Convex Optimization - Tối ưu lồi

1. Tập lồi và hàm lồi

1.1 Convex set - tập hợp lồi:

Là một tập hợp các điểm mà khi nối 2 điểm bất kì thuộc tập đó ta luôn được một đoạn thẳng nằm trong tập đó.

Trong toán học: Tập hợp C đgl tập lồi nếu với bất kì x_1, x_2 thuộc C, điểm x_0 bất kì thỏa mãn:

$$x_0 = \theta x_1 + (1 - \theta)x_2 \text{ cũng nằm trong C } \textit{với } 0 <= \theta <= 1.$$

Example: Các siêu mặt phẳng (hyperplane), nửa không gian (haftspace), …

Note: - Giao của các tập lồi là một tập lồi

1.2 Convex functions - hàm lồi:

Hàm số f(x) được gọi là hàm lồi nếu **dom**f (tập xác định của hàm f(x)) là tập lồi và

$$f(\theta x + (1 - \theta)y) \le \theta f(x) + (1 - \theta)f(y) \qquad \text{với mọi } x, y \in \mathbf{dom}f,\ 0 \le \theta \le 1.$$

Stictly convex function (lồi chặt): **dom**f là tập lồi và

$$f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)$$

*với mọi x, y ∈ **dom**f, x != y, 0 < θ < 1 (chỉ khác với hàm **convex** ở dấu nhỏ hơn).*

Note:

- Nếu một hàm số là Strictly convex và có điểm cực trị thì điểm cực trị đó là duy nhất và cũng là global minimun
- Một hàm f(x) đgl concave (lõm) nếu -f(x) là một hàm lồi
- Tổng của 2 hàm lồi là hàm lồi
- Maximum của các hàm lồi là một hàm lồi

1.3 Kiểm tra tính chất lồi dựa vào đạo hàm

- ***First-order condition:*** Giả sử hàm số f(x) có tập xác định là lồi, có đạo hàm tại mọi điểm trên tập xác định, khi đó hàm f(x) là hàm lồi nếu và chỉ nếu với mọi x, x_0 trên **dom**f, ta có:

$$f(x) >= f(x_0) + \nabla f(x_0)^T(x - x_0)$$

có nghĩa là một hàm số là lồi nếu mặt tiếp tuyến tại một điểm bất kì luôn nằm dưới đồ thị đó

- ***Second-order condition:*** Một hàm số có đạo hàm bậc 2 là convex nếu domf là convex và Hesian (đạo hàm bậc 2: $\nabla^2 f(x)$) của nó là một ma trận nửa xác định dương với mọi x trong tập xác định. Với hàm số 1 biến thì điều kiện này tương đương f''(x) >= 0 với mọi x thuộc tập xác định

2. Bài toán tối ưu lồi:

Bài toán tối ưu lồi (convex optimization problem) là bài toán tối ưu có dạng:

$$\begin{aligned}\mathbf{x}^* &= \arg\min_{\mathbf{x}} f_0(\mathbf{x}) \\ \text{thoả mãn: } f_i(\mathbf{x}) &\leq 0, \quad i = 1, 2, \dots, m \\ h_j(\mathbf{x}) &= \mathbf{a}_j^T\mathbf{x} - b_j = 0, j = 1, \dots,\end{aligned}$$

trong đó $f_0, f_1, \dots, f_m$ là các hàm lồi.

Tập hợp các điểm thỏa mãn ràng buộc được gọi là feasible set. Khi feasible set là tập rỗng thì ta nói bài toán tối ưu vô nghiệm

Một điểm x* đgl optimal point (điểm tối ưu) là nghiệm của bài toán nếu x* là feasible và $f_0(x^*) = p^*$ (p* là giá trị tối ưu của bài toán). Tập hợp tất cả các optimal point gọi là optimal set. Khi optimal set khác rỗng, bài toán được giải và ngược là, optimal value không thể đạt được

Note:

- Local optimum của bài toán tối ưu lồi chính là global optimum của nó

- Theo First-order condition, điều kiện cần và đủ để một điểm x_0 là optimal point là:

$\nabla f(x_0)^T(x - x_0) >= 0$ với mọi x thuộc feasible set

3. Duality (dual problem: bài toán đối ngẫu)

3.1 Giới thiệu

Xét bài toán:

$$\begin{aligned}\mathbf{x} &= \arg\min_{\mathbf{x}} f_0(\mathbf{x}) \\ \text{thoả mãn: } f_1(\mathbf{x}) &= 0\end{aligned}$$

Bài toán này là bài toán tối ưu tổng quát, không nhất thiết phải lồi. Có thể giải bằng phương pháp nhân từ Lagrange. Cụ thể, ta xét hàm số (hàm hỗ trợ):

$$L(x, \lambda) = f_0(x) + \lambda f_1(x)$$

Người ta đã chứng minh được rằng điểm optimal value của bài toán nêu trên thỏa mãn điều kiện

$\nabla_{x,\lambda} L(x, \lambda) = 0$, tương đương với:

$$\begin{aligned}\nabla_{\mathbf{x}} f_0(\mathbf{x}) + \lambda \nabla_{\mathbf{x}} f_1(\mathbf{x}) &= 0 \\ f_1(\mathbf{x}) &= 0\end{aligned}$$

3.2 hàm đối ngẫu Lagrange

Xét bài toán tối ưu tổng quát:

$$\begin{aligned}\mathbf{x}^* &= \arg\min_{\mathbf{x}} f_0(\mathbf{x}) \\ \text{thoả mãn: } f_i(\mathbf{x}) &\leq 0, \quad i = 1, 2, \dots, m \\ h_j(\mathbf{x}) &= 0, \quad j = 1, 2, \dots, p\end{aligned}$$

từ bài toán trên ta xây dựng Lagrangian:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^{p} \nu_j h_j(\mathbf{x})$$

Hàm đối ngẫu Lagrange của bài toán tối ưu là một hàm của các biến đói ngẫu **λ** và **ν**, được định nghĩa là giá trị nhỏ nhất theo **x** của Lagrangian:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^{p} \nu_j h_j(\mathbf{x}) \right)$$

inf là viết tắt của hàm Infinium (lấy cận dưới), nếu Lagrangina không bị chặn dưới, hàm đối ngẫu sẽ lấy giá trị -∞.

Note:

- inf được lấy trên miền xác định của bài toán, không phải feasible set.

- Hàm đối ngẫu của một bài toán tối ưu bất kỳ là một hàm concave, bất kể bài toán ban đầu có phải convex hay không

Nếu p* là optimal value của bài toán thì với các biến đối ngẫu **λ** và **ν** bất kì, ta có:

$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) <= p^* \ (p^* = f_0(x_0))$

Để chứng minh, ta giả sử x_0 là một điểm feasible bất kì thỏa mãn các điều kiện ràng buộc.

$$\mathcal{L}(\mathbf{x}_0, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}_0) + \sum_{i=1}^{m} \underbrace{\lambda_i f_i(\mathbf{x}_0)}_{\leq 0} + \sum_{j=1}^{p} \underbrace{\nu_j h_j(\mathbf{x}_0)}_{=0} \leq f_0(\mathbf{x}_0)$$

Vì điều này đúng với mọi $\mathbf{x}_0$ *feasible*, ta sẽ có tính chất quan trọng sau đây:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq \mathcal{L}(\mathbf{x}_0, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\mathbf{x}_0).$$

3.3 Bài toán đối ngẫu Lagrange

Với mỗi cặp (λ, ν), hàm đối ngẫu Lagrange cho chúng ta một chặn dưới cho optimal value p* của bài toán, ta phải đi tìm giá trị nào là optimal của bài toán:

$$\boldsymbol{\lambda}^*, \boldsymbol{\nu}^* = \arg\max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$
$$\text{thoả mãn: } \boldsymbol{\lambda} \succeq 0$$

