Quadratic Programming

December 14

Unconstrained Quadratic Programming I

- Standard form

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T \boldsymbol{x}$$

  where $\mathbf{Q} \in \mathbf{S}^n$, $\boldsymbol{x} \in \mathbb{R}^n$
- First order optimality condition

$$\nabla f = \mathbf{Q}\boldsymbol{x} + \boldsymbol{p} = 0, \qquad \boldsymbol{x}^\star = -\mathbf{Q}^{-1}\boldsymbol{p}$$

- Convex problems
  - $\mathbf{Q}$ is positive semi-definite
  - any local minimizer is global
- Non-convex problem
  - $\mathbf{Q}$ may be indefinite
  - may have many local minimizers, or can be unbounded from below

Unconstrained Quadratic Programming II

Example:

$$\min_{x,y} x^2 - y^2 \text{ s.t. } \quad x \geq 0$$

- The matrix $\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ is indefinite, i.e., it has both positive and negative eigenvalues.
- The feasible direction $\mathbf{d} = (0, -1)$ satisfies the constraint $x \geq 0$, and the objective function decreases without bound along that direction,

$$\lim_{t \to \infty} \quad (x + 0t)^2 - (y - t)^2 = -\infty$$

for any feasible point (x, y).

Unconstrained Quadratic Programming III

- Least squares problem is a QP

$$\min_{\boldsymbol{x}} \quad \|\boldsymbol{b} - \mathbf{A}\boldsymbol{x}\|_2^2$$

$$\|\boldsymbol{b} - \mathbf{A}\boldsymbol{x}\|_2^2 = \boldsymbol{x}^T \mathbf{A}^T \mathbf{A}\boldsymbol{x} - 2\boldsymbol{b}^T \mathbf{A}\boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{b}$$

QP with Equality Constraints I

- $$\begin{aligned} \min_{\boldsymbol{x}} \quad & \frac{1}{2}\boldsymbol{x}^T \mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T \boldsymbol{x} \\ \text{s.t.} \quad & \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

  Assume $\mathbf{A}$ is $m \times n$, $m < n$, and $\text{rank}(\mathbf{A}) = m$, i.e., constraints are independent
- Lagrangian

  $$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T \mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T \boldsymbol{x} + \boldsymbol{\lambda}^T (\mathbf{A}\boldsymbol{x} - \boldsymbol{b})$$

  and Karush-Kuhn-Tucker (KKT) conditions

  $$\begin{aligned} \nabla_{\boldsymbol{x}} \mathcal{L} &= \mathbf{Q}\boldsymbol{x} + \boldsymbol{p} + \mathbf{A}^T \boldsymbol{\lambda} = 0 \\ \nabla_{\boldsymbol{\lambda}} \mathcal{L} &= \mathbf{A}\boldsymbol{x} - \boldsymbol{b} = 0 \end{aligned}$$

QP with Equality Constraints II

reduce to a linear system of equations

$$\begin{bmatrix} \mathbf{Q} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{p} \\ \boldsymbol{b} \end{bmatrix}$$

- KKT matrix

$$\mathbf{K} = \begin{bmatrix} \mathbf{Q} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{K}^{-1} = \begin{bmatrix} \mathbf{H} & \mathbf{T}^T \\ \mathbf{T} & \mathbf{U} \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{H} &= \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{A}^T(\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A})^{-1}\mathbf{A}\mathbf{Q}^{-1} \\ \mathbf{T} &= (\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A})^{-1}\mathbf{A}\mathbf{Q}^{-1} \\ \mathbf{U} &= -(\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A})^{-1} \end{aligned}$$

QP with Equality Constraints III

- Optimal solution

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

Unsolvable KKT System I

$$\min_{x_1, x_2} \quad \frac{1}{2}x_2^2 + x_1, \quad \text{s.t.} \quad x_2 = 0$$

- The problem is a convex QP with

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{p} = [1, 0]^T, \quad \mathbf{A} = [0, 1], \quad b = 0$$

- KKT system has no solution, while the problem is unbounded from below, $f^\star = -\infty$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$$

Unsolvable KKT System II

If KKT system has no solution, the EQP is either *infeasible* or *unbounded below*

- KKT system has no solution iff

$$\begin{bmatrix} -\boldsymbol{p} \\ \boldsymbol{b} \end{bmatrix} \notin \text{Range}(\mathbf{K}) = \text{Range}(\mathbf{K}^T) = \text{Null}(\mathbf{K})^\perp$$

- Assume $[\boldsymbol{v}^T, \boldsymbol{w}^T]^T \in \text{null}(\mathbf{K})$, and $\begin{bmatrix} -\boldsymbol{p} \\ \boldsymbol{b} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix} \neq 0$

$$\mathbf{K} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}\boldsymbol{v} + \mathbf{A}^T \boldsymbol{w} \\ \mathbf{A}\boldsymbol{v} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- If $\boldsymbol{x}_0$ is feasible, i.e., $\mathbf{A}\boldsymbol{x}_0 = \boldsymbol{b}$, then $\boldsymbol{x} + t\boldsymbol{v}$ is feasible for any t

$$\mathbf{A}(\boldsymbol{x} + t\boldsymbol{v}) = \mathbf{A}\boldsymbol{x} + t\,\mathbf{A}\boldsymbol{v} = \boldsymbol{b}$$

Unsolvable KKT System III

- Now moving $\boldsymbol{x}$ along direction $\boldsymbol{v}$,

$$
\begin{aligned}
f(\boldsymbol{x}_0 + t\boldsymbol{v}) &= f(\boldsymbol{x}_0) + t\left(\boldsymbol{x}_0^T \mathbf{Q}\boldsymbol{v} + \boldsymbol{p}^T \boldsymbol{v}\right) + \frac{1}{2}t^2 \boldsymbol{v}^T \mathbf{Q}\boldsymbol{v} \\
&= f(\boldsymbol{x}_0) + t\left(-\boldsymbol{x}_0^T \mathbf{A}^T \boldsymbol{w} + \boldsymbol{p}^T \boldsymbol{v}\right) + \frac{1}{2}t^2 \boldsymbol{v}^T \mathbf{A}^T \boldsymbol{w} \\
&= f(\boldsymbol{x}_0) - t\left(\boldsymbol{b}^T \boldsymbol{w} - \boldsymbol{p}^T \boldsymbol{v}\right) \quad (\text{since } \mathbf{A}\boldsymbol{x}_0 = \boldsymbol{b} \text{ and } \mathbf{A}\boldsymbol{v} = 0)
\end{aligned}
$$

When $t \to \operatorname{sign}(\boldsymbol{b}^T \boldsymbol{w} - \boldsymbol{p}^T \boldsymbol{v})\infty$,
$f(\boldsymbol{x} + t\boldsymbol{v})$ goes to $-\infty$, i.e., unbounded below

Nonsingularity of KKT Matrix I

KKT matrix is nonsingular if and only if $\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} > 0$ for $\boldsymbol{x} \neq 0$ and $\mathbf{A}\boldsymbol{x} = 0$

- Assume that KKT matrix is nonsingular.
  If there exists $\boldsymbol{x} \neq 0$ in nullspace of $\mathbf{A}$ and nullspace of $\mathbf{Q}$, i.e., $\mathbf{A}\boldsymbol{x} = 0$ and $\mathbf{Q}\boldsymbol{x} = 0$, then the KKT matrix is singular

$$\mathbf{K}\begin{bmatrix}\boldsymbol{x}\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\mathbf{Q} & \mathbf{A}^T\\ \mathbf{A} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\mathbf{Q}\boldsymbol{x}\\ \mathbf{A}\boldsymbol{x}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$$

