Statistics and Machine Learning

Fall, 2005

Lecture 2: Optimization

李育杰

yuh-jye@mail.ntust.edu.tw

National Taiwan University of Science and Technology

You Have Learned (Unconstrained) Optimization in Your High School

Let $f(x) = ax^2 + bx + c, \;\; a \neq 0, \;\; x^* = -\frac{b}{2a}$

Case I: $f''(x^*) = a > 0 \Rightarrow x^* \in \arg\min_{x \in R} f(x)$

Case II: $f''(x^*) = a < 0 \Rightarrow x^* \in \arg\max_{x \in R} f(x)$

For minimization problem (Case I),

$f'(x^*) = 0$ is called the first order optimality condition

$f''(x^*) = a > 0$ is the second order optimality condition

Optimization Examples in this Book

On p.62, Maximum likelihood estimation

On p.68, Maximum a Posteriori estimation

On p.74, Least squares estimates

On p.207, Gradient descent method

On p.246, Backpropagation

Gradient and Hessian

- Let $f : R^n \rightarrow R$ be a differentiable function. The gradient of function f at a point $x \in R^n$ is defined as $\nabla f(x) = [\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n}] \in R^n$
- If $f : R^n \rightarrow R$ is a twice differentiable function. The Hessian matrix of f at a point $x \in R^n$ is defined as

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \in R^{n \times n}$$

Example of Gradient and Hessian

$$f(x) = x_1^2 + x_2^2 - 2x_1 + 4x_2$$

$$= \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\nabla f(x) = [2x_1 - 2, 2x_2 + 4] \qquad \nabla^2 f(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

By letting $\nabla f(x) = 0$, we have

$$x^* = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \in \arg\min_{x \in R^2} f(x)$$

Quadratic Functions (Standard Form)

$$f(x) = \frac{1}{2}x^T H x + p^T x$$

Let $f : R^n \to R$ and $f(x) = \frac{1}{2}x^T H x + p^T x$

where $H \in R^{n\times n}$ is a symmetric matrix and $p \in R^n$

then $\nabla f(x) = Hx + p$

$\nabla^2 f(x) = H$ (Hessian)

Note : If H is positive definite, then $x^* = -H^{-1}p$ is the unique solution of $\min f(x)$

Least-squares Problem

$$\min_{x \in R^n} \left\| Ax - b \right\|_2^2, \quad A \in R^{m \times n}, \quad b \in R^m$$

$$f(x) = (Ax - b)^T (Ax - b)$$
$$= x^T A^T A x - 2b^T A x + b^T b$$

$$\nabla f(x) = 2A^T A x - 2A^T b$$
$$\nabla^2 f(x) = 2A^T A$$

$$x^* = (A^T A)^{-1} A^T b \in \arg \min_{x \in R^n} \left\| Ax - b \right\|_2^2$$

if $A^T A$ is nonsingular matrix $\Rightarrow$ P.D.

Note : x^* is an analytical solution

How to Solve an Unconstrained MP

- Get an initial point and iteratively decrease the obj. function value
- Stop once the stopping criteria satisfied
- Steep decent might not be a good choice
- Newton's method is highly recommended
  - Local and quadratic convergent algorithm
  - Need to choose a good step size to guarantee global convergence

The First Order Taylor Expansion

Let $f : R^n \to R$ be a differentiable function

$$f(x+d) = f(x) + \nabla f(x) \cdot d + \alpha(x,d)\left\|d\right\|$$

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

If $\nabla f(x) d < 0$ and d is small enough

then $f(x+d) < f(x)$.

We call d is a descent direction.

Steep Descent with Exact Line Search

Start with any $x^0 \in R^n$. Having x^i, stop if $\nabla f(x^i) = 0$

Else compute x^{i+1} as follows:

