Lecture #22 ADMM EI 260

- Scaled form ADMM
- Examples
- consensus (distributed) optimization
- Convergence result.

(afternoon)

Exam : 8^{th} December 2021 9:00 am

(20 pts) 24 hr turn in time (9^{th} Dec 2021 9:00 am)

Theory + Programming

ADMM Scaled form :

- Combine linear & quadratic terms in $L_\rho(x, z, y)$:

$$L_\rho(\underline{x}, \underline{z}, \underline{y}) = f(\underline{x}) + g(\underline{z}) + \underline{y}^T(A\underline{x} + B\underline{z} - c) + \frac{\rho}{2}\|A\underline{x} + B\underline{z} - \underline{c}\|_2^2$$

$$L_\rho(\underline{x}, \underline{z}, \underline{u}) = f(\underline{x}) + g(\underline{z}) + \frac{\rho}{2}\|A\underline{x} + B\underline{z} - \underline{c} + \underline{u}\|_2^2 + \text{const.}$$

with $\underline{u}^k = \left(\frac{1}{\rho}\right)\underline{y}^k$

$$\underline{y}^T\underline{r} + \left(\rho/2\right)\|\underline{r}\|_2^2 = \rho/2\ \|\underline{r} + (1/\rho)\,\underline{y}\|_2^2 - \left(\frac{1}{2\rho}\right)\|\underline{y}\|_2^2$$

$$= \rho/2\ \|r\|^2 + \frac{1}{\rho^2}\cdot\frac{\rho}{2}\|y\|^2 + \frac{2}{\rho}\cdot\frac{\rho}{2}\ y^T\underline{x} - \frac{1}{2\rho}\|y\|^2$$

$$= (\rho/2)\ \|\underline{r} + \underline{u}\|_2^2 - (\rho/2)\ \|\underline{u}\|_2^2$$

Scaled - form ADMM:

$$\underline{x}^{k+1} = \arg\min_{x} \; f(\underline{x}) + (\rho/2) \| A\underline{x} + B\underline{z}^{k} - \underline{c} + \underline{u}^{k} \|_2^2$$

$$\underline{z}^{k+1} = \arg\min_{\underline{z}} \; g(\underline{z}) + (\rho/2) \| A\underline{x}^{k+1} + B\underline{z} - \underline{c} + \underline{u}^{k} \|_2^2$$

$$\underline{u}^{k+1} = \underline{u}^{k} + A\underline{x}^{k+1} + B z^{k+1} - c$$

Example:

$$\min_{x} \; f(x) \quad \text{s.t.} \quad \underline{x} \in C$$

$$\equiv \min_{\underline{x}, \underline{z}} \; f(\underline{x}) + \underbrace{g(\underline{z})}_{I_C} \quad \text{s.t.} \quad \underline{x} - \underline{z} = 0$$

ADMM:

$$\underline{x}^{k+1} = \arg\min_{\underline{x}} \; f(x) + \rho/2 \, \| \underline{x} - \underline{z}^{k} + \underline{u}^{k} \|_2^2$$

$$\underline{z}^{k+1} = P_C \left(\underline{x}^{k+1} + \underline{u}^{k} \right)$$

$$\underline{u}^{k+1} = \underline{u}^{k} + \underline{x}^{k+1} - \underline{z}^{k+1}$$

Proximal operator :

$\underline{x}$ update :

$$\underline{x}^+ = \underset{\underline{x}}{\arg\min} \; f(x) + (\rho/2) \, \| \underline{x} - \underline{v} \|_2^2$$

$$= \text{prox}_{f,\rho} (\underline{v})$$

Recall :

- $f = I_C$ [indicator function]

$$\underline{x}^+ = P_C (\underline{v})$$

- $f = \lambda \| \cdot \|_1$

$$\underline{x}^+ = S_{\lambda/\rho} (v_i)$$

(Soft thresholding)

LASSO:

$$\text{minimize} \quad \left(\tfrac{1}{2}\right) \|A\underline{x} - b\|_2^2 + \lambda \|\underline{x}\|_1$$

ADMM form:

$$\text{minimize} \quad \left(\tfrac{1}{2}\right) \|A\underline{x} - b\|_2^2 + \lambda \|\underline{z}\|_1$$

$$\text{Subject to} \quad \underline{x} - \underline{z} = 0$$

ADMM:

x - update:

$$\underline{x}^{k+1} = \arg\min_{x} \; \frac{1}{2} \|A\underline{x} - b\|_2^2 + \left(\tfrac{\rho}{2}\right) \|\underline{x} - \underline{z}^k + \underline{u}^k\|_2^2$$

$$A^T(A\underline{x} - b) + \rho(\underline{x} - \underline{z}^k + \underline{u}^k) = 0$$

