Optimization

Yuh-Jye Lee

Data Science and Machine Intelligence Lab
National Chiao Tung University

March 28, 2017

The Key Idea of Newton's Method

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a twice differentiable function

$$f(x+d) = f(x) + \nabla f(x)^\top d + \frac{1}{2} d^\top \nabla^2 f(x) d + \beta(x,d) \parallel d \parallel$$

where $\lim_{d \to 0} \beta(x,d) = 0$

At i^{th} iteration, use a quadratic function to approximate

$$f(x) \approx f(x^i) + \nabla f(x^i)(x - x^i) + \frac{1}{2}(x - x^i)^\top \nabla^2 f(x^i)(x - x^i)$$

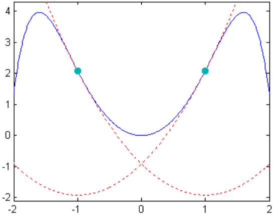
$x^{i+1} = \arg\min \tilde{f}(x)$

Newton's Method

Start with $x^0 \in \mathbb{R}^n$. Having x^i ,stop if $\nabla f(x^i) = 0$
Else compute x^{i+1} as follows:

1. Newton direction: $\nabla^2 f(x^i) d^i = -\nabla f(x^i)$
   Have to solve a system of linear equations here!
2. Updating: $x^{i+1} = x^i + d^i$
   - Converge only when x^0 is close to x^* enough.

Newton's Method with BAD Initial Point



$$f(x) = \frac{-1}{6}x^6 + \frac{1}{4}x^4 + 2x^2$$
$$g_i(x) = f(x^i) + f'(x^i)(x - x^i) + \frac{1}{2}f''(x^i)(x - x^i)^2$$
$$g_1(x) = f(1) + 4(x - 1) + (x - 1)^2$$
$$g_2(x) = f(-1) + 4(x + 1) + (x + 1)^2$$
$$g_1'(-1) = g_2'(1) = 0$$

It can not converge to the optimal solution.

Constrained Optimization Problem

Problem setting: Given function f, g_i, $i = 1, ..., k$ and h_j, $j = 1, ..., m$, defined on a domain $\Omega \subseteq \mathbb{R}^n$,

$$
\begin{aligned}
\min_{x\in\Omega} \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \leq 0, \quad \forall i \\
& h_j(x) = 0, \quad \forall j
\end{aligned}
$$

where $f(x)$ is called the objective function and $g(x) \leq 0$, $h(x) = 0$ are called constrains.

Example I

$$\begin{aligned} \min \quad & f(x) = 2x_1^2 + x_2^2 + 3x_3^2 \\ \text{s.t.} \quad & 2x_1 - 3x_2 + 4x_3 = 49 \end{aligned}$$

<sol>

$$L(x, \beta) = f(x) + \beta(2x_1 - 3x_2 + 4x_3 - 49),\ \beta \in \mathbb{R}$$

$$\begin{aligned} \frac{\partial}{\partial x_1} L(x, \beta) = 0 \quad & \Rightarrow \quad 4x_1 + 2\beta = 0 \\ \frac{\partial}{\partial x_2} L(x, \beta) = 0 \quad & \Rightarrow \quad 2x_2 - 3\beta = 0 \\ \frac{\partial}{\partial x_3} L(x, \beta) = 0 \quad & \Rightarrow \quad 6x_3 + 4\beta = 0 \end{aligned}$$

$$2x_1 - 3x_2 + 4x_3 - 49 = 0 \Rightarrow \beta = -6$$

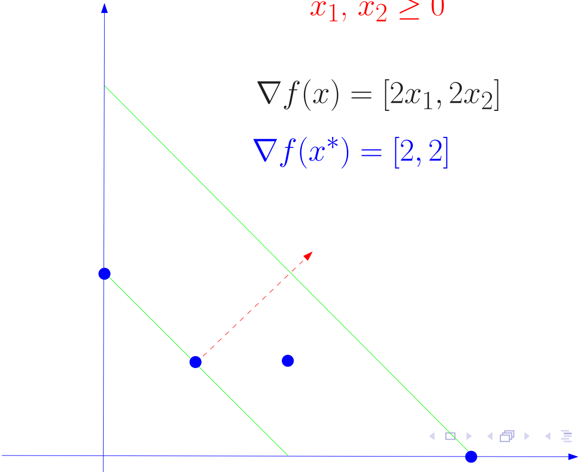
$$\Rightarrow x_1 = 3,\ x_2 = -9,\ x_3 = 4$$