(i) Steep descent direction: $d^i = -\nabla f(x^i)$

(ii) Exact line search: Choose a stepsize $\lambda \in R$ such that

$$\frac{df(x^i + \lambda d^i)}{d\lambda} = f'(x^i + \lambda d^i) = 0$$

(iii) Updating: $x^{i+1} = x^i + \lambda d^i$

MATLAB Code for
Steep Descent with Exact Line Search
(Quadratic Function Only)

```
function [x, f_value, iter] = grdlines(Q,p, x0, esp)


%
% min 0.5*x'Q*x+p'x
% Solving unconstrained minimization via
% steep descent with exact line search
%
```

```
flag =1;
iter = 0;
while flag > esp
        grad = Q*x0+p;
        temp1 = grad'*grad;
        if temp1 < 10^-12
            flag = esp
        else
            stepsize = temp1/(grad'*Q*grad);
            x1 = x0 - stepsize*grad;
            flag = norm(x1-x0);
            x0=x1;
        end;
        iter = iter+1;
 end;
 x = x0;
 f_value = 0.5*x'*Q*x+p'*x;
```

The Key Idea of Newton's Method

Let $f : R^n \to R$ be a twice differentiable function

$$f(x+d) = f(x) + \nabla f(x) \cdot d + \frac{1}{2} x^T \nabla^2 f(x) x + \beta(x,d) \|d\|$$

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

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

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

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

Newton's Method

Start with $x^0 \in R^n$. Having x^i , stop if $\nabla f(x^i) = 0$

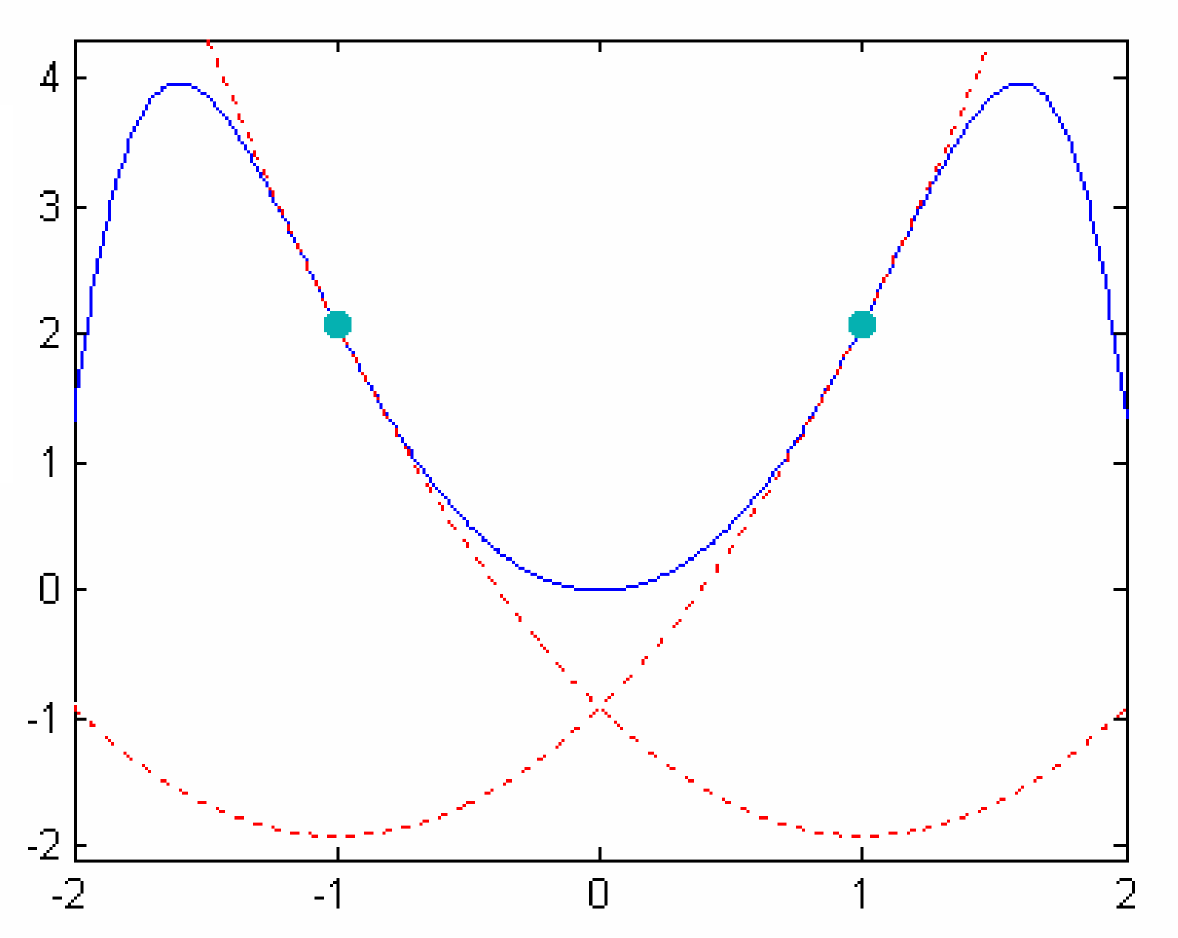
Else compute x^{i+1} as follows:

(i) Newton direction: $\nabla^2 f(x^i) d^i = -\nabla f(x^i).$

Have to solve a system of linear equations here!