  Contradicting the nonsingularity of $\mathbf{K}$.
  It means that $\mathbf{Q}$ and $\mathbf{A}$ have no nontrivial common nullspace
  If $\mathbf{A}\boldsymbol{s} = 0$ for $\boldsymbol{s} \neq 0$ then $\boldsymbol{s}^T\mathbf{Q}\boldsymbol{s} > 0$

Nonsingularity of KKT Matrix II

- Assume that $\text{null}(\mathbf{Q}) \cap \text{null}(\mathbf{A}) = \{0\}$, i.e., $\mathbf{Q}$ and $\mathbf{A}$ have no nontrivial common nullspace.
  If KKT matrix $\mathbf{K}$ is singular, then there exist $\boldsymbol{x}$ and $\boldsymbol{z}$, not both zeros, such that

  $$\mathbf{0} = \mathbf{K}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{bmatrix} = \begin{bmatrix}\mathbf{Q}\boldsymbol{x} + \mathbf{A}^T\boldsymbol{z}\\ \mathbf{A}\boldsymbol{x}\end{bmatrix} = \mathbf{0}$$

  Multiply the first equation by $\boldsymbol{x}^T$

  $$0 = \boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} + \boldsymbol{x}^T\mathbf{A}^T\boldsymbol{z} = \boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} + \mathbf{0}^T\boldsymbol{z} = \boldsymbol{x}^T\mathbf{Q}\boldsymbol{x}$$

  Denote by $\boldsymbol{u}_i$ and $\lambda_i \geq 0$ eigenvectors and eigenvalues of $\mathbf{Q}$, respectively

  $$\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} = \boldsymbol{x}^T\left(\sum_i \lambda_i \boldsymbol{u}_i\boldsymbol{u}_i^T\right)\boldsymbol{x} = \sum_i \lambda_i(\boldsymbol{u}_i^T\boldsymbol{x})^2 = 0$$

Nonsingularity of KKT Matrix III

i.e., $\boldsymbol{x}^T \boldsymbol{u}_i = 0$ for $\lambda_i > 0$.
Hence, $\mathbf{Q}\boldsymbol{x} = \sum_i \lambda_i \boldsymbol{u}_i(\boldsymbol{u}_i^T \boldsymbol{x}) = 0$. Implying that $\boldsymbol{x} \in$ null$(\mathbf{Q}) \cap$ null$(\mathbf{A})$, or $\boldsymbol{x} = 0$ and $\boldsymbol{z}$ is non-zero. Then $\mathbf{A}^T \boldsymbol{z} = -\mathbf{Q}\boldsymbol{x} = 0$ contradicts $rank(\mathbf{A}) = m$, $\mathbf{A}^T$ is of full column rank

Nonsingularity of KKT Matrix IV

Example

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\boldsymbol{x}^T \mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T \boldsymbol{x}, \quad \text{s.t.} \qquad \mathbf{A}\boldsymbol{x} = b \tag{1}$$

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \end{bmatrix}$$

The null space of $\mathbf{A}$ is spanned by the vectors $\mathbf{u} = [1, -1, 0]^T$ and $\mathbf{v} = [1, 0, -1]^T$.

$$\mathbf{A}^T[\boldsymbol{u}, \boldsymbol{v}] = [0, 0]$$

$\mathbf{Q}$ is singular because it has a zero eigenvalue, but positive definite over null$(\mathbf{A})$, because for any nonzero vector $\boldsymbol{x} = \alpha\boldsymbol{u} + \beta\boldsymbol{v}$, we have:

Nonsingularity of KKT Matrix V

$$\boldsymbol{x}^T \mathbf{Q} \boldsymbol{x} = (\alpha \boldsymbol{u} + \beta \boldsymbol{v})^T \mathbf{Q} (\alpha \boldsymbol{u} + \beta \boldsymbol{v}) = \alpha^2 + (\alpha + \beta)^2 > 0$$

KKT matrix is nonsingular

$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

and the KKT system has unique solution

$$\boldsymbol{x}^\star = [001]^T$$

with dual variable $\lambda^\star = 1$.

Elimination of Equality Constraints I

- Denote by $\hat{\boldsymbol{x}}$ any feasible $\mathbf{A}\hat{\boldsymbol{x}} = \boldsymbol{b}$
- Matrix $\mathbf{F}$ of size $n \times (n - m)$ such that range($\mathbf{F}$) = null($\mathbf{A}$), i.e., $\mathbf{AF} = 0$
- Then $\{\boldsymbol{x} : \mathbf{A}\boldsymbol{x} = \boldsymbol{b}\} = \{\mathbf{F}\boldsymbol{z} + \hat{\boldsymbol{x}}, \boldsymbol{z} \in \mathbb{R}^{n-m}\}$
  and the EQP problem becomes

$$\min \quad \frac{1}{2}\boldsymbol{z}^T\mathbf{F}^T\mathbf{QF}\boldsymbol{z} + \hat{\boldsymbol{x}}^T\mathbf{QF}\boldsymbol{z} + \boldsymbol{p}^T\mathbf{F}\boldsymbol{z}$$

  which gives the optimal solution

$$\boldsymbol{z}^\star = -(\mathbf{F}^T\mathbf{QF})^{-1}\left(\mathbf{F}^T\boldsymbol{p} + \mathbf{F}^T\mathbf{Q}\hat{\boldsymbol{x}}\right)$$

Elimination of Equality Constraints II

Example

$$\begin{aligned} \min \quad & \frac{1}{2}(x_1^2 + x_2^2) + 2x_1 + x_2 - x_3 \\ \text{s.t.} \quad & x_2 + x_3 = 1 \end{aligned}$$

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{p} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{A} = [0, 1, 1], \quad \mathbf{F} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}$$

$\hat{\boldsymbol{x}} = [0, 0, 1]^T$ is a feasible point, and $\boldsymbol{z}^\star = -[2, 2]^T$. Hence, the optimal solution

$$\boldsymbol{x}^\star = [-2, -2, 3]^T$$

QP with Nonnegativity Constraints I

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{x} \geq 0 \end{aligned}$$

Optimality KKT conditions

$$\begin{aligned} \mathbf{Q}\boldsymbol{x} + \boldsymbol{p} - \boldsymbol{\lambda} &= 0 \\ \boldsymbol{\lambda} &\geq 0 \\ x_i\lambda_i &= 0 \end{aligned}$$

The gradient $\boldsymbol{g} = \nabla f(\boldsymbol{x}^\star) = \boldsymbol{\lambda} \geq 0$ and $g_i x_i^\star = 0$.

QP with Nonnegativity Constraints II

- Gradient is nonnegative and orthogonal to $\boldsymbol{x}^\star$
- There exists an index set $\mathcal{A} = \{i : g_i > 0\}$, $x_{i\in\mathcal{A}} = 0$ (active set)
  Denote by $\mathcal{I} = \{i : g_i = 0\}$

Question: How to determine the active set $\mathcal{A}$?

- Assume $\boldsymbol{x}$ is a feasible point, and $\mathcal{A}$ is the active set of $\boldsymbol{x}$, and $\mathcal{I}$ the inactive set.
- Step 1: We move $\boldsymbol{x}$ to a new point $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{d}$ such that $\hat{\boldsymbol{x}}_{i\in\mathcal{A}} = 0$.
  The search direction vector $\boldsymbol{d}$ also has $d_i = 0$, $i \in \mathcal{A}$.

