Alternating direction method of multipliers

Liang Zhang

Mar. 21\textsuperscript{th}, 2017


Parts of slides adapted from the 14th lecture of V. Kekatos, Optimization Theory, 2015
Problem of interest

- Constrained (convex) optimization problem

\[
\min_{x,z} \ f(x) + g(z) \\
\text{s.t.} \ Ax + Bz = c
\]

- Augmented Lagrangian

\[
L_\rho(x, z; y) := f(x) + g(z) + y^T(Ax + Bz - c) + \frac{\rho}{2} \|Ax + Bz - c\|_2^2
\]

- ADMM steps

\[
x^{t+1} := \arg\min_x L_\rho(x, z^t; y^t) \\
z^{t+1} := \arg\min_z L_\rho(x^{t+1}, z; y^t) \\
y^{t+1} := y^t + \rho(Ax^{t+1} + Bz^{t+1} - c)
\]
Variants

- Accelerated ADMM
  [Tom Goldstein, Brendan O'Donoghue, Simon Setzer, Richard Baraniuk' 14 ]

- Online ADMM and Bregman divergence
  [Huahua Wang, Arindam Banerjee’12 , ’14]

- Weighted ADMM
  [Qing Ling, Yaohua Liu, Wei Shi, Zhi Tian’16]

- Multi-block ADMM
  - On the expected convergence of randomly permuted ADMM
    [Ruoyu Sun, Zhi-Quan Luo, Yinyu Ye’ 15 ]
  - On the linear convergence of ADMM
    [Mingyi Hong, Zhi-Quan Luo’ 12 ]
Convergence

- **AS1.** The (extended-real-valued) functions $f$ and $g$ are closed, proper and convex
- **AS2.** The unaugmented Lagrangian function has a saddle point

As $t \to \infty$,

1. residual convergence: $\mathbf{A}x^t + \mathbf{B}z^t - \mathbf{c} \to 0$
2. objective convergence: $f(x^t) + g(z^t) \to \mathbf{p}^*$
3. dual variable convergence: $\mathbf{y} \to \mathbf{y}^*$

- Primal variable converges under additional assumptions
- Inexact updates
Application: \( l-1 \) regularization

- ADMM
  - ["scaled form"]
  - \( u := y / \rho \)
  - \( x^{t+1} := \arg\min_x f(x) + \frac{\rho}{2} \|Ax + Bz^t - c + u^t\|_2^2 \)
  - \( z^{t+1} := \arg\min_z g(z) + \frac{\rho}{2} \|Ax^{t+1} + Bz - c + u^t\|_2^2 \)
  - \( u^{t+1} := u^t + Ax^{t+1} + Bz^{t+1} - c \)

- For a general loss function
  - \( \min_{x,z} \{ l(x) + \lambda \|z\|_1 : x - z = 0 \} \)

- ADMM steps
  - Soft thresholding
    - \( S_\alpha(x) := x \cdot \left[1 - \frac{\alpha}{|x|}\right]_+ \)
  - \( x^{t+1} := \arg\min_x l(x) + \frac{\rho}{2} \|x - z^t + u^t\|_2^2 \)
  - \( z^{t+1} := S_{\lambda/\rho}(x^{t+1} + u^t) \)
  - \( u^{t+1} := u^t + x^{t+1} - z^{t+1} \)

- Lasso
  - \( l(x) = \frac{1}{2} \|b - Ax\|_2^2 \)

  \( x \) updates in closed-form (QP)
Consensus setup

\[
\min_x \sum_{i=1}^{N} f_i(x) + r(x) \quad \rightarrow \quad \min_{\{x_i\}, z} \left\{ \sum_{i=1}^{N} f_i(x_i) + r(z) : x_i - z = 0 \ \forall i \right\}
\]

- Local optimization

\[
x_{i+1} := \arg \min_{x_i} f_i(x_i) + \frac{\rho}{2} \|x_i - z^t + u_i^t\|_2^2
\]

- Consensus

\[
z^{t+1} := \arg \min_{z} r(z) + \frac{\rho}{2} \sum_{i=1}^{N} \|x_{i+1}^t - z + u_i^t\|_2^2
\]

\[
u_{i+1}^t := u_i^t + x_{i+1}^t - z^{t+1}
\]

