Multi-block ADMM Methods for Linear Programming

Nico Chaves, Junjie Zhu

December 7, 2016

1 Introduction

The goal of this project is to use alternating direction method of multipliers (ADMM)

$$\begin{aligned} \text{minimize}_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{OPT1}$$

or its dual

$$\begin{aligned} \text{maximize}_{\mathbf{y},\mathbf{s}} \quad & \mathbf{b}^T\mathbf{y} \\ \text{subject to} \quad & A^T\mathbf{y} + \mathbf{s} = \mathbf{c}, \\ & \mathbf{s} \geq \mathbf{0}. \end{aligned} \tag{OPT2}$$

2 Algorithms

Primal ADMM

We re-formulate the problem as:

$$\begin{aligned} \text{minimize}_{\mathbf{x}_1,\mathbf{x}_2} \quad & \mathbf{c}^T\mathbf{x}_1 \\ \text{subject to} \quad & A\mathbf{x}_1 = \mathbf{b} \\ & \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0} \\ & \mathbf{x}_2 \geq \mathbf{0} \end{aligned} \tag{OPT3}$$

anc consider the split Lagrangian function:

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

Primal ADMM Update

$$\nabla_{\mathbf{x}_1} L^P(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = \mathbf{c} - A^T\mathbf{y} - \mathbf{s} + \beta\left(A^T(A\mathbf{x}_1 - \mathbf{b}) + (\mathbf{x}_1 - \mathbf{x}_2)\right)$$

Setting the gradient to 0, we obtain the update step for $\mathbf{x}_1$:

$$\mathbf{x}_1 = \left(A^TA + I\right)^{-1}\left(\frac{1}{\beta}A^T\mathbf{y} + \frac{1}{\beta}\mathbf{s} - \frac{1}{\beta}\mathbf{c} + A^T\mathbf{b} + \mathbf{x}_2\right)$$

For $\mathbf{x}_2$:

$$\nabla_{\mathbf{x}_2} L^P = \mathbf{s} + \beta(\mathbf{x}_2 - \mathbf{x}_1)$$

Setting the gradient to 0, we obtain the update step for $\mathbf{x}_2$:

$$\mathbf{x}_2 = \max\left\{\mathbf{x}_1 - \frac{\mathbf{s}}{\beta}, 0\right\}$$

where the max is computed component-wise to ensure that $\mathbf{x}_2 \geq 0$.

Primal with Block Splitting

For simplicity, we first consider splitting the problem into 2 blocks of equal size as follows:

$$\mathbf{x}_1 = \begin{bmatrix}\mathbf{x}_{1,1}\\ \mathbf{x}_{1,2}\end{bmatrix}$$

$$A = \begin{bmatrix}A_1 & A_2\end{bmatrix}$$

Then, the Lagrangian can be expressed as:

$$\begin{aligned} L^P(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) &= \mathbf{c}_1^T\mathbf{x}_{1,1} + \mathbf{c}_2^T\mathbf{x}_{1,2} - \mathbf{y}^T\left(A_1\mathbf{x}_{1,1} + A_2\mathbf{x}_{1,2} - \mathbf{b}\right) - \mathbf{s}_1^T\left(\mathbf{x}_{1,1} - \mathbf{x}_{2,1}\right) - \mathbf{s}_2^T\left(\mathbf{x}_{1,2} - \mathbf{x}_{2,2}\right) \\ &\quad + \frac{\beta}{2}\left(\left\|A_1\mathbf{x}_{1,1} + A_2\mathbf{x}_{1,2} - \mathbf{b}\right\|^2 + \left\|\mathbf{x}_{1,1} - \mathbf{x}_{2,1}\right\|^2 + \left\|\mathbf{x}_{1,2} - \mathbf{x}_{2,2}\right\|^2\right) \end{aligned}$$

Taking the gradient with respect to the 1st block of $\mathbf{x}_1$:

$$\nabla_{\mathbf{x}_{1,1}} L^P(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = \mathbf{c}_1 - A_1^T\mathbf{y} - \mathbf{s}_1 + \beta\left(A_1^T\left(A_1\mathbf{x}_{1,1} + A_2\mathbf{x}_{1,2} - \mathbf{b}\right) + \left(\mathbf{x}_{1,1} - \mathbf{x}_{2,1}\right)\right)$$

Setting the gradient to 0, we obtain the update for $\mathbf{x}_{1,1}$:

$$\mathbf{x}_{1,1}^{k+1} = \left(A_1^TA_1 + I\right)^{-1}\left(\frac{1}{\beta}A_1^T\mathbf{y} + \frac{1}{\beta}\mathbf{s}_1 - \frac{1}{\beta}\mathbf{c}_1 + A_1^T\mathbf{b} + \mathbf{x}_{2,1}^k - A_1^TA_2\mathbf{x}_{1,2}^k\right)$$

