ADMM: The Variational Inequality Perspective

Hailiang ZHAO @ ZJU-CS

http://hliangzhao.me

October 25, 2022

Outline

Outline

Variational Inequality

In this slide, we introduce the variational inequality viewpoint of ADMM.

Recall that ADMM is for solving the following problem:

$$\min_{\boldsymbol{x},\boldsymbol{y}} \quad f(\boldsymbol{x}) + g(\boldsymbol{y}), \quad s.t. \quad \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} = \boldsymbol{b}. \tag{1}$$

It introduces the Lagrangian function:

$$L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + g(\boldsymbol{y}) + \langle \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b}, \boldsymbol{\lambda} \rangle. \tag{2}$$

And, $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}^*)$ is a saddle point if it satisfies

$$L(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}) \leq L(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\lambda}^*) \leq L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}^*), \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}. \tag{3}$$

Variational Inequality

For the left inequality of (3), we have

$$\langle \boldsymbol{\lambda} - \boldsymbol{\lambda}^*, \mathbf{A}\boldsymbol{x}^* - \mathbf{B}\boldsymbol{y}^* - \boldsymbol{b} \rangle \le 0. \tag{4}$$

For the right inequality of (3), we have

$$\begin{aligned} & f(\boldsymbol{x}) + g(\boldsymbol{y}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) \\ & + \langle \boldsymbol{x} - \boldsymbol{x}^*, \mathbf{A}^T\boldsymbol{\lambda}^* \rangle + \langle \boldsymbol{y} - \boldsymbol{y}^*, \mathbf{B}^T\boldsymbol{\lambda}^* \rangle \ge 0, \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}. \end{aligned} \tag{5}$$

Combing them together, we have

$$\begin{aligned} & f(\boldsymbol{x}) + g(\boldsymbol{y}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*) \\ & + \left\langle \begin{pmatrix} \boldsymbol{x} - \boldsymbol{x}^* \\ \boldsymbol{y} - \boldsymbol{y}^* \\ \boldsymbol{\lambda} - \boldsymbol{\lambda}^* \end{pmatrix}, \begin{pmatrix} \mathbf{A}^T\boldsymbol{\lambda}^* \\ \mathbf{B}^T\boldsymbol{\lambda}^* \\ -(\mathbf{A}\boldsymbol{x}^* + \mathbf{B}\boldsymbol{y}^* - \boldsymbol{b}) \end{pmatrix} \right\rangle \\ & + \langle \boldsymbol{y} - \boldsymbol{y}^*, \mathbf{B}^T\boldsymbol{\lambda}^* \rangle \ge 0, \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}. \end{aligned} \tag{6}$$

Variational Inequality

Denote

$$\boldsymbol{w} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{\lambda} \end{pmatrix}, \boldsymbol{u}(\boldsymbol{w}) = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}, \text{ and } F(\boldsymbol{w}) = \begin{pmatrix} \mathbf{A}^T\boldsymbol{\lambda} \\ \mathbf{B}^T\boldsymbol{\lambda} \\ -(\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b}) \end{pmatrix}.$$

We further define

$$\theta(\boldsymbol{u}) = f(\boldsymbol{x}) + g(\boldsymbol{y}), \tag{7}$$

where $\boldsymbol{u}$ is a simplification of $\boldsymbol{u}(\boldsymbol{w})$. The (6) reduces to

$$\theta(\boldsymbol{u}) - \theta(\boldsymbol{u}^*) + \langle \boldsymbol{w} - \boldsymbol{w}^*, F(\boldsymbol{w}^*) \rangle \geq 0, \forall \boldsymbol{w}. \tag{8}$$

(8) is called the variational inequality of problem (1).

Variational Inequality

We can easily check that

$$\langle \boldsymbol{w} - \hat{\boldsymbol{w}}, F(\boldsymbol{w}) - F(\hat{\boldsymbol{w}}) \rangle = 0, \forall \boldsymbol{w}, \hat{\boldsymbol{w}}. \tag{9}$$

Thus, (8) is equivalent to

$$\theta(\boldsymbol{u}) - \theta(\boldsymbol{u}^*) + \langle \boldsymbol{w} - \boldsymbol{w}^*, F(\boldsymbol{w}) \rangle \geq 0, \forall \boldsymbol{w}. \tag{10}$$

Note that we want to approximate the optimal solution $\boldsymbol{w}^*$. We say that $\tilde{\boldsymbol{w}}$ is an approximate solution of the variational inequality problem (10) with accuracy ϵ if it satisfies

$$\theta(\tilde{\boldsymbol{u}}) - \theta(\boldsymbol{u}) + \langle \tilde{\boldsymbol{w}} - \boldsymbol{w}, F(\boldsymbol{w}) \rangle \leq \epsilon, \forall \boldsymbol{w}. \tag{11}$$

Especially, $\epsilon = 0$ gives (10).

