Large-Scale Optimization

L08. ADMM 1

TU Dortmund, Dr. Sangkyun Lee
Today

ADMM: Alternating Direction Method of Multipliers


For introduction: [Boyd et al., FnT ML, 2010]

Aim:

• Understanding the algorithm
• Convergence (no proof)
• Decomposition

TU Dortmund, Dr. Sangkyun Lee
Dual Ascent

\[
\min_{x \in \mathbb{R}^n} f(x) \quad A \in \mathbb{R}^{m \times n}, \ f : \mathbb{R}^n \to \mathbb{R} \text{ convex}
\]

\[s.t. \ Ax = b\]

Lagrangian \[\mathcal{L}(x; y) = f(x) + y^T(Ax - b) \quad y \in \mathbb{R}^m\]

Dual objective function: \[g(y) = \inf_x \mathcal{L}(x; y)\]

Let \(x^*\) minimizes \(\mathcal{L}(x; y)\) for a given \(y\). Then

\[g(y) = \mathcal{L}(x^*; y) = f(x^*) + y^T(Ax^* - b)\]

\[\therefore \ Ax^* - b \in \partial g(y)\]
Dual Problem

\[
\max_y g(y) \quad \quad \quad \quad \quad Ax^* - b \in \partial g(y), \\
x^* \in \arg\min_x \mathcal{L}(x; y)
\]

Using subgradient ascent:

\[
x^{k+1} \in \arg\min_x \mathcal{L}(x; y^k) \\
y^{k+1} = y^k + \alpha^k (Ax^{k+1} - b)
\]

\[\alpha^k > 0\] is a stepsize
Dual Subgradient Ascent

\[ x^{k+1} \in \text{arg min}_x \ L(x; y^k) \]
\[ y^{k+1} = y^k + \alpha^k (A x^{k+1} - b) \]

When alpha^k is chosen carefully, and with additional assumptions, this procedure can produce
\[ x^k \to x^*, \ y^k \to y^* \]

However, this requires conditions often do not hold in practice.
Dual Decomposition

Suppose that \( f \) is block separable, i.e.,

\[
f(x) = \sum_{i=1}^{N} f_i(x_i), \quad x = (x_1, \ldots, x_N), \quad x_i \in \mathbb{R}^{n_i}
\]

\[
Ax = \sum_{i=1}^{N} A_i x_i, \quad A_i \text{ is the } i\text{th block column submatrix of } A
\]

\[
\mathcal{L}(x; y) = \sum_{i=1}^{N} \mathcal{L}_i(x_i; y) = \sum_{i=1}^{N} (f_i(x_i) + y^T A_i x_i - (1/N)y^T b)
\]

That is, the Lagrangian is also separable in \( x \)
Dual Decomposition

That is, when \( f \) is separable, \( x \)-minimization can be decomposed:

\[
x_{i}^{k+1} \in \arg \min_{x} \mathcal{L}_i(x; y^k), \quad i = 1, 2, \ldots, N
\]

\[
y^{k+1} = y^k + \alpha^k (Ax^{k+1} - b)
\]

Message-Passing Algorithm
- Communication cost?
- Data storage?
Augmented Lagrangian

Lagrangian: \( \mathcal{L}(x; y) = f(x) + y^T(Ax - b) \)

Augmented Lagrangian: \( \rho > 0 \) : penalty parameter

\[
\mathcal{L}_\rho(x; y) = f(x) + y^T(Ax - b) + \frac{\rho}{2} \|Ax - b\|^2
\]

This is the Lagrangian function associated with an equivalent problem:

\[
\min_x f(x) + \frac{\rho}{2} \|Ax - b\|^2
\]

\[s.t. \ Ax = b\]
**MM: Method of Multipliers**

**Dual ascent with augmented Lagrangian:**

\[ x^{k+1} \in \arg \min_x \mathcal{L}_\rho(x; y^k) \]

\[ y^{k+1} = y^k + \rho(Ax^{k+1} - b) \]

- Converges in more general conditions than dual ascent
- But it loses decomposability!
Convergence of MM

\[ x^{k+1} \in \arg \min_x \mathcal{L}_{\rho_k}(x; y^k) \]

\[ y^{k+1} = y^k + \rho_k(Ax^{k+1} - b) \]

\[ \rho_{k+1} \geq \rho_k \]

If \[ z^T \nabla^2 f(x^*) z > 0, \quad \forall z \neq 0, \quad Az = 0 \]

\[ \rho_k \geq \bar{\rho} \] for some sufficiently large \( \bar{\rho} \)

Then \[ y^k \rightarrow y^* \] linearly when \( \{\rho_k\} \) is bounded


superlinearly when \( \rho_k \rightarrow \infty \)

[Bertsekas 1996]
ADMM

\[
\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} f(x) + g(z)
\]

\[s.t. \ Ax + Bz = c\]

Augmented Lagrangian:

