Convex Quadratic Programming Problem

Asst. Prof. Dr.-Ing. Sudchai Boonto

November 22, 2023

Department of Control System and Instrumentation Engineering
King Mongkut's Unniversity of Technology Thonburi
Thailand

Objective

- Understand the Quadratic Programming (QP)

Singular-Value Decomposition (SVD)

Given a matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ of rank r, there exist unitary matrices $\mathbf{U} \in \mathbb{C}^{m \times m}$ and $\mathbf{V} \in \mathbb{C}^{n \times n}$ such that

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^H, \text{ where } \Sigma = \begin{bmatrix} \mathbf{S} & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}, \quad \text{and } \mathbf{S} = \mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_r\},$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

- The matrix decomposition $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^H$ is known as the **singular-value decomposition (SVD)** of $\mathbf{A}$. $\{\cdot\}^H$ is a complex conjugate transpose. If $\mathbf{A}$ is a real-values matrix, then $\mathbf{U}$ and $\mathbf{V}$ become orthogoanl matrices and $\mathbf{V}^H$ becomes $\mathbf{V}^T$.
- The positive scalars σ_i for $i = 1, 2, \ldots, r$ are called the **singular values** of $\mathbf{A}$.
- If $\mathbf{U} = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$ and $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$, vectors u_i and v_i are called the **left** and **right singular vectors** of $\mathbf{A}$. We have

$$\mathbf{A}\mathbf{A}^H = \mathbf{U} \begin{bmatrix} \mathbf{S}^2 & 0 \\ 0 & 0 \end{bmatrix}_{m \times m} \mathbf{U}^H, \quad \mathbf{A}^H\mathbf{A} = V \begin{bmatrix} \mathbf{S}^2 & 0 \\ 0 & 0 \end{bmatrix}_{n \times n} \mathbf{V}^H$$

- See `lecture11/svd_ex.jl`.

The Moore-Penrose pseudo-inverse

The Moore-Penrose pseudo-inverse of a matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ is defined as the matrix $\mathbf{A}^\dagger \in \mathbb{C}^{n \times m}$ that satisfies the following four conditions:

1. $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$
2. $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$
3. $(\mathbf{A}\mathbf{A}^\dagger)^H = \mathbf{A}\mathbf{A}^\dagger$
4. $(\mathbf{A}^\dagger\mathbf{A})^H = \mathbf{A}^\dagger\mathbf{A}$

The example of $\mathbf{A}^\dagger$ is $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. The Moore-Penrose pseudo-inverse of $\mathbf{A}$ can be obtained as

$$\mathbf{A}^\dagger = \mathbf{V}\Sigma^\dagger\mathbf{U}^H, \quad \Sigma^\dagger = \begin{bmatrix} \mathbf{S}^{-1} & 0 \\ 0 & 0 \end{bmatrix}_{n \times m}, \text{ and } \mathbf{S}^{-1} = \mathrm{diag}\{\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_r^{-1}\}$$

$$\mathbf{A}^\dagger = \sum_{i=1}^{r} \frac{v_i u_i^H}{\sigma_i}$$

The Moore-Penrose pseudo-inverse

For an underdetermined system of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \Longrightarrow \mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b} + \mathbf{V}_r\boldsymbol{\phi}$$

Given

$$\mathbf{A} = \begin{bmatrix} 2.8284 & -1 & 1 \\ 2.8284 & 1 & -1 \end{bmatrix}, \quad \mathbf{A}^T\mathbf{A} = \begin{bmatrix} 16 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & -2 & 2 \end{bmatrix} = \mathbf{V}\begin{bmatrix} 16 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}\mathbf{V}^T$$

The nonzero singular values of $\mathbf{A}$ are $\sigma_1 = \sqrt{16} = 4$ and $\sigma_2 = \sqrt{4} = 2$. By using a command `[U1, S1, V1] = svd(A'*A)` of Matlab or `U1, S1, V1 = svd(A'*A)` of Julia to obtain S and `V= V1`. Then using `[U2, S2, V2] = svd(A)` of Matlab or `U2, S2, V2 = svd(A)` of Julia to get `U=U2`.

