ADMM for Nonlinear Convex Problems

Hailiang ZHAO @ ZJU-CS

http://hliangzhao.me

October 25, 2022

Nonlinear Convex Problems

In this slide, we introduce how to extend ADMM to solve the generally convex program with both equality and inequality constraints. Consider problem:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{y}} \quad & f(\boldsymbol{x}) + g(\boldsymbol{y}), \\ s.t. \quad & h_0(\boldsymbol{x}) \le 0, \\ & p_0(\boldsymbol{y}) \le 0, \\ & \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} = \boldsymbol{b}, \end{aligned}$$

where f, g, h_0, and p_0 are convex functions. Define

$$h(\boldsymbol{x}) = \max\{0, h_0(\boldsymbol{x})\}, \tag{1}$$

$$p(\boldsymbol{y}) = \max\{0, p_0(\boldsymbol{y})\}. \tag{2}$$

Transform to a Linear One

Then we can turn the inequality constraints into equality constraints. Thus, we consider the following problem instead:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{y}} \quad & f(\boldsymbol{x}) + g(\boldsymbol{y}), \\ s.t. \quad & h(\boldsymbol{x}) = 0, \\ & p(\boldsymbol{y}) = 0, \\ & \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} = \boldsymbol{b}. \end{aligned}$$

Further we have the augmented Lagrangian function:

$$\begin{aligned} & L_{\rho_1,\rho_2,\beta}(\boldsymbol{x}, \boldsymbol{y}, \gamma, \tau, \boldsymbol{\lambda}) \\ = & f(\boldsymbol{x}) + g(\boldsymbol{y}) + \gamma h(\boldsymbol{x}) + \frac{\rho_1}{2} h^2(\boldsymbol{x}) + \tau p(\boldsymbol{y}) + \frac{\rho_2}{2} p^2(\boldsymbol{y}) \\ & + \langle \boldsymbol{\lambda}, \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b} \rangle + \frac{\beta}{2} \|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{b}\|^2. \end{aligned} \tag{3}$$

ADMM-NC

We thus have the following iterations:

$$\boldsymbol{x}^{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} L_{\rho_1,\rho_2,\beta}(\boldsymbol{x}, \boldsymbol{y}^k, \gamma^k, \tau^k, \boldsymbol{\lambda}^k) \tag{4}$$

$$\boldsymbol{y}^{k+1} = \operatorname*{argmin}_{\boldsymbol{y}} L_{\rho_1,\rho_2,\beta}(\boldsymbol{x}^{k+1}, \boldsymbol{y}, \gamma^k, \tau^k, \boldsymbol{\lambda}^k) \tag{5}$$

$$\gamma^{k+1} = \gamma^k + \rho_1 h(\boldsymbol{x}^{k+1}) \tag{6}$$

$$\tau^{k+1} = \tau^k + \rho_2 p(\boldsymbol{y}^{k+1}) \tag{7}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \beta(\mathbf{A}\boldsymbol{x}^{k+1} + \mathbf{B}\boldsymbol{y}^{k+1} - \boldsymbol{b}). \tag{8}$$

We call (4) $\sim$ (8) ADMM for Nonlinear Constraints (*ADMM-PC*).

Convergence Rate of ADMM-PC

Suppose that f, g, h_0, and p_0 are convex functions. Then for ADMM-PC, we have

$$\begin{aligned}
&|f(\hat{\boldsymbol{x}}^{K+1}) + g(\hat{\boldsymbol{y}}^{K+1}) - f(\boldsymbol{x}^*) - g(\boldsymbol{y}^*)| \\
&\qquad \leq \frac{C}{2(K+1)} + \frac{2\sqrt{C}\|\boldsymbol{\lambda}^*\|}{\sqrt{\beta}(K+1)} \\
&\qquad + \frac{2\sqrt{C}|\gamma^*|}{\sqrt{\rho_1}(K+1)} + \frac{2\sqrt{C}|\tau^*|}{\sqrt{\rho_2}(K+1)},
\end{aligned} \tag{9}$$

$$\|\mathbf{A}\hat{\boldsymbol{x}}^{K+1} + \mathbf{B}\hat{\boldsymbol{y}}^{K+1} - \boldsymbol{b}\| \leq \frac{2\sqrt{C}}{\sqrt{\beta}(K+1)}, \tag{10}$$

$$h(\hat{\boldsymbol{x}}^{K+1}) \leq \frac{2\sqrt{C}}{\sqrt{\rho_1}(K+1)}, \tag{11}$$

$$p(\hat{\boldsymbol{y}}^{K+1}) \leq \frac{2\sqrt{C}}{\sqrt{\rho_2}(K+1)}. \tag{12}$$

Convergence Rate of ADMM-PC (Cont'd)

In (9) $\sim$ (12),

$$\hat{\boldsymbol{x}}^{K+1} = \frac{1}{K+1} \sum_{k=1}^{K+1} \boldsymbol{x}^k, \tag{13}$$

$$\hat{\boldsymbol{y}}^{K+1} = \frac{1}{K+1} \sum_{k=1}^{K+1} \boldsymbol{y}^k, \tag{14}$$

$$\begin{aligned} C = {} & \frac{1}{\beta}\|\boldsymbol{\lambda}^0 - \boldsymbol{\lambda}^*\|^2 + \frac{1}{\rho_1}(\gamma^0 - \gamma^*)^2 \\ & + \frac{1}{\rho_2}(\tau_0 - \tau^*)^2 + \beta\|\mathbf{B}\boldsymbol{y}^0 - \mathbf{B}\boldsymbol{y}^*\|^2. \end{aligned} \tag{15}$$

The proof is similar to the proff we have presented in *ADMM Slide: Part 2.*

ADMM for General Nonlinear Convex Problems

Consider the following general nonlinear convex probelm:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \sum_{i=1}^{m} f_i(\boldsymbol{x}) \\ s.t. \quad & h_i(\boldsymbol{x}) \leq 0, \quad \forall i \in [m], \\ & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned}$$

We can introduce the auxiliary variable $\boldsymbol{z} = \boldsymbol{x}_i, \forall i \in [m]$, then the probelm is transformed to

$$\begin{aligned} \min_{\{\boldsymbol{x}_i\},\boldsymbol{z}} \quad & \sum_{i=1}^{m} f_i(\boldsymbol{x}_i) \\ s.t. \quad & h_i(\boldsymbol{x}_i) \leq 0, \quad \forall i \in [m], \\ & \mathbf{A}'\boldsymbol{z} - \mathbf{I}'\boldsymbol{x}' = \boldsymbol{b}'. \end{aligned}$$

ADMM for General Nonlinear Convex Problems

In the above slide, $\mathbf{A}'\boldsymbol{z} - \mathbf{I}'\boldsymbol{x}' = \boldsymbol{b}'$ is

$$\begin{pmatrix} \mathbf{I} \\ \mathbf{I} \\ \vdots \\ \mathbf{I} \\ \mathbf{A} \end{pmatrix} \boldsymbol{z} - \begin{pmatrix} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \cdots \\ \boldsymbol{x}_m \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \boldsymbol{b} \end{pmatrix}.$$

We can use ADMM-SPU (*The ADMM Slide: Part 3*) to solve it, where we solve the first subproblem with $(\boldsymbol{x}_1^T, ..., \boldsymbol{x}_m^T)^T$, and then solve the second subproblem with $\boldsymbol{z}$. Note that the first subproblem can be decomposed into m subproblems in parallel.

Thus, this is an approach of joint serial and parallel updates.

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.