Outline

Unified Framework in Variational Inequality

With (9), (10) of *ADMM: Part 2* and the convexity of f and g, we have

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^{k+1}) + \langle \mathbf{A}^T \tilde{\boldsymbol{\lambda}}^{k+1}, \boldsymbol{x} - \boldsymbol{x}^{k+1} \rangle \geq 0, \tag{12}$$

$$\begin{aligned} g(\boldsymbol{y}) - g(\boldsymbol{y}^{k+1}) + \langle \mathbf{B}^T \tilde{\boldsymbol{\lambda}}^{k+1} + \beta \mathbf{B}^T (\mathbf{B}\boldsymbol{y}^{k+1} - \mathbf{B}\boldsymbol{y}^k), \\ \boldsymbol{y} - \boldsymbol{y}^{k+1} \rangle \geq 0, \end{aligned} \tag{13}$$

where

$$\tilde{\boldsymbol{w}}^{k+1} = \begin{pmatrix} \tilde{\boldsymbol{x}}^{k+1} \\ \tilde{\boldsymbol{y}}^{k+1} \\ \tilde{\boldsymbol{\lambda}}^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}^{k+1} \\ \boldsymbol{y}^{k+1} \\ \boldsymbol{\lambda}^k + \beta(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^k - \boldsymbol{b}) \end{pmatrix}$$

Unified Framework in Variational Inequality

Add (12) to (13) we have

$$
\begin{aligned}
& f(\boldsymbol{x}) + g(\boldsymbol{y}) - f(\tilde{\boldsymbol{x}}^{k+1}) - g(\tilde{\boldsymbol{y}}^{k+1}) \\
& + \left\langle \begin{pmatrix} \boldsymbol{x} - \tilde{\boldsymbol{x}}^{k+1} \\ \boldsymbol{y} - \tilde{\boldsymbol{y}}^{k+1} \\ \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}^{k+1} \end{pmatrix}, \begin{pmatrix} \mathbf{A}^T \tilde{\boldsymbol{\lambda}}^{k+1} \\ \mathbf{B}^T \tilde{\boldsymbol{\lambda}}^{k+1} \\ -(\mathbf{A}\tilde{\boldsymbol{x}}^{k+1} + \mathbf{B}\tilde{\boldsymbol{y}}^{k+1} - \boldsymbol{b}) \end{pmatrix} \right\rangle \\
& \geq \langle \tilde{\boldsymbol{\lambda}}^{k+1} - \boldsymbol{\lambda}, \mathbf{A}\tilde{\boldsymbol{x}}^{k+1} + \mathbf{B}\tilde{\boldsymbol{y}}^{k+1} - \boldsymbol{b} \rangle \\
& + \beta \langle \mathbf{B}\tilde{\boldsymbol{y}}^{k+1} - \mathbf{B}\boldsymbol{y}^k, \mathbf{B}\tilde{\boldsymbol{y}}^{k+1} - \mathbf{B}\boldsymbol{y} \rangle \\
& = \begin{pmatrix} \mathbf{A}\tilde{\boldsymbol{x}}^{k+1} - \mathbf{A}\boldsymbol{x} \\ \mathbf{B}\tilde{\boldsymbol{y}}^{k+1} - \mathbf{B}\boldsymbol{y} \\ \tilde{\boldsymbol{y}}^{k+1} - \boldsymbol{\lambda} \end{pmatrix}^T \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \frac{1}{\beta}\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}\tilde{\boldsymbol{x}}^{k+1} - \mathbf{A}\boldsymbol{x}^k \\ \mathbf{B}\tilde{\boldsymbol{y}}^{k+1} - \mathbf{B}\boldsymbol{y}^k \\ \tilde{\boldsymbol{y}}^{k+1} - \boldsymbol{\lambda}^k \end{pmatrix} \\
& = (\tilde{\boldsymbol{w}}^{k+1} - \boldsymbol{w})^T \mathbf{P}^T \mathbf{H} \mathbf{M} \mathbf{P} (\tilde{\boldsymbol{w}}^{k+1} - \boldsymbol{w}^k). \qquad (14)
\end{aligned}
$$