The Moore-Penrose pseudo-inverse

We obtain

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.7071 & -0.7071 \\ 0 & -0.7071 & -0.7071 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix},$$

$$\mathbf{U} = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \end{bmatrix}$$

$$\mathbf{A}^\dagger = \frac{v_1 u_1^T}{\sigma_1} + \frac{v_2 u_2^T}{\sigma_2} = \begin{bmatrix} 0.1768 & 0.1768 \\ -0.2500 & 0.2500 \\ 0.2500 & -0.2500 \end{bmatrix}$$

QP Problem

Consider a problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{x}^T\mathbf{p} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}, \qquad \mathbf{A} \in \mathbb{R}^{p\times n} \end{aligned}$$

The solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{V}_r\boldsymbol{\phi} + \mathbf{A}^\dagger\mathbf{b}$, where ϕ is an arbitrary r-dimensional vector with $r = n - p$, $\mathbf{A}^\dagger$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{A}$, $\mathbf{V}_r$ is composed of the last r columns of $\mathbf{V}$, and $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$. Then

$$\underset{\boldsymbol{\phi}}{\text{minimize}} = \frac{1}{2}\boldsymbol{\phi}^T\hat{\mathbf{H}}\boldsymbol{\phi} + \boldsymbol{\phi}^T\hat{\mathbf{p}},$$

where $\hat{\mathbf{h}} = \mathbf{V}_r^T\mathbf{H}\mathbf{V}_r$ and $\hat{\mathbf{p}} = \mathbf{V}_r^T\left(\mathbf{H}\mathbf{A}^\dagger\mathbf{b} + \mathbf{p}\right)$
The global minimizer of the problem is $\mathbf{x}^* = \mathbf{V}_r\boldsymbol{\phi}^* + \mathbf{A}^\dagger\mathbf{b}$, where $\boldsymbol{\phi}^*$ is the solution of the linear system of equations $\hat{\mathbf{H}}\boldsymbol{\phi} = -\hat{\mathbf{p}}$.

QP Problem: Example

Solve the QP problem:

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}_1^2 + \mathbf{x}_2^2\right) + 2\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_3 \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

where $\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$, and $\mathbf{b} = 1$.

$$\mathbf{V}_r = \begin{bmatrix} -0.7071 & -0.707 \\ 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}, \qquad \mathbf{A}^T\left(\mathbf{A}^T\mathbf{A}\right)^{-1} = \begin{bmatrix} 0 & 0.5 & 0.5 \end{bmatrix}^T,$$

$$\hat{\mathbf{H}} = \mathbf{V}_r^T\mathbf{H}\mathbf{V}_r = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}, \quad \hat{\mathbf{p}} = \mathbf{V}^T\left(\mathbf{H}\mathbf{A}^\dagger\mathbf{b} + \mathbf{p}\right) = \begin{bmatrix} -0.2322 \\ -3.7677 \end{bmatrix}$$

$$\mathbf{x}^* = \mathbf{V}_r\boldsymbol{\phi}^* + \mathbf{A}^\dagger\mathbf{b} = \begin{bmatrix} -2 & -2 & 3 \end{bmatrix}^T$$

Quadratic Optimization Problems

The quadratic problem is

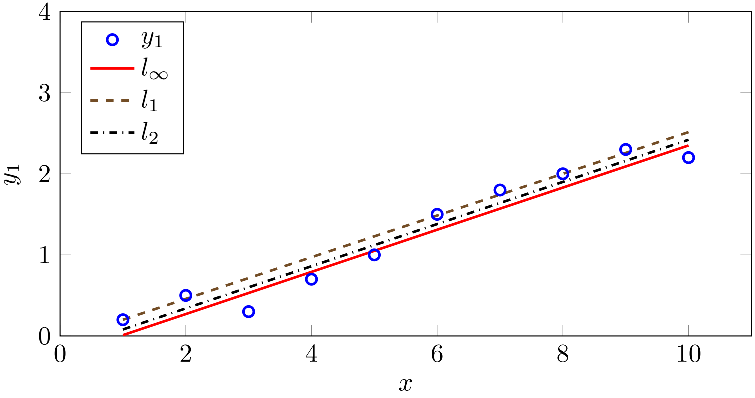
$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{p}^T\mathbf{x} + \mathbf{c} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{C}\mathbf{x} \leq \mathbf{d} \end{aligned}$$

Example: least-squares

$$\text{minimize} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

- analytical solution $\mathbf{x}^* = \mathbf{A}^\dagger\mathbf{b}$
- we can add linear constraints , e.g. $l \leq \mathbf{x} \leq u$