QP with Nonnegativity Constraints III

- Let $\boldsymbol{g} = \mathbf{Q}\boldsymbol{x} + \boldsymbol{p}$ be the gradient of $f(\boldsymbol{x})$. Since

$$f(\boldsymbol{x} + \boldsymbol{d}) = \frac{1}{2}\boldsymbol{d}^T\mathbf{Q}\boldsymbol{d} + (\mathbf{Q}\boldsymbol{x} + \boldsymbol{p})^T\boldsymbol{d} + f(\boldsymbol{x})$$

  $\boldsymbol{d}(\mathcal{I})$ is solution of a QP

$$\min \frac{1}{2}\boldsymbol{d}(\mathcal{I})^T\mathbf{Q}(\mathcal{I},\mathcal{I})\boldsymbol{d}(\mathcal{I}) + \boldsymbol{g}(\mathcal{I})^T\boldsymbol{d}(\mathcal{I})$$

  which has a closed-form update

$$\boldsymbol{d}(\mathcal{I})^\star = -\mathbf{Q}(\mathcal{I},\mathcal{I})^{-1}\boldsymbol{g}(\mathcal{I}) = -\boldsymbol{x}(\mathcal{I}) - \mathbf{Q}(\mathcal{I},\mathcal{I})^{-1}\boldsymbol{p}(\mathcal{I})$$

QP with Nonnegativity Constraints IV

- Step 2: Check primal feasibility: i.e., the new update $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{d} \geq 0$
  Note that

$$\hat{\boldsymbol{x}}(\mathcal{A}) = \boldsymbol{x}(\mathcal{A}) + \boldsymbol{d}(\mathcal{A}) = 0$$

and

$$\begin{aligned}
\hat{\boldsymbol{x}}(\mathcal{I}) &= \boldsymbol{x}(\mathcal{I}) + \boldsymbol{d}(\mathcal{I}) = \boldsymbol{x}(\mathcal{I}) - \mathbf{Q}(\mathcal{I}, \mathcal{I})^{-1} \boldsymbol{g}(\mathcal{I}) \\
&= \boldsymbol{x}(\mathcal{I}) - \mathbf{Q}(\mathcal{I}, \mathcal{I})^{-1} (\mathbf{Q}(I, I)\boldsymbol{x}(\mathcal{I}) + \boldsymbol{p}(\mathcal{I})) \\
&= -\mathbf{Q}(\mathcal{I}, \mathcal{I})^{-1} \boldsymbol{p}(\mathcal{I})
\end{aligned}$$

QP with Nonnegativity Constraints V

- Step 2a: Check dual feasibility, i.e., $\hat{\boldsymbol{g}} = \mathbf{Q}\hat{\boldsymbol{x}} + \boldsymbol{p} \geq 0$
  Since

$$\hat{\boldsymbol{g}}(\mathcal{I}) = \mathbf{Q}(\mathcal{I}, \mathcal{I})\hat{\boldsymbol{x}}(\mathcal{I}) + \boldsymbol{p}(\mathcal{I}) = 0$$

  we need to check only $\hat{\boldsymbol{g}}(\mathcal{A}) = \mathbf{Q}(\mathcal{A}, \mathcal{I})\hat{\boldsymbol{x}}(\mathcal{I}) + \boldsymbol{p}(\mathcal{A}) \geq 0$.
  If there exists $i \in \mathcal{A}$, $\hat{g}(i) < 0$, eliminate i from the active set, $\mathcal{A}$, and add it to the inactive set, $\mathcal{I}$

$$\mathcal{A} = \mathcal{A} \setminus \{i\}, \quad \mathcal{I} = \mathcal{I} \cup \{i\}$$

- Step 2b: if $\hat{\boldsymbol{g}}(\mathcal{A}) \geq 0$, $\hat{\boldsymbol{x}}$ is the optimal solution

QP with Nonnegativity Constraints VI

- Step 3: If $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{d}$ is infeasible, we seek a maximum $0 \le \alpha \le 1$ such that $\boldsymbol{x} + \alpha \boldsymbol{d}$ is feasible

$$\begin{aligned} \max \quad & \alpha \\ \text{s.t.} \quad & \boldsymbol{x} + \alpha \boldsymbol{d} \ge 0 \end{aligned}$$

The constraints are actually only for negative search direction $\boldsymbol{d}(i \in \mathcal{I}) < 0$. This gives the optimal $\alpha^\star$

$$\alpha^\star = \min_{i \in \mathcal{I}, d(i) < 0} \left(\frac{x(i)}{-d(i)} \right)$$

QP with Nonnegativity Constraints VII

Example

$$\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{x}^T \mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{x} \geq 0 \end{aligned}$$

where

$$\mathbf{Q} = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}, \quad \boldsymbol{p} = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}$$

- Take $\boldsymbol{x}_0 = [0, 0, 0]^T$, and the gradient $\boldsymbol{g}_0 = \boldsymbol{p} = [1, -3, 2]^T$.
  - Active set $\mathcal{A} = \{1, 3\}$ and $\mathcal{I} = \{2\}$

QP with Nonnegativity Constraints VIII

- Direction $\boldsymbol{d}(\mathcal{I}) = 1$, and $\boldsymbol{d} = [0, 1, 0]^T$ gives the new update

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \boldsymbol{d} = [0, 1, 0]^T$$

  - $\boldsymbol{x}_1$ is feasible
  - gradient $\boldsymbol{g}_1(S) = [2, 3]$ while $\boldsymbol{g}(\mathcal{I}) = 0$, i.e., $\boldsymbol{g}$ holds dual feasibility
- Hence $[0, 1, 0]$ is the optimal solution

- Take $\boldsymbol{x}_0 = [0, 0, 1]^T$, which gives the gradient $\boldsymbol{g}_0 = [2, -2, 6]^T$.
  - Active set $\mathcal{A} = \{1, 2\}$ and $\mathcal{I} = \{3\}$
  - Direction $\boldsymbol{d}(\mathcal{I}) = -\frac{3}{2}$, $\boldsymbol{d} = [0, 0, \frac{-3}{2}]^T$
  - giving the new update which is infeasible

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \boldsymbol{d} = [0, 0, \frac{-1}{2}]^T$$

QP with Nonnegativity Constraints IX

- Search α such that $\boldsymbol{x} + \alpha \boldsymbol{d} \geq 0$

$$\alpha^\star = \min(\frac{1}{1.5}) = \frac{2}{3}$$

  and the new $\boldsymbol{x}_1 = [0, 0, 0]^T$ (feasible)
- Next steps: ...

QP with Nonnegativity Constraints X

Example

$$\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{x}^T \mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{x} \geq 0 \end{aligned}$$

where

$$\mathbf{Q} = \begin{bmatrix} 3 & 1 & 0 & 1 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 1 & 0 & 1 & 4 \end{bmatrix}, \quad \boldsymbol{p} = \begin{bmatrix} 2 \\ -3 \\ 1 \\ 1 \end{bmatrix}$$

- The unconstraint QP solution $\boldsymbol{x}_{qp} = -\mathbf{Q}^{-1}\boldsymbol{p} = \frac{-1}{18}[21, -25, 5, -2]^T$