Weak duality

Gọi giá trị tối ưu của bài toán đối ngẫu là d*, ta luôn có d* <= p*. Tính chất này gọi là weak duality

Note:

- Nếu bài toán gốc không bị chặn dưới, p* = -∞, d* = -∞. Bài toán đối ngẫu là infeasible
- Nếu bài toán đối ngẫu không bị chặn trên thì d* = +∞, suy ra p* = +∞, bài toán gốc infeasible

Strong duality

Nếu đẳng thức p* = d* xảy ra, thì ta nói có strong duality. Lúc này, bài toán đối ngẫu giúp ta tìm được chính xác giá trị tối ưu của bài toán gốc

Slater's constraint qualification

Xét bài toán:

$$
\begin{aligned}
x &= \arg\min_{\mathbf{x}} f_0(\mathbf{x}) \\
\text{thoả mãn: } & f_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m \\
& \mathbf{A}\mathbf{x} = \mathbf{b}
\end{aligned}
$$

Trong đó f_0, f_2, f_2,… , f_m là các hàm lồi. Một điểm feasible của bài toán được gọi là strictly feasible nếu các $f_i(x) < 0$, với i = 1, 2, …, m. Khi tồn tại một điểm strictly feasible thì strong duality xảy ra.

4. Các điều kiện tối ưu

4.1 Complementary slackness

Giả sử strong duality xảy ra, x* là điểm optimal của bài toán gốc và (λ*, ν*) là cặp điểm optimal của bài toán đối ngẫu. Ta có:

$$
\begin{aligned}
f_0(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\mathbf{x}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}) + \sum_{j=1}^{p} \nu_j^* h_j(\mathbf{x}) \right) \\
&\leq f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) + \sum_{j=1}^{p} \nu_j^* h_j(\mathbf{x}^*) \\
&\leq f_0(\mathbf{x}^*)
\end{aligned}
$$

Khi đẳng thức xảy ra, ta được: $\sum_{i=1}^{m} \lambda_i^* f_i(\mathbf{x}^*) = 0.$ Vì $\lambda_i^* \geq 0, f_i \leq 0$ nên mỗi phần tử $\lambda_i^* f_i(\mathbf{x}^*) \leq 0$

Từ đó suy ra:

$$
\begin{aligned}
\lambda_i^* > 0 &\Rightarrow f_i(\mathbf{x}^*) = 0 \\
f_i(\mathbf{x}^*) < 0 &\Rightarrow \lambda_i^* = 0
\end{aligned}
$$

4.2 Các điều kiện tối ưu KKT

Điều kiện KKT cho bài toán không lồi

Giả sử strong duality xảy ra, gọi x*, và (λ*, ν*) là bất kỳ primal và dual optimal point. Vì x* tối uw hàm khả vi L(x, λ*, ν*) nên ta có đạo hàm của Lagrangian tại x* bằng 0. Điều kiện KKT (Karush-Kuhn-Tucker) nói rằng x*, λ*, ν* phải thỏa mãn các điều kiện sau để là nghiệm của primal problem và dual problem

$$
\begin{aligned}
f_i(\mathbf{x}^*) &\leq 0, i = 1, 2, \ldots, m \\
h_j(\mathbf{x}^*) &= 0, j = 1, 2, \ldots, p \\
\lambda_i^* &\geq 0, i = 1, 2, \ldots, m \\
\lambda_i^* f_i(\mathbf{x}^*) &= 0, i = 1, 2, \ldots, m \\
\nabla_{\mathbf{x}} f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_{\mathbf{x}} f_i(\mathbf{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla_{\mathbf{x}} h_j(\mathbf{x}^*) &= 0
\end{aligned}
$$

Điều kiện KKT cho bài toán lồi

Với các bài toán lồi và strong duality xảy ra, các điều kiện KKt phía trên cũng là điều kiện đủ.

Note: Với một bài toán lồi và điều kiện Slater thỏa mãn thì các điều kiện KKT là các điều kiện cần và đủ của nghiệm