Unified Framework in Variational Inequality

In the above slide,

$$\mathbf{H} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\beta}\mathbf{I} \end{pmatrix}, \mathbf{M} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \beta\mathbf{I} & \mathbf{I} \end{pmatrix}, \text{ and } \mathbf{P} = \begin{pmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

From (11) of *ADMM: Part 2* we have $\mathbf{P}\boldsymbol{w}^{k+1} = \mathbf{P}\boldsymbol{w}^k - \mathbf{M}(\mathbf{P}\boldsymbol{w}^k - \mathbf{P}\tilde{\boldsymbol{w}}^{k+1})$. Now, we can *summarize ADMM to the unified variational inequality framework:*

1. *Predict* $\tilde{\boldsymbol{w}}^{k+1}$ *satisfying*

$$\theta(\boldsymbol{u}) - \theta(\tilde{\boldsymbol{u}}^{k+1}) + \langle \boldsymbol{w} - \tilde{\boldsymbol{w}}^{k+1}, F(\tilde{\boldsymbol{w}}^{k+1}) \rangle \geq (\tilde{\boldsymbol{w}}^{k+1} - \boldsymbol{w})^T \mathbf{P}^T \mathbf{H}\mathbf{M}(\mathbf{P}\tilde{\boldsymbol{w}}^{k+1} - \boldsymbol{\xi}^k), \forall \boldsymbol{w}. \quad (15)$$

2. *Correct* $\boldsymbol{\xi}^{k+1} (:= \mathbf{P}\boldsymbol{w}^{k+1})$ *by*

$$\boldsymbol{\xi}^{k+1} = \boldsymbol{\xi}^k - \mathbf{M}(\boldsymbol{\xi}^k - \mathbf{P}\tilde{\boldsymbol{w}}^{k+1}). \quad (16)$$

Cast LADMM to the Framework

Similarly, Linearized ADMM (including LADMM-PS, etc.) can be casted into the above framework, too. We have:

$$
\begin{aligned}
& f(\boldsymbol{x}) + g(\boldsymbol{y}) - f(\tilde{\boldsymbol{x}}^{k+1}) - g(\tilde{\boldsymbol{y}}^{k+1}) \\
& + \left\langle \begin{pmatrix} \boldsymbol{x} - \tilde{\boldsymbol{x}}^{k+1} \\ \boldsymbol{y} - \tilde{\boldsymbol{y}}^{k+1} \\ \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}^{k+1} \end{pmatrix}, \begin{pmatrix} \mathbf{A}^T \tilde{\boldsymbol{\lambda}}^{k+1} \\ \mathbf{B}^T \tilde{\boldsymbol{\lambda}}^{k+1} \\ -(\mathbf{A}\tilde{\boldsymbol{x}}^{k+1} + \mathbf{B}\tilde{\boldsymbol{y}}^{k+1} - \boldsymbol{b}) \end{pmatrix} \right\rangle \\
& \geq (\tilde{\boldsymbol{w}}^{k+1} - \boldsymbol{w})^T \mathbf{H}\mathbf{M}(\tilde{\boldsymbol{w}}^{k+1} - \boldsymbol{w}^k),
\end{aligned} \tag{17}
$$

where $\tilde{\boldsymbol{w}}$ is the same as (14),

$$
\mathbf{H} = \begin{pmatrix} \beta\|\mathbf{A}\|^2\mathbf{I} - \beta\mathbf{A}^T\mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta\|\mathbf{B}\|^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\beta}\mathbf{I} \end{pmatrix}, \mathbf{M} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \beta\mathbf{B} & \mathbf{I} \end{pmatrix}.
$$

We further have

$$
\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - \mathbf{M}(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^{k+1}). \tag{18}
$$