QP with Nonnegativity Constraints XI

- Use $\boldsymbol{x}_0 = \max(0, \boldsymbol{x}_{qp}) = \frac{1}{18}[21, 0, 5, 0]^T$ with the active set $\mathcal{A} = \{1, 3\}$ and $\mathcal{I} = \{2, 4\}$
- Direction vector

$$\boldsymbol{d}(\mathcal{I}) = -\boldsymbol{x}_0(\mathcal{I}) - \mathbf{Q}(\mathcal{I}, \mathcal{I})^{-1}\boldsymbol{p}(\mathcal{I}) = \frac{1}{36}\begin{bmatrix} -14 \\ -13 \end{bmatrix}$$

  and the full vector $\boldsymbol{d} = \frac{1}{36}[0, -14, 0, -13]^T$
- New update $\boldsymbol{x}_1 = \boldsymbol{x}_0 + \boldsymbol{d} = [0, 1, 0, -0.25]^T$ is infeasible
  - Seek $\boldsymbol{x}_1 = \boldsymbol{x} + \alpha\boldsymbol{d}$ where

$$\alpha^* = \min(\frac{25}{18} : \frac{14}{36}, \frac{1}{9} : \frac{13}{36}) = \frac{4}{13}$$

    Hence $\boldsymbol{x}_1 = [0, \frac{33}{26}, 0, 0]^T$, and the associated active set $\mathcal{A} = \{1, 3, 4\}$, and inactive set $\mathcal{I} = \{2\}$
  - Direction vector $\boldsymbol{d}(\mathcal{I}) = -\boldsymbol{x}_1(\mathcal{I}) - \mathbf{Q}(\mathcal{I}, \mathcal{I})^{-1}\boldsymbol{p}(\mathcal{I}) = \frac{-7}{26}$
- New update $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \boldsymbol{d} = [0, 1, 0, 0]^T$ which is feasible
- Since $\boldsymbol{g}_2(\mathcal{A}) = [3, 1, 1]^T > 0$ holds the dual feasibility, $\boldsymbol{x}_2$ is the optimal solution.

QP with Nonnegativity Constraints XII

Algorithm 1: Nonnegative Quadratic Programming

Data: $\mathbf{Q} \in \mathbf{S}_+^n$, $\boldsymbol{p} \in \mathbb{R}^n$

Result: $\boldsymbol{x} = \arg\min \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{x}$ s.t. $\boldsymbol{x} \geq 0$

Find a feasible $\boldsymbol{x}_0$, e.g., $\boldsymbol{x}_0 = \max(0, -\mathbf{Q}^{-1}\boldsymbol{p})$;

$\mathcal{A} = \{i : x_0(i) = 0\}$ and $\mathcal{I} = \{i : x_0(i) > 0\}$;

repeat

  $\boldsymbol{x}_{k+1}(\mathcal{I}) = -\mathbf{Q}(\mathcal{I}, \mathcal{I})^{-1}\ \boldsymbol{p}(\mathcal{I})$;

  if $\boldsymbol{x}_{k+1} \geq 0$ **then**

    Gradient $\boldsymbol{g}(\mathcal{A}) = \mathbf{Q}(\mathcal{A}, \mathcal{I})\boldsymbol{x}_{k+1}(\mathcal{I}) + \boldsymbol{p}(\mathcal{A})$

    if $\exists i \in \mathcal{A}, \boldsymbol{g}(i) < 0$ **then**

      Eliminate i from $\mathcal{A} = \mathcal{A} \setminus \{i\}$, $\mathcal{I} = \mathcal{I} \cup \{i\}$

    else

      Return the optimal $\boldsymbol{x}_{k+1}$

    end

  else

    $\boldsymbol{d} = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$, $\alpha = \min_{i \in \mathcal{I}, d(i)<0}(\frac{x_i}{-d_i})$, $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha\boldsymbol{d}$

    Update $\mathcal{A}$ and $\mathcal{I}$

  end

until;

QP with Inequality Constraints I

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{x} \\
\text{s.t.} \quad & \mathbf{A}\boldsymbol{x} \le \boldsymbol{b}
\end{aligned}
$$

- Lagrangian $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(\mathbf{A}\boldsymbol{x} - \boldsymbol{b})$
- KKT conditions

$$
\begin{aligned}
\mathbf{Q}\boldsymbol{x}^\star + \mathbf{A}^T\boldsymbol{\lambda} &= -\boldsymbol{p} \\
\boldsymbol{\lambda}^\star &\ge 0 \\
\mathbf{A}\boldsymbol{x}^\star &\le \boldsymbol{b} \\
\lambda_i(\boldsymbol{a}_i^T\boldsymbol{x}^\star - b_i) &= 0, \quad i = 1, \ldots, m
\end{aligned}
$$

$\boldsymbol{a}_i^T$ are rows of $\mathbf{A}$ (of size $m \times n$).

QP with Inequality Constraints II

- Define active set $\mathcal{A} = \{i \mid \boldsymbol{a}_i^T \boldsymbol{x} = b_i\}$. Assume that $\boldsymbol{a}_i$ are linearly independent.
- Note that optimal dual $\lambda_i^\star = 0$ for $i \notin \mathcal{A}$. Hence if $\mathcal{A}$ is known,
  - We may delete all inactive inequality constraints and corresponding zero Lagrange multipliers
  - Reduce the QP problem to QP with equality constraints

Active Set Method for Inequality constrained QP I

Basic algorithm

- Pick up a subset $\mathcal{A}_k$ of $\{1, 2, \ldots, m\}$
- Find $\boldsymbol{x}_{k+1} = \arg\min f(\boldsymbol{x})$ s.t. $\boldsymbol{a}_i^T \boldsymbol{x} = b_i$ for all $i \in \mathcal{A}_k$
- If $\boldsymbol{x}_{k+1}$ does not solve the considered QP, adjust $\mathcal{A}_k$ to form $\mathcal{A}_{k+1}$ and repeat

Questions

- How to verify if $\boldsymbol{x}_{k+1}$ solve QP?
- How to form a new $\mathcal{A}_{k+1}$?

Active Set Method for Inequality constrained QP I

- Assume that we know a feasible point $\boldsymbol{x}_k$ (how to find?)
- Define the active set $\mathcal{A}_k$ at the current iterate
- Keep the constraints active for $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}$ and find $\boldsymbol{d}$

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & \frac{1}{2}(\boldsymbol{x}+\boldsymbol{d})^T \mathbf{Q}(\boldsymbol{x}+\boldsymbol{d}) + \boldsymbol{p}^T(\boldsymbol{x}+\boldsymbol{d}) \\ \text{s.t.} \quad & \boldsymbol{a}_i^T(\boldsymbol{x}+\boldsymbol{d}) = b_i, \quad i \in \mathcal{A}_k \end{aligned}$$

or a reduced equality constrained QP

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & \frac{1}{2}\boldsymbol{d}^T \mathbf{Q}\boldsymbol{d} + \boldsymbol{g}_k^T \boldsymbol{d} \\ \text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{d} = 0, \quad i \in \mathcal{A}_k \end{aligned}$$

where $\boldsymbol{g}_k = \mathbf{Q}\boldsymbol{x}_k + \boldsymbol{p}$ is gradient of $f(\boldsymbol{x}_k)$

Active Set Method for Inequality constrained QP II

- The optimal $\boldsymbol{d}^\star$ and the associated dual $\boldsymbol{\lambda}^\star$ are solution of a KKT system

$$\begin{bmatrix} \mathbf{Q} & \mathbf{A}_{\mathcal{A}_k}^T \\ \mathbf{A}_{\mathcal{A}_k} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{d}^\star \\ \boldsymbol{\lambda}^\star \end{bmatrix} = \begin{bmatrix} -\boldsymbol{g}_k \\ \mathbf{0} \end{bmatrix}$$

  where $\mathbf{A}_{\mathcal{A}_k} = [\boldsymbol{a}_i]^T$, $i \in \mathcal{A}_k$

Question: What if $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}^\star$ is not feasible?