Example II

$$\min_{x \in R^2} x_1^2 + x_2^2 \qquad \begin{aligned} x_1 + x_2 &\leq 4 \\ -x_1 - x_2 &\leq -2 \\ x_1, x_2 &\geq 0 \end{aligned}$$

$$\nabla f(x) = [2x_1, 2x_2]$$

$$\nabla f(x^*) = [2, 2]$$



Definitions and Notation

- Feasible region:

$$\mathcal{F} = \{x \in \Omega \mid g(x) \leq 0, h(x) = 0\}$$

where $g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_k(x) \end{bmatrix}$ and $h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix}$

- A solution of the optimization problem is a point $x^* \in \mathcal{F}$ such that $\nexists x \in \mathcal{F}$ for which $f(x) < f(x^*)$ and x^* is called a global minimum.

Definitions and Notation

- A point $\bar{x} \in \mathcal{F}$ is called a local minimum of the optimization problem if $\exists \varepsilon > 0$ such that

$$f(x) \geq f(\bar{x}), \qquad \forall x \in \mathcal{F} \text{ and } \|x - \bar{x}\| < \varepsilon$$

- At the solution x^*, an inequality constraint $g_i(x)$ is said to be active if $g_i(x^*) = 0$, otherwise it is called an inactive constraint.
- $g_i(x) \leq 0 \Leftrightarrow g_i(x) + \xi_i = 0$, $\xi_i \geq 0$ where ξ_i is called the slack variable

Definitions and Notation

- Remove an inactive constraint in an optimization problem will NOT affect the optimal solution
  - Very useful feature in SVM
- If $\mathcal{F} = \mathbb{R}^n$ then the problem is called unconstrained minimization problem
  - Least square problem is in this category
  - SSVM formulation is in this category
  - Difficult to find the global minimum without convexity assumption

The Most Important Concepts in Optimization(minimization)

- A point is said to be an *optimal solution* of a unconstrained minimization if there exists no decent direction
  $\Longrightarrow \nabla f(x^*) = 0$
- A point is said to be an optimal solution of a constrained minimization if there exists no feasible decent direction
  $\Longrightarrow$ KKT conditions
    - There might exist decent direction but move along this direction will leave out the feasible region

Minimum Principle

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex and *continuously differentiable* function $\mathcal{F} \subseteq \mathbb{R}^n$ be the feasible region.

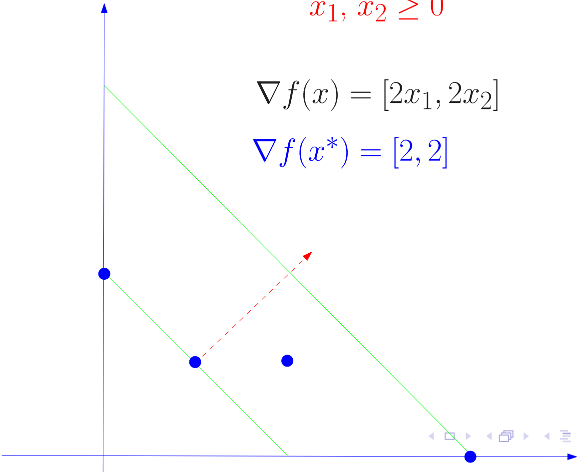
$$x^* \in \arg\min_{x \in \mathcal{F}} f(x) \iff \nabla f(x^*)(x - x^*) \geq 0 \quad \forall x \in \mathcal{F}$$

Example:

$$\min (x-1)^2 \quad \text{s.t.} \quad a \leq x \leq b$$

Example II

$$\min_{x \in R^2} x_1^2 + x_2^2 \quad \begin{array}{c} x_1 + x_2 \leq 4 \\ -x_1 - x_2 \leq -2 \\ x_1, x_2 \geq 0 \end{array}$$



$$\nabla f(x) = [2x_1, 2x_2]$$

$$\nabla f(x^*) = [2, 2]$$

Linear Programming Problem

- An optimization problem in which the objective function and all constraints are linear functions is called a linear programming problem

$$
\begin{aligned}
\text{(LP)} \qquad \min \quad & p^\top x \\
\text{s.t.} \quad & Ax \leq b \\
& Cx = d \\
& L \leq x \leq U
\end{aligned}
$$

Linear Programming Solver in MATLAB

X=LINPROG(f,A,b) attempts to solve the linear programming problem:

$$\min_x \quad f'*x \quad \text{subject to: } A*x <= b$$

X=LINPROG(f,A,b,Aeq,beq) solves the problem above while additionally satisfying the equality constraints Aeq*x = beq.