Cast ADMM-PC to the Framework

For the multi-block problems defined in (1) of *The ADMM Slide: Part 3*, denote

$$\boldsymbol{w} = \begin{pmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_m \\ \boldsymbol{\lambda} \end{pmatrix}, \boldsymbol{u} = \begin{pmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_m \end{pmatrix}, F(\boldsymbol{w}) = \begin{pmatrix} \mathbf{A}_1^T \boldsymbol{\lambda} \\ \vdots \\ \mathbf{A}_m^T \boldsymbol{\lambda} \\ -(\sum_{i=1}^m \mathbf{A}_i \boldsymbol{x}_i - \boldsymbol{b}) \end{pmatrix},$$

$$\theta(\boldsymbol{u}) = \sum_{i=1}^m f_i(\boldsymbol{x}_i), \tag{19}$$

and

$$\tilde{\boldsymbol{\lambda}}^{k+1} = \boldsymbol{\lambda}^k + \beta(\sum_{i=1}^m \boldsymbol{\xi}_i^k - \boldsymbol{b}). \tag{20}$$

Recall that $\boldsymbol{\xi}_i = \mathbf{A}_i \boldsymbol{x}_i$.

Cast ADMM-PC to the Framework

For the prediction precedure of ADMM-PC ((28), (29) of *The ADMM Slide: Part 3*), we have

$$
\begin{aligned}
&\sum_{i=1}^{m} f_i(\boldsymbol{x}_i) - \sum_{i=1}^{m} f_i(\tilde{\boldsymbol{x}}^{k+1}) \\
&+ \left\langle \begin{pmatrix} \boldsymbol{x}_1 - \tilde{\boldsymbol{x}}_1^{k+1} \\ \vdots \\ \boldsymbol{x}_m - \tilde{\boldsymbol{x}}_m^{k+1} \\ \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}^{k+1} \end{pmatrix}, \begin{pmatrix} \mathbf{A}_1^T \tilde{\boldsymbol{\lambda}}^{k+1} \\ \vdots \\ \mathbf{A}_m^T \tilde{\boldsymbol{\lambda}}^{k+1} \\ -(\sum_{i=1}^{m} \mathbf{A}_i \tilde{\boldsymbol{x}}_i^{k+1} - \boldsymbol{b}) \end{pmatrix} \right\rangle \\
&\geq (\tilde{\boldsymbol{w}}^{k+1} - \boldsymbol{w})^T \mathbf{P}^T \mathbf{H} \mathbf{M} (\mathbf{P} \tilde{\boldsymbol{w}}^{k+1} - \boldsymbol{\xi}^k),
\end{aligned} \tag{21}
$$

where

$$
\mathbf{H} = \begin{pmatrix} \beta \mathbf{L}\mathbf{L}^T & \mathbf{0} \\ \mathbf{0} & \frac{1}{\beta}\mathbf{I} \end{pmatrix}, \mathbf{M} = \begin{pmatrix} \mathbf{L}^{-T} & \mathbf{0} \\ \beta(\mathbf{I}, \ldots, \mathbf{I}) & \mathbf{I}. \end{pmatrix} \tag{22}
$$

Cast ADMM-PC to the Framework

In (21) we also denote (cont'd)

$$\mathbf{P} = \begin{pmatrix} \mathbf{A}_1 & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{A}_m & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I} \end{pmatrix} \tag{23}$$

and

$$\tilde{\boldsymbol{w}} = (\tilde{\boldsymbol{x}}_1, ..., \tilde{\boldsymbol{x}}_m^T, \tilde{\boldsymbol{\lambda}}^T)^T \tag{24}$$

$$\boldsymbol{\xi} = (\boldsymbol{\xi}_1^T, ..., \boldsymbol{\xi}_m^T, \boldsymbol{\lambda}^T)^T. \tag{25}$$

Further we have

$$\boldsymbol{\xi}^{k+1} = \boldsymbol{\xi}^k - \mathbf{M}(\boldsymbol{\xi}^k - \mathbf{P}\tilde{\boldsymbol{w}}^{k+1}). \tag{26}$$

Unified Convergence Analysis

In the above slides, we cast all the ADMMs into the framework of variational inequality. Thus, the convergent results of the ADMMs can also be expressed in the variational inequality way.

The reuslt can be found at Theorem 3.17 of the ADMM book.

Outline

References

1. Lin, Zhouchen, Huan Li, and Cong Fang. *Alternating Direction Method of Multipliers for Machine Learning.* Springer Nature, 2022.
2. Boyd, Stephen, Stephen P. Boyd, and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.