Quadratic Programming with Equality Constraints

Lei Zhao, Wu-Sheng Lu

The standard form of Quadratic Programming (CP) with equality constraints:

$$\begin{aligned}&\textbf{minimize} \quad \frac{1}{2}\boldsymbol{x^T H x} + \boldsymbol{x^T p} + k \\ &\textbf{subject to: } \boldsymbol{Ax} = \boldsymbol{b}\end{aligned} \tag{1}$$

where $\boldsymbol{A} \in R^{p\times n}$, $\boldsymbol{x} \in R^{n\times 1}$. We assume that the Hessian $\boldsymbol{H}$ is symmetric and positive semi-definite, Rank($\boldsymbol{A}$) $= p$, $p < n$. Because the constraints are linear equations, it gives us the idea that we can search for the optimal solution $\boldsymbol{x}^* \in R^{n\times 1}$ from all the solutions defined by the constraint equations. We can transform the wanted variable $\boldsymbol{x} \in R^{n\times 1}$ to another variable $\hat{\boldsymbol{\varphi}} \in R^{n\times 1}$, and then we can easily reduce several elements in $\hat{\boldsymbol{\varphi}}$ and only focus on a variable $\boldsymbol{\varphi} \in R^{(n-p)\times 1}$ with smaller dimensions.

To get the solutions of

$$\boldsymbol{Ax} = \boldsymbol{b} \tag{2}$$

we need to get the Moore-Penrose pseudo-inverse of $\boldsymbol{A}$ which is denoted by $\boldsymbol{A}^+$. Because we need to have infinite feasible points in the feasible region defined by these linear equations, matrix $\boldsymbol{A}$ is always flat, which means $p < n$, all the rows in $\boldsymbol{A}$ are also linearly independent, i.e., Rank($\boldsymbol{A}$) $= p$, we call $\boldsymbol{A}$ is full row rank, the pseudo-inverse $\boldsymbol{A}^+ \in R^{n\times p}$ is a right inverse of $\boldsymbol{A}$:

$$\begin{aligned}&\boldsymbol{A}^+ = \boldsymbol{A^T}(\boldsymbol{AA^T})^{-1} \\ &\boldsymbol{AA}^+ = \boldsymbol{AA^T}(\boldsymbol{AA^T})^{-1} = \boldsymbol{I_p}\end{aligned} \tag{3}$$

All the feasible points defined by the linear equality constraint can be characterized by

$$\boldsymbol{x} = \boldsymbol{A}^+\boldsymbol{b} + (\boldsymbol{I_n} - \boldsymbol{A}^+\boldsymbol{A})\hat{\boldsymbol{\varphi}} \tag{4}$$

where $\hat{\boldsymbol{\varphi}} \in R^{n\times 1}$ is an arbitrary n-dimensional parameter vector. Because $\boldsymbol{A}, \boldsymbol{b}$ are constant, we have transformed the variable from $\boldsymbol{x}$ to new variable $\hat{\boldsymbol{\varphi}}$. Then, based on SVD of $\boldsymbol{A}$, we can continue to transform $\hat{\boldsymbol{\varphi}}$ to a lower dimensional variable $\boldsymbol{\varphi} \in R^{(n-p)\times 1}$.

$$\boldsymbol{A} = \boldsymbol{U\Sigma V^T} \tag{5}$$

where $\boldsymbol{U} \in R^{p\times p}$, $\boldsymbol{V} \in R^{n\times n}$, and $\boldsymbol{\Sigma} \in R^{p\times n}$, which can be denoted by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{S} & \boldsymbol{0} \end{bmatrix} \quad \text{and} \quad \boldsymbol{S} = \text{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_p\} \tag{6}$$

The pseudo-inverse $\boldsymbol{A}^+$ can be denoted by

$$\boldsymbol{A}^+ = \boldsymbol{A^T}(\boldsymbol{AA^T})^{-1} = \boldsymbol{V\Sigma^T U^T}(\boldsymbol{U\Sigma\Sigma^T U^T})^{-1} = \boldsymbol{V\Sigma^T}(\boldsymbol{\Sigma\Sigma^T})^{-1}\boldsymbol{U^T} = \boldsymbol{V}\begin{bmatrix} \boldsymbol{S}^{-1} \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{U^T} \tag{7}$$

Thus,

$$\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{V}\begin{bmatrix} \boldsymbol{S}^{-1} \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{U^T A} = \boldsymbol{V}\begin{bmatrix} \boldsymbol{S}^{-1} \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{U^T U}\begin{bmatrix} \boldsymbol{S} & \boldsymbol{0} \end{bmatrix}\boldsymbol{V^T} = \boldsymbol{V}\begin{bmatrix} \boldsymbol{I_p} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}\boldsymbol{V^T} \tag{8}$$

and

$$\boldsymbol{I_n} - \boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{I_n} - \boldsymbol{V}\begin{bmatrix} \boldsymbol{I_p} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}\boldsymbol{V^T} = \boldsymbol{V}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I_{n-p}} \end{bmatrix}\boldsymbol{V^T} = \sum_{\boldsymbol{i=p+1}}^{\boldsymbol{n}} \boldsymbol{v_i v_i^T} = \boldsymbol{V_r V_r^T} \tag{9}$$