Quadratic Optimization Problems



$$\arg\operatorname{minimize} \|\hat{f}(k) - y_1(k)\|_2^2.$$

Quadratic Optimization Problems

Consider

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{x}^T\mathbf{p} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

Assume $\mathbf{A} \in \mathbb{R}^{p \times n}$ is of full row rank and $p < n$. By using the first-order necessary conditions $\nabla f_0(\mathbf{x}^*) + \mathbf{A}^T\boldsymbol{\lambda} = 0$, we have

$$\begin{aligned} \mathbf{H}\mathbf{x}^* + \mathbf{p} + \mathbf{A}^T\boldsymbol{\lambda}^* &= 0 \\ \mathbf{A}\mathbf{x}^* - \mathbf{b} &= 0 \end{aligned}$$

$$\begin{bmatrix} \mathbf{H} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} -\mathbf{p} \\ \mathbf{b} \end{bmatrix} \qquad (1)$$

If $\mathbf{H}$ is positive definite and $\mathbf{A}$ is of full row rank, then the system matrix (1) is nonsingular and the solution $\mathbf{x}^*$ is the unique global minimizer of the problem.

Quadratic Optimization Problems

The block matrix inversion lemma is

$$\begin{bmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{C} & \mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \Delta^{-1} & -\Delta^{-1}\mathbf{D}\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{C}\Delta^{-1} & \mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{C}\Delta^{-1}\mathbf{D}\mathbf{B}^{-1} \end{bmatrix}$$

where

$$\Delta = \mathbf{A} - \mathbf{D}\mathbf{B}^{-1}\mathbf{C}$$

Then we have

$$\begin{aligned} \boldsymbol{\lambda}^* &= -(\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{H}^{-1}\mathbf{p} + \mathbf{b}) \\ \mathbf{x}^* &= -\mathbf{H}^{-1}(\mathbf{A}^T\boldsymbol{\lambda}^* + \mathbf{p}) \end{aligned}$$

Sequential Quadratic Optimization Problems (SQP)

Equality Constrained SQP

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = \mathbf{x}^T \mathbf{H} \mathbf{x} + \mathbf{p}^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{h}(\mathbf{x}) = 0 \end{aligned}$$

The Lagrangian is $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(x)$. At the minimum point, we have

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) &= \nabla f(\mathbf{x}) + \mathbf{J}_h^T \boldsymbol{\lambda} = 0 \\ \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) &= h(\mathbf{x}) = 0 \end{aligned}$$

The quadratic approximation:

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x} + \Delta \mathbf{x}, \boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}) &= \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) + \nabla^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \Delta \mathbf{x} + \frac{\partial}{\partial \mathbf{x}} \left(\nabla_\lambda \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \right)^T \Delta \boldsymbol{\lambda} \\ &= \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) + \mathbf{H}_{\mathcal{L}} \Delta \mathbf{x} + \mathbf{J}_h^T \Delta \mathbf{x} \end{aligned}$$

Sequential Quadratic Optimization Problems (SQP)

$$\begin{aligned}
\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\mathbf{x}+\Delta\mathbf{x},\boldsymbol{\lambda}+\Delta\boldsymbol{\lambda}) &= \nabla_{\boldsymbol{\lambda}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) + \nabla_{\boldsymbol{\lambda}}\left(\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda})\right)\Delta\mathbf{x} + \nabla^2_{\mathbf{x}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda})\Delta\boldsymbol{\lambda} \\
&= h(\mathbf{x}) + \mathbf{J}_h\Delta\mathbf{x}
\end{aligned}$$

$$\begin{bmatrix} \mathbf{H}_{\mathcal{L}} & \mathbf{J}_h^T \\ \mathbf{J}_h & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) \\ -h(\mathbf{x}) \end{bmatrix}$$

where

$$\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k\Delta\mathbf{x}_k \\
\boldsymbol{\lambda}_{k+1} &= \boldsymbol{\lambda}_k + \Delta\boldsymbol{\lambda}_k
\end{aligned}$$

By using slack variables,

$$\begin{aligned} \text{minimize} \quad & f(\tilde{\mathbf{x}}) = \frac{1}{2}\tilde{\mathbf{x}}^T H \tilde{\mathbf{x}} + \tilde{\mathbf{x}}^T \mathbf{p} \\ \text{subject to} \quad & \mathbf{A}\tilde{\mathbf{x}} \leq \mathbf{b} \end{aligned}$$