By symmetry, the update step for the 2nd block of $\mathbf{x}_1$ is:

$$\mathbf{x}_{1,2}^{k+1} = \left(A_2^TA_2 + I\right)^{-1}\left(\frac{1}{\beta}A_2^T\mathbf{y} + \frac{1}{\beta}\mathbf{s}_2 - \frac{1}{\beta}\mathbf{c}_2 + A_2^T\mathbf{b} + \mathbf{x}_{2,2}^k - A_2^TA_1\mathbf{x}_{1,1}^{k+1}\right)$$

Note that here we use $\mathbf{x}_{1,1}^{k+1}$ to update $\mathbf{x}_{1,2}$. If we use a randomized update order, then we may end up updating $\mathbf{x}_{1,2}$ first instead.

Block Splitting (for a general number of blocks)

We can easily extend the above result to a general number of blocks. Let U denote the set of blocks which have already been updated at the current iteration. To update block i of $\mathbf{x}_1$, we compute:

$$\mathbf{x}_{1,i}^{k+1} = \left(A_i^TA_i + I\right)^{-1}\left(\frac{1}{\beta}A_i^T\mathbf{y} + \frac{1}{\beta}\mathbf{s}_i - \frac{1}{\beta}\mathbf{c}_i + A_i^T\mathbf{b} + \mathbf{x}_{2,i}^k - \sum_{j\neq i, j\in U} A_i^TA_j\mathbf{x}_{1,j}^{k+1} - \sum_{j\neq i, j\notin U} A_i^TA_j\mathbf{x}_{1,j}^k\right)$$

where A_i refers to the i^{th} block of columns of A.

Dual ADMM

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

Dual ADMM Update

$$\nabla_{\mathbf{y}} L^d(\mathbf{y}, \mathbf{s}, \mathbf{x}) = -\mathbf{b} - A\mathbf{x} + \beta A \left(A^T \mathbf{y} + \mathbf{s} - \mathbf{c}\right)$$

Setting the gradient to 0, we obtain the update step for $\mathbf{y}$:

$$\begin{aligned}
A\left(A^T \mathbf{y} + \mathbf{s} - \mathbf{c}\right) &= \frac{1}{\beta}\left(A\mathbf{x} + \mathbf{b}\right) \\
AA^T \mathbf{y} &= \frac{1}{\beta}\left(A\mathbf{x} + \mathbf{b}\right) - A\mathbf{s} + A\mathbf{c} \\
\mathbf{y} &= \left(AA^T\right)^{-1}\left(\frac{1}{\beta}\left(A\mathbf{x} + \mathbf{b}\right) - A\mathbf{s} + A\mathbf{c}\right)
\end{aligned}$$

For $\mathbf{s}$:

$$\nabla_{\mathbf{s}} L^d(\mathbf{y}, \mathbf{s}, \mathbf{x}) = -\mathbf{x} + \beta\left(A^T \mathbf{y} + \mathbf{s} - \mathbf{c}\right)$$

Setting the gradient to 0, we obtain the update step for s:

$$\mathbf{s} = \max\left\{\frac{1}{\beta}\mathbf{x} - A^T \mathbf{y} + \mathbf{c}, 0\right\}$$

Dual ADMM With Block Splitting

For simplicity, we first consider splitting the problem into 2 blocks of equal size as follows:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$$

$$A^T = \begin{bmatrix} A_1^T & A_2^T \end{bmatrix}$$

Then the Lagrangian can be expressed as:

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

Differentiating with respect to $\mathbf{y}_1$:

$$\nabla_{\mathbf{y}_1} L^d(\mathbf{y}, \mathbf{s}, \mathbf{x}) = -\mathbf{b}_1 - A_1\mathbf{x} + \beta A_1\left(A_1^T \mathbf{y}_1 + A_2^T \mathbf{y}_2 + \mathbf{s} - \mathbf{c}\right)$$

Setting the gradient to 0, we obtain the update for $\mathbf{y}_1$:

$$\begin{aligned}
A_1\left(A_1^T \mathbf{y}_1^{k+1} + A_2^T \mathbf{y}_2^k + \mathbf{s} - \mathbf{c}\right) &= \frac{1}{\beta}\left(A_1\mathbf{x} + \mathbf{b}_1\right) \\
A_1 A_1^T \mathbf{y}_1^{k+1} &= \frac{1}{\beta}\left(A_1\mathbf{x} + \mathbf{b}_1\right) - A_1\left(A_2^T \mathbf{y}_2^k + \mathbf{s} - \mathbf{c}\right) \\
\mathbf{y}_1^{k+1} &= \left(A_1 A_1^T\right)^{-1}\left(\frac{1}{\beta}\left(A_1\mathbf{x} + \mathbf{b}_1\right) - A_1\left(A_2^T \mathbf{y}_2^k + \mathbf{s} - \mathbf{c}\right)\right)
\end{aligned}$$

By symmetry, the update for $\mathbf{y}_2$ is:

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

assuming that we update $\mathbf{y}_2$ after $\mathbf{y}_1$.

