Optimization

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

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

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.

Least-Square Problem

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

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

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

Lagrangian Function

For the problem

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

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

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

For a fixed $\alpha \geq 0$ and a fixed β, if $\Bar{x}\in\mathop{\mathrm{arg~min}}\{\mathcal{L}(x,\alpha,\beta)| x\in\mathbb{R}^n\}$ then $$\left.\frac{\partial\mathcal{L}(x,\alpha,\beta)}{\partial x}\right|_{x=\Bar{x}} =\nabla f(\Bar{x}) + \alpha^T\nabla g(\Bar{x})+\beta^T\nabla h(\Bar{x})=0$$

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 $(\Bar{\alpha}\geq0,\Bar{\beta})$ 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. *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$$

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 $$\begin{aligned} \mathbf{Stationarity \quad} & \nabla f(\Bar{x}) + \Bar{\alpha}^T\nabla g(\Bar{x}) + \Bar{\beta}^T\nabla h(\Bar{x})=0\\ \mathbf{Complmentary~Slackness \quad} & \Bar{\alpha}^T g(\Bar{x})=0\\ \mathbf{Primal~Feasibility \quad} & h(\Bar{x})=0,~g(\Bar{x})\leq 0\\ \mathbf{Dual~Feasibility \quad} & \Bar{\alpha}\geq 0

\end{aligned}$$

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 $\Bar{x}\in\mathrm{arg~min}\{\mathcal{L}(x,\alpha,\beta)|x\in\Omega\}$, 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

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

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

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

Consider the problem

$$\min \mathbf{0}^Tx \quad \text{s.t.} \quad A^TAx = A^Tb$$

Then for

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

and the dual problem

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

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 $A^TAx = A^Tb$ must have a solution. Otherwise the dual problem won't have optimal solution.

Quadratic Problem

The primal problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^TQx + p^Tx \quad \text{s.t.} \quad Ax \le b$$

For

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

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

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