Regularized Equality-Constrained Quadratic Programming

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

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

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

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

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 compactly expressed:

$$\begin{bmatrix} Q & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ p^\star \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} \tag{3}$$

On the other hand the Lagrangian of the regularized problem (2) is

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

which has the KKT conditions

$$\begin{bmatrix} Q + \lambda I & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x_\lambda^\star \\ p_\lambda^\star \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} \tag{4}$$

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

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

$\square$

4 Suboptimality via Pseudoinverses

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 rather than hiding far away from zero inside $\ker Q$. 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 $n \times q$ matrix 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

$$\frac{1}{2}(x^\star_\lambda)^\top Q x^\star_\lambda - \frac{1}{2}(x^\star)^\top Q x^\star \leq O\left(\frac{\lambda q_0}{q_0 - \lambda}\right)$$

where q_0 is the smallest non-zero eigenvalue of Q and $\lambda < q_0$.

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

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

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 $\hat{x} \in \operatorname{ran} A^\top = (\ker A)^\perp$, there is a further simplification as $Z^\top \hat{x} = 0$:

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

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

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

Which simplifies to

$$
\frac{1}{2}(x^\star_\lambda)^\top Q x^\star_\lambda - \frac{1}{2}(x^\star)^\top Q x^\star \leq \left(\left\|Q^{1/2}\hat{x}\right\| + \frac{1}{2}\left\|Q^{1/2}Zu\right\| + \frac{1}{2}\left\|Q^{1/2}Zy\right\|\right) \cdot \left\|Q^{1/2}Z(y-u)\right\| \tag{5}
$$

We now apply to Theorem 2 to $Q^{1/2}Zy$: identifying the mapping $V \mapsto Z^\top QZ$ and $X \mapsto \sqrt{\lambda}Z$, first show that X lies in the column space of V. For this, consider the eigen decomposition

$$
V = FDF^\top, \qquad F = \begin{bmatrix} F_1 & F_2 \end{bmatrix}, \qquad D = \operatorname{diag}(d_1, \dots, d_k, 0, \dots, 0), \qquad X = \sqrt{\lambda}F_1
$$

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

$$
VV^\dagger X = (FDF^\top)(FD^\dagger F^\top)\sqrt{\lambda}F_1 = \sqrt{\lambda}F_1 = X
$$

thus, Theorem 2 applies and

$$
\begin{aligned}
Q^{1/2}Zy &= Q^{1/2}Z\big(Z^\top(Q+\lambda I)Z\big)^\dagger Z^\top Q\hat{x} \\
&= Q^{1/2}Z\big[V^\dagger - V^\dagger X(I + X^\top V^\dagger X)^{-1}X^\top V^\dagger\big]Z^\top Q\hat{x}
\end{aligned}
$$

With this, we now bound the terms that appear in (5). For the easy one:

$$
\left\|Q^{1/2}\hat{x}\right\| = \left\|Q^{1/2}A^\dagger b\right\| \leq \lambda_{\max}(Q)^{1/2}\frac{1}{\sigma_{\min}(A)}\|b\| \tag{6}
$$

For the inside second term, accounting for the fact that Z is a projection,

$$
\left\|Q^{1/2}Zu\right\| = \left\|Q^{1/2}ZV^\dagger Z^\top QA^\dagger b\right\| \leq \lambda_{\max}(Q)^{3/2}\frac{1}{q_0}\frac{1}{\sigma_{\min}(A)}\|b\| \tag{7}
$$

where q_0 is the smallest non-zero eigenvalue of Q. For the third inside term:

$$\begin{aligned}
&\left\|Q^{1/2}Zy\right\| \\
&\quad = \left\|Q^{1/2}Z\left[V^\dagger - V^\dagger X(I+X^\top V^\dagger X)^{-1}X^\top V^\dagger\right]Z^\top QA^\dagger b\right\| \\
&\quad \le \left\|Q^{1/2}ZV^\dagger Z^\top QA^\dagger b\right\| + \left\|Q^{1/2}ZV^\dagger X(I+X^\top V^\dagger X)^{-1}X^\top V^\dagger Z^\top QA^\dagger b\right\| \\
&\quad \le \left\|Q^{1/2}ZV^\dagger Z^\top QA^\dagger b\right\| + \|Q\|^{3/2}\left\|V^\dagger\right\|^2 \cdot \|X\|^2 \cdot \left\|A^\dagger b\right\| \cdot \left\|(I+X^\top V^\dagger X)^{-1}\right\| \\
&\quad \le \frac{\lambda_{\max}(Q)^{3/2}\cdot\|b\|}{q_0\sigma_{\min}(A)} + \frac{\lambda_{\max}(Q)^{3/2}\cdot\|b\|\lambda}{q_0^2\sigma_{\min}(A)} \cdot \left\|(I+X^\top V^\dagger X)^{-1}\right\| \\
&\quad \le \frac{\lambda_{\max}(Q)^{3/2}\cdot\|b\|}{q_0\sigma_{\min}(A)} + \frac{\lambda_{\max}(Q)^{3/2}\cdot\|b\|\lambda}{q_0^2\sigma_{\min}(A)} \sum_{l=0}^{\infty} \left\|X^\top V^\dagger X\right\|^l \\
&\quad \le \frac{\lambda_{\max}(Q)^{3/2}\cdot\|b\|}{q_0\sigma_{\min}(A)} + \frac{\lambda_{\max}(Q)^{3/2}\cdot\|b\|\lambda}{q_0^2\sigma_{\min}(A)} \sum_{l=0}^{\infty} (\lambda/q_0)^l \\
&\quad \le \frac{\lambda_{\max}(Q)^{3/2}\cdot\|b\|}{q_0\sigma_{\min}(A)} + \frac{\lambda_{\max}(Q)^{3/2}\cdot\|b\|\lambda}{q_0^2\sigma_{\min}(A)} \left(1-\frac{\lambda}{q_0}\right)^{-1}
\end{aligned}$$

which requires $\lambda < q_0$ for convergence. Finally for the right outer term:

$$\left\|Q^{1/2}Z(y-u)\right\| = \left\|Q^{1/2}ZV^\dagger X(I+X^\top V^\dagger X)^{-1}X^\top V^\dagger Z^\top QA^\dagger b\right\| \le \frac{\lambda_{\max}(Q)^{3/2}\|b\|\lambda}{q_0^2\sigma_{\min}(A)}\left(1-\frac{\lambda}{q_0}\right)^{-1}$$

Putting these together, we find that

$$\begin{aligned}
&\frac{1}{2}(x^\star_\lambda)^\top Qx^\star_\lambda - \frac{1}{2}x^\star Qx^\star \\
&\quad \le \left(\left\|Q^{1/2}\hat{x}\right\| + \frac{1}{2}\left\|Q^{1/2}Zu\right\| + \frac{1}{2}\left\|Q^{1/2}Zy\right\|\right) \cdot \left\|Q^{1/2}Z(y-u)\right\| \\
&\quad \le \max\{\lambda_{\max}(Q)^2, \lambda_{\max}(Q)^3\} \cdot \max\{q_0^{-2}, q_0^{-3}, q_0^{-4}\} \cdot \left(\frac{\|b\|}{\sigma_{\min}(A)}\right)^2 O\left(\frac{\lambda q_0}{q_0-\lambda}\right)
\end{aligned}$$

□

5 Suboptimality via Interior Points

The analysis of interior point methods, in particular log-barrier functions in [1] Section 11.2 gives yet another way to bound optimality gap.

Consider an optimization problem of form

$$\begin{aligned}
\text{minimize} \quad & f(x) + \frac{1}{t}\phi(x) \\
\text{subject to} \quad & Ax = b
\end{aligned} \tag{8}$$

where

$$f(x) = \frac{1}{2} x^\top Q x, \qquad \phi(x) = \|x\|^2 - M$$

where $M > 0$ is a large constant This problem is equivalent to ℓ^2 regularized constrained quadratic minimization, and in fact that M constant is not needed, but nevertheless this allows us to achieve theoretical guarantees. Intuitively, as $t \to \infty$, the solution $x^\star(t)$ converges to the optimal $x^\star$ of the original problem (1).

Additionally impose that $\|x^\star(t)\|^2 \leq M$ for all $t > 0$.

