Optimization Methods (CS1.404), Spring 2024
Lecture 22

Naresh Manwani

Machine Learning Lab, IIIT-H

April 8th, 2024

Second Order Necessary Conditions

Theorem

Let $\mathbf{x}^*$ be a local minimum of the optimization problem described below.

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ s.t. \quad & h_j(\mathbf{x}) \leq 0,\ j = 1\ldots l \\ & e_i(\mathbf{x}) = 0;\ i = 1\ldots m \end{aligned}$$

where $f(\mathbf{x}) \in \mathbb{C}^2(\mathbb{R}^n)$, $h_j \in \mathbb{C}^2(\mathbb{R}^n)$, $j = 1\ldots l$ and $e_i \in \mathbb{C}^2(\mathbb{R}^n)$, $i = 1\ldots m$. Suppose that $\mathbf{x}^*$ is regular, which means $\nabla h_j(\mathbf{x}^*)$, $j \in I(\mathbf{x}^*)$ and $\nabla e_i(\mathbf{x}^*)$, $i \in \{1,\ldots,m\}$ are linearly independent, where $I(\mathbf{x}^*) = \{j \in \{1,\ldots,l\} \mid h_j(\mathbf{x}^*) = 0\}$.

1. Then there exist $\boldsymbol{\lambda}^* = [\lambda_1^* \ \ldots \ \lambda_l^*]^\top \in \mathbb{R}_+^l$ and $\boldsymbol{\mu}^* = [\mu_1^* \ \ldots \ \mu_m^*]^\top \in \mathbb{R}^m$, such that

$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\mathbf{x}^*) = \mathbf{0}$$

$$\lambda_j^* h_j(\mathbf{x}^*) = 0,\ j = 1\ldots l$$

2. and $\mathbf{y}^\top[\nabla^2 f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j^2(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i^2(\mathbf{x}^*)]\mathbf{y} \geq 0$ for all $\mathbf{y} \in \hat{T}(\mathbf{x}^*)$ where

$$\hat{T}(\mathbf{x}^*) = \{\mathbf{y} \in \mathbb{R}^n \mid \nabla h_j(\mathbf{x}^*)^\top \mathbf{y} = 0,\ j \in I(\mathbf{x}^*);\ \nabla e_i(\mathbf{x}^*)^\top \mathbf{y} = 0,\ i = 1\ldots m\}.$$

Proof

- Proof of part 1, we have already seen it. To prove part 2, we note that because $\mathbf{x}^*$ is a regular local minimizer of f on the set $\{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0,\ j = 1 \dots l;\ e_i(\mathbf{x}) = 0,\ i = 1 \dots m\}$, it is also a a regular minimizer of f on the set $\hat{S} = \{\mathbf{x} \in \mathbb{R}^n \mid h_j(\mathbf{x}) \leq 0,\ j \in I(\mathbf{x}^*);\ e_i(\mathbf{x}) = 0,\ i = 1 \dots m\}$.
- Note that the latter set only contains equality constraints. Therefore, from Lagrange's theorem, there exist vectors $\boldsymbol{\lambda}^* \in \mathbb{R}^l_+$ and $\boldsymbol{\mu}^* \in \mathbb{R}^m$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\mathbf{x}^*) = \mathbf{0}$$

  where for all $j \notin I(\mathbf{x}^*)$, we have $\lambda_j^* = 0$.
- Suppose $\mathbf{y} \in \hat{T}(\mathbf{x}^*)$ where $\hat{T}(\mathbf{x}^*)$ is tangent space to $\hat{S}$ at $\mathbf{x}^*$.
- Because $h_j \in \mathbb{C}^2(\mathbb{R}^n),\ j = 1 \dots l$ and $e_i \in \mathbb{C}^2(\mathbb{R}^n),\ i = 1 \dots m$, there exists a twice continuously differentiable curve $\{\mathbf{x}(t) \mid t \in (a, b)\}$ on $\hat{S}$ such that $\mathbf{x}(t^*) = \mathbf{x}^*$ and $\mathbf{x}'(t^*) = \mathbf{y}$ for some $t^* \in (a, b)$.
- Since $\mathbf{x}^* = \mathbf{x}(t^*)$ is local minimizer of f, thus t^* is local minimizer of function $\phi(t) = f(\mathbf{x}(t))$. From second order necessary condition for unconstrained minimization, we get $\frac{d^2\phi}{dt^2}(t^*) \geq 0$. Using, $\frac{d\phi}{dt}(t) = \nabla f(\mathbf{x}(t))^\top \mathbf{x}'(t)$, we get