Note that the active constraints are preserved by assumption

$$\begin{aligned} \boldsymbol{a}_i^T(\boldsymbol{x}_k + \boldsymbol{d}^\star) &= \boldsymbol{a}_i^T \boldsymbol{x}_k + \boldsymbol{a}_i^T \boldsymbol{d} \\ &= b_i + 0, \quad i \in \mathcal{A}_k \end{aligned}$$

However, $\boldsymbol{x}_{k+1}$ may not fulfill the inactive inequality constraints
i.e., $\boldsymbol{a}_i^T(\boldsymbol{x}_k + \boldsymbol{d}^\star) > b_i$ for some $i \notin \mathcal{A}_k$

New Constraints to the Active Set I

Instead of moving to $\boldsymbol{x}_k + \boldsymbol{d}^\star$, we seek $\boldsymbol{x}_{k+1} = \boldsymbol{x} + \alpha \boldsymbol{d}^\star$ with a maximum step length α

$$\begin{aligned} \max \quad & \alpha \\ \text{s.t.} \quad & \mathbf{A}(\boldsymbol{x}_k + \alpha \boldsymbol{d}^\star) \le \boldsymbol{b} \end{aligned}$$

We can actually eliminate inequality constraints

- for $i \in \mathcal{A}_k$ since $\boldsymbol{a}_i^T(\boldsymbol{x}_k + \alpha \boldsymbol{d}^\star) = b_i + \alpha 0 = b_i$
- for $j \notin \mathcal{A}_k$ such that $\boldsymbol{a}_j^T \boldsymbol{d}^\star < 0$

Therefore, we solve α for

$$\begin{aligned} \max \quad & \alpha \\ \text{s.t.} \quad & \alpha\, \boldsymbol{a}_i^T \boldsymbol{d}^\star \le b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k \end{aligned}$$

where $i \notin \mathcal{A}_k$ and $\boldsymbol{a}_i^T \boldsymbol{d}^\star > 0$

New Constraints to the Active Set II

This gives the optimal $\alpha = \min_{i:\boldsymbol{a}_i^T \boldsymbol{d}^\star > 0} \left(\frac{b_i - \boldsymbol{a}_i^T \boldsymbol{x}_k}{\boldsymbol{a}_i^T \boldsymbol{d}^\star} \right)$

The new active set $\mathcal{A}_{k+1} = \mathcal{A}_{k+1} \cup \{l\}$ where l is with the new constraint $\boldsymbol{a}_l^T \boldsymbol{x} = b_l$.

Remove Constraints from the Active Set I

- Assume $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}^\star$ is feasible, i.e., $\mathbf{A}\boldsymbol{x}_{k+1} \leq \boldsymbol{b}$
- The associated dual $\boldsymbol{\lambda}^*$ may not hold dual feasibility, i.e., $\lambda_i^\star < 0$ for some i
    - Eliminate i from the active set $\mathcal{A}_k$ and form $\mathcal{A}_{k+1} = \mathcal{A}_k \setminus \{i\}$
    - Repeat the procedure, solve a new EQP
- If $\boldsymbol{\lambda}^\star \geq 0$, then $\boldsymbol{x}_{k+1}$ is the optimal solution

Examples for QP with Inequality Constraints I

Example

$$\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{x}^T \mathbf{Q}\boldsymbol{x} + \boldsymbol{p}^T \boldsymbol{x} \\ \text{s.t.} \quad & \mathbf{A}\boldsymbol{x} \leq \boldsymbol{b} \end{aligned}$$

where

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad \boldsymbol{p} = \begin{bmatrix} -4 \\ -6 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} -2 & 1 \\ 1 & 1 \\ 0 & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix}$$

- Step 1: Start with $\boldsymbol{x}_1 = [0, 0]^T$ and the associated active set $\mathcal{A}_1 = \{1, 3\}$

Examples for QP with Inequality Constraints II

- Gradient $\boldsymbol{g}_1 = [-4, -6]^T$. Solve KKT system

$$\begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 2 & 1 & -1 \\ -2 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 0 \\ 0 \end{bmatrix}$$

  we obtain the optimal solution

$$\boldsymbol{d}^\star = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{\lambda}^\star = \begin{bmatrix} -2 \\ -8 \end{bmatrix}$$

  which gives $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \boldsymbol{d}^\star = [0, 0]^T$.
- Since $\lambda_1^\star < 0$, violates the dual feasibility condition, we eliminate 1 from $\mathcal{A}_1$, and form a new active set $\mathcal{A}_2 = \{3\}$

Examples for QP with Inequality Constraints III

- Step 2: Solve the KKT system

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 0 \end{bmatrix}$$

  - get the optimal $\boldsymbol{d}^\star$ and $\boldsymbol{\lambda}^*$

$$\boldsymbol{d}^\star = \begin{bmatrix} 4 \\ 0 \end{bmatrix}, \qquad \lambda^\star = \begin{bmatrix} -6 \end{bmatrix}, \qquad \boldsymbol{x}_3 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

  - $\boldsymbol{x}_3$ is a feasible point, $\mathbf{A}\boldsymbol{x}_3 \leq \boldsymbol{b}$
  - However, since $\lambda^\star < 0$, we eliminate 3 from the current active set $\mathcal{A}_4 = \{\}$.

Examples for QP with Inequality Constraints IV

- Step 3: Solve the KKT system

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

  - get the optimal $\boldsymbol{d}^\star$

$$\boldsymbol{d}^\star = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \qquad \boldsymbol{x}_4 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

  - $\boldsymbol{x}_4$ is infeasible
  - Solve $\alpha = 0$ gives $\boldsymbol{x}_4 = \boldsymbol{x}_3$
  - Refine the active set $\mathcal{A}_5 = \{2, 3\}$.

Examples for QP with Inequality Constraints V

- Step 4: Solve KKT system

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 0 \\ 0 \end{bmatrix}$$

  - we obtain the optimal solution

$$\boldsymbol{d}^\star = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{\lambda}^\star = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$$

    which gives $\boldsymbol{x}_5 = \boldsymbol{x}_4 = [4, 0]^T$.
  - Since $\lambda_2^\star < 0$, violates the dual feasibility condition, we eliminate 3 from $\mathcal{A}_5$, and form a new active set $\mathcal{A}_6 = \{2\}$

Examples for QP with Inequality Constraints VI

- Step 5: Solve KKT system

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 0 \end{bmatrix}$$

  - we obtain the optimal solution

$$\boldsymbol{d}^\star = \begin{bmatrix} -2 \\ 2 \end{bmatrix}, \qquad \boldsymbol{\lambda}^\star = \begin{bmatrix} 2 \end{bmatrix}$$

    which gives $\boldsymbol{x}_6 = \boldsymbol{x}_5 + \boldsymbol{d} = [2, 2]^T$: feasible.
  - Since $\lambda^\star > 0$, we obtain the optimal solution $\boldsymbol{x}^\star = [2, 2]$.