Theorem 4. *Let $M > 0$ be sufficiently large such that the problem*

$$\text{minimize} \quad f(x), \qquad \text{subject to} \quad Ax = b, \quad \|x\|^2 \leq M \tag{9}$$

is feasible at $x^\star(t)$ for all $t > 0$, the optimal (central path) solutions to (8) as a function of t. Let $p^\star$ be the optimal value of (9), then the suboptimality is bounded by

$$f(x^\star(t)) - p^\star \leq \frac{M - \|x^\star(t))\|^2}{t}$$

Proof. Note that (8) is minimized at $x^\star(t)$ for each $t > 0$, with optimality conditions

$$\nabla f(x^\star(t)) + \frac{1}{t} \nabla \phi(x^\star(t)) + A^\top \gamma = 0, \qquad \phi(x^\star(t)) \leq 0, \qquad Ax^\star(t) = b$$

We show that every $x^\star(t)$ corresponds to a dual feasible point of (9), and hence a lower bound on $p^\star$. To see this, examine the Lagrangian of (9):

$$L(x, \lambda, \nu) = f(x) + \lambda \phi(x) + \nu^\top (Ax - b)$$

and relating (8) with (9) by using the mappings $\hat{\lambda} \mapsto 1/t$ and $\hat{\nu} \mapsto \gamma$,

$$L(x, \hat{\lambda}, \hat{\nu}) = f(x) + \hat{\lambda} \phi(x) + \hat{\nu}^\top (Ax - b) = f(x) + \frac{1}{t} \phi(x) + \gamma^\top (Ax - b)$$

Thus, $x^\star(t)$ is a feasible point of (9) (since $\phi(x^\star(t)) \leq 0$ and $Ax^\star(t) = b$ by conditions of (8)) that minimizes the Lagrangian for points $\hat{\lambda}, \hat{\nu}$. As $\hat{\lambda} = 1/t > 0$, conclude that $\hat{\lambda}, \hat{\nu}$ are dual-*feasible* (not necessarily optimal) points of (9). Consequently, the Largange dual function value $g(\hat{\lambda}, \hat{\nu})$ is finite, and the desired suboptimality bound is achieved:

$$\begin{aligned} g(\hat{\lambda}, \hat{\nu}) &= f(x^\star(t)) + \hat{\lambda} \phi(x^\star(t)) + \hat{\nu}^\top (Ax^\star(t) - b) \\ &= f(x^\star(t)) + \hat{\lambda} \phi(x^\star(t)) \\ &= f(x^\star(t)) + \frac{1}{t} (\|x^\star(t)\|^2 - M) \leq p^\star \end{aligned}$$

□

For *sufficiently* large enough $M > 0$, we also have that $f(x^\star) = p^\star$ is also the optimal value of (1).

So how large of an M is needed? Since $\|x^\star(t)\| \leq \|x^\star\|$ due to regularization, it suffices to take $M \geq \|x^\star\|^2$. To bound $\|x^\star\|$, leveraging the results from previous sections, where $\hat{x} = A^\dagger b$ and $Z = I - A^\dagger A$ such that solutions are of form $\hat{x} - Zu$

$$\begin{aligned}
\|x^\star\|^2 &= \left\|\hat{x} - Z(Z^\top Q Z)^\dagger Z^\top Q \hat{x}\right\|^2 \\
&= \|\hat{x}\|^2 + 2 \cdot \|\hat{x}\| \cdot \left\|Z(Z^\top Q Z)^\dagger Z^\top Q \hat{x}\right\| + \left\|Z(Z^\top Q Z)^\dagger Z^\top Q \hat{x}\right\|^2 \\
&\leq \|\hat{x}\|^2 + 2 \cdot \|\hat{x}\|^2 \cdot \left(\frac{q_{\text{max}}}{q_{\text{min}}}\right) + \|\hat{x}\|^2 \cdot \left(\frac{q_{\text{max}}}{q_{\text{min}}}\right)^2 \\
&= \|\hat{x}\|^2 \cdot \left(1 + \frac{q_{\text{max}}}{q_{\text{min}}}\right)^2 \\
&= \left(\frac{\|b\|}{\sigma_{\text{min}}(A)}\right)^2 \left(1 + \frac{q_{\text{max}}}{q_{\text{min}}}\right)^2 \leq M
\end{aligned}$$

where $q_{\text{max}} = \lambda_{\text{max}}(Q)$ and q_{min} is the smallest non-zero eigenvalue of Q.

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.