- For \( r(z) = 0 \), updates further simplified

- Hence, \( \bar{u}^{t+1} = 0 \Rightarrow z^{t+2} = \bar{x}^{t+1} \)

- No need to keep \( z \) variable
Lasso: consensus setup

- Each agent $i$ manages its own training examples (rows)

$$\min_{x} \frac{1}{2} \| Ax - b \|_2^2 + \lambda \| x \|_1 \quad \Rightarrow \quad \min_{\{x_i\}, z} \sum_{i} \| A_i x_i - b_i \|_2^2 + \lambda \| z \|_1$$

subject to $x_i - z = 0$

- Local opt.

\[ x_{i}^{t+1} := \arg\min_{x_i} \frac{1}{2} \| A_i x_i - b_i \|_2^2 + \frac{\rho}{2} \| x_i - z^t + u_i^t \|_2^2 \]

- Consensus

\[ z^{t+1} := S_{\lambda/\rho N}(\overline{x}^{t+1} + \bar{u}^t) \]

\[ u_{i}^{t+1} := u_i^t + x_{i}^{t+1} - z^{t+1} \]

- For logistic and SVM regression

- Change the loss function and $x$ variable updates
Sharing setup

\[
\min_{\{x_i\}} \sum_{i=1}^{N} f_i(x_i) + g\left(\sum_{i=1}^{N} x_i\right) \rightarrow \min_{\{x_i, z_i\}} \sum_{i=1}^{N} f_i(x_i) + g\left(\sum_{i=1}^{N} z_i\right)
\]

s.t. \(x_i - z_i = 0 \ \forall i\)

- \(x\) and \(u\) updates are decoupled across \(i\)
- Oftentimes, \(z\) update can be solved efficiently
Trick for sharing setup

- Introduce $\bar{z} = \sum_{i=1}^{N} \frac{1}{N} z_i$ in $z$ updates

$$\{z_i^{t+1}, \bar{z}^{t+1}\} := \arg\min_{\{z_i, \bar{z}\}} \left\{ g\left(N\bar{z}\right) + \frac{\rho}{2} \sum_{i=1}^{N} \|z_i - u_i^t - x_i^{t+1}\|_2^2 : \bar{z} = \sum_{i=1}^{N} \frac{1}{N} z_i \right\}$$

- Let $a_i^t = u_i^t + x_i^{t+1}$, and solve w.r.t. $\{z_i\}$, $z_i^{t+1} = a_i^t + \bar{z} - \bar{a}^t$

- Then solve w.r.t. $\bar{z}$ via $\bar{z}^{t+1} = \arg\min_{\bar{z}} g\left(N\bar{z}\right) + \frac{\rho}{2} \sum_{i=1}^{N} \|\bar{z} - \bar{a}\|_2^2$

- In $u$ updates $u_i^{t+1} = u_i^t + x_i^{t+1} - z_i^{t+1} = \bar{a}^t - \bar{z}^{t+1}$ Same across $i$

- ADMM steps

$$x_i^{t+1} := \arg\min_{x_i} f_i(x_i) + \frac{\rho}{2} \|x_i - x_i^t + \bar{x}^t - \bar{z}^t + u^t\|_2^2$$

$$z^{t+1} := \arg\min_{\{z\}} g\left(N\bar{z}\right) + \frac{N\rho}{2} \sum_{i=1}^{N} \|\bar{z} - u^t - x_i^{t+1}\|_2^2$$

$$u^{t+1} := u^t + \bar{x}^{t+1} - \bar{z}^{t+1}$$

{\{z_i\} updates inexplicitly}
Learning in the sharing setup

- Each agent $i$ is assigned a subset of features (columns) $\mathbf{g}$

$$\min_{\{\mathbf{x}_i\}} \left( l \left( \sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i - \mathbf{b} \right) + \sum_{i=1}^{N} r_i(\mathbf{x}_i) \right) \rightarrow \min_{\{\mathbf{x}_i, \mathbf{z}_i\}} \left( l \left( \sum_{i=1}^{N} \mathbf{z}_i - \mathbf{b} \right) + \sum_{i=1}^{N} r_i(\mathbf{x}_i) \right)$$

s.t. $\mathbf{A}_i \mathbf{x}_i - \mathbf{z}_i = 0 \ \forall i$

- Note that all dual variables are equal

- Use $\bar{\mathbf{A}} \mathbf{x}^t := \sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i^t / N$ instead of mean $\mathbf{x}$ variable