Examples for QP with Inequality Constraints I

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = (x_1 - 1)^2 + (x_2 - 2.5)^2 \\
\text{s.t.} \quad & x_1 - 2x_2 + 2 \geq 0 \\
& -x_1 - 2x_2 + 6 \geq 0 \\
& -x_1 + 2x_2 + 2 \geq 0 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

- Step 1: initial $\boldsymbol{x}_1 = [2, 0]^T$ with active set $\mathcal{A}_1 = \{3, 5\}$
  - gradient $\boldsymbol{g}_1 = [2, -5]^T$

Examples for QP with Inequality Constraints II

- Solve KKT system

$$\begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & -2 & -1 \\ 1 & -2 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \\ 0 \\ 0 \end{bmatrix}$$

- obtain $\boldsymbol{d}^\star = [0,0]^T$ and $\boldsymbol{\lambda}^\star = [-2,-1]^T$
- $\boldsymbol{x}_2 = [2,0]^T$ remains feasible
- $\boldsymbol{\lambda} < 0$ does not hold dual feasibility. Eliminate one constraint and form $\mathcal{A}_2 = \{5\}$

- Step 2:
  - Solve KKT system

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \\ \lambda \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \\ 0 \end{bmatrix}$$

Examples for QP with Inequality Constraints III

- obtain $\boldsymbol{d}^\star = [-1, 0]^T$ and $\lambda^\star = -5$
- $\boldsymbol{x}_3 = [1, 0]$: feasible
- $\lambda^\star = -5$: infeasible (dual). Eliminate the constraint (5) and form $\mathcal{A}_3 = \{\}$

- Step 3
  - Solve KKT system

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$$

  - obtain $\boldsymbol{d}^\star = [0, 2.5]^T$
  - $\boldsymbol{x}_4 = [1, 2.5]^T$: infeasible
  - Seek a new feasible $\boldsymbol{x}_4 = \boldsymbol{x}_3 + \alpha \boldsymbol{d}$

$$\alpha = \min_{\boldsymbol{a}_i^T \boldsymbol{d} > 0} \left(\frac{b_i - \boldsymbol{a}_i^T \boldsymbol{x}_3}{-\boldsymbol{a}_i^T \boldsymbol{d}} \right) = \min(0.6, 1) = 0.6$$

Move to $\boldsymbol{x}_4 = [1, 1.5]^T$ and the new active set $\mathcal{A}_4 = \{1\}$

Examples for QP with Inequality Constraints IV

- Step 4
  - Solve KKT system

$$\begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & 2 \\ -1 & 2 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

  - obtain $\boldsymbol{d}^\star = [0.4, 0.2]^T$ and $\lambda^\star = 0.8$
  - $\boldsymbol{x}_5 = [1.4, 1.7]^T$: primal feasible
  - $\lambda^\star = 0.8 > 0$: dual feasible
- $\boldsymbol{x}^\star = [1.4, 1.7]$: optimal solution

Solving QP via Linear Programming I

Suppose $\mathcal{D}$ is a compact (closed and bounded) convex set

$$\min f(\mathbf{x}) \quad \text{s.t.} \quad \boldsymbol{x} \in \mathcal{D}$$

For QP, since $f(\boldsymbol{x})$ is convex, we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + (\boldsymbol{y} - \boldsymbol{x})^T \nabla f(\boldsymbol{x}), \quad \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D},$$

Let $\boldsymbol{x}^\star$ be the optimal solution. The best lower bound with respect to a given point $\boldsymbol{x}$ is given by

$$\begin{aligned}
f(\boldsymbol{x}^*) &\geq f(\boldsymbol{x}) + (\boldsymbol{x}^* - \boldsymbol{x})^T \nabla f(\boldsymbol{x}) \\
&\geq \min_{\boldsymbol{y} \in D} \left\{ f(\boldsymbol{x}) + (\boldsymbol{y} - \boldsymbol{x})^T \nabla f(\boldsymbol{x}) \right\} \\
&= f(\boldsymbol{x}) - \boldsymbol{x}^T \nabla f(\boldsymbol{x}) + \min_{\boldsymbol{y} \in D} \boldsymbol{y}^T \nabla f(\boldsymbol{x})
\end{aligned}$$

Frank-Wolfe (1956) algorithm minimizes the latter optimization.

Solving QP via Linear Programming II

- Minimize the linear approximation of the given problem around $\boldsymbol{x}_k$

$$\min_{\boldsymbol{s}} \nabla f(\boldsymbol{x}_k)^T \boldsymbol{s} \quad \text{s.t.} \quad \boldsymbol{s} \in \mathcal{D}$$

- Update $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha(\boldsymbol{s} - \boldsymbol{x}_k)$

The FW algorithm is very easy to implement, e.g., using LP, does not require projections.

Sequential quadratic programming I

Newton's method for unconstrained problem

- Necessary condition for minimum is $\nabla f(\boldsymbol{x}^\star) = 0$

$$
\begin{aligned}
& \nabla_{\boldsymbol{x}} f(\boldsymbol{x}_k + \boldsymbol{d}_k) \approx \nabla_{\boldsymbol{x}} f(\boldsymbol{x}_k) + \nabla^2_{\boldsymbol{x}} f(\boldsymbol{x}_k)\boldsymbol{d}_k = 0 \\
\Longleftrightarrow \quad & \nabla^2_{\boldsymbol{x}} f(\boldsymbol{x}_k)\,\boldsymbol{d}_k = -\nabla_x f(\boldsymbol{x}_k)
\end{aligned}
$$

Sequential Quadratic Programming (SQP)

$$
\min_{\boldsymbol{x} \in \mathbb{R}^n} \quad f(\boldsymbol{x}) \quad \text{s.t.} \quad g(\boldsymbol{x}) = 0
$$

- uses similar idea as Newton's method: at each step, build a QP

Sequential quadratic programming II

- Lagrangian $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T g(\boldsymbol{x})$

$$\nabla \mathcal{L}(\boldsymbol{x}_k + \boldsymbol{d}_k, \boldsymbol{\lambda}_k + \boldsymbol{\mu}_k) \approx \nabla L(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) + \nabla^2 \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) \begin{bmatrix} \boldsymbol{d}_k \\ \boldsymbol{\mu}_k \end{bmatrix} = 0$$

- Find steps $\boldsymbol{d}_k$ and $\boldsymbol{\mu}_k$

$$\nabla^2 \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) \begin{bmatrix} \boldsymbol{d}_k \\ \boldsymbol{\mu}_k \end{bmatrix} = -\nabla \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k)$$

where

$$\nabla^2 \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) = \begin{bmatrix} \nabla^2_{xx} L(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) & \nabla_{\boldsymbol{x}} g(\boldsymbol{x}_k) \\ \nabla_{\boldsymbol{x}} g(\boldsymbol{x}_k)^T & 0 \end{bmatrix}$$

Sequential quadratic programming III

- Optimal $\boldsymbol{d}_k^\star$ is solution to a QP

$$
\begin{aligned}
\min_{\boldsymbol{d}} \quad & \frac{1}{2}\,\boldsymbol{d}^T\,\nabla^2_{xx} L(\boldsymbol{x}_k,\lambda_k)\,\boldsymbol{d} + \boldsymbol{d}^T\,\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}_k,\lambda_k) \\
\text{s.t.} \quad & \nabla_{\boldsymbol{x}} g(\boldsymbol{x}_k)^T\,\boldsymbol{d} + g(\boldsymbol{x}_k) = 0
\end{aligned}
$$

  - quadratic approximation of the Lagrangian
  - linearization of the constraints

SQP I

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x}) \quad \text{s.t.} \quad g(\boldsymbol{x}) \leq 0$$

- Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T g(\boldsymbol{x})$$

- KKT conditions

$$\begin{aligned}
\nabla_x \mathcal{L}(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star) &= 0 \\
g_i(\boldsymbol{x}^\star) &\leq 0 \\
\boldsymbol{\lambda}^\star &\geq 0 \\
\lambda_i^\star g_i(\boldsymbol{x}^\star) &= 0, \quad i = 1, 2, \ldots, m
\end{aligned}$$

