Alternating Direction Method of Multipliers (ADMM)

Dr. Dingzhu Wen

School of Information Science and Technology (SIST)
ShanghaiTech University

wendzh@shanghaitech.edu.cn

April 8, 2024

Overview

1. ADMM
2. ADMM for Consensus Optimization

ADMM: Basic Ideas

Optimization Problem (with convex f and g):

$$\min_{\mathbf{x}\in\mathbb{R}^N} \; f(\mathbf{x})+g(\mathbf{z}),$$
$$\text{s.t.} \;\; \mathbf{Ax}+\mathbf{Bz}=\mathbf{c}.$$

Augemented Lagrange function:

$$\mathcal{L}(\mathbf{x},\mathbf{z},\boldsymbol{\lambda})=f(\mathbf{x})+g(\mathbf{z})+\boldsymbol{\lambda}^T\left(\mathbf{Ax}+\mathbf{Bz}-\mathbf{c}\right)+\frac{\rho}{2}\left\|\mathbf{Ax}+\mathbf{Bz}-\mathbf{c}\right\|^2.$$

ADMM in one ieration:

$$\begin{aligned}
&\mathbf{x}-\text{minimization}: && \mathbf{x}_{k+1}=\arg\min_{\mathbf{x}} \; \mathcal{L}(\mathbf{x},\mathbf{z}_k,\boldsymbol{\lambda}_k),\\
&\mathbf{z}-\text{minimization}: && \mathbf{z}_{k+1}=\arg\min_{\mathbf{z}} \; \mathcal{L}(\mathbf{x}_{k+1},\mathbf{z},\boldsymbol{\lambda}_k),\\
&\text{Dual update}: && \boldsymbol{\lambda}_{k+1}=\boldsymbol{\lambda}_k+\rho(\mathbf{Ax}_{k+1}+\mathbf{Bz}_{k+1}-\mathbf{c}).
\end{aligned}$$

Qeustion: How to show the convergence of ADMM?

ADMM: Basic Ideas

Properties

- Perform one cycle of block-coordinate descent in each step,
- Can perform the $(\mathbf{x}; \mathbf{z})$ minimizations inexactly (**Why?**),
- If we minimized over $\mathbf{x}$ and $\mathbf{z}$ jointly, reduces to method of multipliers,
- Applications: Compressed sensing, image processing, matrix completion, sparse principal components analysis, etc.

ADMM and Optimality Conditions

Optimality Conditions

- Primal feasibility: $\mathbf{Ax} + \mathbf{Bz} - \mathbf{c} = \mathbf{0}$.
- Dual feasibility: $\nabla f(\mathbf{x}) + \mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0}, \quad \nabla g(\mathbf{z}) + \mathbf{B}^T\boldsymbol{\lambda} = \mathbf{0}$.

Optimality is guaranteed:

- Since $\mathbf{z}_{k+1}$ minimizes $\mathcal{L}(\mathbf{x}_{k+1}, \mathbf{z}, \boldsymbol{\lambda}_k)$, we have

$$\begin{aligned}\mathbf{0} =& \nabla g(\mathbf{z}_{k+1}) + \mathbf{B}^T\boldsymbol{\lambda}_k + \rho\mathbf{B}^T(\mathbf{Ax}_{k+1} + \mathbf{Bz}_{k+1} - \mathbf{c}),\\ =& g(\mathbf{z}_{k+1}) + \mathbf{B}^T\boldsymbol{\lambda}_{k+1}.\end{aligned}$$

- So, with ADMM dual variable update, $(\mathbf{x}_{k+1}, \mathbf{z}_{k+1}, \mathbf{y}_{k+1})$ satisfies second dual feasibility condition.
- Primal and first dual feasibility are achieved as $k \to +\infty$.

