Summer Course Project IV: ADMM for Conic Optimization

Yinyu Ye

June 2, 2018

1 Introduction

Consider solving the linear program

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & f^p(\mathbf{x}) \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{1}$$

Or solve

$$\begin{aligned} \text{minimize}_{\mathbf{y},\mathbf{s}} \quad & f^d(\mathbf{y}) \\ \text{s.t.} \quad & A^T\mathbf{y} + \mathbf{s} = \mathbf{c}, \\ & \mathbf{s} \geq \mathbf{0}; \end{aligned} \tag{2}$$

The augmented Lagrangian function for problem (1) would be

$$L^p(\mathbf{x}, \mathbf{y}) = f^p(\mathbf{x}) - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) + \frac{\beta}{2}\|A\mathbf{x} - \mathbf{b}\|^2, \tag{3}$$

where β is a positive parameter. And the one for problem (2) is

$$L^d(\mathbf{y}, \mathbf{s}, \mathbf{x}) = f^d(\mathbf{y}) - \mathbf{x}^T(A^T\mathbf{y} + \mathbf{s} - \mathbf{c}) + \frac{\beta}{2}\|A^T\mathbf{y} + \mathbf{s} - \mathbf{c}\|^2. \tag{4}$$

2 ADMM for solving Problem (1)

The **Augmented Lagrangian Method (ALM)** for the primal would be: starting from any $\mathbf{x}^0 \geq 0$ and $\mathbf{y}^0$, do the iterative update:

- Update variable $\mathbf{x}$:

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

- Update multiplier $\mathbf{y}$:

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

However, the computation of new $\mathbf{x}$ is still too much work – it is minimization over the nonnegative cone. We now reformulate problem (1) as

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}_1,\mathbf{x}_2} & f^p(\mathbf{x}_1) \\ \text{s.t.} & A\mathbf{x}_1 = \mathbf{b} \\ & \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0}; \\ & \mathbf{x}_2 \geq \mathbf{0}, \end{array} \tag{5}$$

and consider the split augmented Lagrangian function:

$$L^p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = f^p(\mathbf{x}_1) - \mathbf{y}^T(A\mathbf{x}_1 - \mathbf{b}) - \mathbf{s}^T(\mathbf{x}_1 - \mathbf{x}_2) + \frac{\beta}{2}\left(\|A\mathbf{x}_1 - \mathbf{b}\|^2 + \|\mathbf{x}_1 - \mathbf{x}_2\|^2\right). \tag{6}$$

Then the **Alternating Direction Method with Multipliers (ADMM)** would be: starting from any $\mathbf{x}_1^0$, $\mathbf{x}_2^0 \geq \mathbf{0}$, and multiplier $(\mathbf{y}^0, \mathbf{s}^0)$, do the iterative update:

- Update variable $\mathbf{x}_1$:

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

- Update variable $\mathbf{x}_2$:

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

- Update multipliers $\mathbf{y}$ and $\mathbf{s}$:

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

You may now find out that the updates of $\mathbf{x}_1$ and $\mathbf{x}_2$ become much easy! The update of $\mathbf{x}_1$ is a unconstrained minimization problem (often it has a close-form solution); and the update of $\mathbf{x}_2$, although still over the nonnegative cone, has a simple close-form solution.

Question 1: Write out the explicit formula for updating of $\mathbf{x}_1$ and $\mathbf{x}_2$. Implement the spliting ADMM in your favorite language or platform, and try it on some LP and/or convex QP problems. How does it perform? Does the update order of $\mathbf{x}_1$ and $\mathbf{x}_2$ make a difference?

Let $A' = (AA^T)^{-1/2}A$ and $\mathbf{b}' = (AA^T)^{-1/2}\mathbf{b}$, and consider

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}_1,\mathbf{x}_2} & f(\mathbf{x}_1) \\ \text{s.t.} & A'\mathbf{x}_1 = \mathbf{b}' \\ & \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0}; \\ & \mathbf{x}_2 \geq \mathbf{0}, \end{array} \tag{7}$$

This problem is equivalent to the original problem but the constraint matrix is preconditioned. Apply the same ADMM and try it on the preconditioned formulation (7), and compare its performance with that on solving (5).