(ii) Updating: $x^{i+1} = x^i + d^i$

- Converge only when x^0 is close to x^* enough.



$$f(x) = -\frac{1}{6}x^6 + \frac{1}{4}x^4 + 2x^2$$

$$g(x) = f(x^i) + f'(x^i)(x - x^i) + \frac{1}{2}f''(x^i)(x - x^i)$$

It can not converge to the optimal solution.

Constrained Optimization Problem

Problem setting: Given functions $f, \; g_i, \; i = 1, \ldots, \; k$ and $h_j, \; j = 1, \ldots, \; m$, defined on a domain $\Omega \subseteq R^n$,

$$\min_{x \in \Omega} \; f(x)$$

$$\textit{subject to} \quad \begin{aligned} g_i(x) &\leqslant 0, \quad \forall i \\ h_j(x) &= 0, \quad \forall j \end{aligned}$$

where $f(x)$ is called the objective function and $\mathbf{g}(x) \leqslant \mathbf{0}, \; \mathbf{h}(x) = \mathbf{0}$ are called constraints.

Example

$$\min \quad f(x) = 2x_1^2 + x_2^2 + 3x_3^2$$

s.t. $\quad 2x_1 - 3x_2 + 4x_3 = 49$

<sol>

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

$$\frac{\partial}{\partial x_1}\mathcal{L}(x, \beta) = 0 \quad \Rightarrow \quad 4x_1 + 2\beta = 0$$

$$\frac{\partial}{\partial x_2}\mathcal{L}(x, \beta) = 0 \quad \Rightarrow \quad 2x_2 - 3\beta = 0$$

$$\frac{\partial}{\partial x_3}\mathcal{L}(x, \beta) = 0 \quad \Rightarrow \quad 6x_3 + 4\beta = 0$$

