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

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

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

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

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,

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

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

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

In other words:

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

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

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

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

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

□

4 Optimal Value Deviation

One challenge with merely bounding $\|x^\star - x^\star_\lambda\|$ is that this norm difference may grow unbounded. Consider an instance where the objective Q has a non-trivial kernel: we may have

$$\|\Pi_{\ker Q} x^\star\| \gg \|\Pi_{\ker Q} x^\star_\lambda\|$$

where $\Pi_{\ker Q}$ is the projection onto $\ker Q$; this is possible because the regularized problems would force $\Pi_{\ker Q} x^\star_\lambda$ to be close to the origin. Instead, another way of examining sub-optimality is to consider the difference

$$\frac{1}{2}(x^\star_\lambda)^\top Q x^\star_\lambda - \frac{1}{2}(x^\star)^\top Q x^\star > 0$$

which we know holds because $x^\star_\lambda$ is no more optimal than $x^\star$.

Theorem 2 ([2])**.** *If X is an $n \times q$ matirx contained in the column-space of an $n \times n$ symmetrical matrix V, then*

$$(V + XX^\top)^\dagger = V^\dagger - V^\dagger X(I + X^\top V^\dagger X)^{-1} X^\top V^\dagger$$

where "to be in the column-space" means the same as $VV^\dagger X = X$, and in a more general case $(I - VV^\dagger)X \neq 0$.

Theorem 3. *Define*

$$Z = I - A^\dagger A, \qquad \hat{x} = A^\dagger b$$

so that solutions take form $\hat{x} + Zv$, then a bound on the optimality gap is

$$\begin{aligned}
&\frac{1}{2}(x^\star_\lambda)^\top Q x^\star_\lambda - \frac{1}{2}(x^\star)^\top Q x^\star \\
&\qquad \leq \left\|Q^{1/2}\hat{x}\right\|^2 \cdot \left\|Q^{1/2}Z\right\|^2 \cdot \frac{\lambda m_0}{1 - \lambda m_0} + \left\|Q^{1/2}\hat{x}\right\|^4 \cdot \left\|Q^{1/2}Z\right\|^4 \cdot \left(\frac{\lambda m_0}{1 - \lambda m_0}\right)^2
\end{aligned}$$

where m_0 is the smallest non-zero eigenvalue of V.

Proof. Let $\hat{x} = A^\dagger b$, and $\Lambda = Q + \lambda I$, then

$$x^\star = \hat{x} - Zu, \qquad x^\star = \hat{x} - Zy$$

where

$$Z = I - A^\dagger A, \qquad u = (Z^\top M Z)^\dagger Z^\top Q\hat{x}, \qquad y = (Z^\top \Lambda Z)^\dagger Z^\top \Lambda \hat{x}$$

which can be derived from examining the optimality conditions of the quadratic problem. However, because $\hat{x} \in \ker A^\top$, there is a further simplification:

$$y = (Z^\top \Lambda Z)^\dagger Z^\top (Q + \lambda I)\hat{x} = (Z^\top \Lambda Z)^\dagger Z^\top Q\hat{x}$$

Comparing the objective values attained by $x^\star_\lambda$ and $x^\star$, we have

$$\begin{aligned}
&\frac{1}{2}(x^\star_\lambda)^\top Q x^\star_\lambda - \frac{1}{2}(x^\star)^\top Q x^\star \\
&\quad = \frac{1}{2}(\hat{x} + Zy)^\top Q(\hat{x} + Zy) - \frac{1}{2}(\hat{x} + Zu)^\top Q(\hat{x} + Zu) \\
&\quad = (\hat{x})^\top QZ(y-u) + \frac{1}{2}y^\top Z^\top QZy - \frac{1}{2}u^\top Z^\top QZu \\
&\quad = (\hat{x})^\top QZ(y-u) + \frac{1}{2}\left(\|Q^{1/2}Zy\| - \|Q^{1/2}Zu\|\right) \\
&\quad \leq (\hat{x})^\top QZ(y-u) + \frac{1}{2}\left\|Q^{1/2}Z(y-u)\right\|^2 && \text{(Reverse triangle inequality)} \\
&\quad \leq \left\|Q^{1/2}\hat{x}\right\| \cdot \left\|Q^{1/2}Z(y-u)\right\| + \frac{1}{2}\left\|Q^{1/2}Z(y-u)\right\|^2
\end{aligned}$$