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{H}\mathbf{x} + \mathbf{x}^T \mathbf{p} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & -\mathbf{x} \leq 0 \end{aligned} \tag{2}$$

We have $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{x}^T\mathbf{p} - \boldsymbol{\mu}^T\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$ and $\mathbf{g}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{\lambda},\boldsymbol{\mu}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ if and only if $\nabla\mathcal{L} = 0$ or $-\mathbf{A}^T\boldsymbol{\lambda} + \boldsymbol{\mu} - \mathbf{H}\mathbf{x} = \mathbf{p}$ The dual problem becomes

$$\begin{aligned} \text{maximize} \quad & h(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = -\frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} - \boldsymbol{\lambda}^T\mathbf{b} \\ \text{subject to} \quad & -\mathbf{A}^T\boldsymbol{\lambda} + \boldsymbol{\mu} - \mathbf{H}\mathbf{x} = \mathbf{p} \\ & \boldsymbol{\mu} \geq 0 \end{aligned} \tag{3}$$

Interior-Point Methods for Convex QP Problems

The KKT conditions

$$\begin{aligned} \mathbf{A}\mathbf{x} - \mathbf{b} &= 0 \text{ for } \mathbf{x} \geq 0 \\ -\mathbf{A}^T\boldsymbol{\lambda} + \boldsymbol{\mu} - \mathbf{H}\mathbf{x} - \mathbf{p} &= 0 \text{ for } \boldsymbol{\mu} \geq 0 \\ \mathbf{X}\boldsymbol{\mu} &= 0, \end{aligned} \tag{4}$$

where $\mathbf{X} = \mathrm{diag}\{x_1, x_2, \ldots, x_n\}$.
Let $W = \{\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}\}$ be a feasible for the problems (2). The duality gap can be obtained for $\{\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}\}$ as

$$\begin{aligned} \delta(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - h(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{x}^T\mathbf{p} + \boldsymbol{\lambda}^T\mathbf{b} \\ &= \mathbf{x}^T(-\mathbf{A}^T\boldsymbol{\lambda} + \boldsymbol{\mu}) + \boldsymbol{\lambda}^T\mathbf{b} = \mathbf{x}^T\boldsymbol{\mu} \end{aligned} \tag{5}$$

Setting $\mathbf{w}(\tau) = \{\mathbf{x}(\tau), \boldsymbol{\lambda}(\tau), \boldsymbol{\mu}(\tau)\}$ that satisfies the KKT condition, the last line of (4) is changed to

$$\mathbf{X}\boldsymbol{\mu} = \tau\mathbf{e}$$
$$\mathbf{e} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T \in \mathbb{R}^n$$
$$\delta[\mathbf{x}(\tau), \boldsymbol{\lambda}(\tau), \boldsymbol{\mu}(\tau)] = \mathbf{x}^T(\tau)\boldsymbol{\mu}(\tau) = n\tau$$

Hence the duality gap approaches zero linearly as $\tau \to 0$.

Let $\mathbf{w}_k = \{\mathbf{x}_k, \boldsymbol{\lambda}_k, \boldsymbol{\mu}_k\}$ be such that $\mathbf{x}_k$ is strictly feasible for the primal problem (2) and $\{\boldsymbol{\lambda}_k, \boldsymbol{\mu}_k\}$ is strictly feasible for the dual problem (3). The increment set is $\boldsymbol{\delta}_w = \{\boldsymbol{\delta}_x, \boldsymbol{\delta}_\lambda, \boldsymbol{\delta}_\mu\}$. We need $\mathbf{w}_{k+1} = \{\mathbf{x}_{k+1}, \boldsymbol{\lambda}_{k+1}, \boldsymbol{\mu}_{k+1}\} = \mathbf{w}_k + \boldsymbol{\delta}_w$ remains strictly feasible and approaches the central path.