$$\begin{aligned}\frac{d^2\phi}{dt^2}(t^*) &= \frac{d}{dt}[\nabla f(\mathbf{x}^*)^\top \mathbf{x}'(t^*)] = \mathbf{x}'(t^*)^\top \nabla^2 f(\mathbf{x}^*)\mathbf{x}'(t^*) + \nabla f(\mathbf{x}^*)^\top \mathbf{x}''(t^*) \\ &= \mathbf{y}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{y} + \nabla f(\mathbf{x}^*)^\top \mathbf{x}''(t^*) \geq 0 \qquad (1)\end{aligned}$$

- We know that $\forall t \in (a,b)$, $e_i(\mathbf{x}(t)) = 0$, $i = 1 \ldots m$ and $h_j(\mathbf{x}(t)) = 0$, $\forall j \in I(\mathbf{x}^*)$. Thus, we have $\frac{d^2}{dt^2}\left[\sum_{j=1}^{l} \lambda_j^* h_j(\mathbf{x}(t)) + \sum_{i=1}^{m} \mu_i^* e_i(\mathbf{x}(t))\right] = 0$.

$$\frac{d^2}{dt^2}\left[\sum_{j=1}^{l} \lambda_j^* h_j(\mathbf{x}(t)) + \sum_{i=1}^{m} \mu_i^* e_i(\mathbf{x}(t))\right] = \frac{d}{dt}\left[\sum_{j=1}^{l} \lambda_j^* \frac{d}{dt} h_j(\mathbf{x}(t)) + \sum_{i=1}^{m} \mu_i^* \frac{d}{dt} e_i(\mathbf{x}(t))\right]$$

$$= \frac{d}{dt}\left[\sum_{j=1}^{l} \lambda_j^* \nabla h_j(\mathbf{x}(t))^\top \mathbf{x}'(t) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\mathbf{x}(t))^\top \mathbf{x}'(t)\right]$$

$$= \sum_{j=1}^{l} \lambda_j^* \frac{d}{dt}\{\nabla h_j(\mathbf{x}(t))^\top \mathbf{x}'(t)\} + \sum_{i=1}^{m} \mu_i^* \frac{d}{dt}\{\nabla e_i(\mathbf{x}(t))^\top \mathbf{x}'(t)\}$$

$$= \sum_{j=1}^{l} \lambda_j^* \left[\mathbf{x}'(t)^\top \nabla^2 h_j(\mathbf{x}(t))\mathbf{x}'(t) + \nabla h_j(\mathbf{x}(t))^\top \mathbf{x}''(t)\right]$$

$$+ \sum_{i=1}^{m} \mu_i^* \left[\mathbf{x}'(t)^\top \nabla^2 e_i(\mathbf{x}(t))\mathbf{x}'(t) + \nabla e_i(\mathbf{x}(t))\mathbf{x}''(t)\right] = 0$$

Proof - Continue

- In particular, the above is also true at $t = t^*$. Thus, using $\mathbf{x}(t^*) = \mathbf{x}^*$ and $\mathbf{x}'(t^*) = \mathbf{y}$.

$$\sum_{j=1}^{l} \lambda_j^* \left[\mathbf{y}^\top \nabla^2 h_j(\mathbf{x}^*)\mathbf{y} + \nabla h_j(\mathbf{x}^*)^\top \mathbf{x}''(t^*) \right] + \sum_{i=1}^{m} \mu_i^* \left[\mathbf{y}^\top \nabla^2 e_i(\mathbf{x}^*)\mathbf{y} + \nabla e_i(\mathbf{x}^*)\mathbf{x}''(t^*) \right] = 0 \tag{2}$$

- Adding eq.(1) and eq.(2), we get

$$\mathbf{y}^\top \left[\nabla^2 f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla^2 h_j(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla^2 e_i(\mathbf{x}^*) \right] \mathbf{y} + \left[\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\mathbf{x}^*) \right]^\top \mathbf{x}''(t^*) \geq 0$$

- But, by Lagrange theorem, $\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\mathbf{x}^*) = \mathbf{0}$. Therefore,

$$\mathbf{y}^\top \left[\nabla^2 f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla^2 h_j(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla^2 e_i(\mathbf{x}^*) \right] \mathbf{y} \geq 0$$

- Which proves the result.