From here,

$$\begin{aligned}
\left\|Q^{1/2}Z(y-u)\right\| &= \left\|Q^{1/2}Z\Big[\left(Z^\top QZ\right)^\dagger Z^\top Q\hat{x} - \left(Z^\top \Lambda Z\right)^\dagger Z^\top Q\hat{x}\Big]\right\| \\
&= \left\|Q^{1/2}Z\Big[\left(Z^\top QZ\right)^\dagger Z^\top Q\hat{x} - \left(Z^\top (Q + \lambda I)Z\right)^\dagger Z^\top Q\hat{x}\Big]\right\|
\end{aligned}$$

We now apply Theorem 2: identifying the mapping $V \mapsto Z^\top QZ$ and $X \mapsto \sqrt{\lambda}Z^\top$. First we show that X lies in the column space of V. Consider the eigenvalue decomposition

$$V = F^\top DF, \qquad D = \mathrm{diag}(d_1, \ldots, d_k, 0, \ldots, 0), \qquad X = \sqrt{\lambda}F^\top I,$$

where F is orthogonal and D is diagonal. Then $D^\dagger = \mathrm{diag}(d_1^{-1}, \ldots, d_k^{-1}, 0, \ldots 0)$ and so

$$VV^\dagger X = (F^\top DF)(F^\top D^\dagger F)F^\top \sqrt{\lambda} I = \sqrt{\lambda}F^\top = X$$

Now, the above norm-bound can be re-written as follows,

$$\begin{aligned}
&\left\|Q^{1/2}Z(y-u)\right\| \\
&\quad = \left\|Q^{1/2}Z\big[V^\dagger - \big(V^\dagger - V^\dagger X(I + X^\top V^\dagger X)^{-1}X^\top V^\dagger\big)\big]Z^\top Q\hat{x}\right\| \\
&\quad = \left\|Q^{1/2}Z\big(V^\dagger X(I + X^\top V^\dagger X)^{-1}X^\top V^\dagger\big)Z^\top Q\hat{x}\right\| \\
&\quad \leq \left\|Q^{1/2}Z\right\| \cdot \left\|V^\dagger X\right\|^2 \cdot \left\|Z^\top Q\hat{x}\right\| \cdot \left\|(I + X^\top V^\dagger X)^{-1}\right\| \\
&\quad \leq \left\|Q^{1/2}Z\right\| \cdot \left\|V^\dagger X\right\|^2 \cdot \left\|Z^\top Q\hat{x}\right\| \cdot \left(\sum_{l=0}^{\infty} \left\|X^\top V^\dagger X\right\|^l\right) \\
&\quad \leq \left\|Q^{1/2}Z\right\| \cdot \left\|V^\dagger X\right\|^2 \cdot \left\|Z^\top Q\hat{x}\right\| \cdot \left(\sum_{l=0}^{\infty} \lambda^l \left\|V^\dagger\right\|^l\right) \\
&\quad \leq \left\|Q^{1/2}Z\right\| \cdot \left\|V^\dagger X\right\|^2 \cdot \left\|Z^\top Q\hat{x}\right\| \cdot (1 - \lambda m_0)^{-1} \\
&\quad \leq \left\|Q^{1/2}Z\right\| \cdot \left\|Z^\top Q\hat{x}\right\| \cdot \frac{\lambda m_0}{1 - \lambda m_0} \\
&\quad \leq \left\|Q^{1/2}Z\right\|^2 \cdot \left\|Q^{1/2}\hat{x}\right\| \cdot \frac{\lambda m_0}{1 - \lambda m_0}
\end{aligned}$$

where m_0 is the smallest non-zero eigenvalue of V, and so

$$\begin{aligned}
&\frac{1}{2}(x_\lambda^\star)^\top Q x_\lambda^\star - \frac{1}{2}(x^\star)^\top Q x^\star \\
&\quad \leq \left\|Q^{1/2}\hat{x}\right\| \cdot \left\|Q^{1/2}Z(y-u)\right\| + \frac{1}{2}\left\|Q^{1/2}Z(y-u)\right\|^2 \\
&\quad \leq \left\|Q^{1/2}\hat{x}\right\|^2 \cdot \left\|Q^{1/2}Z\right\|^2 \cdot \frac{\lambda m_0}{1-\lambda m_0} + \left\|Q^{1/2}\hat{x}\right\|^4 \cdot \left\|Q^{1/2}Z\right\|^4 \cdot \left(\frac{\lambda m_0}{1-\lambda m_0}\right)^2
\end{aligned}$$

and this vanishes as $\lambda \to 0$.

$\square$

References

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

[2] Pavel Kovanic. “On the pseudoinverse of a sum of symmetric matrices with applications to estimation”. In: *Kybernetika* 15.5 (1979), pp. 341–348.