Regularized Equality Constrained Quadratic Optimization

Anton Xue and Nikolai Matni

1 Introduction

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

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

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

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

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*

$$Q = \operatorname{diag}(q_1, q_2, \dots, q_n), \quad q_1 \ge \cdots \ge q_n \ge 0.$$

Then

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

Proof. Our angle of attack is to seek a bound $\big\|\Gamma - \Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma\big\|$. For this, note that both $\Gamma \succ 0$ and $\Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma \succ 0$ and so one technique is to find a matrix G such that

$$G \prec \Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma \quad\implies\quad \big\|\Gamma - \Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma\big\| \le \|\Gamma - G\|$$

We can iteratively construct G. Let a singular value decomposition of A be

$$A = U\Sigma V^\top = U\begin{bmatrix}\Sigma_1 & 0\end{bmatrix}\begin{bmatrix}V_1^\top \\ V_2^\top\end{bmatrix} = U\Sigma_1 V_1^\top, \qquad \Sigma_1 = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_m)$$

Then first upper-bounding the inside of the inverse in order to lower-bound the overall inverse,

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

Then for the inverse we would have

$$\begin{aligned}\Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma &\preceq \Gamma A^\top\big[(q_n+\lambda)U\Sigma_1^{-2}U^\top\big]A\Gamma \\ &= (q_n+\lambda)\Gamma V_1\Sigma_1U^\top\big[U\Sigma_1^{-2}U^\top\big]U\Sigma_1V_1^\top\Gamma \\ &= (q_n+\lambda)\Gamma\Gamma\end{aligned}$$

Putting these together:

$$\begin{aligned}\Gamma - \Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma &\preceq \Gamma[I - (q_n+\lambda)\Gamma] \\ &\preceq \Gamma\left\{I - (q_n+\lambda)\begin{bmatrix}\frac{1}{q_1+\lambda} & & \\ & \ddots & \\ & & \frac{1}{q_n+\lambda}\end{bmatrix}\right\} \\ &\preceq \Gamma\left(1 - \frac{q_n+\lambda}{q_1+\lambda}\right)\end{aligned}$$

Applying sub-multiplicativity of norms we arrive at

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

$\square$

References

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