If w_k satisfy the KKT and central path condition with τ_{k+1}, we have

$$\begin{aligned} \mathbf{A}(\mathbf{x}_k + \boldsymbol{\delta}_x) - \mathbf{b} &= 0 \\ -\mathbf{A}^T(\boldsymbol{\lambda}_k + \boldsymbol{\delta}_{\boldsymbol{\lambda}}) + (\boldsymbol{\mu}_k + \boldsymbol{\delta}_\mu) - \mathbf{H}(\mathbf{x}_k + \boldsymbol{\delta}_x) - \mathbf{p} &= 0 \\ (\mathbf{X} + \Delta\mathbf{X})(\boldsymbol{\mu}_k + \boldsymbol{\delta}_\mu) &= \tau_{k+1}\mathbf{e} \end{aligned}$$

$$\begin{aligned} \mathbf{A}\boldsymbol{\delta}_x &= 0 \\ -\mathbf{H}\boldsymbol{\delta}_x - \mathbf{A}^T\boldsymbol{\delta}_\lambda + \boldsymbol{\delta}_\mu &= 0 \\ \Delta\mathbf{X}\boldsymbol{\mu}_k + \mathbf{X}\boldsymbol{\delta}_\mu + \Delta\mathbf{X}\boldsymbol{\delta}_\mu &= \tau_{k+1} - \mathbf{X}\boldsymbol{\mu}_k \\ \mathbf{M}\boldsymbol{\delta}_x + \mathbf{X}\boldsymbol{\delta}_\mu &= \tau_{k+1} - \mathbf{X}\boldsymbol{\mu}_k \end{aligned} \tag{6}$$

where $\Delta\mathbf{X} = \mathrm{diag}\{(\boldsymbol{\delta}_x)_1, (\boldsymbol{\delta}_x)_2, \ldots, (\boldsymbol{\delta}_x)_n\}$, $\mathbf{M} = \mathrm{diag}\{(\boldsymbol{\mu}_k)_1, (\boldsymbol{\mu}_k)_2, \ldots, (\boldsymbol{\mu}_k)_n\}$, and $\Delta\mathbf{X}\boldsymbol{\delta}_\mu$ is neglected.

Solving (6) we have

$$\begin{aligned}\boldsymbol{\delta}_\lambda &= -Yy \\ \boldsymbol{\delta}_x &= -\Gamma\mathbf{X}\mathbf{A}^T\boldsymbol{\delta}_\lambda - y \\ \boldsymbol{\delta}_\mu &= \mathbf{H}\boldsymbol{\delta}_x + \mathbf{A}^T\boldsymbol{\delta}_\lambda\end{aligned} \tag{7}$$

where $\Gamma = (\mathbf{M} + \mathbf{XH})^{-1}, \mathbf{Y} = (\mathbf{A}\Gamma\mathbf{X}\mathbf{A}^T)^{-1}\mathbf{A}$ and $y = \Gamma(\mathbf{X}\boldsymbol{\mu}_k - \tau_{k+1}\mathbf{e})$

Since $\mathbf{x}_k > 0$ and $\boldsymbol{\mu}_k > 0$, matrices $\mathbf{X}$ and $\mathbf{M}$ are positive definite. Therefore $\mathbf{X}^{-1}\mathbf{M} + \mathbf{H}$ is also positive definite and the inverse of matrix

$$\mathbf{M} + \mathbf{XH} = \mathbf{X}(\mathbf{X}^{-1}\mathbf{M} + \mathbf{H})$$

Since $\mathbf{A}$ is of full row rank, $\mathbf{A}\Gamma\mathbf{X}\mathbf{A}^T = \mathbf{A}(\mathbf{X}^{-1}\mathbf{M} + \mathbf{H})^{-1}\mathbf{A}^T$

Interior-Point Methods for Convex QP Problems

Primal-dual path-following algorithm for the Convex QP problem

1. Input $\mathbf{A}, \mathbf{b}, \mathbf{c}$, and $\mathbf{w}_0 = \{\mathbf{x}_0, \boldsymbol{\lambda}_0, \boldsymbol{\mu}_0\}$. Set $k = 0$ and $\rho > \sqrt{n}$ (n is a dimension of $\mathbf{x}$), and initialize the tolerance ε for the duality gap.
2. If $\boldsymbol{\mu}_k^T x_k \leq \varepsilon$, output solution $\mathbf{w}^* = \mathbf{w}_k$ and stop; otherwise, continue with Step 3
3. Set $\tau_{k+1} = \frac{\boldsymbol{\mu}_k^T \mathbf{x}_k}{n+\rho}$ and compute $\boldsymbol{\delta}_w = \{\boldsymbol{\delta}_x, \boldsymbol{\delta}_\lambda, \boldsymbol{\delta}_\mu\}$ using (7).
4. compute step size α_k as follow:

$$\alpha_k = (1 - 10^{-6})\alpha_{\max} \quad \alpha_{\max} = \min(\alpha_p, \alpha_d)$$

where

$$\alpha_p = \min_{i \text{ with } (\boldsymbol{\delta}_x)_i < 0} \left[-\frac{(\mathbf{x}_k)_i}{(\boldsymbol{\delta}_x)_i} \right], \quad \alpha_d = \min_{i \text{ with } (\boldsymbol{\delta}_\mu)_i < 0} \left[-\frac{(\boldsymbol{\mu}_k)_i}{(\boldsymbol{\delta}_\mu)_i} \right]$$

- The previous algorithm requires a strictly feasible $\mathbf{w}_0$, which might be difficult to obtain particularly for large-scale problems.
- Let $\mathbf{w}_k = \{\mathbf{x}_k, \boldsymbol{\lambda}_k, \boldsymbol{\mu}_k\}$ be such that $\mathbf{x}_k > 0$ and $\boldsymbol{\mu}_k > 0$ but may not satisfy the central path condition. We need to find the next iterate

$$\mathbf{w}_{k+1} = \mathbf{x}_k + \alpha_k \boldsymbol{\delta}_w$$

such that $\mathbf{x}_{k+1} > 0$ and $\boldsymbol{\mu}_{k+1} > 0$ and that $\boldsymbol{\delta}_w = \{\boldsymbol{\delta}_x, \boldsymbol{\delta}_\lambda, \boldsymbol{\delta}_\mu\}$ satisfies the equations

$$-\mathbf{H}(x_k + \boldsymbol{\delta}_x) - \mathbf{p} - \mathbf{A}^T(\boldsymbol{\lambda}_k + \boldsymbol{\delta}_\lambda) + (\boldsymbol{\mu}_k + \boldsymbol{\delta}_\mu) = 0$$
$$\mathbf{A}(\mathbf{x}_k + \boldsymbol{\delta}_x) = \mathbf{b}$$
$$\mathbf{M}\boldsymbol{\delta}_x + \mathbf{X}\boldsymbol{\delta}_\mu = \tau_{k+1}\mathbf{e} - \mathbf{X}\boldsymbol{\mu}_k$$

Interior-Point Methods for Convex QP Problems

$$
\begin{aligned}
-\mathbf{H}\boldsymbol{\delta}_x - \mathbf{A}^T\boldsymbol{\delta}_\lambda + \boldsymbol{\delta}_\mu &= \mathbf{r}_d \\
\mathbf{A}\boldsymbol{\delta}_x &= \mathbf{r}_p \\
\mathbf{M}\boldsymbol{\delta}_x + \mathbf{X}\boldsymbol{\delta}_\mu &= \tau_{k+1}\mathbf{e} - \mathbf{X}\boldsymbol{\mu}_k
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{r}_d &= \mathbf{H}\mathbf{x}_k + \mathbf{p} + \mathbf{A}^T\boldsymbol{\lambda}_k - \boldsymbol{\mu}_k \\
\mathbf{r}_p &= \mathbf{b} - \mathbf{A}\mathbf{x}_k
\end{aligned}
$$

The solution of $\mathbf{w}$ can be obtained as

$$
\begin{aligned}
\boldsymbol{\delta}_\lambda &= -\mathbf{Y}_0(\mathbf{A}\mathbf{y}_d + \mathbf{r}_p) \\
\boldsymbol{\delta}_x &= -\Gamma\mathbf{X}\mathbf{A}^T\boldsymbol{\delta}_\lambda - y_d \\
\boldsymbol{\delta}_\mu &= \mathbf{H}\boldsymbol{\delta}_x + \mathbf{A}^T\boldsymbol{\delta}_\lambda + \mathbf{r}_d
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
\Gamma &= (\mathbf{M} + \mathbf{X}\mathbf{H})^{-1} \\
\mathbf{Y}_0 &= (\mathbf{A}\Gamma\mathbf{X}\mathbf{A}^T)^{-1} \\
y_d &= \Gamma\left[\mathbf{X}(\boldsymbol{\mu}_k + \mathbf{r}_d) - \tau_{k+1}\mathbf{e}\right]
\end{aligned}
$$