Dual Block Splitting (for a general number of blocks)

We can easily extend the above result to a general number of blocks. Let U denote the set of blocks which have already been updated at the current iteration. To update block i of $\mathbf{y}$:

$$\mathbf{y}_i^{k+1} = \left(A_i A_i^T\right)^{-1} \left(\frac{1}{\beta} \left(A_i \mathbf{x} + \mathbf{b}_i\right) - \sum_{j \neq i, j \in U} A_i A_j^T \mathbf{y}_j^{k+1} - \sum_{j \neq i, j \notin U} A_i A_j^T \mathbf{y}_j^k - A_i \left(\mathbf{s} - \mathbf{c}\right) \right)$$

where A_i refers to rows of A.

3 Interior-Point ADMM

Primal

We can use the previous formulation in (OPT3) with $\mathbf{x}_1$ and $\mathbf{x}_2$ using the barrier function:

$$\begin{aligned} \text{minimize}_{\mathbf{x}_1, \mathbf{x}_2} \quad & \mathbf{c}^T \mathbf{x}_1 + \mu \sum_j \ln((\mathbf{x}_2)_j) \\ \text{subject to} \quad & A\mathbf{x}_1 = \mathbf{b}, \\ & \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0}, \\ & \mathbf{x}_2 > \mathbf{0} \end{aligned} \tag{OPT4}$$

$$L_\mu^p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = \mathbf{c}^T \mathbf{x}_1 - \mu \sum_j \ln\left(x_{1,j}\right) - \mathbf{y}^T \left(A\mathbf{x}_1 - \mathbf{b}\right) - \mathbf{s}^T \left(\mathbf{x}_1 - \mathbf{x}_2\right) + \frac{\beta}{2} \left(\left\| A\mathbf{x}_1 - \mathbf{b} \right\|^2 + \left\| \mathbf{x}_1 - \mathbf{x}_2 \right\|^2 \right)$$

TODO: everything past here is not complete.

Update

‘

$$\nabla_{\mathbf{x}_1} L_\mu^P(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = \mathbf{c} - \mu \left(\frac{1}{\mathbf{x}_1} \right) - A^T \mathbf{y} - \mathbf{s} + \beta \left(A^T \left(A\mathbf{x}_1 - \mathbf{b}\right) + \left(\mathbf{x}_1 - \mathbf{x}_2\right) \right)$$

Setting the gradient to 0, we obtain the update step for x_1:

$$\begin{aligned} 0 &= \mathbf{c} - \mu \left(\frac{1}{\mathbf{x}_1} \right) - A^T \mathbf{y} - \mathbf{s} + \beta \left(A^T \left(A\mathbf{x}_1 - \mathbf{b}\right) + \left(\mathbf{x}_1 - \mathbf{x}_2\right) \right) \\ &= \end{aligned}$$

Dual Update

We can use the previous formulation in (OPT2) using the barrier function:

$$\begin{aligned} \text{maximize}_{\mathbf{y}, \mathbf{s}} \quad & \mathbf{b}^T \mathbf{y} + \mu \sum_j \ln(s_j) \\ \text{subject to} \quad & A^T \mathbf{y} + \mathbf{s} = \mathbf{c}, \\ & \mathbf{s} > \mathbf{0}. \end{aligned} \tag{OPT5}$$

$$L_{\mu}^{d}(\mathbf{y},\mathbf{s},\mathbf{x}) = -\mathbf{b}^{T}\mathbf{y} - \mu \sum_{j} \ln\left(s_{j}\right) - \mathbf{x}^{T}\left(A^{T}\mathbf{y} + \mathbf{s} - \mathbf{c}\right) + \frac{\beta}{2}\left\|A^{T}\mathbf{y} + \mathbf{s} - \mathbf{c}\right\|^{2}$$

Update for y:

$$\nabla_{\mathbf{y}} L_{\mu}^{d}(\mathbf{y},\mathbf{s},\mathbf{x}) = -\mathbf{b} - A\mathbf{x} + \beta A\left(A^{T}\mathbf{y} + \mathbf{s} - \mathbf{c}\right)$$

Setting to 0, we get the update:

$$\mathbf{y} = A^{-T}\left(\frac{1}{\beta}A^{-1}\left(A\mathbf{x} + \mathbf{b}\right) - \mathbf{s} + \mathbf{c}\right)$$

Update for s:

$$\nabla_{\mathbf{s}} L_{\mu}^{d}(\mathbf{y},\mathbf{s},\mathbf{x}) = -\mu\left(\frac{1}{\mathbf{s}}\right) - \mathbf{x} + \beta\left(A^{T}\mathbf{y} + \mathbf{s} - \mathbf{c}\right)$$