Large-Scale Optimization
Today

ADMM: Alternating Direction Method of Multipliers
[Boyd et al., FnT ML, 2010]

Aim:

• Fast convergence
• Parallelization (decomposability)
Dual Ascent

\[
\min_{x \in \mathbb{R}^n} \ f(x) \\
\text{s.t.} \ Ax = b
\]

\[
A \in \mathbb{R}^{m \times n}, \ f : \mathbb{R}^n \to \mathbb{R} \text{ convex}
\]

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

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

Use subgradient ascent:

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

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

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

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

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

When \( \alpha^k \) is chosen carefully, and with additional assumptions, this procedure will produce

\[ x^k \to x^*, \ y^k \to y^* \]

However, the requires conditions often do not hold in practice.
Suppose that $f$ is block separable, i.e.,

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

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

$$\mathcal{L}(x; y) = \sum_{i=1}^{N} \mathcal{L}_i(x_i; y) = \sum_{i=1}^{N} \left( f_i(x_i) + y^T A_i x_i - \frac{1}{N} y^T b \right)$$

This is also separable in $x$.
That is, when \( f \) is separable, \( x \)-minimization can be parallelized:

\[
\begin{align*}
x_i^{k+1} & \in \arg \min_x \mathcal{L}_i(x; y^k), \quad i = 1, 2, \ldots, N \\
y^{k+1} & = y^k + \alpha^k (Ax^{k+1} - b)
\end{align*}
\]

- 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. \( Ax = b \)
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 Method of Multipliers

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

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

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

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

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

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

\[ \text{superlinearly when } \rho_k \rightarrow \infty \]

[Bertsekas 1996]
ADMM

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

\[
A \in \mathbb{R}^{p \times n}, \quad B \in \mathbb{R}^{p \times m}
\]

s.t. \quad Ax + Bz = c

Augmented Lagrangian:

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

“Alternating direction” method of multipliers

\[
\begin{align*}
\text{Coordinate-Wise Minimization} \\
x^{k+1} &\in \arg \min_x \mathcal{L}_\rho(x, z^k; y^k) \\
z^{k+1} &\in \arg \min_z \mathcal{L}_\rho(x^{k+1}, z; y^k) \\
y^{k+1} &= y^k + \rho (Ax^{k+1} + Bz^{k+1} - b)
\end{align*}
\]
MM vs. ADMM

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

**MM**

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

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

**ADMM**

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

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

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

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.

$L_0(x^*, z^*, y) \leq L_0(x^*, z^*, y^*) \leq L_0(x, z, y^*), \ \forall x, z, y$

With assumption 1, this implies that

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

$y^*$ is a dual solution

There is no duality gap
Convergence of ADMM

Under assumptions 1 & 2,

Residual \( r^k := Ax^k + Bz^k - c \rightarrow 0 \text{ as } k \rightarrow \infty \)

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

Dual variable \( y^k \rightarrow y^* \text{ as } k \rightarrow \infty \)

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

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

A global variable \( x \) is shared across \( f_i \)'s

A simple reformulation (global consensus problem):

\[
\min_{x_i \in \mathbb{R}^n, z \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(x_i) \\
\text{s.t. } x_i - z = 0, \quad i = 1, 2, \ldots, N.
\]
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:

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

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

$$y_i^{k+1} = y_i^k + \rho (x_i^{k+1} - z^{k+1})$$
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:

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

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

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

Simplified ADMM:

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

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

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

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

Each function access local data