3 ADMM for solving Problem (2)

The ADMM for the dual is straightforward: starting from any $\mathbf{y}^0$, $\mathbf{s}^0 \geq \mathbf{0}$, and multiplier $\mathbf{x}^0$, do the iterative update:

- Update variable $\mathbf{y}$:
$$\mathbf{y}^{k+1} = \arg\min_{\mathbf{y}} L^d(\mathbf{y}, \mathbf{s}^k, \mathbf{x}^k);$$
- Update slack variable $\mathbf{s}$:
$$\mathbf{s}^{k+1} = \arg\min_{\mathbf{s} \geq \mathbf{0}} L^d(\mathbf{y}^{k+1}, \mathbf{s}, \mathbf{x}^k);$$
- Update multipliers $\mathbf{x}$:
$$\mathbf{x}^{k+1} = \mathbf{x}^k - \beta(A^T\mathbf{y}^{k+1} + \mathbf{s}^{k+1} - \mathbf{c}).$$

Note that the updates of $\mathbf{y}$ is a least-squares problem with a fixed matrix (which needs to be factorized once), and the update of $\mathbf{s}$ has a simple close form. Also note that $\mathbf{x}$ would be eventually non-positive.

Question 2: Write out the explicit formula for updating of $\mathbf{y}$ and $\mathbf{s}$. Implement the ADMM in your favorite language or platform, and try it on some LP and/or convex QP problems. How does it perform? Does the update order of $\mathbf{y}$ and $\mathbf{s}$ make a difference?

4 Interior-Point ADMM

Now solving the linear program with the logarithmic barrier function

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}} & f^p(\mathbf{x}) - \mu \sum_j \ln(x_j) \\ \text{s.t.} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} > \mathbf{0}; \end{array} \tag{8}$$

or problem in the format

$$\begin{array}{ll} \text{minimize}_{\mathbf{y},\mathbf{s}} & f^d(\mathbf{y}) + \mu \sum_j \ln(s_j) \\ \text{s.t.} & A^T\mathbf{y} + \mathbf{s} = \mathbf{c}, \ \mathbf{s} > \mathbf{0} \\ & \mathbf{x} \geq \mathbf{0}; \end{array} \tag{9}$$

where μ is a fixed positive constant.

The augmented Lagrangian function for problem (8) would be

$$L^p_\mu(\mathbf{x},\mathbf{y}) = f^p(\mathbf{x}) - \mu\sum_j \ln(x_j) - \mathbf{y}^T(A\mathbf{x}-\mathbf{b}) + \frac{\beta}{2}\|A\mathbf{x}-\mathbf{b}\|^2; \tag{10}$$

and the one for problem (9) would be

$$L^d_\mu(\mathbf{y},\mathbf{s},\mathbf{x}) = f^(\mathbf{y})\mathbf{y} - \mu\sum_j \ln(s_j) - \mathbf{x}^T(A^T\mathbf{y}+\mathbf{s}-\mathbf{c}) + \frac{\beta}{2}\|A^T\mathbf{y}+\mathbf{s}-\mathbf{c}\|^2, \tag{11}$$

Question 3: Apply ADMM for problems (8) and (9). Again, you may split $\mathbf{x}$ in problem (8) into $\mathbf{x}_1$ and $\mathbf{x}_2$ to simplify the update. How do they perform? Again try your implementation on solving the preconditioned formulation (7) with barrier.

Now, we gradually reduced μ as an outer iteration. That is, we start some $\mu = \mu^0$ and apply the ADMM to compute an approximate optimizer, with its multiplier, for problem (8) or problem (9). Now set $\mu = \mu^1 = \gamma\mu^0$ where $0 < \gamma < 1$. Then we use the approximate optimizer and multiplier as the initial point to start ADMM for (8) and (9) with the new μ.

Question 4: Implement the Outer-Iteration process described above, and try different β and γ to see how it performs.

5 Multi-Block ADMM

Question 5: What about to further split variables $\mathbf{x}$ in the primal and/or $\mathbf{y}$ in the dual, and apply the fixed order or random permutation order in each update cycle.