SQP II

- Iterative update $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k$ and $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \boldsymbol{\mu}_k$

$$
\begin{aligned}
\nabla_x \mathcal{L}(\boldsymbol{x}_{k+1}, \boldsymbol{\lambda}_{k+1}) \approx \nabla_x \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) + \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) \boldsymbol{d}_x + \nabla^2_{x,\lambda} \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) \boldsymbol{\mu}_k &= 0 \\
g_i(\boldsymbol{x}_{k+1}) = g_i(\boldsymbol{x}_k) + \boldsymbol{d}_k^T \nabla_x g_i(\boldsymbol{x}_k) \leq 0, \quad i &= 1, 2, \ldots, m \\
\lambda_{k+1}(i) \left(g_i(\boldsymbol{x}_k) + \boldsymbol{d}_k^T \nabla_x g_i(\boldsymbol{x}_k) \right) = 0, \quad i &= 1, 2, \ldots, m
\end{aligned}
$$

Denote by $\mathbf{G}_k = [\nabla_x g_1(\boldsymbol{x}), \nabla_x g_2(\boldsymbol{x}), \ldots, \nabla_x g_m(\boldsymbol{x})]$ of size $n \times m$ the Jacobian of constraint functions $g(\boldsymbol{x})$.
Note that $\nabla^2_{x,\lambda} \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) = \mathbf{G}_k$

- Simplified KKT conditions

$$
\begin{aligned}
\nabla^2_{xx} \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) \boldsymbol{d}_x + \mathbf{G}_k \boldsymbol{\mu}_k &= -\nabla_x \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) \\
\mathbf{G}_k^T \boldsymbol{d}_k &\leq -g(\boldsymbol{x}_k) \\
\lambda_{k+1}(i) \left(\mathbf{G}_k^T \boldsymbol{d}_k + g_i(\boldsymbol{x}_k) \right)(i) = 0, \quad i &= 1, 2, \ldots, m
\end{aligned}
$$

SQP III

- Equivalent QP with inequality constraints

$$
\begin{aligned}
\min \quad & \frac{1}{2}\boldsymbol{d}^T \nabla^2_{xx}\mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k)\boldsymbol{d} + \boldsymbol{d}^T \nabla_x \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) \\
\text{s.t.} \quad & \mathbf{G}_k^T \boldsymbol{d} + g(\boldsymbol{x}_k) \leq 0
\end{aligned}
$$

$$
\nabla^2_{xx}\mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) = \nabla^2_{xx} f(\boldsymbol{x}_k) + \sum_{i=1}^{m} \lambda_{k,i} \nabla^2_{xx} g(\boldsymbol{x}_k)
$$

$$
\nabla_x \mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k) = \nabla_x f(\boldsymbol{x}_k) + \sum_{i=1}^{m} \lambda_{k,i} \nabla_x g(\boldsymbol{x}_k)
$$

- Disadvantage of SQP
  - Need to compute second derivatives
  - Hessian may not be positive definite

SQP IV

- Powell (1978) suggested to use the Broiden, Fletcher, Goldfarb, and Shanno (BFGF) formula

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\boldsymbol{\gamma}\boldsymbol{\gamma}^T}{\boldsymbol{\gamma}^T \boldsymbol{d}_k} - \frac{\mathbf{H}_k \boldsymbol{d}_k \boldsymbol{d}_k^T \mathbf{H}_k}{\boldsymbol{d}_k^T \mathbf{H}_k \boldsymbol{d}_k}$$

  where $\boldsymbol{\gamma} = \nabla\mathcal{L}(\boldsymbol{x}_{k+1}, \boldsymbol{\lambda}_{k+1}) - \nabla\mathcal{L}(\boldsymbol{x}_k, \boldsymbol{\lambda}_k)$, $\mathbf{H}_k$ approximates the Hessian of the Lagrangian at $\boldsymbol{x}_k, \boldsymbol{\lambda}_k$
  $\boldsymbol{\gamma}$ can be replaced by $\hat{\boldsymbol{\gamma}}$ to keep the approximate Hessian $\mathbf{H}_k$ positive definite

$$\hat{\boldsymbol{\gamma}} = \theta\boldsymbol{\gamma} + (1-\theta)\mathbf{H}_k \boldsymbol{d}_k$$

  where $0 \le \theta \le 1$.
- Summary
  - For the first iteration, use the identity matrix as the Hessian of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$

SQP V

- Solve for the optimum to the QP problem which approximates the original problem. Obtain the dual variables
- Execute line search and check if the penalty function is reduced

$$\phi = f + \sum_i \lambda_i |g_i|$$

- Evaluate the Lagrangian gradient at the new point. Update the Lagrangian Hessian using the BFGS update.

Example for SPQ I

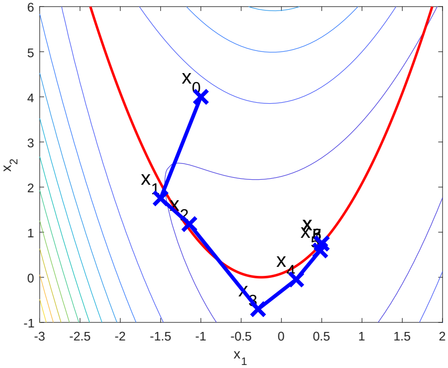
$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = x_1^4 - 2x_1^2x_2 + x_2^2 + x_1^2 - 2x_1 + 5 \\ \text{s.t.} \quad & -(x_1 + 0.25)^2 + 0.75x_2 \geq 0 \end{aligned}$$

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 - 4x_1x_2 + 4x_1^3 - 2 \\ -2x_1^2 + 2x_2 \end{bmatrix}$$

$$\nabla g(\boldsymbol{x}) = \begin{bmatrix} -2x_1 - 1/2 \\ 3/4 \end{bmatrix}$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 12x_1^2 - 4x_2 + 2 & -4x_1 \\ -4x_1 & 2 \end{bmatrix}$$

$$\nabla^2 g(\boldsymbol{x}) = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}$$



Example for SPQ II

- **Step 1:** start with the initial point $\boldsymbol{x}_0 = [-1, 4]^T$. Gradients $\nabla f(\boldsymbol{x}_0) = [8, 6]^T$, $\nabla g(\boldsymbol{x}_0) = [1.5, 0.75]^T$, $g(\boldsymbol{x}_0) = 2.4375$
  - Use identity matrix as the Hessian $\mathbf{H}_0 = \nabla^2_{xx}\mathcal{L} = \mathbf{I}_2$
  - Solve the constrained QP

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & \frac{1}{2}\boldsymbol{d}^T\mathbf{H}_0\boldsymbol{d} + \boldsymbol{d}^T\nabla f(\boldsymbol{x}_0) \\ \text{s.t.} \quad & g(\boldsymbol{x}_0) + \boldsymbol{d}^T\nabla g(\boldsymbol{x}_0) \geq 0 \end{aligned}$$

    and obtain

$$\boldsymbol{d} = [-0.5, -2.25]^T, \quad \boldsymbol{x}_1 = \boldsymbol{x}_0 + \boldsymbol{d} = [-1.5, 1.7]^T, \quad \lambda = 5$$

  - Check if the new point $\boldsymbol{x}_1$ reduces the penalty function

$$\phi(\boldsymbol{x}_0) = f(\boldsymbol{x}_0) = 17 \geq \phi(\boldsymbol{x}_1) = f(\boldsymbol{x}_1) + \lambda|g(\boldsymbol{x}_1)| = 11.75$$

    accept the new point $\boldsymbol{x}_1$. Note that $g(\boldsymbol{x}_1) = -0.25$ violates the constraint.

