Regularized Equality Constrained Quadratic Optimization

Anton Xue

1 Introduction

We look at regularized approximations of equality-constrained quadratic programming. In particular, what is $\|x^\star - x_\lambda^\star\|$, where $x^\star$ is the optimal solution to the original problem, while $x_\lambda^\star$ is the solution to the regularized problem.

2 Background

Stephen Boyd and Lieven Vandenberghe [1].

2.1 Equality Constrained Quadratic Programming

A quadratic program with equality constraints [1] is

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} x^\top M x && (1) \\ \text{subject to} \quad & Ax = b && (2) \end{aligned}$$

with variable in $x \in \mathbb{R}^n$, where $M \succeq 0$ and the constraint $A \in \mathbb{R}^{m \times n}$ is, for our purposes, fat and full rank.

The ℓ^2 regularized version for $\lambda > 0$ looks like

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} x^\top M x + \lambda x^\top x && (3) \\ \text{subject to} \quad & Ax = b && (4) \end{aligned}$$

2.2 Karush-Kuhn-Tucker Conditions

The Lagrangian for (1) is

$$L(x, p) = \frac{1}{2} x^\top M x + p^\top (Ax - b)$$

for which the optimality conditions [1] are

$$Mx^\star + A^\top p^\star = b, \qquad Ax^\star = b$$

or more compcatctly expressed:

$$\begin{bmatrix} M & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ p^\star \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} \tag{5}$$

On the other hand the Lagrangian of the regularized problem (3) is

$$L_\lambda(x, p) = \frac{1}{2} x^\top (M + \lambda I) x + p^\top (Ax - b)$$

which has the KKT conditions

$$\begin{bmatrix} M + \lambda I & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star_\lambda \\ p^\star_\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} \tag{6}$$

2.3 Matrix Inversion

Consider a symmetric matrix

$$M = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$$

when $A \succ 0$, define the Schur complement $S = C - B^\top A^{-1} B$. Then

$$\begin{bmatrix} A & B \\ B^\top C & \end{bmatrix}^{-1} \begin{bmatrix} A^{-1} + A^{-1} B S^{-1} B^\top A^{-1} & -A^{-1} B S^{-1} \\ -S^{-1} B^\top A^{-1} & S^{-1} \end{bmatrix}$$

3 Bounding Norms

Noting that by linearity

$$\begin{bmatrix} M & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ p^\star \end{bmatrix} + \begin{bmatrix} \lambda I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ p^\star \end{bmatrix} = \begin{bmatrix} M + \lambda I & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ p^\star \end{bmatrix} = \begin{bmatrix} \lambda x^\star \\ b \end{bmatrix}$$

and so through subtracting equations,

$$\begin{bmatrix} M + \lambda I & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star - x^\star_\lambda \\ p^\star - p^\star_\lambda \end{bmatrix} = \begin{bmatrix} \lambda x^\star \\ 0 \end{bmatrix}$$

Given our assumptions on $M + \lambda I \succ 0$ and A is full rank. For ease of notation, define $\Lambda = M + \lambda I$. The Schur complement is then $S = -A\Lambda^{-1}A^\top$, and

$$\begin{bmatrix} x^\star - x^\star_\lambda \\ p^\star - p^\star_\lambda \end{bmatrix} = \begin{bmatrix} \Lambda^{-1} + \Lambda^{-1} A^\top S^{-1} A \Lambda^{-1} & -\Lambda^{-1} A^\top S^{-1} \\ -S^{-1} A \Lambda^{-1} & S^{-1} \end{bmatrix} \begin{bmatrix} \lambda x^\star \\ 0 \end{bmatrix}$$

In other words:

$$x^\star - x^\star_\lambda = \left(\Lambda^{-1} - \Lambda^{-1} A^\top (A \Lambda^{-1} A^\top)^{-1} A \Lambda^{-1}\right) \lambda x^\star$$

To simplify notation slightly, use $\Gamma = \Lambda^{-1}$ because it looks like an upside-down L. Then

$$x^\star - x^\star_\lambda = \left(\Gamma - \Gamma A^\top (A \Gamma A^\top)^{-1} A \Gamma)\right) \lambda x^\star$$

References

[1] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.