Numerical Optimization

CHAPTER 13. ADMM

TU Dortmund, Dr. Sangkyun Lee

Today

ADMM: Alternating Direction Method of Multipliers

[Glowinski, R. & Marroco, A., 1975]

[D. Gabay & B. Mercier, 1976]

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

Aim:

- Understanding the algorithm
- Convergence
- Consensus formulation

Dual Ascent

$$\min_{x \in \mathbb{R}^n} \; f(x)$$
$$s.t. \;\; Ax = b$$

$A \in \mathbb{R}^{m \times n}$, $f : \mathbb{R}^n \to \mathbb{R}$ 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

$$\max_y \quad g(y)$$

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

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

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

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

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.

Augmented Lagrangian

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

Augmented Lagrangian: $\rho > 0$: penalty parameter

$$\mathcal{L}_\rho(x; y) = f(x) + y^T(Ax - b) + \frac{\rho}{2}\|Ax - b\|_2^2$$

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

ADMM

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

$$A \in \mathbb{R}^{p \times n}, \;\; B \in \mathbb{R}^{p \times m}$$

$$s.t. \;\; Ax + Bz = c$$

Augmented Lagrangian:

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

"Alternating direction" method of multipliers

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

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

Coordinate-Wise Minimization

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

Dual Ascent

MM vs. ADMM

MM

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

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

ADMM

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

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

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

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

$$\Leftrightarrow \quad \text{epi}\, f = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \le t\}$$
$$\text{epi}\, g = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : g(x) \le t\}$$

are closed nonempty convex sets

This implies that x-update and z-update are <u>solvable</u>, 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 $r^k := Ax^k + Bz^k - c \to 0 \text{ as } k \to \infty$

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

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

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

A global variable x is shared across f_i's

A simple reformulation (global consensus problem):

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

$$\text{s.t. } \ x_i - z = 0, \quad i = 1, 2, \dots, N.$$

ADMM for Global Consensus

Augmented Lagrangian:

$$\mathcal{L}_\rho(x_1, \dots, x_N, z; y) = \sum_{i=1}^{N} \left(f_i(x_i) + y_i^T (x_i - z) + \frac{\rho}{2} \|x_i - z\|_2^2 \right)$$

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:

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

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

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

$$\bar{x} := \frac{1}{N} \sum_{i=1}^{N} x_i$$

$\Rightarrow$ 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:

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

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

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

Simplified ADMM:

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

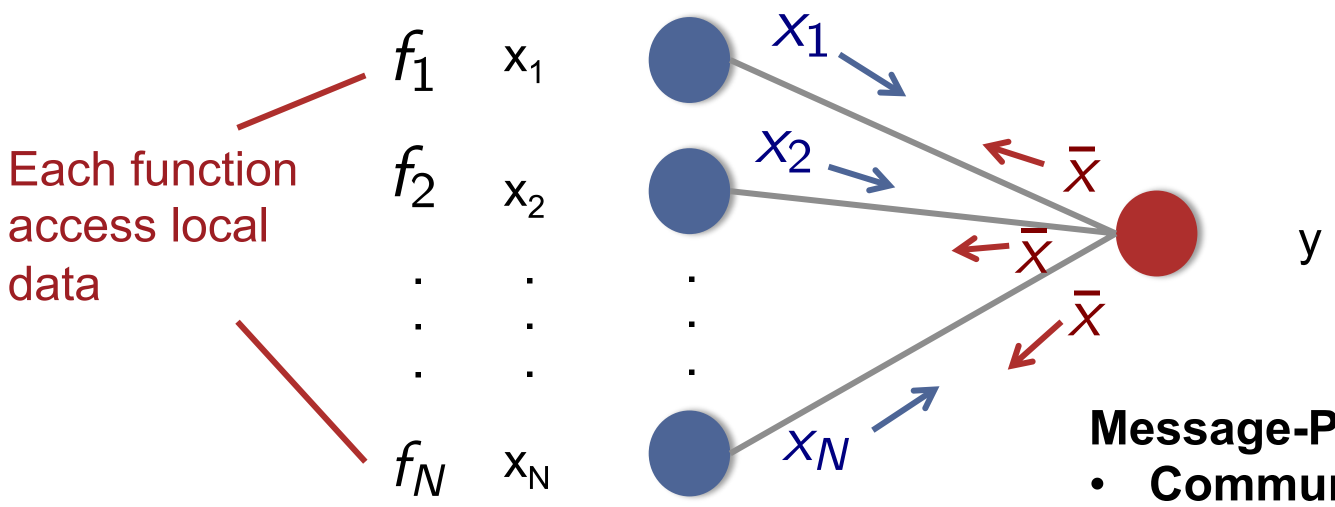
$$y_i^{k+1} = y_i^k + \rho(x_i^{k+1} - \bar{x}^{k+1})$$

ADMM for Global Consensus

Simplified ADMM:

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

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



Message-Passing Algorithm

- **Communication cost?**
- **Data storage?**