First-Order Methods for Linear Programming

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

(LY: Chapter 14)

The Alternating Direction Method with Multipliers

We consider

$$\min \quad f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) \quad \text{s.t.} \quad A_1\mathbf{x}_1 + A_2\mathbf{x}_2 = \mathbf{b},\ \mathbf{x}_1 \in X_1,\ \mathbf{x}_2 \in X_2;$$

where X_1 and X_2 are (simple) convex sets.

Define its Augmented Lagrangian

$$L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) - \mathbf{y}^T(A_1\mathbf{x}_1 + A_2\mathbf{x}_2 - \mathbf{b}) + \frac{\beta}{2}\|A_1\mathbf{x}_1 + A_2\mathbf{x}_2 - \mathbf{b}\|^2.$$

Then, for any given $(\mathbf{x}_1^k, \mathbf{x}_2^k, \mathbf{y}^k)$, we compute a new iterate pair

$$\mathbf{x}_1^{k+1} = \arg\min_{\mathbf{x}_1 \in X_1} L(\mathbf{x}_1, \mathbf{x}_2^k, \mathbf{y}^k),\ \mathbf{x}_2^{k+1} = \arg\min_{\mathbf{x}_2 \in X_2} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2, \mathbf{y}^k)$$

and

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \beta(A_1\mathbf{x}_1^{k+1} + A_2\mathbf{x}_2^{k+1} - \mathbf{b}).$$

Again, the iterates converge for any $\beta > 0$ with the same speed as the SDM.

The Splitting to Handle Inequalities

We consider

$$\min \quad f(\mathbf{x}_1) \quad \text{s.t.} \quad A\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{b},\ \mathbf{x}_2 \geq \mathbf{0}.$$

$$L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = f(\mathbf{x}_1) - \mathbf{y}^T(A\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{b}) + \frac{\beta}{2}\|A\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{b}\|^2.$$

Then, for any given $(\mathbf{x}_1^k, \mathbf{x}_2^k, \mathbf{y}^k)$, we compute a new iterate pair

$$\mathbf{x}_1^{k+1} = \arg\min_{\mathbf{x}_1} L(\mathbf{x}_1, \mathbf{x}_2^k, \mathbf{y}^k)$$

$$\mathbf{x}_2^{k+1} = \arg\min_{\mathbf{x}_2 \geq \mathbf{0}} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2, \mathbf{y}^k)$$

and

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \beta(A\mathbf{x}_1^{k+1} + \mathbf{x}_2^{k+1} - \mathbf{b}).$$

Note that the solution of $\mathbf{x}_2$ can be computed in a close form!

Linear Programming

$$
\begin{array}{lll}
(LP) & \text{minimize} & \mathbf{c} \bullet \mathbf{x} \\
& \text{subject to} & A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0},
\end{array}
$$

We consider an equivalent problem:

$$
\begin{array}{lll}
(LP) & \text{minimize} & \mathbf{c} \bullet \mathbf{x}_1 \\
& \text{subject to} & A\mathbf{x}_1 = \mathbf{b},\ \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0},\ \mathbf{x}_2 \geq \mathbf{0},
\end{array}
$$

The ADMM for LP

Consider its Augmented Lagrangian

$$L(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y},\mathbf{s}) = \mathbf{c}^T\mathbf{x}_1 - \mathbf{y}^T(A\mathbf{x}_1-\mathbf{b}) - \mathbf{s}^T(\mathbf{x}_1-\mathbf{x}_2) + \frac{\beta}{2}\|A\mathbf{x}_1-\mathbf{b}\|^2 + \frac{\beta}{2}\|\mathbf{x}_1-\mathbf{x}_2\|^2.$$

Then, for any given $(\mathbf{x}_1^k,\mathbf{x}_2^k,\mathbf{y}^k,\mathbf{s}^k)$, we compute a new iterate pair

$$\mathbf{x}_1^{k+1} = \arg\min_{\mathbf{x}_1} L(\mathbf{x}_1,\mathbf{x}_2^k,\mathbf{y}^k,\mathbf{s}^k)$$

$$\mathbf{x}_2^{k+1} = \arg\min_{\mathbf{x}_2\geq\mathbf{0}} L(\mathbf{x}_1^{k+1},\mathbf{x}_2,\mathbf{y}^k,\mathbf{s}^k)$$

and

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \beta(A\mathbf{x}_1^{k+1}-\mathbf{b}) \quad \text{and} \quad \mathbf{s}^{k+1} = \mathbf{s}^k - \beta(\mathbf{x}_1^{k+1}-\mathbf{x}_2^{k+1}).$$

The minimization over $\mathbf{x}_1$ is a unconstrained optimization, and the minimization over $\mathbf{x}_2$ can be computed in a close form!