Second Order Sufficiency Conditions

Theorem

Consider the optimization problem described below.

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ s.t. \quad & h_j(\mathbf{x}) \leq 0, \;\; j = 1 \ldots l \\ & e_i(\mathbf{x}) = 0; \;\; i = 1 \ldots m \end{aligned}$$

where $f(\mathbf{x}) \in \mathbb{C}^2(\mathbb{R}^n)$, $h_j \in \mathbb{C}^2(\mathbb{R}^n)$, $j = 1 \ldots l$ and $e_i \in \mathbb{C}^2(\mathbb{R}^n)$, $i = 1 \ldots m$. Suppose there exist a feasible point $\mathbf{x}^*$, $\boldsymbol{\lambda}^* = [\lambda_1 \; \lambda_2 \; \ldots \; \lambda_l]^\top \in \mathbb{R}^l_+$ and $\boldsymbol{\mu} = [\mu_1 \; \mu_2 \; \ldots \; \mu_m]^\top \in \mathbb{R}^m$, such that

1. $\lambda_j h_j(\mathbf{x}^*) = 0, \;\; j = 1 \ldots l$ and

$$\nabla f(\mathbf{x}^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_i \nabla e_i(\mathbf{x}^*) = \mathbf{0}$$

2. Also, for all $\mathbf{y} \in \tilde{T}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, $\mathbf{y} \neq \mathbf{0}$, we have $\mathbf{y}^\top \nabla^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\mathbf{y} > 0$. where

$$\tilde{T}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \{\mathbf{y} \in \mathbb{R}^n \mid \nabla h_j(\mathbf{x}^*)^\top \mathbf{y} = 0, \;\; j \in \hat{I}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*); \;\; \nabla e_i(\mathbf{x}^*)^\top \mathbf{y} = 0, \;\; i = 1 \ldots m\}.$$

for $\hat{I}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \{j \in \{1, \ldots, l\} \mid h_j(\mathbf{x}^*) = 0, \lambda_j^* > 0\}$.

Then $\mathbf{x}^*$ is a local minimizer.

Test Positive Definiteness in a Subspace

- In the second-order sufficiency conditions requires that $\mathbf{d}^\top \nabla^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\mathbf{d} > 0$ for all $\mathbf{d} \in \tilde{T}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, $\mathbf{d} \neq \mathbf{0}$, where

  $$\tilde{T}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \{\mathbf{d} \in \mathbb{R}^n \mid \nabla h_j(\mathbf{x}^*)^\top \mathbf{d} = 0,\ j \in \hat{I};\ \nabla e_i(\mathbf{x}^*)^\top \mathbf{d} = 0,\ i = 1 \ldots m\}.$$

  for $\hat{I} = \{j \in \{1, \ldots, l\} \mid h_j(\mathbf{x}^*) = 0, \lambda_j^* > 0\}$.

- Let $Q = \nabla^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ and $A = \begin{bmatrix} \nabla e_1(\mathbf{x}^*)^\top \\ \vdots \\ \nabla e_m(\mathbf{x}^*)^\top \\ \nabla h_{j_1}(\mathbf{x}^*)^\top \\ \vdots \\ \nabla h_{j_{|\hat{I}|}}(\mathbf{x}^*)^\top \end{bmatrix}$.

- Then, second-order sufficiency conditions requires that $\mathbf{d}^\top Q \mathbf{d} > 0$, $\forall \mathbf{d} \neq \mathbf{0}$ such that $A\mathbf{d} = \mathbf{0}$. (In this case, the subspace is the null space of matrix A.) This test itself might be a nonconvex optimization problem.