$$2x_1 - 3x_2 + 4x_3 - 49 = 0$$

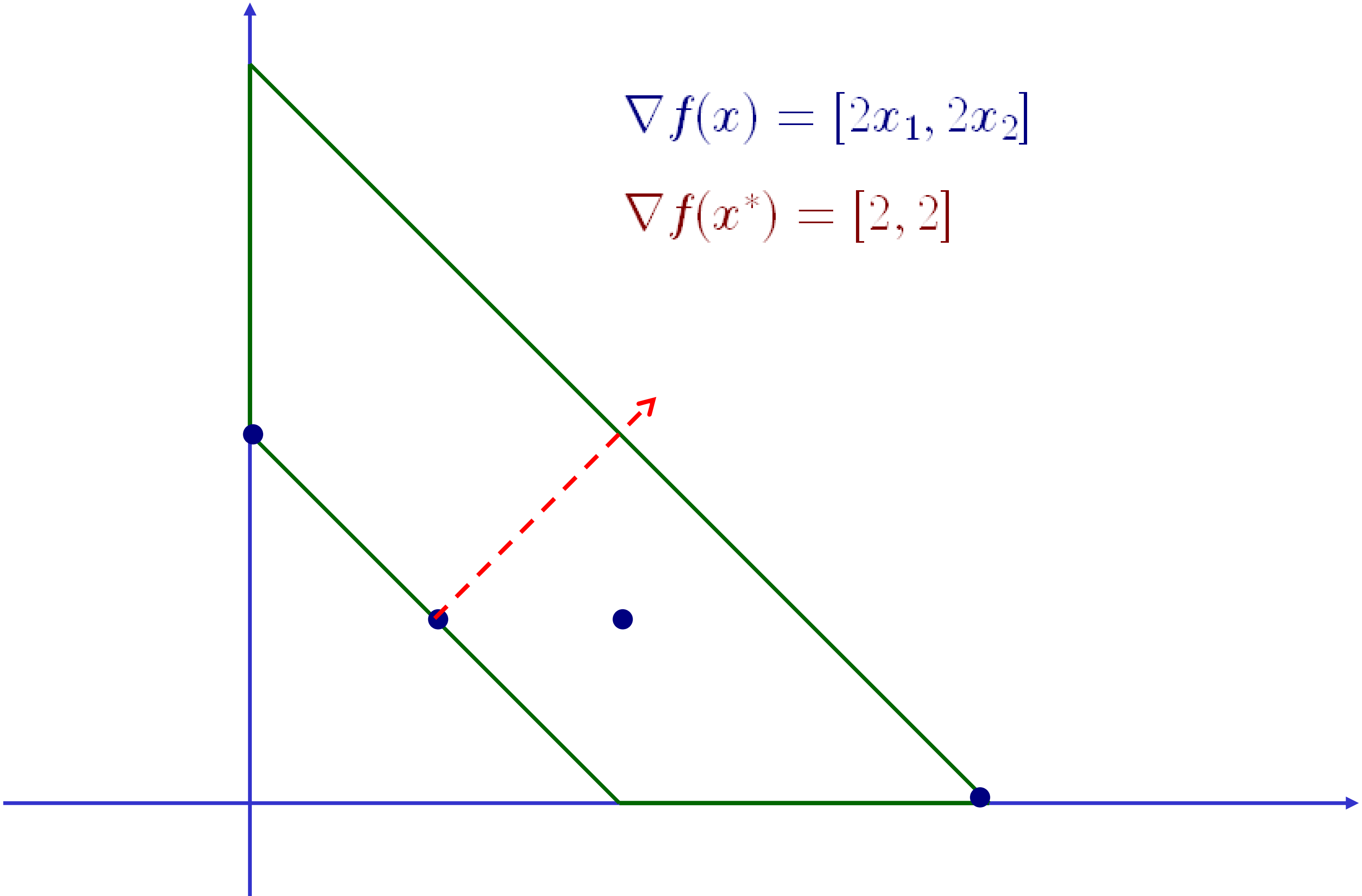
$$\Rightarrow x_1 = 3, \quad x_2 = -9, \quad x_3 = -4$$

$$\min_{x \in R^2} \; x_1^2 + x_2^2$$

$$x_1 + x_2 \leqslant 4$$
$$-x_1 - x_2 \leqslant -2$$
$$x_1, \; x_2 \geqslant 0$$

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

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



Definitions and Notation

- Feasible region:

$$\mathcal{F} = \{x \in \Omega | \ \mathbf{g}(x) \leqslant \mathbf{0}, \ \ \mathbf{h}(x) = \mathbf{0}\}$$

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

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

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

- 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 $\mathcal{F} = R^n$ 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 ➡ $\nabla f(x^*) = 0$

- A point is said to be an optimal solution of a constrained minimization if there exists no feasible decent direction ➡ *KKT* conditions
  - There might exist decent direction but move along this direction will leave out the feasible region

Minimum Principle

Let $f : R^n \to R$ be a convex and differentiable function

$\mathfrak{F} \subseteq R^n$ be the feasible region.

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

Example:

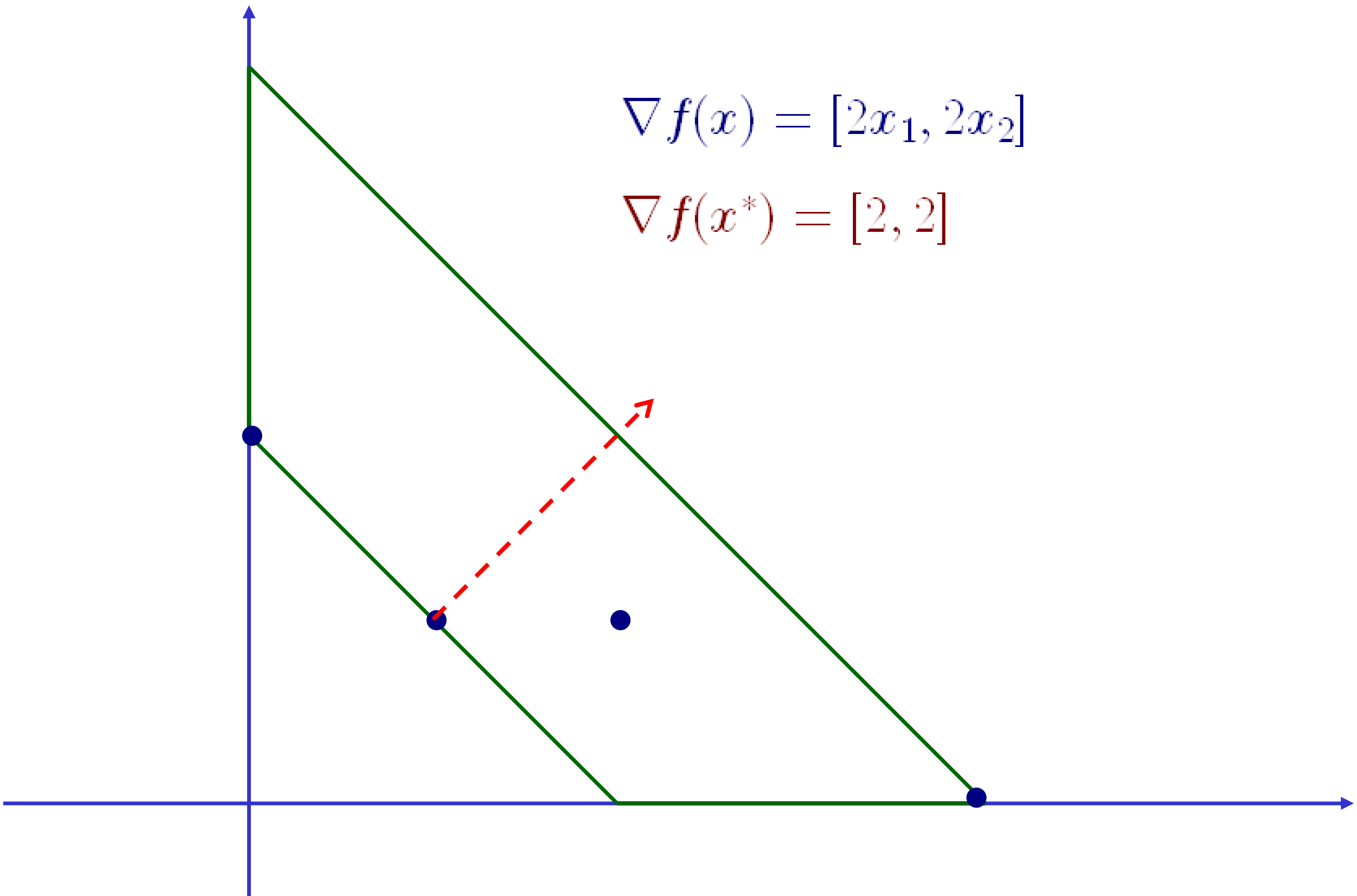
$$\min \ (x-1)^2 \quad s.t. \ a \leq x \leq b$$

$$\min_{x \in R^2} \; x_1^2 + x_2^2$$

$$x_1 + x_2 \leqslant 4$$
$$-x_1 - x_2 \leqslant -2$$
$$x_1, \; x_2 \geqslant 0$$

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

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



Linear Programming Problem

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

(LP) $\min \quad p^T x$

s.t. $Ax \leqslant b$

$$Cx = d$$

$$L \leqslant x \leqslant U$$

Linear Programming Solver in MATLAB

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

min f'*x   subject to: A*x <= b
 x

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

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

$$\min_{x,s} \quad 1^T s$$

s.t. $\quad -s \leqslant Ax - b \leqslant s$

Or

$$\min_{x,s} \quad \sum_{i=1}^{m} s_i$$

s.t. $\quad -s_i \leqslant A_i x - b_i \leqslant s_i \quad \forall \; i$

$$\min_{x,s} \quad [0 \; \cdots \; 0 \;\; 1 \; \cdots \; 1] \begin{bmatrix} x \\ s \end{bmatrix}$$

s.t. $\quad \begin{bmatrix} A & -I \\ -A & -I \end{bmatrix}_{2m \times (n+m)} \begin{bmatrix} x \\ s \end{bmatrix} \leqslant \begin{bmatrix} b \\ -b \end{bmatrix}$