X=LINPROG(f,A,b,Aeq,beq,LB,UB) defines a set of lower and upper bounds on the design variables, X, so that the solution is in the range LB <= X <= UB.
Use empty matrices for LB and UB if no bounds exist. Set LB(i) = -Inf if X(i) is unbounded below; set UB(i) = Inf if X(i) is unbounded above.

Linear Programming Solver in MATLAB

X=LINPROG(f,A,b,Aeq,beq,LB,UB,X0) sets the starting point to X0. This option is only available with the active-set algorithm. The default interior point algorithm will ignore any non-empty starting point.

You can type “help linprog” in MATLAB to get more information!

L_1-Approximation: $\min_{x \in \mathbb{R}^n} \|Ax - b\|_1$

$$\|z\|_1 = \sum_{i=1}^{m} |z_i|$$

$$\begin{aligned} &\min_{x,s} \mathbf{1}^\top s \\ &\text{s.t. } -s \leq Ax - b \leq s \end{aligned}$$

Or

$$\begin{aligned} &\min_{x,s} \sum_{i=1}^{m} s_i \\ &\text{s.t. } -s_i \leq A_i x - b_i \leq s_i \quad \forall i \end{aligned}$$

$$\begin{aligned} &\min_{x,s} \begin{bmatrix} 0 & \cdots & 0 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} \\ &\text{s.t. } \begin{bmatrix} A & -I \\ -A & -I \end{bmatrix}_{2m \times (n+m)} \begin{bmatrix} x \\ s \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix} \end{aligned}$$

Chebyshev Approximation: $\min_{x \in \mathbb{R}^n} \|Ax - b\|_\infty$

$$\|z\|_\infty = \max_{1 \leq i \leq m} |z_i|$$

$$\begin{aligned} \min_{x,\gamma} \quad & \gamma \\ \text{s.t.} \quad & -\mathbf{1}\gamma \leq Ax - b \leq \mathbf{1}\gamma \end{aligned}$$

$$\begin{aligned} \min_{x,s} \; & \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ \gamma \end{bmatrix} \\ \text{s.t.} \; & \begin{bmatrix} A & -\mathbf{1} \\ -A & -\mathbf{1} \end{bmatrix}_{2m \times (n+1)} \begin{bmatrix} x \\ \gamma \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix} \end{aligned}$$

Quadratic Programming Problem

- If the objective function is convex quadratic while the constraints are all linear then the problem is called convex quadratic programming problem

$$
\begin{aligned}
\text{(QP)} \qquad \min \quad & \frac{1}{2} x^\top Q x + p^\top x \\
\text{s.t.} \quad & Ax \leq b \\
& Cx = d \\
& L \leq x \leq U
\end{aligned}
$$

Quadratic Programming Solver in MATLAB

X=QUADPROG(H,f,A,b) attempts to solve the quadratic programming problem:

$$\min_{x} \quad 0.5*x'*H*x+f'*x \quad \text{subject to: } A*x <= b$$

X=QUADPROG(H,f,A,b,Aeq,beq) solves the problem above while additionally satisfying the equality constraints Aeq*x=beq.

X=QUADPROG(H,f,A,b,Aeq,beq,LB,UB) defines a set of lower and upper bounds on the design variables, X, so that the solution is in the range LB <= X <= UB.
Use empty matrices for LB and UB if no bounds exist. Set LB(i) = -Inf if X(i) is unbounded below; set UB(i) = Inf if X(i) is unbounded above.

Quadratic Programming Solver in MATLAB

X=QUADPROG(H,f,A,b,Aeq,beq,LB,UB,X0) sets the starting point to X0.

You can type "help quadprog" in MATLAB to get more information!