Test Positive Definiteness in a Subspace

- Consider any vector $\mathbf{u} \in \mathbb{R}^n$ can be decomposed into two orthogonal components: (a) one which lies in the null space of matrix A, (b) one which lies in the space spanned by the rows of A.
    - If we project $\mathbf{u}$ in the row space of A, we can get the component of $\mathbf{u}$ which lies in the row space of A. The corresponding projection matrix is $P = A^\top(AA^\top)^{-1}A$.
    - Thus, the component of $\mathbf{u}$ in the null space of A is $\mathbf{u} - A^\top(AA^\top)^{-1}A\mathbf{u} = [I - A^\top(AA^\top)^{-1}A]\mathbf{u}$.
- Thus, $\mathbf{d}$ is in the null space of matrix A if and only if $\mathbf{d} = (I - A^\top(AA^\top)^{-1}A)\mathbf{u} = P_A\mathbf{u}$ for some $\mathbf{u} \in \mathbb{R}^n$.
- Thus, the test becomes whether or not

$$\mathbf{u}^\top P_A Q P_A \mathbf{u} > 0, \ \forall \mathbf{u} \in \mathbb{R}^n.$$

- That is, we just need to test positive definiteness of matrix $P_A Q P_A$ as usual.

Dual Problem

- Consider the optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x})$$
$$s.t. \;\; h_j(\mathbf{x}) \le 0, \; j = 1\dots l$$
$$e_i(\mathbf{x}) = 0; \; i = 1\dots m$$

where $f(\mathbf{x})$, h_j, $j = 1\dots l$ and e_i, $i = 1\dots m$ are sufficiently smooth functions over $\mathbb{R}^n$.

- This problem is referred as primal problem. Let p^* be the optimal value of the above problem.
- The Lagrangian of the problem is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{j=1}^{l} \lambda_j h_j(\mathbf{x}) + \sum_{i=1}^{m} \mu_i e_i(\mathbf{x})$$

where $\boldsymbol{\lambda} = [\lambda_1 \;\dots\; \lambda_l]^\top \in \mathbb{R}^l_+$ are nonnegative Lagrange multipliers associated with the inequality constraints and $\boldsymbol{\mu} = [\mu_1 \;\dots\; \mu_m]^\top \in \mathbb{R}^m$ are the Lagrange multipliers associated with the equality constraints.

Dual Problem

- The dual objective function $g : \mathbb{R}^l_+ \times \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is defined to be

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{x}} \ \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

- Note that above minimization problem can be unbounded, i.e., there may be values $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ for which $g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\infty$.
- We define the domain of dual function as

$$dom(g) = \{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^l_+ \times \mathbb{R}^m \mid g(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty\}$$

- The **Dual Problem** is defined as

$$\begin{aligned} g^* = \max \ & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ s.t. \ & (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in dom(g) \end{aligned}$$

Theorem: Convexity of the Dual Problem

Domain of dual function g is convex and g is a concave function over the $dom(g)$.

Example 1: Linear Programming

- Consider the linear programming problem

$$\min_{\mathbf{x}} \mathbf{c}^\top \mathbf{x}$$
$$s.t. \; A\mathbf{x} \leq \mathbf{b}$$

where $\mathbf{c} \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}$, and $\mathbf{b} \in \mathbb{R}^m$. We assume that the problem is feasible (which means, constraint set is nonempty).

- The Lagrangian function is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (A\mathbf{x} - \mathbf{b})$$

where $\boldsymbol{\lambda} = [\lambda_1 \; \ldots \; \lambda_m]^\top \in \mathbb{R}^m_+$ are nonnegative Lagrange multipliers associated with the inequality constraints

- The dual objective function is

$$\begin{aligned} g(\boldsymbol{\lambda}) &= \min_{\mathbf{x}} \; \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \min_{\mathbf{x}} \; \mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (A\mathbf{x} - \mathbf{b}) \\ &= \min_{\mathbf{x}} \; (\mathbf{c} + A^\top \boldsymbol{\lambda})^\top \mathbf{x} - \mathbf{b}^\top \boldsymbol{\lambda} \\ &= \begin{cases} -\mathbf{b}^\top \boldsymbol{\lambda}, & \mathbf{c} + A^\top \boldsymbol{\lambda} = \mathbf{0} \\ -\infty, & \text{else} \end{cases} \end{aligned}$$

