1 Optimization

1.1 Hessian Matrix

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function, its Hessian matrix at $x \in \mathbb{R}^n$ is defined as

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 ff}{\partial x_1 x_3} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 x_3} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \frac{\partial^2 f}{\partial x_n x_3} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \in \mathbb{R}^{n \times n}$$

1.1.1 Quadratic Function

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $f(x) = \frac{1}{2} x^T H x + p^T x$ where $H \in \mathbb{R}^{n \times n}$ is a symmetric matrix and $p \in \mathbb{R}^n$. Then

$$\begin{aligned} \nabla f(x) &= Hx + p \\ \nabla^2 f(x) &= H \end{aligned}$$

If H is positive definite, then $x^* = -H^{-1} p$ is a unique solution.

1.1.2 Least-Square Problem

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

$$\begin{aligned} f(x) &= (Ax - b)(Ax - b)^T \\ &= x^T A^T A x - 2 b^T A x + b^T b \\ \nabla f(x) &= 2 A^T A x - 2 A^T b \\ \nabla^2 f(x) &= 2 A^T A \\ x^* &= (A^T A)^{-1} A^T b \end{aligned}$$

1.1.3 Newton's Method

Compute d^i satisfies

$$\nabla^2 f(x^i) d^i = -\nabla f(x^i)$$

Update

$$x^{i+1} = x^i + d^i$$

Until $\nabla f(x^i) = 0$.

1.2 Lagrangian Function

For the problem

$$
\begin{aligned}
&\min_{x\in\Omega} f(x) \\
&s.t. \quad g_i(x) \leq 0,\ h_i(x) = 0,\ \forall\ i = 1, \ldots, m
\end{aligned}
$$

Write $g(x) = (g_1(x), \ldots, g_m(x))^T$ and $h(x) = (h_1(x), \ldots, h_m(x))^T$. Let

$$\mathcal{L}(x, \alpha, \beta) = f(x) + \alpha^T g(x) + \beta^T h(x) \text{ and } \alpha \in \mathbb{R}^m \geq 0$$

For a fixed $\alpha \geq 0$ and a fixed β, if $\bar{x} \in \arg\min\{\mathcal{L}(x, \alpha, \beta) | x \in \mathbb{R}^n\}$ then

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

1.2.1 Duality

Consider

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

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

$$\max_{\alpha,\beta} \theta(\alpha, \beta) \text{ s.t. } \alpha \geq 0$$

For any $\bar{x} \in \Omega$ and $(\bar{\alpha} \geq 0, \bar{\beta})$ be solutions of the above problems, respectively, since $\bar{\alpha}^T g(\bar{x}) \leq 0$ and $h(\bar{x}) = 0$, we have $f(\bar{x}) \geq \theta(\bar{\alpha}, \bar{\beta})$.

Theorem 1.1. *If the equality happens, the $\bar{x}$ and $(\bar{\alpha} \geq 0, \bar{\beta})$ solve the primal and dual problem, respectively. In this case,*

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

Furthermore, for these $\bar{x}$ and $(\bar{\alpha}, \bar{\beta})$,

$$
\begin{aligned}
f(\bar{x}) &= \theta(\bar{\alpha}, \bar{\beta}) \\
&= \inf_{x\in\Omega} \{f(x) + \bar{\alpha}^T g(x) + \bar{\beta}^T h(x)\} \\
&\leq f(\bar{x}) + \bar{\alpha}^T g(\bar{x}) + \bar{\beta}^T h(\bar{x}) \\
&= f(\bar{x}) + \bar{\alpha}^T g(\bar{x}) \\
&\leq f(\bar{x})
\end{aligned}
$$

This implies that

$$\bar{\alpha}^T g(\bar{x}) = 0$$

1.2.2 Karush-Kuhn-Tucker Condition(KKT Condition)

This is a summary of solve both primal form and dual form. Find $\bar{x} \in \Omega$, $\bar{\alpha}, \bar{\beta} \in \mathbb{R}^m$ such that

Stationarity	$\nabla f(\bar{x}) + \bar{\alpha}^T \nabla g(\bar{x}) + \bar{\beta}^T \nabla h(\bar{x}) = 0$
Complmentary Slackness	$\bar{\alpha}^T g(\bar{x}) = 0$
Primal Feasibility	$h(\bar{x}) = 0,\ g(\bar{x}) \leq 0$
Dual Feasibility	$\bar{\alpha} \geq 0$

1.2.3 Dual Linear Problem

For the primal linear problem

$$\min_{x \in \mathbb{R}^n} p^T x \quad \text{s.t.} \quad Ax \geq b,\ x \geq 0$$

Consider

$$\max_{\alpha_1, \alpha_2 \geq \mathbf{0}} \min_{x \in \mathbb{R}^n} \mathcal{L}(x, \alpha, \beta) = p^T x + \alpha_1^T (b - Ax) + \alpha_2^T (-x)$$

When $\bar{x} \in \arg\min\{\mathcal{L}(x, \alpha, \beta) | x \in \Omega\}$, the gradient $\nabla \mathcal{L}(\bar{x}, \alpha, \beta)$ vanish

$$p - A^T \alpha_1 - \alpha_2 = 0$$

Then we have the dual problem

$$\max_{\alpha_1, \alpha_2 \geq \mathbf{0}} b^T \alpha_1 \quad \text{s.t.} \quad p - A^T \alpha_1 - \alpha_2 = 0$$

Since α_2 is a slack variable, it's equivalent to

$$\max b^T \alpha \quad \text{s.t.} \quad A^T \alpha \leq p,\ \alpha \geq 0$$

1.2.4 Least Square Problem

For $\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$ where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. It's obvious that

$$x^* \in \left\{ \arg\min\{\|Ax - b\|_2^2 \right\} \Rightarrow A^T A x = A^T b$$

Consider the problem

$$\min \mathbf{0}^T x \quad \text{s.t.} \quad A^T A x = A^T b$$

Then for

$$\max_{\alpha} \min \mathbf{0}^T x + \alpha^T (A^T b - A^T A x) \quad \text{s.t.} \quad \alpha \in \mathbb{R}^m$$

and the dual problem

$$\max_{\alpha} b^T A \alpha \quad \text{s.t.} \quad (A^T A)^T \alpha = \mathbf{0}$$

The constraint has a trivial solution $\alpha = \mathbf{0}$ and the objective function has value 0. The objective function of the primal problem and the dual problem have the same value. This implies that $A^T A x = A^T b$ must have a solution. Otherwise the dual problem won't have optimal solution.

1.2.5 Quadratic Problem

The primal problem

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

For

$$\max_{\alpha \geq 0} \min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + p^T x + \alpha^T (Ax - b) \quad \text{s.t.} \quad \alpha \geq 0$$

the gradient needs vanishing

$$Qx + p + A^T \alpha = 0 \Rightarrow x = -Q^{-1}(A^T \alpha + p)$$

Substitute back and we have the dual form

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