Solving Nonnegative Constrained Least Squares

$$(CLS) \quad \text{minimize}_{\mathbf{u}} \quad \mathbf{b} \bullet (\mathbf{u} - \mathbf{a}) + \frac{1}{2}\|\mathbf{u} - \mathbf{a}\|^2$$
$$\text{subject to} \quad \mathbf{u} \geq \mathbf{0}.$$

$$\mathbf{b} + (\mathbf{u} - \mathbf{a}) - \mathbf{v} = \mathbf{0}, \; \mathbf{v} \geq \mathbf{0}, \; \mathbf{u} \bullet \mathbf{v} = 0;$$

or

$$\mathbf{u} - \mathbf{v} = \mathbf{a} - \mathbf{b}, \; \mathbf{u} \bullet \mathbf{v} = 0,$$

where $\mathbf{a} - \mathbf{b}$ is given. This has a close form:

$$\mathbf{u} = (\mathbf{a} - \mathbf{b})^{+} \quad \text{and} \quad \mathbf{v} = -(\mathbf{a} - \mathbf{b})^{-1}.$$

The Interior-Point ADMM for LP?

Consider its Augmented Lagrangian with the barrier function on $\mathbf{x}_2$:

$$\begin{aligned} L^B(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}, \mathbf{s}, \mu) &= \mathbf{c}^T\mathbf{x}_1 - \mathbf{y}^T(A\mathbf{x}_1 - \mathbf{b}) - \mathbf{s}^T(\mathbf{x}_1 - \mathbf{x}_2) \\ &\quad + \frac{\beta}{2}\|A\mathbf{x}_1 - \mathbf{b}\|^2 + \frac{\beta}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|^2 - \mu \cdot B(\mathbf{x}_2). \end{aligned}$$

Then, for any given $(\mathbf{x}_1^k, \mathbf{x}_2^k > \mathbf{0}, \mathbf{y}^k, \mathbf{s}^k, \mu^k > 0)$, we compute a new iterate pair

$$\mathbf{x}_1^{k+1} = \arg\min_{\mathbf{x}_1} L^B(\mathbf{x}_1, \mathbf{x}_2^k, \mathbf{y}^k, \mathbf{s}^k, \mu^k)$$

$$\mathbf{x}_2^{k+1} = \arg\min_{\mathbf{x}_2} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2, \mathbf{y}^k, \mathbf{s}^k, \mu^k)$$

and

$$\mu^{k+1} = \gamma\mu^k, \quad \mathbf{y}^{k+1} = \mathbf{y}^k - \beta(A\mathbf{x}_1^{k+1} - \mathbf{b}) \quad \text{and} \quad \mathbf{s}^{k+1} = \mathbf{s}^k - \beta(\mathbf{x}_1^{k+1} - \mathbf{x}_2^{k+1}).$$

The minimizations over $\mathbf{x}_1$ and $\mathbf{x}_2$ are unconstrained optimization.

ADMM with more than two blocks

$$\begin{array}{ll} \min & f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + f_3(\mathbf{x}_3) \\ \text{s.t.} & A_1\mathbf{x}_1 + A_2\mathbf{x}_2 + A_3\mathbf{x}_3 = \mathbf{b}, \\ & \mathbf{x}_1 \in X_1,\ \mathbf{x}_2 \in X_2,\ \mathbf{x}_3 \in X_3. \end{array}$$

$$\begin{aligned} L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2, \mathbf{y}) &= f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + f_3(\mathbf{x}_3) - \mathbf{y}^T(A_1\mathbf{x}_1 + A_2\mathbf{x}_2 + A_3\mathbf{x}_3 - \mathbf{b}) \\ &+ \frac{\beta}{2}\|A_1\mathbf{x}_1 + A_2\mathbf{x}_2 + A_3\mathbf{x}_3 - \mathbf{b}\|^2. \end{aligned}$$

$$\begin{aligned} \mathbf{x}_1^{k+1} &= \arg\min_{\mathbf{x}_1 \in X_1} L(\mathbf{x}_1, \mathbf{x}_2^k, \mathbf{x}_3^k, \mathbf{y}^k), \\ \mathbf{x}_2^{k+1} &= \arg\min_{\mathbf{x}_2 \in X_2} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2, \mathbf{x}_3^k, \mathbf{y}^k) \\ \mathbf{x}_3^{k+1} &= \arg\min_{\mathbf{x}_3 \in X_3} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \mathbf{x}_3, \mathbf{y}^k), \\ \mathbf{y}^{k+1} &= \mathbf{y}^k - \beta(A_1\mathbf{x}_1^{k+1} + A_2\mathbf{x}_2^{k+1} + A_3\mathbf{x}_3^{k+1} - \mathbf{b}) \end{aligned}$$

Divergent Example of the 3-Block Extended ADMM

Consider

$$\begin{array}{ll} \min & 0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 \\ \text{s.t.} & A_1 x_1 + A_2 x_2 + A_3 x_3 = \mathbf{0}, \\ & \mathbf{x}_1 \in R,\ \mathbf{x}_2 \in R,\ \mathbf{x}_3 \in R; \end{array}$$

where

$$A = (A_1, A_2, A_3) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix}.$$

Thus the extended ADMM with $\beta = 1$ can be specified as

$$\begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & 6 & 0 & 0 & 0 & 0 \\ 5 & 7 & 9 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \\ \mathbf{y}^{k+1} \end{pmatrix} = \begin{pmatrix} 0 & -4 & -5 & 1 & 1 & 1 \\ 0 & 0 & -7 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1^{k} \\ x_2^{k} \\ x_3^{k} \\ \mathbf{y}^{k} \end{pmatrix}.$$