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

$$\left\| z \right\|_\infty = \max_{1 \leqslant i \leqslant m} \left| z_i \right|$$

$$\min_{x,\gamma} \quad \gamma$$

s.t. $\quad -1\gamma \leqslant Ax - b \leqslant 1\gamma$

$$\min_{x,\gamma} \quad [0 \ \cdots \ 0 \ \ 1] \begin{bmatrix} x \\ \gamma \end{bmatrix}$$

s.t.

$$\begin{bmatrix} A & -1 \\ -A & -1 \end{bmatrix}_{2m \times (n+1)} \begin{bmatrix} x \\ \gamma \end{bmatrix} \leqslant \begin{bmatrix} b \\ -b \end{bmatrix}$$

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

(QP)

$$
\begin{aligned}
\min \quad & \tfrac{1}{2}x^T Q x + p^T x \\
\text{s.t.} \quad & Ax \leqslant b \\
& Cx = d \\
& L \leqslant x \leqslant U
\end{aligned}
$$

Quadratic Programming Solver in MATLAB

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

```
min 0.5*x'*H*x + f'*x   subject to:  A*x <= b
 x
```

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

$$\min_{w,b,\xi_A,\xi_B} \quad C(1^T\xi_A + 1^T\xi_B) + \frac{1}{2}\left\|w\right\|_2^2$$

$$(Aw + 1b) + \xi_A \geqslant 1$$

$$(Bw + 1b) - \xi_B \leqslant -1$$

$$\xi_A \geqslant 0, \xi_B \geqslant 0$$

Farkas' Lemma

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

either

$$Ax \geqslant \mathbf{0}, \quad b'x < 0 \;\; has \;\; a \;\; solution$$

or

$$A'\alpha = b, \quad \alpha \geqslant \mathbf{0} \;\; has \;\; a \;\; solution$$

but never both.

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

MP: $\min_{x \in \Omega} f(x)$ such that

$$\mathbf{g}(x) \leqslant \mathbf{0}$$

KTSP: Find $\overline{x} \in \Omega, \ \overline{\alpha} \in R^m$ such that

$$\nabla f(\overline{x}) + \overline{\alpha}' \nabla \mathbf{g}(\overline{x}) = 0$$

$$\overline{\alpha}' \mathbf{g}(\overline{x}) = 0$$

$$\mathbf{g}(\overline{x}) \leqslant \mathbf{0}, \quad \overline{\alpha} \geqslant \mathbf{0}$$

Lagrangian Function

$$\mathcal{L}(x,\alpha) = f(x) + \alpha'\mathbf{g}(x)$$

Let $\mathcal{L}(x,\alpha) = f(x) + \alpha'\mathbf{g}(x)$ and $\alpha \geqslant \mathbf{0}$

- If $f(x), \mathbf{g}(x)$ are convex then $\mathcal{L}(x,\alpha)$ is convex.
- For a fixed $\alpha \geqslant \mathbf{0}$, if $\bar{x} \in \arg\min\{\mathcal{L}(x,\alpha) \mid x \in R^n\}$ then

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

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

KTSP with Equality Constraints?

(Assume $\mathbf{h}(x) = \mathbf{0}$ are linear functions)

$$\mathbf{h}(x) = \mathbf{0} \Leftrightarrow \mathbf{h}(x) \leqslant \mathbf{0} \ \ \mathbf{and} \ -\mathbf{h}(x) \leqslant \mathbf{0}$$

KTSP: Find $\overline{x} \in \Omega, \ \ \overline{\alpha} \in R^k, \ \ \overline{\beta}_+, \overline{\beta}_- \in R^m$ such that

$$\nabla f(\overline{x}) + \overline{\alpha}' \nabla \mathbf{g}(\overline{x}) + (\overline{\beta}_+ - \overline{\beta}_-)' \nabla \mathbf{h}(\overline{x}) = 0$$

$$\overline{\alpha}' \mathbf{g}(\overline{x}) = 0, \ \ (\overline{\beta}_+)' \mathbf{h}(\overline{x}) = 0, \ (\overline{\beta}_-)'(-\mathbf{h}(\overline{x})) = 0$$

$$\mathbf{g}(\overline{x}) \leqslant \mathbf{0}, \ \mathbf{h}(\overline{x}) = 0$$

$$\overline{\alpha} \geqslant, \ \ \overline{\beta}_+, \overline{\beta}_- \geqslant \mathbf{0}$$