Nonfeasible-initialization Primal-dual path-following algorithm for the Convex QP problem

1. Input $\mathbf{A}, \mathbf{b}, \mathbf{c}$, and $\mathbf{w}_0 = \{\mathbf{x}_0, \boldsymbol{\lambda}_0, \boldsymbol{\mu}_0\}$. Set $k = 0$ and $\rho > \sqrt{n}$ (n is a dimension of x), and initialize the tolerance ε for the duality gap.
2. If $\boldsymbol{\mu}_k^T \mathbf{x}_k \leq \varepsilon$, output solution $\mathbf{w}^* = \mathbf{w}_k$ and stop; otherwise, continue with Step 3
3. Set $\tau_{k+1} = \frac{\boldsymbol{\mu}_k^T x_k}{n+\rho}$ and compute $\boldsymbol{\delta}_w = \{\boldsymbol{\delta}_x, \boldsymbol{\delta}_\lambda, \boldsymbol{\delta}_\mu\}$ using (8)
4. compute step size α_k as follow:

$$\alpha_k = (1 - 10^{-6})\alpha_{\max} \quad \alpha_{\max} = \min(\alpha_p, \alpha_d)$$

where

$$\alpha_p = \min_{i \text{ with } (\boldsymbol{\delta}_x)_i < 0} \left[-\frac{(\mathbf{x}_k)_i}{(\boldsymbol{\delta}_x)_i} \right], \quad \alpha_d = \min_{i \text{ with } (\boldsymbol{\delta}_\mu)_i < 0} \left[-\frac{(\boldsymbol{\mu}_k)_i}{(\boldsymbol{\delta}_\mu)_i} \right]$$

Interior-Point Methods for Convex QP Problems

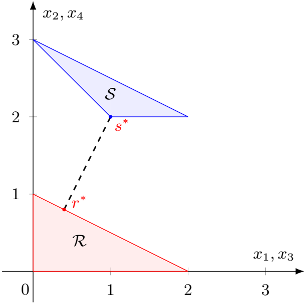
Solve the convex QP problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) = \frac{1}{2}x^T \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix} \mathbf{x} + \mathbf{x}^T \begin{bmatrix} -8 \\ -6 \\ -6 \end{bmatrix} \\ \text{subject to} \quad & x_1 + x_2 + x_3 = 3 \\ & \mathbf{x} \geq 0 \end{aligned}$$

- Using `qp_path_sf.m` with $\mathbf{x}_0 = [1\ 1\ 1]^T$, $\boldsymbol{\lambda}_0 = 7$, $\boldsymbol{\mu}_0 = [3\ 1\ 1]^T$, which is combined as a strictly feasible point $\mathbf{w}_0$. Using 14 iterations, the solution is $\mathbf{x}^* = [0.50\ 1.25\ 1.25]^T$.
- Using `gp_path_nf.m` with $\mathbf{x}_0 = [1\ 2\ 2]^T$, $\boldsymbol{\lambda}_0 = 1$, $\boldsymbol{\mu}_0 = [0.2\ 0.2\ 0.2]^T$, which is not a strictly feasible point w_0. Using 16 iteration, the solution is $\mathbf{x}^* = [0.50\ 1.25\ 1.25]^T$.

Interior-Point Methods for Convex QP Problems

Solve the shortest-distance between triangle $\mathcal{R}$ and $\mathcal{S}$ shown in Figure below and the point $r^* \in \mathcal{R}$ and $s^* \in \mathcal{S}$ that yield the minimum distance.



Solution: Let $\mathbf{r} = [x_1\ x_2]^T \in \mathcal{R}$ and $\mathbf{s} = [x_3\ x_4]^T \in \mathcal{S}$. The square of the distance between r and s is given by

$$(x_1 - x_3)^2 + (x_2 - x_4)^2 = \mathbf{x}^T \mathbf{H} \mathbf{x}, \qquad \mathbf{H} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{x} = [x_1\ x_2\ x_3\ x_4]^T$$

Interior-Point Methods for Convex QP Problems

The constraints of this problem are

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_1 + 2x_2 \leq 2, \quad x_4 \geq 2, \quad x_3 + x_4 \geq 3, \quad x_3 + 2x_4 \leq 6$$

The problem can be formulated as the QP problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq b \end{aligned} \qquad \mathbf{A} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 2 \\ -2 \\ -3 \\ 6 \end{bmatrix}$$

- We need to convert the problem into the form of (3). By letting $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$ where $\mathbf{x}^+ \geq 0$ and $\mathbf{x}^- \geq 0$, and then introducing slack vector $\eta \geq 0$.