Example for SPQ III

- **Step 2:** Hessian $\mathbf{H}_1 = \begin{bmatrix} 32 & 6 \\ 6 & 2 \end{bmatrix}$, $\nabla f(\boldsymbol{x}_1) = \begin{bmatrix} -8 \\ -1 \end{bmatrix}$ and $\nabla g(\boldsymbol{x}_1) = \begin{bmatrix} 2.5 \\ 0.75 \end{bmatrix}$, $g(\boldsymbol{x}_1) = -0.25$
  - Solve the constrained QP

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & \frac{1}{2}\boldsymbol{d}^T \mathbf{H}_1 \boldsymbol{d} + \boldsymbol{d}^T \nabla f(\boldsymbol{x}_1) \\ \text{s.t.} \quad & g(\boldsymbol{x}_1) + \boldsymbol{d}^T \nabla g(\boldsymbol{x}_1) \geq 0 \end{aligned}$$

    and obtain

$$\boldsymbol{d} = [0.3571, -0.5714]^T, \quad \boldsymbol{x}_2 = \boldsymbol{x}_1 + \boldsymbol{d} = [-1.1429, 1.1786]^T, \quad \lambda = 0$$

  - Check if the new point $\boldsymbol{x}_2$ reduces the penalty function

$$\phi(\boldsymbol{x}_2) = 8.6081 \leq \phi(\boldsymbol{x}_1) = 11.75$$

    accept the new point $\boldsymbol{x}_2$.

Example for SPQ IV

- **Step 3:** Hessian $\mathbf{H}_2 = \begin{bmatrix} 12.9592 & 4.5714 \\ 4.5714 & 2 \end{bmatrix}$, $\nabla f(\boldsymbol{x}_2) = \begin{bmatrix} -4.8688 \\ -0.2551 \end{bmatrix}$ and $\nabla_x g(\boldsymbol{x}_2) = \begin{bmatrix} 1.7857 \\ 0.75 \end{bmatrix}$, $g(\boldsymbol{x}_2) = 0.0867$
  - Solve the constrained QP

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & \frac{1}{2}\boldsymbol{d}^T \mathbf{H}_2 \boldsymbol{d} + \boldsymbol{d}^T \nabla f(\boldsymbol{x}_2) \\ \text{s.t.} \quad & g(\boldsymbol{x}_2) + \boldsymbol{d}^T \nabla g(\boldsymbol{x}_2) \geq 0 \end{aligned}$$

    and obtain $\boldsymbol{d} = [1.7073, -3.7749]^T$, $\boldsymbol{x}_3 = \boldsymbol{x}_2 + \boldsymbol{d} = [0.5645, -2.5963]^T$, $\lambda = 0$
  - Check if the new point $\boldsymbol{x}_3$ reduces the penalty function

$$\phi(\boldsymbol{x}_3) = 12.6865 \geq \phi(\boldsymbol{x}_2) = 8.6081$$

    reject $\boldsymbol{x}_3$, and execute linesearch which yields

$$\alpha^* = 0.4995, \quad \boldsymbol{x}_3 = \boldsymbol{x}_2 + \alpha^* \boldsymbol{d} = [-0.2901, -0.7070]^T$$

Example for SPQ V

- **Step 4:** Hessian $\mathbf{H}_3 = \begin{bmatrix} 5.8375 & 1.1602 \\ 1.1602 & 2.0000 \end{bmatrix}$, $\nabla f(\boldsymbol{x}_3) = \begin{bmatrix} -3.4980 \\ -1.5822 \end{bmatrix}$ and $\nabla g(\boldsymbol{x}_3) = \begin{bmatrix} 0.0801 \\ 0.7500 \end{bmatrix}$, $g(\boldsymbol{x}_3) = -0.5318$
  - Solve the constrained QP

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & \frac{1}{2}\boldsymbol{d}^T\mathbf{H}_3\boldsymbol{d} + \boldsymbol{d}^T\nabla f(\boldsymbol{x}_3) \\ \text{s.t.} \quad & g(\boldsymbol{x}_3) + \boldsymbol{d}^T\nabla g(\boldsymbol{x}_3) \geq 0 \end{aligned}$$

    and obtain $\boldsymbol{d} = [0.4735, 0.6585]^T$, $\boldsymbol{x}_4 = \boldsymbol{x}_3 + \boldsymbol{d} = [0.1835, -0.0484]^T$, $\lambda = 0.3790$
  - Check if $\boldsymbol{x}_4$ reduces the penalty function

$$\phi(\boldsymbol{x}_4) = 4.7584 \leq \phi(\boldsymbol{x}_3) = 6.2901$$

    accept $\boldsymbol{x}_4$

Example for SPQ VI

- **Step 5:** Hessian $\mathbf{H}_4 = \begin{bmatrix} 3.3558 & -0.7339 \\ -0.7339 & 2.0000 \end{bmatrix}$, $\nabla f(\boldsymbol{x}_4) = \begin{bmatrix} -1.5728 \\ -0.1642 \end{bmatrix}$ and $\nabla g(\boldsymbol{x}_4) = \begin{bmatrix} -0.8670 \\ 0.7500 \end{bmatrix}$, $g(\boldsymbol{x}_4) = -0.2242$
  - Solve the constrained QP

$$\begin{aligned} \min_{\boldsymbol{d}} \quad & \frac{1}{2}\boldsymbol{d}^T \mathbf{H}_4 \boldsymbol{d} + \boldsymbol{d}^T \nabla f(\boldsymbol{x}_4) \\ \text{s.t.} \quad & g(\boldsymbol{x}_4) + \boldsymbol{d}^T \nabla g(\boldsymbol{x}_4) \geq 0 \end{aligned}$$

    and obtain $\boldsymbol{d} = [0.2980, 0.6435]^T$, $\boldsymbol{x}_5 = \boldsymbol{x}_4 + \boldsymbol{d} = [0.4815, 0.5950]^T$, $\lambda = 1.2053$
  - Check if $\boldsymbol{x}_5$ reduces the penalty function

$$\phi(\boldsymbol{x}_5) = 4.5078 \leq \phi(\boldsymbol{x}_4) = 4.7584$$

    accept $\boldsymbol{x}_5$
- Continue the update, the optimal solution $\boldsymbol{x}_7 = [0.5, 0.75]^T$ with $f(\boldsymbol{x}^\star) = 4.5$

Example for SQP I

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) = \frac{1}{2}((x_1 - x_3)^2 + (x_2 - x_4)^2) \\
\text{s.t.} \quad & -[x_1, x_2]\begin{bmatrix} \frac{1}{4} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [x_1, x_2]\begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \frac{3}{4} \geq 0 \\
& -\frac{1}{8}[x_3, x_4]\begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} x_3 \\ x_4 \end{bmatrix} + \frac{1}{2}[x_3, x_4]\begin{bmatrix} 11 \\ 13 \end{bmatrix} - \frac{35}{2} \geq 0
\end{aligned}
$$

$\boldsymbol{d}$	$\boldsymbol{x}_k$	$f(\boldsymbol{x}_k)$
	$[1, 0.5, 2, 3]$	3.6250
$[1.0000, 0.7500, 0.2353, -0.4412]$	$[2.0000, 1.2500, 2.2353, 2.5588]$	0.8842
$[-0.0549, -0.3140, 0.1967, -0.0470]$	$[1.9451, 0.9360, 2.4320, 2.5118]$	1.3602
$[0.0960, -0.0773, 0.1119, -0.0298]$	$[2.0410, 0.8587, 2.5439, 2.4820]$	1.4440
$[0.0038, -0.0060, 0.0011, 0.0036]$	$[2.0448, 0.8527, 2.5449, 2.4856]$	1.4582
$[0, 0, 0, 0]$	$[2.0447, 0.8527, 2.5449, 2.4856]$	1.4583