KTSP with Equality Constraints

KTSP: Find $\overline{x} \in \Omega,\ \overline{\alpha} \in R^k,\ \overline{\beta} \in R^m$ such that

$$\nabla f(\overline{x}) + \overline{\alpha}'\nabla \mathbf{g}(\overline{x}) + \overline{\beta}\nabla \mathbf{h}(\overline{x}) = 0$$

$$\overline{\alpha}'\mathbf{g}(\overline{x}) = 0,\ \mathbf{g}(\overline{x}) \leqslant \mathbf{0},\ \ \mathbf{h}(\overline{x}) = 0$$

$$\overline{\alpha} \geqslant \mathbf{0}$$

- Let $\overline{\beta} = \overline{\beta}_+ - \overline{\beta}_-$ and $\overline{\beta}_+, \overline{\beta}_- \geqslant \mathbf{0}$ then $\overline{\beta}$ is free variable

Generalized Lagrangian Function

$$\mathcal{L}(x,\alpha,\beta) = f(x) + \alpha'\mathbf{g}(x) + \beta'\mathbf{h}(x)$$

Let $\mathcal{L}(x,\alpha,\beta) = f(x) + \alpha'\mathbf{g}(x) + \beta'\mathbf{h}(x)$ and $\alpha \geqslant \mathbf{0}$

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

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

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

Lagrangian Dual Problem

$$\max_{\alpha,\beta} \min_{x\in\Omega} \mathcal{L}(x,\alpha,\beta)$$

subject to $\alpha \geqslant \mathbf{0}$

Lagrangian Dual Problem

$$\max_{\alpha,\beta} \min_{x\in\Omega} \mathcal{L}(x,\alpha,\beta)$$

subject to $\alpha \geqslant \mathbf{0}$

⬍

$$\max_{\alpha,\beta} \ \theta(\alpha,\beta)$$

subject to $\alpha \geqslant \mathbf{0}$

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

Weak Duality Theorem

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

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

Corollary: $\sup\{\theta(\alpha, \beta) \mid \alpha \geqslant \mathbf{0}\}$

$$\leqslant \inf\{f(x) \mid \mathbf{g}(x) \leqslant \mathbf{0}, \ \mathbf{h}(x) = \mathbf{0}\}$$