Standard Support Vector Machine

$$\min_{w,b,\xi_A,\xi_B} C(\mathbf{1}^\top \xi_A + \mathbf{1}^\top \xi_B) + \frac{1}{2}\|w\|_2^2$$

$$(Aw + \mathbf{1}b) + \xi_A \geq \mathbf{1}$$
$$(Bw + \mathbf{1}b) - \xi_B \leq -\mathbf{1}$$
$$\xi_A \geq 0, \xi_B \geq 0$$

Farkas' Lemma

For any matrix $A \in \mathbb{R}^{m \times n}$ and any vector $b \in \mathbb{R}^n$, either

$$Ax \leq 0, \quad b^\top x > 0 \;\text{ has a solution}$$

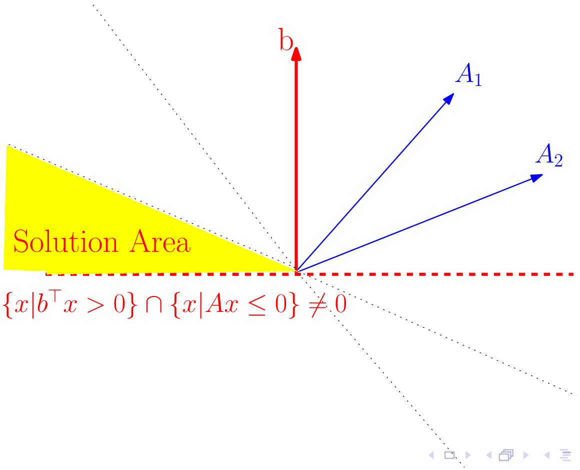
or

$$A^\top \alpha = b, \quad \alpha \geq 0 \;\text{ has a solution}$$

but never both.

Farkas' Lemma
$Ax \le \mathbf{0}$, $b^\top x > 0$ has a solution

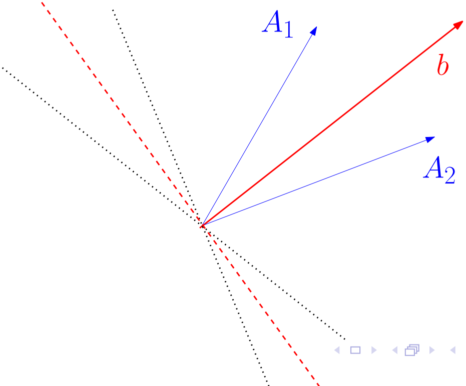
b is NOT in the cone generated by A_1 and A_2



Farkas' Lemma
$A^\top \alpha = b$, $\alpha \geq 0$ has a solution
b is in the cone generated by A_1 and A_2

$$\{x | b^\top x > 0\} \cap \{x | Ax \leq 0\} = \emptyset$$

$$\{x | b^\top > 0\} \cap \{x | Ax \leq 0\} = \emptyset$$



Minimization Problem *vs.* Kuhn-Tucker Stationary-point Problem

MP:

$$\begin{aligned} \min_{x \in \Omega} \quad & f(x) \\ s.t. \quad & g(x) \leq 0 \end{aligned}$$

KTSP:

$$\begin{aligned} \text{Find} \quad & \bar{x} \in \Omega,\ \bar{\alpha} \in \mathbb{R}^m \ \text{ such that} \\ & \nabla f(\bar{x}) + \bar{\alpha}^\top \nabla g(\bar{x}) = 0 \\ & \bar{\alpha}^\top g(\bar{x}) = 0 \\ & g(\bar{x}) \leq 0 \\ & \bar{\alpha} \geq 0 \end{aligned}$$

Lagrangian Function $\mathcal{L}(x, \alpha) = f(x) + \alpha^\top g(x)$

Let $\mathcal{L}(x, \alpha) = f(x) + \alpha^\top g(x)$ and $\alpha \geq 0$

- If $f(x)$, $g(x)$ are convex the $\mathcal{L}(x, \alpha)$ is convex.
- For a fixed $\alpha \geq 0$, if $\bar{x} \in \arg\min\{\mathcal{L}(x, \alpha) | x \in \mathbb{R}^n\}$ then
$$\left.\frac{\partial \mathcal{L}(x, \alpha)}{\partial x}\right|_{x=\bar{x}} = \nabla f(\bar{x}) + \alpha^\top \nabla g(\bar{x}) = 0$$
- Above result is a sufficient condition if $\mathcal{L}(x, \alpha)$ is convex.