Or equivalently,

$$\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \mathbf{y}^{k+1} \end{pmatrix} = M \begin{pmatrix} x_2^{k} \\ x_3^{k} \\ \mathbf{y}^{k} \end{pmatrix},$$

where

$$M = \frac{1}{162} \begin{pmatrix} 144 & -9 & -9 & -9 & 18 \\ 8 & 157 & -5 & 13 & -8 \\ 64 & 122 & 122 & -58 & -64 \\ 56 & -35 & -35 & 91 & -56 \\ -88 & -26 & -26 & -62 & 88 \end{pmatrix}.$$

The matrix $M = V\mathrm{Diag}(\mathbf{d})V^{-1}$, where

$$\mathbf{d} = \begin{pmatrix} 0.9836 + 0.2984i \\ 0.9836 - 0.2984i \\ 0.8744 + 0.2310i \\ 0.8744 - 0.2310i \\ 0 \end{pmatrix}.$$

Note that spectral radius of $\rho(M) = |d_1| = |d_2| > 1$, and

$$V = \begin{bmatrix} 0.1314 + 0.2661i & 0.1314 - 0.2661i & 0.1314 - 0.2661i & 0.1314 + 0.2661i & 0 \\ 0.0664 - 0.2718i & 0.0664 + 0.2718i & 0.0664 + 0.2718i & 0.0664 - 0.2718i & 0 \\ -0.2847 - 0.4437i & -0.2847 + 0.4437i & 0.2847 - 0.4437i & 0.2847 + 0.4437i & 0.5774 \\ 0.5694 & 0.5694 & -0.5694 & -0.5694 & 0.5774 \\ -0.4270 + 0.2218i & -0.4270 - 0.2218i & 0.4270 + 0.2218i & 0.4270 - 0.2218i & 0.5774 \end{bmatrix},$$

Take the initial point $(x_2^0, x_3^0, \mathbf{y}^0)$ as $V(:,1) + V(:,2) \in R^5$. Then

$$
\begin{aligned}
\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \mathbf{y}^{k+1} \end{pmatrix} &= V\mathrm{Diag}(\mathbf{d}^{k+1})V^{-1} \begin{pmatrix} x_2^0 \\ x_3^0 \\ \mathbf{y}^0 \end{pmatrix} \\
&= V\mathrm{Diag}(\mathbf{d}^{k+1}) \begin{pmatrix} 1 \\ 1 \\ \mathbf{0} \end{pmatrix} \\
&= V \begin{pmatrix} (0.9836 + 0.2984i)^{k+1} \\ (0.9836 - 0.2984i)^{k+1} \\ \mathbf{0} \end{pmatrix},
\end{aligned}
$$

which is divergent.

In fact, it is divergent for every $\beta > 0$.

Strong Convexity Helps?

Consider the following example

$$
\begin{array}{ll}
\min & 0.05x_1^2 + 0.05x_2^2 + 0.05x_3^2 \\
\text{s.t.} & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}.
\end{array}
$$

- the matrix M in the extended ADMM ($\beta = 1$) has

$$\rho(M) = 1.0087 > 1$$

- able to find a proper initial point such that the extended ADMM diverges
- even for strongly convex programming, the extended ADMM is not necessarily convergent for a given $\beta > 0$

The Small Step-size Variant of ADMM

In the direct extension of ADMM, the Lagrangian multiplier is updated as

$$\mathbf{y}^{k+1} := \mathbf{y}^k - \gamma\beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - \mathbf{b}),$$

with a positive step-size γ.

Convergence is proved:

- One-Block; (Augmented Lagrangian Method) for all $\gamma \in (0, 2)$ (Hestenes '69, Powell '69).
- Two-Block; (Alternating Direction Method of Multipliers) for all $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ (Glowinski, '84).
- Three-Block; for γ sufficiently small provided additional conditions on the problem (Hong and Luo '12).

Question: Is there a problem-data-independent interval for γ such that the method converges?

A Numerical Study (Ongoing)

Consider the linear system

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1+\gamma \\ 1 & 1+\gamma & 1+\gamma \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}.$$

Table 1: The radius of the linear mapping

γ	1	0.1	1e-2	1e-3	1e-4	1e-5	1e-6	1e-7
$\rho(M)$	1.0278	1.0026	1.0001	>1	>1	>1	>1	>1

Thus, there seems no practical problem-data-independent γ such that the small-step size variant works.

An Open Problem

Is there a ”simple correction” of the extended ADMM for the multi-block convex minimization problems?

A Possible Answer: Independent and uniform random permutation in each iteration!

It works for the example, and it works in general?