Weak Duality Theorem

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

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

Saddle Point of Lagrangian

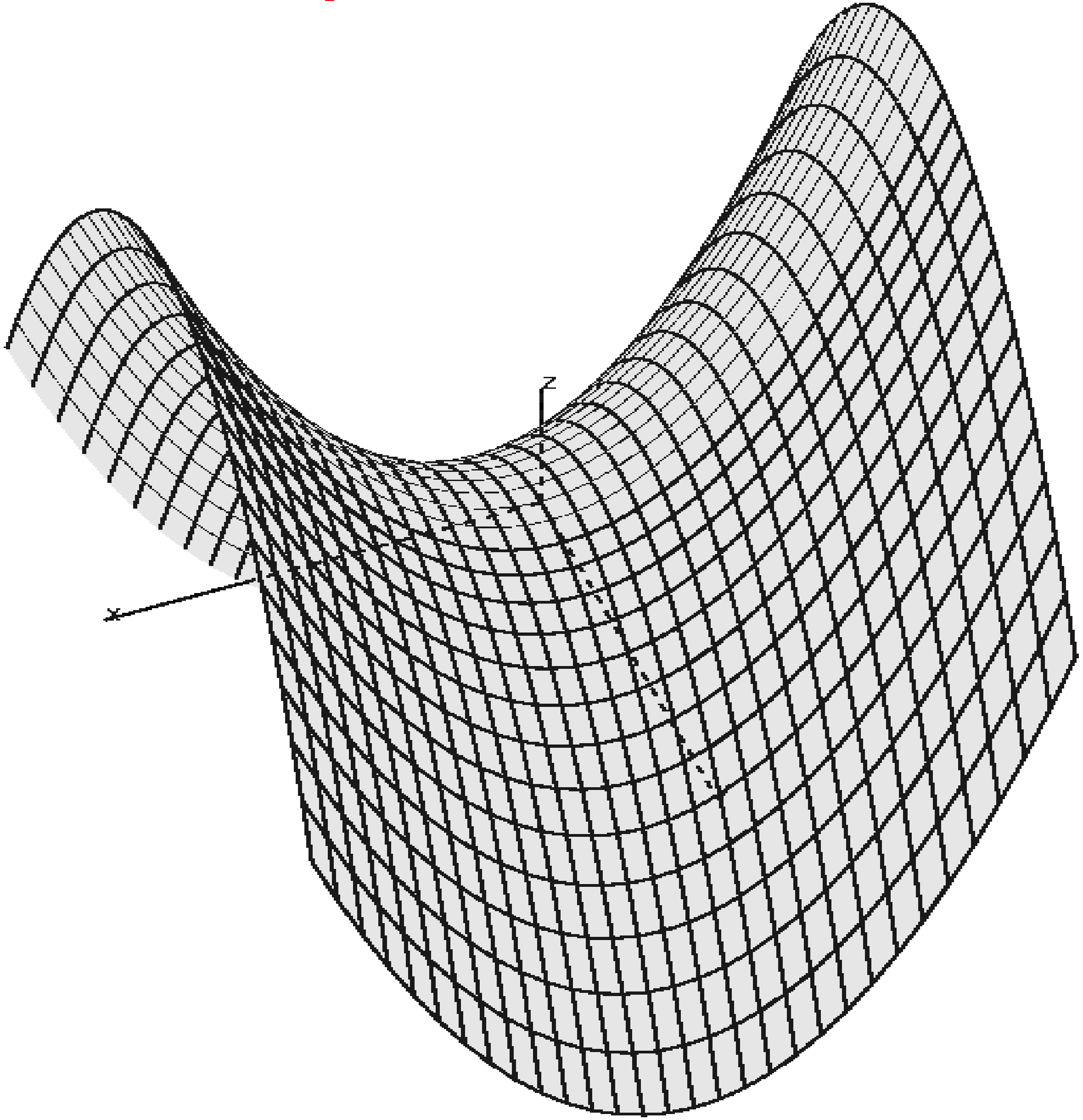
Let $x^* \in \Omega, \alpha^* \geqslant \mathbf{0}, \quad \beta^* \in R^m$ satisfying

$$\mathcal{L}(x^*, \alpha, \beta) \leqslant \mathcal{L}(x^*, \alpha^*, \beta^*) \leqslant \mathcal{L}(x, \alpha^*, \beta^*),$$

$\forall \ x \in \Omega, \quad \alpha \geqslant \mathbf{0}.$ 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

Primal LP

$$\min_{x \in R^n} \quad p'x$$

subject to $\quad Ax \geqslant b, \quad x \geqslant \mathbf{0}$

Dual LP

$$\max_{\alpha \in R^m} \quad b'\alpha$$

subject to $\quad A'\alpha \leqslant p, \quad \alpha \geqslant \mathbf{0}$

※ All duality theorems hold and work perfectly!

Lagrangian Function of Primal LP

$$\mathcal{L}(x,\alpha) = p'x + \alpha_1'(b - Ax) + \alpha_2'(-x)$$

$$\max_{\alpha_1,\alpha_2 \geqslant 0} \min_{x \in R^n} \mathcal{L}(x,\alpha_1,\alpha_2) \Longleftrightarrow$$

$$\max_{\alpha_1,\alpha_2 \geqslant 0} p'x + \alpha_1'(b - Ax) + \alpha_2'(-x)$$

$$p - A'\alpha_1 - \alpha_2 = 0$$

$$(\nabla_x \mathcal{L}(x,\alpha_1,\alpha_2) = 0)$$

Application of LP Duality

LSQ-Normal Equation Always Has a Solution

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

consider $\min_{x \in R^n} ||Ax - b||_2^2$

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

Claim: $A'Ax = A'b$ always has a solution.

Dual Problem of Strictly Convex Quadratic Program

Primal QP

$$\min_{x \in R^n} \frac{1}{2}x'Qx + p'x$$

subject to $Ax \leqslant b$

With *strictly convex* assumption, we have

Dual QP

$$\max \quad -\frac{1}{2}(p' + \alpha' A)Q^{-1}(A'\alpha + p) - \alpha' b$$

subject to $\alpha \geqslant \mathbf{0}$