First order methods to solve for the linear system

$$(A^TA + \rho I)\,\underline{x} = A^T\underline{b} + \rho(\underline{z}^k - \underline{u}^k)$$

$$\Rightarrow \underline{x}^{k+1} = \underbrace{[A^TA + \rho I]}_{\text{pre-cache}}{}^{-1}\left(A^T\underline{b} + \rho(\underline{z}^k - \underline{u}^k)\right)$$

$\rightarrow$ matrix inversion lemma

z - update:

$$\underline{z}^{k+1} = \arg\min_{\underline{z}} \lambda\|\underline{z}\|_1 + (\rho/2)\,\|\underline{x}^{k+1} + \underline{z} + \underline{u}^k\|_2^2$$

$$= S_{\lambda/\rho}(\underline{x}^{k+1} + \underline{u}^k)$$

u - update:

$$\underline{u}^{k+1} = \underline{u}^k + \underline{x}^{k+1} - \underline{z}^{k+1}.$$

Graphical LASSO:

Precision matrix } Partial correlation

$$\underset{\Theta \succ 0}{\text{minimize}} \quad -\log\det(\Theta) + \mathrm{tr}(S\Theta) + \lambda \|\Theta\|_1$$

S: Empirical covariance matrix $= \frac{1}{N} XX^T = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T$

$X = [x_1, \dots x_N]; \; x_i \sim \mathcal{N}(0, \Theta^{-1})$

ADMM form:

$$\text{minimize} \quad \mathrm{Tr}(S\Theta) - \log\det(\Theta) + \mathbb{I}_{S_f}(\Theta) + \lambda \|Z\|_1$$

s.t. $\quad X - Z = 0$

$$S_f := \{ X : X \succ 0 \}$$

Θ - update:

$$\Theta^{k+1} = \underset{\Theta}{\arg\min} \left[\mathrm{Tr}(S\Theta) - \log\det(\Theta) + \left(\frac{\rho}{2}\right) \|\Theta - Z^k + U^k\|_F^2 \right]$$

First-order optimality condition:

$$S - \Theta^{-1} + \rho\left(\Theta - Z^k + U^k\right) = 0$$

$$\rho\Theta - \Theta^{-1} = \rho\left(Z^k - U^k\right) - S$$

Suppose: $\left.\begin{array}{l}\rho(Z - U) - S = Q \Lambda Q^T \\ \rho\tilde{\Theta} - \tilde{\Theta}^{-1} = \Lambda\end{array}\right\}$ $Q^T Q = I$

$$Q^T \Theta Q = \tilde{\Theta}$$

Find positive numbers $\tilde{\Theta}_{ii}$ that satisfy

$$\rho \tilde{\Theta}_{ii} - \tilde{\Theta}_{ii}^{-1} = \lambda_i$$

$$\Rightarrow \quad \tilde{\Theta}_{ii} = \frac{\lambda_i + \sqrt{\lambda_i^2 + 4\rho}}{2\rho}$$

Z-update:

$$Z^{k+1} = S_{\lambda/\rho}\left(\Theta^{k+1} + U^k\right)$$

U-update:

$$U^{k+1} = U^k + \left(X^{k+1} - Z^{k+1}\right)$$

Consensus optimization:

$$\underset{\underline{x}}{\text{minimize}} \quad \sum_{i=1}^{N} f_i(\underline{x})$$

Suppose we have 'N' clients

$$\text{minimize} \quad \sum_{i=1}^{N} f_i(\underline{x}_i)$$

$$\text{s.t.} \quad \underline{x}_i = \underline{z} \qquad i = 1, \dots, N$$

Constraints:

$$\underbrace{\begin{bmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_N \end{bmatrix}}_{} = \underbrace{\begin{bmatrix} I \\ \vdots \\ I \end{bmatrix}}_{} \underline{z}$$

$$I\,\underline{x} - B\,\underline{z} = 0$$

ADMM update:

$$\underline{x}^{k+1} = \underset{\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}}{\arg\min} \left\{ \sum_{i=1}^{N} f_i(\underline{x}_i) + \rho/2 \sum_{i=1}^{N} \| \underline{x}_i - \underline{z}^k + \underline{u}_i^k \|_2^2 \right\}$$

"Computed in parallel"

$$\underline{z}^{k+1} = \underset{\underline{z}}{\arg\min} \left\{ \frac{\rho}{2} \sum_{i=1}^{N} \| \underline{x}_i^{k+1} - \underline{z} + \underline{u}_i^k \|_2^2 \right\}$$