More precisely, consider solving the dual and matrix $A = [A_1;\ A_2]$, $\mathbf{b} = [\mathbf{b}_1;\mathbf{b}_2]$, and $\mathbf{y} = [\mathbf{y}_1;\mathbf{y}_2]$, then the augmented Lagrangian function

$$L^d(\mathbf{y}_1,\mathbf{y}_2,\mathbf{s},\mathbf{x}) = -\mathbf{b}_1^T\mathbf{y}_1 - \mathbf{b}_2^T\mathbf{y}_2 - \mathbf{x}^T(A_1^T\mathbf{y}_1 + A_2^T\mathbf{y}_2 + \mathbf{s} - \mathbf{c}) + \frac{\beta}{2}\|A_1^T\mathbf{y}_1 + A_2^T\mathbf{y}_2 + \mathbf{s} - \mathbf{c}\|^2. \tag{12}$$

Starting from any $\mathbf{y}_1^0$, $\mathbf{y}_2^0$, $\mathbf{s}^0 \geq \mathbf{0}$, and multiplier $\mathbf{x}^0$, do the iterative update:

- Update variable $\mathbf{y}_1$:

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

- Update variable $\mathbf{y}_2$:

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

- Update slack variable $\mathbf{s}$:

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

- Update multipliers $\mathbf{x}$:

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

Note that the least-squares problem for each individual block $\mathbf{y}_i$ involves a smaller matrix $(A_iA_i^T)$.

One can also consider to reformulate the dual as

$$\begin{array}{ll} \text{maximize}_{\mathbf{y},\mathbf{s},\mathbf{u}_1,\mathbf{u}_2} & \mathbf{b}^T\mathbf{y} \\ \text{s.t.} & A_1^T\mathbf{y}_1 - \mathbf{u}_1 = \mathbf{0}, \quad (\mathbf{v}_1) \\ & A_2^T\mathbf{y}_2 - \mathbf{u}_2 = \mathbf{0}, \quad (\mathbf{v}_2) \\ & \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{s} = \mathbf{c}, \quad (\mathbf{x}) \\ & \mathbf{s} \geq \mathbf{0}; \end{array} \tag{13}$$

with the multiplier $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{x}$ for the three sets of the equality constraints. The augmented Lagrangian function becomes

$$\begin{array}{rl} L^d(\mathbf{y}_1,\mathbf{y}_2,\mathbf{u}_1,\mathbf{u}_2,\mathbf{s},\mathbf{v}_1,\mathbf{v}_2,\mathbf{x}) & = -\mathbf{b}_1^T\mathbf{y}_1 - \mathbf{b}_2^T\mathbf{y}_2 - \mathbf{v}_1^T(A_1^T\mathbf{y}_1 - \mathbf{u}_1) - \mathbf{v}_2^T(A_2^T\mathbf{y}_2 - \mathbf{u}_2) - \mathbf{x}^T(\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{s} - \mathbf{c}) \\ & + \frac{\beta}{2}\left(\|A_1^T\mathbf{y}_1 - \mathbf{u}_1\|^2 + \|A_2^T\mathbf{y}_2 - \mathbf{u}_2\|^2 + \|\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{s} - \mathbf{c}\|^2\right). \end{array} \tag{14}$$

Note that $\mathbf{y}_i$, $i = 1, 2$, and $\mathbf{s} \geq \mathbf{0}$ can be independently and in parallel, and $\mathbf{u}_i$, $i = 1, 2$, can be updated jointly with a close form(?). This is essentially a two-block ADMM and guaranteed to be convergent.

6 ADMM for SDP Cones

Consider solving the SDP problem

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}} & C \bullet X \\ \text{s.t.} & \mathcal{A}X = \mathbf{b}, \\ & X \succeq \mathbf{0}; \end{array} \tag{15}$$

or its dual

$$\begin{array}{ll} \text{maximize}_{\mathbf{y},S} & \mathbf{b}^T\mathbf{y} \\ \text{s.t.} & \mathcal{A}^T\mathbf{y} + S = C, \\ & S \succeq \mathbf{0}; \end{array} \tag{16}$$

Repeat Questions 1-5 for SDP optimization, where the primal problem (15) is analog to (15), and the dual problem (16) is analog to (2).

References

[1] Chen, He, Ye, and Yuan. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Mathematical Programming, 155 (1-2), 2016, 57-79.

[2] D. Davis and W. Yin. Convergence rate analysis of several splitting schemes. http://www.math.ucla.edu/ wotaoyin/papers/convergence_rate_splitting.html

[3] D. Luenberger and Y. Ye. Linear and Nonlinear Programming. http://web.stanford.edu/class/msande310/310trialtext.pdf

[4] Sun, Luo, and Ye. On the expected convergence of randomly permuted ADMM. 2015, https://arxiv.org/abs/1503.06387