Regularized Equality Constrained Quadratic Optimization

Anton Xue and Nikolai Matni

1 Introduction

We look at regularized approximations of equality-constrained quadratic programming. In particular, how much does the optimal solution deviate when a regularizer is introduced?

2 Background

Stephen Boyd and Lieven Vandenberghe [1].

2.1 Equality Constrained Quadratic Programming

A quadratic program with equality constraints [1] is

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

with variable in $x \in \mathbb{R}^n$, where $Q \succeq 0$ 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

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

2.2 Karush-Kuhn-Tucker Conditions

The Lagrangian for (1) is of form

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

for which the optimality conditions [1] are

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

or more compcatctly expressed:

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

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

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

which has the KKT conditions

$$\begin{bmatrix} Q + \lambda I & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x_\lambda^\star \\ p_\lambda^\star \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} \tag{6}$$

2.3 Matrix Inversion

Consider a symmetric matrix

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

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

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

and so through subtracting equations,

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

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

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

In other words:

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

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

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

Theorem 1. *Without loss of generality, assume that*

$$Q = \begin{bmatrix} q_1 & & \\ & \ddots & \\ & & q_n \end{bmatrix}, \qquad q_1 \geq \cdots \geq q_n > 0.$$

Then for $\lambda \leq q_n$,

$$\|x^\star - x^\star_\lambda\| \leq \left(\frac{\lambda}{q_n+\lambda} - \frac{\lambda}{q_1+\lambda}\right)\|x^\star\|$$

Proof. First note that

$$\Lambda = \begin{bmatrix} q_1+\lambda & & \\ & \ddots & \\ & & q_n+\lambda \end{bmatrix}, \qquad \Gamma = \Lambda^{-1} = \begin{bmatrix} \frac{1}{q_1+\lambda} & & \\ & \ddots & \\ & & \frac{1}{q_n+\lambda} \end{bmatrix}$$

We seek the bound the RHS of the inequality

$$\|x^\star - x^\star_\lambda\| \leq \big\|\Gamma - \Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma\big\| \cdot \lambda\|x^\star\|$$

The primary challenge here is correctly *lower bounding* the $\Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma$ term. For this we *upper bound* the $A\Gamma A^\top$ term. Letting

$$A = U\Sigma V^\top = U\begin{bmatrix}\Sigma_1 & 0\end{bmatrix}\begin{bmatrix}V_1^\top \\ V_2^\top\end{bmatrix}, \qquad \Sigma_1 = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{bmatrix}$$

be an SVD of A; an upper bound of $A\Gamma A^\top$ is

$$A\Gamma A^\top = U\Sigma_1 V_1^\top \begin{bmatrix} \frac{1}{q_1+\lambda} & & \\ & \ddots & \\ & & \frac{1}{q_n+\lambda} \end{bmatrix} V_1\Sigma_1 U^\top \preceq \frac{1}{q_n+\lambda}U\Sigma_1^2U^\top$$

Consequently, a lower bound for $\Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma$ is

$$\Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma \succeq \Gamma V_1\Sigma_1U^\top\left(\frac{1}{q_n+\lambda}U\Sigma_1^2U^\top\right)^{-1}U\Sigma_1V_1^\top\Gamma = (q_n+\lambda)\Gamma V_1V_1^\top\Gamma \succeq \frac{q_n+\lambda}{q_1+\lambda}\Gamma$$

Then using this,

$$\Gamma - \Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma \preceq \Gamma - \frac{q_n+\lambda}{q_1+\lambda}\Gamma \preceq \left(\frac{1}{q_n+\lambda} - \frac{1}{q_1+\lambda}\right)I$$

for which we derive the desired bound

$$\|x^\star - x^\star_\lambda\| \leq \big\|\Gamma - \Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma\big\| \cdot \lambda\|x^\star\| \leq \left(\frac{\lambda}{q_n+\lambda} - \frac{\lambda}{q_1+\lambda}\right)\|x^\star\|$$

$\square$

References

[1] Stephen P Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.