ,\ -\rho w_i - \beta + 1 \leq v_i \leq -\rho w_i + \beta + 1 \\ \frac{1}{\rho}(\beta + 1 - v_i) & ,\ v_i > -\rho w_i + \beta + 1\,. \end{cases}$$

Problem in *y*. Problem (4) is equivalent to

$$\begin{array}{ll} \underset{y}{\arg\min} & \frac{\rho}{2}\|y\|^2 - \rho\, {x^{k+1}}^\top y - {\lambda^k}^\top y \\ \text{s.t.} & Ay = b \end{array} \qquad \Longleftrightarrow \qquad \begin{array}{ll} \underset{y}{\arg\min} & \|y\|^2 - \frac{2}{\rho}(\lambda^k + \rho\, x^{k+1})^\top y \\ \text{s.t.} & Ay = b \end{array}$$

$$\Longleftrightarrow \qquad \begin{array}{ll} \underset{y}{\arg\min} & \|y - \frac{1}{\rho}(\lambda^k + \rho\, x^{k+1})\|^2 \\ \text{s.t.} & Ay = b \end{array}$$

Defining $z := (1/\rho)(\lambda^k + \rho\, x^{k+1})$, this is equivalent to projecting a point onto $\{y \,:\, Ay = b\}$:

$$\begin{array}{ll} \underset{y}{\text{minimize}} & \frac{1}{2}\|y - z\|^2 \\ \text{subject to} & Ay = b\,, \end{array} \tag{7}$$

which has the closed-form solution

$$y^\star = z - A^\top (AA^\top)^{-1}(Az - b)\,. \tag{8}$$

For large-scale problems, or if we only have access to the operations Ax and $A^\top y$ but not to the full matrix A, (8) can be computed via the conjugate gradient method.

References

[1] J. Mota, N. Deligiannis, M. Rodrigues, “Compressed Sensing with Prior Information: Optimal Strategies, Geometry, and Bounds,” *submitted to IEEE Transaction on Information Theory*, preprint: `http://arxiv.org/abs/1408.5250` 2014

[2] R. Glowinski and A. Marrocco, “Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité, d’une classe de problèmes de dirichelet non linéaires,” *Revue Française d’Automatique, Informatique, et Recherche Opérationelle*, vol. 9, no. 2, pp. 41-76, 1975.

[3] D. Gabay and B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite element approximations,” *Computers and Mathematics with Applications*, vol. 2, no. 1, pp. 17-40, 1976

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2011