- The dual problem is

$$\max \; -\mathbf{b}^\top \boldsymbol{\lambda}$$
$$s.t. \; \mathbf{c} + A^\top \boldsymbol{\lambda} = \mathbf{0}$$
$$\boldsymbol{\lambda} \geq \mathbf{0}$$

Example 2: Strictly Convex Quadratic Programming

- Consider the linear programming problem

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \mathbf{c}^\top \mathbf{x}$$
$$s.t.\ A\mathbf{x} \leq \mathbf{b}$$

  where $Q \in \mathbb{R}^{n\times n}$ is positive definite, $\mathbf{c} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$.

- The Lagrangian function is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \mathbf{c}^\top \mathbf{x} + \boldsymbol{\lambda}^\top (A\mathbf{x} - \mathbf{b}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + (A^\top \boldsymbol{\lambda} + \mathbf{c})^\top \mathbf{x} - \mathbf{b}^\top \boldsymbol{\lambda}$$

  where $\boldsymbol{\lambda} = [\lambda_1\ \ldots\ \lambda_m]^\top \in \mathbb{R}^m_+$ are nonnegative Lagrange multipliers.

- To find the dual function, we minimize the Lagrangian with respect to $\mathbf{x}$. The minimizer is attained at the stationary point which is the solution to

$$\nabla_{\mathbf{x}}\, \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = Q\mathbf{x}^* + A^\top \boldsymbol{\lambda} + \mathbf{c} = \mathbf{0} \Rightarrow \mathbf{x}^* = -Q^{-1}(A^\top \boldsymbol{\lambda} + \mathbf{c})$$

- Using $g(\boldsymbol{\lambda}) = \min_{\mathbf{x}}\ \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda})$, we obtain

$$g(\boldsymbol{\lambda}) = \frac{1}{2}(A^\top \boldsymbol{\lambda} + \mathbf{c})^\top Q^{-1}QQ^{-1}(A^\top \boldsymbol{\lambda} + \mathbf{c}) - (A^\top \boldsymbol{\lambda} + \mathbf{c})^\top Q^{-1}(A^\top \boldsymbol{\lambda} + \mathbf{c}) - \mathbf{b}^\top \boldsymbol{\lambda}$$
$$= -\frac{1}{2}\boldsymbol{\lambda}^\top AQ^{-1}A^\top \boldsymbol{\lambda} - (AQ^{-1}\mathbf{c} + \mathbf{b})^T \boldsymbol{\lambda} - \mathbf{c}^\top Q^{-1}\mathbf{c}$$

- The dual problem is

$$\max\ -\frac{1}{2}\boldsymbol{\lambda}^\top AQ^{-1}A^\top \boldsymbol{\lambda} - (AQ^{-1}\mathbf{c} + \mathbf{b})^T \boldsymbol{\lambda} - \mathbf{c}^\top Q^{-1}\mathbf{c}$$
$$s.t.\ \boldsymbol{\lambda} \geq \mathbf{0}$$

Weak Duality Theorem

Theorem

Consider the primal problem

$$\begin{aligned} p^* = \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ s.t. \quad & h_j(\mathbf{x}) \leq 0,\ j = 1\dots l \\ & e_i(\mathbf{x}) = 0;\ i = 1\dots m \end{aligned}$$

and dual problem

$$\begin{aligned} d^* = \max \quad & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ s.t. \quad & (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in dom(g) \end{aligned}$$

where $g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{x}} \ \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$. Then,

$$d^* \leq p^*$$

Example

- Consider the problem

$$\min\ x_1^2 - 3x_2^2$$
$$s.t.\ x_1 = x_2^3$$

- Substituting $x_1 = x_2^3$ into the objective function, the resulting unconstrained optimization problem is $\min_{x_2}\ x_2^6 - 3x_2^2$.
- The stationary points are $x_2 = 0, \pm 1$. Thus, the candidates for optimal solution are $(0,0), (1,1), (-1,-1)$.
- It is can be easily verified that the optimal solutions are $(1,1)$ and (-1,-1) with optimal value $p^* = -2$.
- Let us consider the dual problem. The Lagrangian is

$$\mathcal{L}(x_1, x_2, \mu) = x_1^2 - 3x_2^2 + \mu(x_1 - x_2^3) = x_1^2 + \mu x_1 - 3x_2^2 - \mu x_2^3$$

- Obviously, for any value of $\mu \in \mathbb{R}$, $\min_{x_1,x_2}\ \mathcal{L}(x_1, x_2, \mu) = -\infty$.
- Hence, the dual optimal value is $d^* = -\infty$, which is an extremely poor lower bound on the primal optimal value $p^* = -2$.