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

“Alternating direction” method of multipliers

\[
x^{k+1} \in \text{arg min}_x \mathcal{L}_\rho(x, z^k; y^k)
\]

\[
z^{k+1} \in \text{arg min}_z \mathcal{L}_\rho(x^{k+1}, z; y^k)
\]

\[
y^{k+1} = y^k + \rho(Ax^{k+1} + Bz^{k+1} - b)
\]
MM vs. ADMM

\[
(x^{k+1}, z^{k+1}) \in \arg \min_x \mathcal{L}_\rho(x, z; y^k)
\]

\[
y^{k+1} = y^k + \rho(Ax^{k+1} + Bz^{k+1} - b)
\]

MM

\[
x^{k+1} \in \arg \min_x \mathcal{L}_\rho(x, z^k; y^k)
\]

\[
z^{k+1} \in \arg \min_z \mathcal{L}_\rho(x^{k+1}, z; y^k)
\]

\[
y^{k+1} = y^k + \rho(Ax^{k+1} + Bz^{k+1} - b)
\]

ADMM
Convergence of ADMM

Assumptions:

1. \( f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}, \ g : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\} \) are closed, proper, and convex

\[ \iff \quad \text{epi } f = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq t \} \]

\[ \text{epi } g = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : g(x) \leq t \} \]

are closed nonempty convex sets

This implies that x-update and z-update are solvable, i.e., minimizers exist (but may not be unique)
Assumption 2:

$L_0$ has a saddle point, i.e. there exists $(x^*, z^*, y^*)$ s.t.

\[ L_0(x^*, z^*, y) \leq L_0(x^*, z^*, y^*) \leq L_0(x, z, y^*), \quad \forall x, z, y \]

With assumption 1, this implies that

$(x^*, z^*)$ is a primal solution of min \( f(x) + g(z) \) s.t. \( Ax + Bz = c \)

\( y^* \) is a dual solution

There is no duality gap
Convergence of ADMM

Under assumptions 1 & 2,

Residual \[ r^k := Ax^k + Bz^k - c \to 0 \text{ as } k \to \infty \]

Objective \[ f(x^k) + g(z^k) \to f^* + g^* \text{ as } k \to \infty \]

Dual variable \[ y^k \to y^* \text{ as } k \to \infty \]

Primal variables need not converge to optimal values, although such results can be shown under additional assumptions.
Convergence of ADMM

\[
\begin{align*}
\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^m} & \quad f(x) + g(z) \\
\text{s.t.} & \quad Ax + Bz = c \\
& \quad x \in X, \ z \in Z
\end{align*}
\]

For primal convergence, we need additionally:
- X and Z are polyhedral sets
- A and B have full column-rank

\[
\{(x^k, z^k)\} \rightarrow \text{a single limit point } (x^*, z^*), \text{ which solves the problem}
\]

\[
\{y^k\} \text{ has a unique limit point } y^*, \text{ which solves the dual problem}
\]

[Mota et al., arxiv:1112.2295, 2011]
Convergence of ADMM

General convex case

• Sublinear convergence $O(1/k)$
• [He & Yuan, SIAM J Numerical Analysis, 2012]

Strongly convex case

• Linear convergence
• [Deng & Yin, Rice Univ. Tech rep, TR12-14, 2012]
Global Variable Consensus

$$\min_{x \in \mathbb{R}^n} f(x) = \sum_{i=1}^{N} f_i(x) \quad f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\} \text{ convex}$$

A global variable $x$ is shared across $f_i$’s

A simple reformulation (global consensus problem):

$$\min_{x_i \in \mathbb{R}^n, z \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(x_i)$$

s.t. $x_i - z = 0, \quad i = 1, 2, \ldots, N.$
ADMM for Global Consensus

Augmented Lagrangian:

\[ L_\rho(x_1, \ldots, x_N, z; y) = \sum_{i=1}^{N} \left( f_i(x_i) + y_i^T (x_i - z) + \frac{\rho}{2} \|x_i - z\|_2^2 \right) \]

ADMM:

\[ x_i^{k+1} = \underset{x_i}{\operatorname{arg\,min}} \left( f_i(x_i) + (y_i^k)^T (x_i - z^k) + \frac{\rho}{2} \|x_i - z^k\|_2^2 \right) \]

\[ z^{k+1} = \frac{1}{N} \sum_{i=1}^{N} \left( x_i^{k+1} + \frac{1}{\rho} y_i^k \right) \]

\[ y_i^{k+1} = y_i^k + \rho(x_i^{k+1} - z^{k+1}) \]
Simplification

ADMM:

\[ x_{i}^{k+1} = \arg \min_{x_{i}} \left( f_{i}(x_{i}) + (y_{i}^{k})^{T}(x_{i} - z^{k}) + \frac{\rho}{2}\|x_{i} - z^{k}\|^{2}_{2} \right) \]

