

# 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

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

$$\tag{2}$$

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

$$\begin{aligned} & \text{minimize} && \frac{1}{2}x^\top Qx + \lambda x^\top x \\ & \text{subject to} && Ax = b \end{aligned} \tag{3}$$

$$\tag{4}$$

### 2.2 Karush-Kuhn-Tucker Conditions

The Lagrangian for (1) is of form

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

for which the optimality conditions [1] are

$$Qx^* + A^\top p^* = b, \quad Ax^* = b$$

or more compactly expressed:

$$\begin{bmatrix} Q & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ p^* \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} \quad (5)$$

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

$$\begin{bmatrix} Q + \lambda I & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x_\lambda^* \\ p_\lambda^* \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} \quad (6)$$

## 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

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

and so through subtracting equations,

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

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  $S = -\Lambda^{-1}A^\top$ , and

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

In other words:

$$x^* - x_\lambda^* = (\Lambda^{-1} - \Lambda^{-1}A^\top(A\Lambda^{-1}A^\top)^{-1}A\Lambda^{-1})\lambda x^*$$

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

$$x^* - x_\lambda^* = (\Gamma - \Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma)\lambda x^*$$

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

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

Then for  $\lambda \leq q_n$ ,

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

*Proof.* First note that

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

We seek the bound the RHS of the inequality

$$\|x^* - x_\lambda^*\| \leq \|\Gamma - \Gamma A^\top(A\Gamma A^\top)^{-1}A\Gamma\| \cdot \lambda \|x^*\|$$

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

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

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

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

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

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

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

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

□

## 4 Optimal Value Deviation

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

$$\|\Pi_{\ker Q} x^*\| \gg \|\Pi_{\ker Q} x_\lambda^*\|$$

where  $\Pi_{\ker Q}$  is the projection onto  $\ker Q$ ; this is possible because the regularized problems would force  $\Pi_{\ker Q} x_\lambda^*$  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

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

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

**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*

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

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

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

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

$$x^* = \hat{x} - Zu, \quad x_\lambda^* = \hat{x} - Zy$$

where

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

which can be derived from examining the optimality conditions of the quadratic problem. However, because  $\hat{x} \in \text{ran } A^\top = (\ker A)^\perp$ , 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  $x_\lambda^\star$  and  $x^\star$ , we have

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

Which simplifies to

$$\frac{1}{2}(x_\lambda^\star)^\top Q x_\lambda^\star - \frac{1}{2}(x^\star)^\top Q x^\star \leq \left(\|Q^{1/2}\hat{x}\| + \frac{1}{2}\|Q^{1/2}Zu\| + \frac{1}{2}\|Q^{1/2}Zy\|\right) \cdot \|Q^{1/2}Z(y - u)\| \quad (7)$$

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

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

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

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

With this, we now bound the terms that appear in (7). For the easy one:

$$\|Q^{1/2}\hat{x}\| = \|Q^{1/2}A^\dagger b\| \leq \lambda_{\max}(Q)^{1/2} \frac{1}{\sigma_{\min}(A)} \|b\| \quad (8)$$

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

$$\|Q^{1/2}Zu\| = \|Q^{1/2}ZV^\dagger Z^\top QA^\dagger b\| \leq \lambda_{\max}(Q)^{3/2} \frac{1}{q_0 \sigma_{\min}(A)} \|b\| \quad (9)$$

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

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

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

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

Putting these together, we find that

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

□

## 5 Interior Point Inspired

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

Consider an optimization problem of form

$$\begin{aligned}
& \text{minimize} && f(x) + \frac{1}{t}\phi(x) \\
& \text{subject to} && Ax = b
\end{aligned} \tag{10}$$

where

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

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

**Theorem 4.** *Let  $p^*$  be the optimal value for the original (quadratic linear equality constrained) problem and  $x^*(t)$  be the optimal (central-path) solutions to (10) as a function of  $t$ , then the optimality gap is*

$$f(x^*(t)) - p^* \leq \frac{M - \|x^*(t)\|^2}{t}, \quad M = \kappa(Q + (1/t)I)^2 \cdot \|A^\dagger b\|^2 + \varepsilon$$

for small  $\varepsilon > 0$ .

*Proof.* Based on the formulation of (10), the optimality conditions are

$$\nabla f(x) + \frac{1}{t}\nabla\phi(x) + A^\top\nu = 0, \quad Ax = b$$

which we identify with the Lagrangian of the following problem:

$$\text{minimize } f(x), \quad \text{subject to } Ax = b, \quad \|x\|^2 \leq M \quad (11)$$

with Lagrangian

$$L(x, \lambda, \nu) = f(x) + \lambda(\|x\|^2 - M) + \nu^\top(Ax - b)$$

which suggests the mapping  $\lambda \mapsto 1/t$ .

Strong duality holds for (11) since the problem is strictly feasible. This is because  $\hat{x} = A^\dagger b$  is the least-norm solution of  $Ax = b$ , and examining the ellipsoidal level sets of  $x \mapsto x^\top(Q + (1/t)I)x$ : major axis at  $\hat{x}$  has radius no more than  $\|\hat{x}\| \cdot \kappa(Q + (1/t)I)$ . As the optimal solution  $x^*(t)$  must be at a *smaller* level set (given this is a minimization problem), it likewise has a norm strictly smaller than  $\|\hat{x}\| \cdot \kappa(Q + (1/t)I) + \delta$  for any small perturbation  $\delta > 0$ .

Because strong duality holds, letting  $p^*$  be the optimal solution to (10) and  $g(\lambda, \nu)$  the Lagrange dual function of (11),

$$f(x^*(t), \lambda, \nu^*) = f(x^*(t)) + \lambda(\|x^*(t)\|^2 - M) \leq g(\lambda, \nu^*) \leq p^*$$

and taking  $\lambda \mapsto 1/t$  this simplifies to the desired result

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

□

## 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.