ADMM with Scaled Dual Variables

Combine linear and quadratic terms in augmented Lagrangian

$$\begin{aligned}\mathcal{L}(\mathbf{x},\mathbf{z},\boldsymbol{\lambda}) &= f(\mathbf{x}) + g(\mathbf{z}) + \boldsymbol{\lambda}^T(\mathbf{Ax}+\mathbf{Bz}-\mathbf{c}) + \frac{\rho}{2}\|\mathbf{Ax}+\mathbf{Bz}-\mathbf{c}\|^2, \\ &= f(\mathbf{x}) + g(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{Ax}+\mathbf{Bz}-\mathbf{c}+\boldsymbol{\mu}\|^2 + \text{constant},\end{aligned}$$

with $\boldsymbol{\mu}_k = \boldsymbol{\lambda}_k/\rho$.

ADMM (scaled dual form):

$$\begin{aligned}\mathbf{x}-\text{minimization}: &\quad \mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \; f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{Ax}+\mathbf{Bz}_k-\mathbf{c}+\boldsymbol{\mu}_k\|^2, \\ \mathbf{z}-\text{minimization}: &\quad \mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} \; g(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{Ax}_{k+1}+\mathbf{Bz}-\mathbf{c}+\boldsymbol{\mu}_k\|^2, \\ \text{Dual update}: &\quad \boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k + (\mathbf{Ax}_{k+1}+\mathbf{Bz}_{k+1}-\mathbf{c}).\end{aligned}$$

ADMM: Proximal Operator

consider $\mathbf{x}$-update when $\mathbf{A} = \mathbf{I}$:

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \quad f(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{v}\|^2 = \mathrm{Prox}_{f,\rho}(\mathbf{v}).$$

Some Special Cases:

- f is an indicator function of a convex set $\mathbb{C}$, $\mathbf{x}_{k+1} = \Pi_{\mathbb{C}}(\mathbf{v})$ (Prohection onto $\mathbb{C}$),
- $f = \alpha \|\|_1$, $\mathbf{x}_{k+1} = S_{\alpha/\rho}(\mathbf{v})$.

ADMM for Consensus Optimization

Optimization Problem: $\min_{\mathbf{x}\in\mathbb{R}^N} \sum_{i=1}^{M} f(\mathbf{x})$.

Decomposition: Form M copies of the $\mathbf{x}$:

$$\min_{\mathbf{x}_i,\mathbf{z}\in\mathbb{R}^N} \sum_{i=1}^{M} f(\mathbf{x}_i),$$
$$\mathbf{x}_i = \mathbf{z}, \quad 1 \leq i \leq M.$$

- $\{\mathbf{x}_i\}$ are local variables,
- $\mathbf{z}$ is the global variable,
- $\{\mathbf{x}_i = \mathbf{z}\}$are **consistency** or **consensus** constraints,
- Can add regularization using a $g(\mathbf{z})$ term.

ADMM for Consensus Optimization

Augmented Lagrange Function:

$$\mathcal{L}(\mathbf{x}_i, \mathbf{z}, \boldsymbol{\lambda}_i) = \sum_{i=1}^{M} \left[f(\mathbf{x}_i) + \boldsymbol{\lambda}_i^T (\mathbf{x}_i - \mathbf{z}) + \frac{\rho}{2} \left\| \mathbf{x}_i - \mathbf{z} \right\|^2 \right].$$

ADMM:

$$\mathbf{x}_{i,k+1} = \arg\min_{\mathbf{x}_i} \quad f(\mathbf{x}_i) + \boldsymbol{\lambda}_{i,k}^T (\mathbf{x}_i - \mathbf{z}_k) + \frac{\rho}{2} \left\| \mathbf{x}_i - \mathbf{z}_k \right\|^2, \ \forall i,$$

$$\mathbf{z}_{k+1} = \frac{1}{M} \sum_{i=1}^{M} \left(\mathbf{x}_{i,k+1} + \frac{1}{\rho} \boldsymbol{\lambda}_{i,k} \right),$$

$$\boldsymbol{\lambda}_{i,k+1} = \boldsymbol{\lambda}_{i,k} + \rho (\mathbf{x}_{i,k+1} - \mathbf{z}_{i,k+1}), \ \forall i.$$

Thank you!

wendzh@shanghaitech.edu.cn