KTSP with Equality Constraints?
(Assume $h(x) = 0$ are linear functions)

$$h(x) = 0 \;\Leftrightarrow\; h(x) \leq 0 \text{ and } -h(x) \leq 0$$

KTSP:

Find $\quad \bar{x} \in \Omega, \bar{\alpha} \in \mathbb{R}^k, \bar{\beta}_+, \bar{\beta}_- \in \mathbb{R}^m$ such that

$$\nabla f(\bar{x}) + \bar{\alpha}^\top \nabla g(\bar{x}) + (\bar{\beta}_+ - \bar{\beta}_-)^\top \nabla h(\bar{x}) = 0$$

$$\bar{\alpha}^\top g(\bar{x}) = 0,\; (\bar{\beta}_+)^\top h(\bar{x}) = 0,\; (\bar{\beta}_-)^\top (-h(\bar{x})) = 0$$

$$g(\bar{x}) \leq 0,\; h(\bar{x}) = 0$$

$$\bar{\alpha} \geq 0,\; \bar{\beta}_+,\; \bar{\beta}_- \geq 0$$

KTSP with Equality Constraints

KTSP:

$$\begin{aligned} \text{Find} \quad & \bar{x} \in \Omega, \bar{\alpha} \in \mathbb{R}^k, \bar{\beta} \in \mathbb{R}^m \ \text{ such that} \\ & \nabla f(\bar{x}) + \bar{\alpha}^\top \nabla g(\bar{x}) + \bar{\beta} \nabla h(\bar{x}) = 0 \\ & \bar{\alpha}^\top g(\bar{x}) = 0, \ g(\bar{x}) \leq 0, \ h(\bar{x}) = 0 \\ & \bar{\alpha} \geq 0 \end{aligned}$$

- Let $\bar{\beta} = \bar{\beta}_+ - \bar{\beta}_-$ and $\bar{\beta}_+, \bar{\beta}_- \geq 0$
  then $\bar{\beta}$ is free variable

Generalized Lagrangian Function $\mathcal{L}(x, \alpha, \beta) = f(x) + \alpha^\top g(x) + \beta^\top h(x)$

Let $\mathcal{L}(x, \alpha, \beta) = f(x) + \alpha^\top g(x) + \beta^\top h(x)$ and $\alpha \geq 0$

- If $f(x)$, $g(x)$ are convex and $h(x)$ is linear then $\mathcal{L}(x, \alpha, \beta)$ is convex.
- For fixed $\alpha \geq 0$, if $\bar{x} \in \arg\min\{\mathcal{L}(x, \alpha, \beta) | x \in \mathbb{R}^n\}$ then

$$\left.\frac{\partial \mathcal{L}(x, \alpha, \beta)}{\partial x}\right|_{x=\bar{x}} = \nabla f(\bar{x}) + \alpha^\top \nabla g(\bar{x}) + \beta^\top \nabla h(\bar{x}) = 0$$

- Above result is a sufficient condition if $\mathcal{L}(x, \alpha, \beta)$ is convex.

Lagrangian Dual Problem

$$
\begin{aligned}
\max_{\alpha,\beta} \min_{x \in \Omega} \quad & \mathcal{L}(x, \alpha, \beta) \\
s.t. \quad & \alpha \geq 0
\end{aligned}
$$

Lagrangian Dual Problem

$$\max_{\alpha,\beta} \min_{x\in\Omega} \quad \mathcal{L}(x,\alpha,\beta)$$
$$s.t. \quad \alpha \geq 0$$

$$\Updownarrow$$

$$\max_{\alpha,\beta} \quad \theta(\alpha,\beta)$$
$$s.t. \quad \alpha \geq 0$$
$$\text{where} \quad \theta(\alpha,\beta) = \inf_{x\in\Omega} \mathcal{L}(x,\alpha,\beta)$$

Weak Duality Theorem

Let $\bar{x} \in \Omega$ be a feasible solution of the primal problem and (α, β) a feasible sulution of the *dual* problem. then $f(\bar{x}) \geq \theta(\alpha, \beta)$

$$\theta(\alpha, \beta) = \inf_{x \in \Omega} \mathcal{L}(x, \alpha, \beta) \leq \mathcal{L}(\tilde{x}, \alpha, \beta)$$