\[ z^{k+1} = \frac{1}{N} \sum_{i=1}^{N} \left( x_{i}^{k+1} + \frac{1}{\rho}y_{i}^{k} \right) \Rightarrow z^{k+1} = \bar{x}^{k+1} + \frac{1}{\rho} \bar{y}^{k} \]

\[ y_{i}^{k+1} = y_{i}^{k} + \rho(x_{i}^{k+1} - z^{k+1}) \]

\[ \Rightarrow \text{by averaging, } \bar{y}^{k+1} = \bar{y}^{k} + \rho(\bar{x}^{k+1} - z^{k+1}) \]

\[ \therefore \bar{y}^{k+1} = 0 \]

\[ \therefore z^{k+1} = \bar{x}^{k+1} \]
**Simplification**

ADMM:

\[
x_i^{k+1} = \arg \min_{x_i} \left( f_i(x_i) + (y_i^k)^T(x_i - z^k) + \frac{\rho}{2} \|x_i - z^k\|_2^2 \right)
\]

\[
z^{k+1} = \frac{1}{N} \sum_{i=1}^{N} \left( x_i^{k+1} + \frac{1}{\rho} y_i^k \right) \quad z^{k+1} = \bar{X}^{k+1}
\]

\[
y_i^{k+1} = y_i^k + \rho(x_i^{k+1} - z^{k+1})
\]

Simplified ADMM:

\[
x_i^{k+1} = \arg \min_{x_i} \left( f_i(x_i) + (y_i^k)^T(x_i - \bar{X}^k) + \frac{\rho}{2} \|x_i - \bar{X}^k\|_2^2 \right)
\]

\[
y_i^{k+1} = y_i^k + \rho(x_i^{k+1} - \bar{X}^{k+1})
\]
ADMM for Global Consensus

Simplified ADMM:

\[
x_i^{k+1} = \arg \min_{x_i} \left( f_i(x_i) + (y_i^k)^T (x_i - \bar{x}^k) + \frac{\rho}{2} \| x_i - \bar{x}^k \|_2^2 \right)
\]

\[
y_i^{k+1} = y_i^k + \rho (x_i^{k+1} - \bar{x}^{k+1})
\]

Each function access local data.

TU Dortmund, Dr. Sangkyun Lee
ADMM : Consensus + Regularization

Augmented Lagrangian:

\[
L_\rho(x_1, \ldots, x_N, z; y) = \sum_{i=1}^{N} \left( f_i(x_i) + g(z) + y_i^T (x_i - z) + \frac{\rho}{2} \| x_i - z \|_2^2 \right)
\]

\[
x_i^{k+1} = \arg \min_{x_i} \left( f_i(x_i) + (y_i^k)^T (x_i - z^k) + \frac{\rho}{2} \| x_i - z^k \|_2^2 \right)
\]

\[
z^{k+1} = \arg \min_{z} \left( g(z) + \sum_{i=1}^{N} \left( -(y_i^k)^T z + \frac{\rho}{2} \| x_i^{k+1} - z \|_2^2 \right) \right)
\]

\[
y_i^{k+1} = y_i^k + \rho (x_i^{k+1} - z^{k+1})
\]
\[ x_i^{k+1} = \arg \min_{x_i} \left( f_i(x_i) + (y_i^k)^T (x_i - z^k) + \frac{\rho}{2} \| x_i - z^k \|^2_2 \right) \]

\[ z^{k+1} = \arg \min_z \left( g(z) + \sum_{i=1}^{N} (-(y_i^k)^T z + \frac{\rho}{2} \| x_i^{k+1} - z \|^2_2) \right) \]

\[ y_i^{k+1} = y_i^k + \rho(x_i^{k+1} - z^{k+1}) \]

\[ z^{k+1} = \arg \min_z \left( g(z) + \frac{N\rho}{2} \| z - \bar{x}^{k+1} - \frac{1}{\rho} \bar{y}^k \|^2_2 \right) \]

\[ = \text{prox}_{\frac{1}{N\rho} g} \left( \bar{x}^{k+1} + \frac{1}{\rho} \bar{y}^k \right) \]

Further simplification is not possible unless \( g(z) = \text{const.} \)
Ex.

\[ z^{k+1} = \text{prox}_{\frac{1}{N\rho}g} \left( \tilde{x}^{k+1} + \frac{1}{\rho} \tilde{y}^k \right) \]

\[ g(z) = \lambda \|z\|_1 \quad \quad \quad \quad \quad \quad z^{k+1} = S_{\frac{\lambda}{N\rho}} \left( \tilde{x}^{k+1} + \frac{1}{\rho} \tilde{y}^k \right) \]

\[ g(z) = \begin{cases} 
0 & z \geq 0 \\
\infty & \text{o.w.} 
\end{cases} \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad z^{k+1} = \left( \tilde{x}^{k+1} + \frac{1}{\rho} \tilde{y}^k \right)_+ \]