where $\boldsymbol{V_r} = [\boldsymbol{v_{p+1}}, \boldsymbol{v_{p+2}}, \boldsymbol{\cdots}, \boldsymbol{v_n}]$. All feasible points defined by the linear equality constraints becomes

$$\boldsymbol{x} = \boldsymbol{A^+ b} + \boldsymbol{V_r V_r^T}\hat{\boldsymbol{\varphi}} = \boldsymbol{A^+ b} + \boldsymbol{V_r\varphi} = \boldsymbol{x_s} + \boldsymbol{V_r\varphi} \tag{10}$$

where we define

$$\begin{aligned} \boldsymbol{\varphi} &= \boldsymbol{V_r^T}\hat{\boldsymbol{\varphi}} \in R^{(n-p)\times 1} \\ \boldsymbol{x_s} &= \boldsymbol{A^+ b} \end{aligned} \tag{11}$$

We put $\boldsymbol{x} = \boldsymbol{x_s} + \boldsymbol{V_r\varphi}$ back to the original objective function, the constrained QP problem w.r.t. $\boldsymbol{x} \in R^{n\times 1}$ becomes an unconstrained QP problem w.r.t. $\boldsymbol{\varphi} \in R^{(n-p)\times 1}$.

$$\begin{aligned} \textbf{minimize} \quad & \frac{1}{2}(\boldsymbol{x_s} + \boldsymbol{V_r\varphi})^{\boldsymbol{T}}\boldsymbol{H}(\boldsymbol{x_s} + \boldsymbol{V_r\varphi}) + (\boldsymbol{x_s} + \boldsymbol{V_r\varphi})^{\boldsymbol{T}}\boldsymbol{p} + k \\ &= \frac{1}{2}\boldsymbol{\varphi^T V_r^T H V_r \varphi} + \boldsymbol{\varphi^T}(\boldsymbol{V_r^T H x_s} + \boldsymbol{V_r^T p}) + \frac{1}{2}\boldsymbol{x_s^T H x_s} + \boldsymbol{p^T x_s} + k \\ &= \frac{1}{2}\boldsymbol{\varphi^T}\hat{\boldsymbol{H}}\boldsymbol{\varphi} + \boldsymbol{\varphi^T}\hat{\boldsymbol{p}} + \hat{k} \end{aligned} \tag{12}$$

where

$$\boldsymbol{\hat{H}} = \boldsymbol{V_r^T H V_r}$$
$$\boldsymbol{\hat{p}} = \boldsymbol{V_r^T H x_s} + \boldsymbol{V_r^T p} \tag{13}$$
$$\hat{k} = \frac{1}{2}\boldsymbol{x_s^T H x_s} + \boldsymbol{p^T x_s} + k$$

Because $\boldsymbol{V_r} \in R^{n\times(n-p)}$, thus $\boldsymbol{\hat{H}} \in R^{(n-p)\times(n-p)}$. We shrink the corresponding Hessian matrix after transforming the variable $\boldsymbol{x} \in R^{n\times 1}$ to $\boldsymbol{\varphi} \in R^{(n-p)\times 1}$. Now this problem becomes an unconstrained problem w.r.t. $\boldsymbol{\varphi}$, to get the optimal solution, we need to compute the gradient and set it to $\mathbf{0}$:

$$\nabla_{\varphi}(\frac{1}{2}\boldsymbol{\varphi^T \hat{H} \varphi} + \boldsymbol{\varphi^T \hat{p}} + \hat{k}) = \boldsymbol{\hat{H}\varphi} + \boldsymbol{\hat{p}} = \mathbf{0} \tag{14}$$

If $\boldsymbol{H}$ is P.D., then $\boldsymbol{\hat{H}}$ is also P.D, that means all the eigenvalues are positive and they are invertible, we can get the unique global minimizer:

$$\boldsymbol{x^*} = \boldsymbol{x_s} + \boldsymbol{V_r \varphi^*} = \boldsymbol{A^+ b} + \boldsymbol{V_r \varphi^*} \tag{15}$$

where

$$\boldsymbol{\varphi^*} = -\boldsymbol{\hat{H}^{-1}\hat{p}} \tag{16}$$

If $\boldsymbol{H}$ is P.S.D, then $\boldsymbol{\hat{H}}$ maybe P.S.D, they are not invertible, how to compute $\boldsymbol{\varphi}^*$? By setting the gradient to $\mathbf{0}$, actually we are trying to solve

$$\boldsymbol{\hat{H}\varphi^*} = -\boldsymbol{\hat{p}} \tag{17}$$

If vector $\boldsymbol{\hat{p}}$ locates in the space spanned by the columns of $\boldsymbol{\hat{H}}$, i.e., $\boldsymbol{\hat{p}}$ can be expressed as a linear combination of the columns of $\boldsymbol{\hat{H}}$, then, there are infinite global minimizers. Otherwise, if $\boldsymbol{\hat{p}}$ can not be represented by the columns of $\boldsymbol{\hat{H}}$, there will be no minimizer.