Geometric Interpretation

- We can give a simple geometric interpretation of the dual function in terms of the set

$$\mathcal{G} = \{(h_1(\mathbf{x}), \ldots, h_l(\mathbf{x}), e_1(\mathbf{x}), \ldots, e_m(\mathbf{x}), f(\mathbf{x})) \in \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R} \mid \mathbf{x} \in \mathbb{R}^n\}$$

which is the set of values taken on by the constraint and objective functions.

- The optimal value p^* of primal problem is easily expressed in terms of G

$$p^* = \inf\{t \mid (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G}, \mathbf{u} \leq \mathbf{0}, \mathbf{v} = \mathbf{0}\}$$

- To evaluate the dual function at $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, we minimize the affine function

$$(\boldsymbol{\lambda}, \boldsymbol{\mu}, 1)^\top (\mathbf{u}, \mathbf{v}, t) = \sum_{j=1}^{l} \lambda_j h_j + \sum_{i=1}^{m} \mu_i e_i + f$$

over $(\mathbf{u}, \mathbf{v}, t) \in \mathcal{G}$.

- Thus, we have

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf \{(\boldsymbol{\lambda}, \boldsymbol{\mu}, 1)^\top (\mathbf{u}, \mathbf{v}, t) \mid (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G}\}$$

- In particular, we see that if the infimum is finite, then the inequality $(\boldsymbol{\lambda}, \boldsymbol{\mu}, 1)^\top (\mathbf{u}, \mathbf{v}, t) \geq g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ defines a supporting hyperplane to $\mathcal{G}$.

- Now suppose $\boldsymbol{\lambda} \geq \mathbf{0}$, Then, we see that $t \geq (\boldsymbol{\lambda}, \boldsymbol{\mu}, 1)^\top(\mathbf{u}, \mathbf{v}, t)$ if $\mathbf{u} \leq \mathbf{0}$ and $\mathbf{v} = \mathbf{0}$.
- Therefore,

$$
\begin{aligned}
p^* &= \inf\{t \mid (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G}, \mathbf{u} \leq \mathbf{0}, \mathbf{v} = \mathbf{0}\} \\
&\geq \inf\{(\boldsymbol{\lambda}, \boldsymbol{\mu}, 1)^\top(\mathbf{u}, \mathbf{v}, t) \mid (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G}, \mathbf{u} \leq \mathbf{0}, \mathbf{v} = \mathbf{0}\} \\
&\geq \inf\{(\boldsymbol{\lambda}, \boldsymbol{\mu}, 1)^\top(\mathbf{u}, \mathbf{v}, t) \mid (\mathbf{u}, \mathbf{v}, t) \in \mathcal{G}\} \\
&= g(\boldsymbol{\lambda}, \boldsymbol{\mu})
\end{aligned}
$$

- Thus, we have weak duality.

Geometric Interpretation with one inequality constraint

- This graph assumes $\lambda \geq 0$.
- We first see that the optimal value p^* is given by the tangent horizontal line that indicates the minimum value when $u \leq 0$ (when all constraints are satisfied).
- For a given λ, the line $\lambda u + t = g(\lambda)$ provides the lower bound on the objective value for each $x \in \mathbb{R}^n$.
- The line has slope $-\lambda$. Since we defined $\lambda \geq 0$, $-\lambda \leq 0$. The line is always tangent to at least one point on the boundary of $\mathcal{G}$.
- We may compute u from our constraint $h_1(\mathbf{x})$, and we may also compute $g(\lambda)$ by minimizing $(\lambda, 1)^\top(u, t)$.