Augmented Lagrangian Method

Dr. Dingzhu Wen

School of Information Science and Technology (SIST)
ShanghaiTech University

wendzh@shanghaitech.edu.cn

April 8, 2024

Overview

Problems with Linear Constraints

Optimization Problem:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \ f(\mathbf{x}), \quad \text{s.t.} \ \mathbf{A}\mathbf{x} = \mathbf{b},$$

where f is smooth.
KKT conditions are necessary for idetifying the optimal solution

$$\nabla f(x_*) = -\mathbf{A}^T \lambda_*,$$
$$\mathbf{A}\mathbf{x}_* = \mathbf{b}.$$

When f is smooth and convex, these conditions are also sufficient. (In fact, it's enough for f to be convex on the null space of A.)

Problems with Linear Constraints

Define the Lagrangian function:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x}_* - \mathbf{b}).$$

KKT conditions:

$$\nabla\mathcal{L}(\mathbf{x}_*, \boldsymbol{\lambda}_*) = \begin{bmatrix} \nabla\mathcal{L}_{\mathbf{x}}(\mathbf{x}_*, \boldsymbol{\lambda}_*) \\ \nabla\mathcal{L}_{\boldsymbol{\lambda}}(\mathbf{x}_*, \boldsymbol{\lambda}_*) \end{bmatrix} = \mathbf{0}$$

Suppose now that f is convex but not smooth. First-order optimality conditions (necessary and sufficient) are that there exists $\boldsymbol{\lambda}_*$ such that

$$-\mathbf{A}^T\lambda_* \in \partial f(x_*), \qquad \mathbf{A}\mathbf{x}_* = \mathbf{b}.$$

Augmented Lagrangian Methods

Optimization Problem:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \quad f(\mathbf{x}), \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b},$$

with f proper, lower semi-continuous, and convex,
The augmented Lagrangian is (with $\rho > 0$)

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}; \rho) = f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} - \mathbf{b}\right\|_2^2.$$

Basic augmented Lagrangian (a.k.a. method of multipliers) is

$$\mathbf{x}_k = \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_{k-1}; \rho);$$
$$\boldsymbol{\lambda}_k = \boldsymbol{\lambda}_{k-1} + \rho(\mathbf{A}\mathbf{x}_k - \mathbf{b}).$$

A Favorite Derivation

The problem can be re-written as

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} \; f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}),$$

Obviously, the max w.r.t. $\boldsymbol{\lambda}$ will be $+\infty$, unless $\mathbf{A}\mathbf{x} = \mathbf{b}$. So this is equivalent to the original problem.

This equivalence is not very useful, computationally: $\max_{\boldsymbol{\lambda}}$ function is highly nonsmooth w.r.t. $\mathbf{x}$. Smooth it by adding a "proximal point" term, penalizing deviations from a prior estimate $\bar{\boldsymbol{\lambda}}$:

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} \; f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) - \frac{1}{2\rho} \left\| \boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}} \right\|_2^2 .$$

Maximization w.r.t. $\boldsymbol{\lambda}$ is now trivial (a concave quadratic), yielding

$$\boldsymbol{\lambda} = \bar{\boldsymbol{\lambda}} + \rho(\mathbf{A}\mathbf{x} - \mathbf{b}).$$

A Favorite Derivation

Insert $\boldsymbol{\lambda} = \bar{\boldsymbol{\lambda}} + \rho(\mathbf{A}\mathbf{x} - \mathbf{b})$ leads to

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) + \bar{\boldsymbol{\lambda}}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \mathcal{L}(\mathbf{x}, \bar{\boldsymbol{\lambda}}; \rho).$$

Hence, the augmented Lagrangian process can be viewed as:

- $\min_{\mathbf{x}} \quad \mathcal{L}(\mathbf{x}, \bar{\boldsymbol{\lambda}}; \rho)$ to get new x.
- Shift the "prior" on $\boldsymbol{\lambda}$ by updating to the latest max: $\bar{\boldsymbol{\lambda}} + \rho(\mathbf{A}\mathbf{x} - \mathbf{b})$.
- repeat until convergence.

Can also increase ρ (to sharpen the effect of the prox term), if needed.

Ineqaulity Linear Constraints & Nonlinear Constraints

Optimization Problem

$$\min_{\mathbf{x}\in\mathbb{R}^N} \; f(\mathbf{x}),$$
$$\text{s.t.} \;\; \mathbf{A}\mathbf{x} \geq \mathbf{b}.$$

Apply the same reasoning to the constrained min-max formulation:

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}\geq\mathbf{0}} \; f(\mathbf{x}) - \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x}-\mathbf{b}).$$

After the prox-term is added, can find the minimal $\boldsymbol{\lambda}$ in closed form (as for prox-operators). Leads to update formula:

$$\boldsymbol{\lambda} = \max\left\{\bar{\boldsymbol{\lambda}} + \rho(\mathbf{A}\mathbf{x}-\mathbf{b}), \mathbf{0}\right\}.$$

This derivation extends immediately to nonlinear constraints $\mathbf{c}(\mathbf{x}) = \mathbf{0}$ and $\mathbf{c}(\mathbf{x}) \geq \mathbf{0}$.

"Explicit" Constraints, Inequality Constraints

Optimization Problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^N} \quad & f(\mathbf{x}), \\ \text{s.t.} \quad & \mathbf{c}(\mathbf{x}) \geq \mathbf{0}. \end{aligned}$$

There may be other constraints on $\mathbf{x}$ (such as $\mathbf{x}^2$) that we prefer to handle explicitly in the subproblem.

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^N} \quad & f(\mathbf{x}), \\ \text{s.t.} \quad & \mathbf{c}(\mathbf{x}) = \mathbf{s}, \\ & \mathbf{s} \geq 0. \end{aligned}$$

Thank you!

wendzh@shanghaitech.edu.cn