Corollary:

$$\sup\{\theta(\alpha, \beta) | \alpha \geq 0\} \leq \inf\{f(x) | g(x) \leq 0,\ h(x) = 0\}$$

Weak Duality Theorem

Corollary

If $f(x^*) = \theta(\alpha^*, \beta^*)$ where $\alpha^* \geq \mathbf{0}$ and $g(x^*) \leq \mathbf{0}$, $h(x^*) = \mathbf{0}$,then x^* and (α^*, β^*) solve the *primal* and *dual* problem respectively. In this case,

$$\mathbf{0} \leq \alpha \perp g(x) \leq \mathbf{0}$$

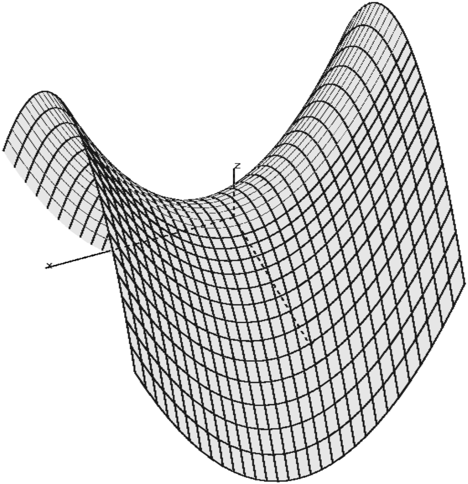
Saddle Point of Lagrangian

Let $x^* \in \Omega$,$\alpha^* \geq \mathbf{0}$, $\beta^* \in \mathbb{R}^m$ satisfying

$$\mathcal{L}(x^*, \alpha, \beta) \leq \mathcal{L}(x^*, \alpha^*, \beta^*) \leq \mathcal{L}(x, \alpha^*, \beta^*) \ , \forall x \in \Omega \ , \ \alpha \geq \mathbf{0}$$

Then (x^*, α^*, β^*) is called The saddle point of the Lagrangian function

Saddle Point of $f(x, y) = x^2 - y^2$



Dual Problem of Linear Program

$$\textit{Primal}\ \text{LP} \qquad \min_{x \in \mathbb{R}^n} \quad p^\top x \qquad \text{subject to} \quad Ax \geq b\ ,\ x \geq \mathbf{0}$$

$$\textit{Dual}\ \text{LP} \qquad \max_{\alpha \in \mathbb{R}^m} \quad b^\top \alpha \qquad \text{subject to} \quad A^\top \alpha \leq p\ ,\ \alpha \geq \mathbf{0}$$

- All duality theorems hold and work perfectly!

Lagrangian Function of Primal LP
$\mathcal{L}(x, \alpha) = p^\top x + \alpha_1^\top (b - Ax) + \alpha_2^\top (-x)$

$$\max_{\alpha_1, \alpha_2 \geq \mathbf{0}} \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \alpha_1, \alpha 2)$$

$$\Updownarrow$$

$$\begin{aligned} \max_{\alpha_1, \alpha_2 \geq \mathbf{0}} \quad & p^\top x + \alpha_1^\top (b - Ax) + \alpha_2^\top (-x) \\ \text{subject to} \quad & p - A^\top \alpha_1 - \alpha_2 = \mathbf{0} \\ & (\nabla_x \mathcal{L}(x, \alpha_1, \alpha_2) = \mathbf{0}) \end{aligned}$$

Application of LP Duality
LSQ − NormalEquation Always Has a Solution

For any matrix $A \in \mathbb{R}^{mxn}$ and any vector $b \in \mathbb{R}^{m}$,
consider $\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$

$$x^* \in \arg\min\{\|Ax - b\|_2^2\} \Leftrightarrow A^\top A x^* = A^\top b$$

Claim : $A^\top Ax = A^\top b$ always has a solution.

Dual Problem of Strictly Convex Quadratic Program

Primal QP

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} x^\top Q x + p^\top x \\ \text{s.t.} \quad & Ax \leq b \end{aligned}$$

With *strictlyconvex* assumption, we have

Dual QP

$$\begin{aligned} \max \quad & -\frac{1}{2}(p^\top + \alpha^\top A) Q^{-1} (A^\top \alpha + p) - \alpha^\top b \\ \text{s.t.} \quad & \alpha \geq \mathbf{0} \end{aligned}$$