Changing to Standard form, we have

$$\begin{aligned} \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{x}^T\mathbf{p} \quad &\Rightarrow \frac{1}{2}(\mathbf{x}^+ - \mathbf{x}^-)^T\mathbf{H}(\mathbf{x}^+ - \mathbf{x}^-) + (\mathbf{x}^+ - \mathbf{x}^-)^T\mathbf{p} \\ \mathbf{A}\mathbf{x} \le \mathbf{b} \quad &\Rightarrow \mathbf{A}(\mathbf{x}^+ - \mathbf{x}^-) + \boldsymbol{\eta} = \mathbf{b} \end{aligned}$$

Then the problem is changed to

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\hat{\mathbf{x}}^T\hat{\mathbf{H}}\hat{\mathbf{x}} + \hat{\mathbf{x}}^T\hat{\mathbf{p}} \\ \text{subject to} \quad & \hat{\mathbf{A}}\hat{\mathbf{x}} = \mathbf{b} \\ & \hat{\mathbf{x}} \ge 0 \end{aligned}$$

$$\mathbf{H} \in \mathbb{R}^{4\times4}, \quad \mathbf{A} \in \mathbb{R}^{6\times4}$$
$$\mathbf{p} \in \mathbb{R}^{6\times1}, \quad \boldsymbol{\eta} \in \mathbb{R}^{6\times1}$$
$$\mathbf{x} \in \mathbb{R}^{4\times1}$$

$$\hat{\mathbf{H}} = \begin{bmatrix} \mathbf{H} & -\mathbf{H} & 0_{4\times6} \\ -\mathbf{H} & \mathbf{H} & 0_{4\times6} \\ 0_{6\times4} & 0_{6\times4} & 0_{6\times6} \end{bmatrix}, \quad \hat{\mathbf{p}} = \begin{bmatrix} \mathbf{p} \\ -\mathbf{p} \\ 0_{6\times1} \end{bmatrix}, \quad \hat{\mathbf{x}} = \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \\ \boldsymbol{\eta} \end{bmatrix}, \quad \hat{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & -\mathbf{A} & \mathbf{I}_{6\times6} \end{bmatrix}$$

- We use nonfeasible-initialization method with $\mathbf{x}_0 = \text{ones}\{14,1\}$, $\boldsymbol{\lambda}_0 = \text{ones}\{6,1\}$,$\boldsymbol{\mu}_0 = \text{ones}\{14,1\}$, where ones$\{m,1\}$ represents a colmun vector of dimension m whose elements are all equal to one.
- Setting $\varepsilon = 10^{-5}$ and $\rho = n + 20\sqrt{n}$, the algorithm using `gp_path_nf.m` takes 11 iterations to converge to $\hat{x}^*$.
- The solution of $\mathbf{x}$ is

$$\mathbf{x}^* = \hat{\mathbf{x}}^*[1:4] - \hat{\mathbf{x}}^*[5:8] = \begin{bmatrix} 0.4 \\ 0.8 \\ 1.0 \\ 2.0 \end{bmatrix}$$

- The shortest distance is

$$\text{shortest distance} = \sqrt{(\mathbf{x}^*)^T \mathbf{H} \mathbf{x}^*} = 1.3416 \text{ unit}$$

Reference

1. Joaquim R. R. A. Martins, Andrew Ning, "*Engineering Design Optimization*," Cambridge University Press, 2021.
2. Mykel J. kochenderfer, and Tim A. Wheeler, "*Algorithms for Optimization*," The MIT Press, 2019.
3. Ashok D. Belegundu, Tirupathi R. Chandrupatla, "*Optimization Concepts and Applications in Engineering*," Cambridge University Press, 2019.
4. Laurent Lessarn , "*Introduction to Optimization*," Lecture Note, University of Wisconsin–Madison.
5. Andreas Antoniou and Wu-Sheng Lu, "*Practical Optimization: Algorithms and Engineering Applications*," 2^{nd} edition, Springer, 2021.
6. Stephen Boyd an dLieven Vandenberghe "*Convex Optimization*," Cambridge University Press, 2004.