\[
\begin{align*}
\mathbf{x}_i^{t+1} &:= \arg\min_{\mathbf{x}_i} r_i(\mathbf{x}_i) + \frac{\rho}{2} \| \mathbf{A}_i \mathbf{x}_i - \mathbf{A}_i \mathbf{x}_i^t + \bar{\mathbf{A}} \mathbf{x}^t - \bar{\mathbf{z}}^t + \mathbf{u}^t \|_2^2 \\
\bar{\mathbf{z}}^{t+1} &:= \arg\min_{\bar{\mathbf{z}}} l \left( N \bar{\mathbf{z}} \right) + \frac{N \rho}{2} \| \bar{\mathbf{z}} - \bar{\mathbf{A}} \mathbf{x}^{t+1} - \mathbf{u}^t \|_2^2 \\
\mathbf{u}^{t+1} &:= \mathbf{u}^t + \bar{\mathbf{A}} \mathbf{x}^{t+1} - \bar{\mathbf{z}}^{t+1}
\end{align*}
\]
Lasso: sharing setup

\[
\begin{align*}
\min_{\{x_i, z_i\}} & \quad \| \sum_{i=1}^{N} z_i - b \|_2^2 + \sum_{i=1}^{N} \| x_i \|_1 \\
\text{s.to} & \quad A_i x_i - z_i = 0 \quad \forall i
\end{align*}
\]

\[
\begin{align*}
x_{i}^{t+1} & := \arg \min_{x_i} \frac{\rho}{2} \| A_i x_i - A_i x_i^t + \bar{A}x^t - \bar{z}^t + u^t \|_2^2 + \lambda \| x_i \|_1 \\
\bar{z}^{t+1} & := \frac{1}{N + \rho} \left( b + \rho \bar{A} x^{t+1} + \rho u^t \right) \\
u^{t+1} & := u^t + \bar{A}x^{t+1} - \bar{z}^{t+1}
\end{align*}
\]

- \( x \)-updates are local Lasso problems of smaller dimensions
Network optimization

- Motivation
  - Lack of central coordinator (data fusion)
  - Agents communicate with single-hop neighbors

\[
\begin{align*}
\min_{\mathbf{x}} & \quad \sum_{i=1}^{N} f_i(\mathbf{x}) \\
\min_{\{x_i\},\{z_{ij}\}} & \quad \sum_{i=1}^{N} f_i(x_i) \\
\text{s.to} & \quad x_i = z_{ij} \quad \forall \ (i,j) \in \mathcal{E}
\end{align*}
\]

\[
\begin{align*}
\min_{\{x_i\}} & \quad \sum_{i=1}^{N} f_i(x_i) \\
\text{s.to} & \quad x_i = x_j \quad \forall \ (i,j) \in \mathcal{E}
\end{align*}
\]

\[N \times \text{variables} \quad |\mathcal{E}| \times \text{z variables}\]

Iterates

\[
\min_{\{x_i\},\{z_{ij}\}} \sum_{i=1}^{N} f_i(x_i)
\]

s.t. \(x_i - z_{ij} = 0\) \(\forall (i, j) \in \mathcal{E}\)

\[
\begin{align*}
\{x_{i}^{t+1}\} &:= \arg \min_{x_i} f_i(x_i) + \frac{\rho}{2} \sum_{j \in \mathcal{N}_i} \|x_i - z_{ij}^t + u_{ij}^t\|_2^2 \\
\{z_{ij}^{t+1}\} &:= \arg \min_{z_{ij}} \|x_{i}^{t+1} - z_{ij} + u_{ij}^t\|_2^2 + \|x_{j}^{t+1} - z_{ij} + u_{ji}^t\|_2^2 \\
u_{ij}^{t+1} &:= u_{ij}^t + x_{i}^{t+1} - z_{ij}^{t+1}
\end{align*}
\]

- Note that
- Variable \(z\) can be updated inexplicitly

\[
\begin{align*}
u_{ij}^{t+1} + u_{ji}^{t+1} &= 0 \\
z_{ij}^{t+1} &= \frac{x_{i}^{t+1} + x_{j}^{t+1}}{2}
\end{align*}
\]
Thank you!