(gather all local updates)

$$\underline{u}_i^{k+1} = \underline{u}_i^k + \underline{x}_i^{k+1} - \underline{z}^{k+1} \qquad i = 1, \ldots N$$

(broadcast to update local multipliers)

Splitting across Examples:

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix} \qquad \underline{b} = \begin{bmatrix} \underline{b}_1 \\ \vdots \\ \underline{b}_N \end{bmatrix} \in \mathbb{R}^m ; \qquad m = \sum_{i=1}^{N} m_i$$

i^{th} client has data (A_i, b_i)

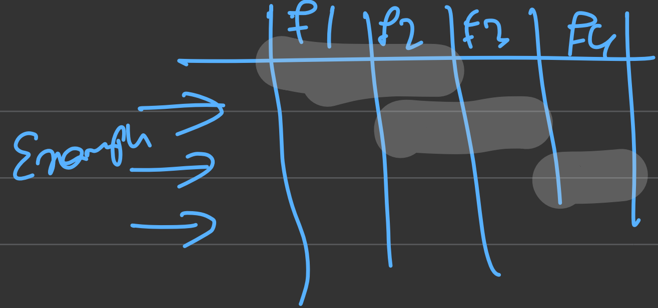
$$\text{minimize} \quad \sum_{i=1}^{N} l_i(A_i \underline{x}_i - b_i) + r(\underline{z})$$

$$\text{Subject to} \quad \underline{x}_i - \underline{z} = 0 \quad ; i = 1, \ldots, N$$

LASSO: $l_i(A_i \underline{x} - b_i) = \frac{1}{2} \| A_i x_i - b_i \|_2^2$

$$r(\underline{z}) = \lambda \| \underline{z} \|_1$$

Splitting across features:



partion the parameter vector

$$\underline{x} = [\underline{x}_1, \dots x_N] \; ; \; \underline{x}_i \in \mathbb{R}^{n_i}$$

$$\sum_{i=1}^{N} n_i = N$$

$$A : [A_1, \dots A_N] \qquad A_i \in \mathbb{R}^{m \times n_i}$$

$$r(\underline{x}) = \sum_{i=1}^{N} r_i(\underline{x}_i)$$

$$\Rightarrow \quad A\underline{x} = \sum_{i=1}^{N} A_i \underline{x}_i$$

partial prediction

$$\underset{\underline{x}}{\text{minimize}} \quad l\left(\sum_{i=1}^{N} A_i x_i - \underline{b}\right) + \sum_{i=1}^{N} r_i(\underline{x_i})$$

ADMM form:

$$\text{minimize} \quad l\left(\sum_{i=1}^{N} z_i - b\right) + \sum_{i=1}^{N} r_i(\underline{x_i})$$

$$\text{subject to} \quad A_i \underline{x_i} - \underline{z_i} = 0 \qquad i = 1, \ldots, N$$

Scaled-form ADMM:

$$\underline{x_i}^{k+1} = \underset{\underline{x_i}}{\arg\min} \left(r_i(x_i) + \rho/2 \, \| A_i \underline{x_i} - \underline{z_i}^k + \underline{u_i}^k \|_2^2 \right)$$

$$\underline{z}^k = \underset{z}{\arg\min} \left(l\left(\sum_{i=1}^{N} z_i - b\right) + \sum_{i=1}^{N} (\rho/2) \, \| A_i \underline{x_i}^{k+1} - \underline{z_i}^k + \underline{u_i}^k \|_2^2 \right)$$

$$\underline{u_i}^{k+1} = \underline{u_i}^k + A_i \underline{x_i}^{k+1} - \underline{z_i}^{k+1}$$

Convergence rate : $O(1/k)$

iteration complexity : $O(1/\varepsilon)$

Convergence of ADMM:

Suppose f and g are closed convex functions; and r is any constant so that $r \geq 2\|\underline{y}^*\|_2$.

Then,

$$F(\underline{x}^k, \underline{z}^k) - F^{opt} \leq \frac{\|\underline{z}^0 - z^*\|^2_{\rho B^T B} + \frac{(r + \|\underline{y}^0\|_2)^2}{\rho}}{2(k+1)}$$

$$\|A\underline{x}^{(k)} + B\underline{z}^{(k)} - \underline{c}\|_2 \leq \frac{\|\underline{z}^0 - z^*\|^2_{\rho B^T B} + (r + \|\underline{y}^0\|_2)^2}{r(k+1)}$$

$$\underline{x}^{(k)} = \frac{1}{k+1}\sum_{t=1}^{k+1} \underline{x}^t, \quad \underline{z}^{(k)} = \frac{1}{k+1}\sum_{t=1}^{k+1} z^t; \quad \|z\|_C^2 = \underline{z}^T C \underline{z}$$