We also could use the SVD on $\boldsymbol{\hat{H}}$ to explain what are the solutions when $\boldsymbol{\hat{H}}$ is P.S.D. We use $\boldsymbol{\Lambda}$ to denote the diagonal matrix that contains all the non-zero eigenvalues of $\boldsymbol{\hat{H}}$. Since $\boldsymbol{\hat{H}}$ is square, its SVD goes like:

$$\boldsymbol{\hat{H}} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\boldsymbol{U^T} \tag{18}$$

And the vectors $\boldsymbol{\varphi}$ and $\boldsymbol{\hat{p}}$ can also be separated into two parts corresponding to the non-zero eigenvalues and zero eigenvalues:

$$\boldsymbol{\varphi} = \begin{bmatrix} \boldsymbol{\varphi}_1 \\ \boldsymbol{\varphi}_2 \end{bmatrix} \quad \hat{\boldsymbol{p}} = \begin{bmatrix} \hat{\boldsymbol{p}}_1 \\ \hat{\boldsymbol{p}}_2 \end{bmatrix} \tag{19}$$

Then, the unconstrained objective function becomes to

$$\textbf{minimize} \quad \frac{1}{2}\boldsymbol{\varphi}^{\boldsymbol{T}}\boldsymbol{U} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{\varphi} + \boldsymbol{\varphi}^{\boldsymbol{T}}\hat{\boldsymbol{p}} + \hat{k} \tag{20}$$

and we can continue to transform the variable:

$$\bar{\boldsymbol{\varphi}} = \boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{\varphi} = \boldsymbol{U}^{\boldsymbol{T}} \begin{bmatrix} \boldsymbol{\varphi}_1 \\ \boldsymbol{\varphi}_2 \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{\varphi}}_1 \\ \bar{\boldsymbol{\varphi}}_2 \end{bmatrix} \quad \bar{\boldsymbol{p}} = \boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{p} = \boldsymbol{U}^{\boldsymbol{T}} \begin{bmatrix} \hat{\boldsymbol{p}}_1 \\ \hat{\boldsymbol{p}}_2 \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{p}}_1 \\ \bar{\boldsymbol{p}}_2 \end{bmatrix} \tag{21}$$

the objective function becomes to

$$\begin{aligned} \textbf{minimize} \quad & \frac{1}{2}\bar{\boldsymbol{\varphi}}^{\boldsymbol{T}} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \bar{\boldsymbol{\varphi}} + \bar{\boldsymbol{\varphi}}^{\boldsymbol{T}}\bar{\boldsymbol{p}} + \hat{k} \\ & = \frac{1}{2}[\bar{\boldsymbol{\varphi}}_{\boldsymbol{1}}^{\boldsymbol{T}} \ \bar{\boldsymbol{\varphi}}_{\boldsymbol{2}}^{\boldsymbol{T}}] \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\varphi}}_1 \\ \bar{\boldsymbol{\varphi}}_2 \end{bmatrix} + [\bar{\boldsymbol{\varphi}}_{\boldsymbol{1}}^{\boldsymbol{T}} \ \bar{\boldsymbol{\varphi}}_{\boldsymbol{2}}^{\boldsymbol{T}}] \begin{bmatrix} \bar{\boldsymbol{p}}_1 \\ \bar{\boldsymbol{p}}_2 \end{bmatrix} + \hat{k} \\ & = \frac{1}{2}\bar{\boldsymbol{\varphi}}_{\boldsymbol{1}}^{\boldsymbol{T}}\boldsymbol{\Lambda}\bar{\boldsymbol{\varphi}}_{\boldsymbol{1}} + \bar{\boldsymbol{\varphi}}_{\boldsymbol{1}}^{\boldsymbol{T}}\bar{\boldsymbol{p}}_{\boldsymbol{1}} + \bar{\boldsymbol{\varphi}}_{\boldsymbol{2}}^{\boldsymbol{T}}\bar{\boldsymbol{p}}_{\boldsymbol{2}} + \hat{k} \end{aligned} \tag{22}$$

The trick part is in $\bar{\boldsymbol{\varphi}}_{\boldsymbol{2}}^{\boldsymbol{T}}\bar{\boldsymbol{p}}_{\boldsymbol{2}}$, if we regard this term as constant, we can obtain an unique global minimizer $\bar{\boldsymbol{\varphi}}_{\boldsymbol{1}}^{*}$ by setting the gradient to $\boldsymbol{0}$:

$$\bar{\boldsymbol{\varphi}}_{\boldsymbol{1}}^{*} = -\boldsymbol{\Lambda}^{-\boldsymbol{1}}\bar{\boldsymbol{p}}_{\boldsymbol{1}} \tag{23}$$

However, there is no constraint on $\bar{\boldsymbol{\varphi}}_{\boldsymbol{2}}$, this linear part could go to infinity, which leads to no minimizer:

$$-\infty \leq \bar{\boldsymbol{\varphi}}_{\boldsymbol{2}}^{\boldsymbol{T}}\bar{\boldsymbol{p}}_{\boldsymbol{2}} \leq +\infty \tag{24}$$

This explanation only wants to show what the solution looks like when H is P.S.D.