Numerical Optimization

Unit 9: Constrained Optimization Problems

Che-Rung Lee

Department of Computer Science
National Tsing Hua University

November 22, 2022

Problem formulation

General formulation

$$\begin{aligned} \min_{\vec{x}} \quad & f(\vec{x}) \\ \text{s.t.} \quad & c_i(\vec{x}) = 0, \quad i \in \mathcal{E} \\ & c_i(\vec{x}) \geq 0, \quad i \in \mathcal{I}. \end{aligned} \tag{1}$$

- $\mathcal{E}$ is the index set for equality constraints; $\mathcal{I}$ is the index set for inequality constraints.
- $\Omega = \{\vec{x} | c_i(\vec{x}) = 0, i \in \mathcal{E}$ and $c_j(\vec{x}) \geq 0, j \in \mathcal{I}\}$ is the set of feasible solutions.
- The function $f(\vec{x})$ and $c_i(\vec{x})$ can be linear or nonlinear.

Example 1

Example

$$\min_{x_1, x_2} f(x_1, x_2) = x_1 + x_2$$

$$\text{s.t. } c(x_1, x_2) = x_1^2 + x_2^2 - 2 = 0.$$

- The optimal solution is at $\vec{x}^* = (x_1^*, x_2^*) = (-1, -1)$
- The gradient of c is $\nabla c = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$, and $\nabla c(\vec{x}^*) = \begin{pmatrix} -2 \\ -2 \end{pmatrix}$
- The gradient of $\nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Properties of the optimal solution in Example 1

Suppose $\vec{s}$ is a valid direction to move, and $\|\vec{s}\| \leq \epsilon$ is small.

1. $f(\vec{x}^* + \vec{s}) \geq f(\vec{x}^*)$

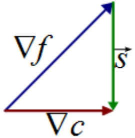
$$f(\vec{x}^* + \vec{s}) = f(\vec{x}^*) + \nabla f(\vec{x}^*)^T \vec{s} + O(\|\vec{s}\|^2) \Rightarrow \nabla f(\vec{x}^*)^T \vec{s} \geq 0,$$

2. $\vec{c}(\vec{x}^*) = \vec{c}(\vec{x}^* + \vec{s}) = 0$

$$\vec{c}(\vec{x}^* + \vec{s}) \approx c(\vec{x}^*) + \nabla c(\vec{x}^*)^T \vec{s} = 0 \Rightarrow \nabla c(\vec{x}^*)^T \vec{s} = 0$$

3. From 1. and 2., we can infer that ∇f must be parallel to ∇c at the optimal solution. (why?)

If ∇f is not parallel to ∇c, there will be an $\vec{s}$ that makes $\nabla f^T \vec{s} < 0$ and $\nabla c^T \vec{s} = 0$.



Example 2

Example

$$\min_{x_1, x_2} f(x_1, x_2) = x_1 + x_2$$

$$\text{s.t. } c(\vec{x}) = 2 - x_1^2 - x_2^2 \geq 0$$

In general, there are two cases for inequality constraints.

1. If $\vec{x}^*$ is inside the circle , then $\nabla f(\vec{x}^*) = 0$. (why?)
2. If $\vec{x}^*$ is on the circle , then $c(\vec{x}^*) = 0$, which goes back to the equality constraint.
3. From 1. and 2., we can conclude that $\nabla f(\vec{x}^*) = \lambda \nabla c(\vec{x}^*)$ for some scalar λ.
    - In the first case, $\lambda = 0$.
    - In the second case, λ is the scaling factor of $\nabla f(\vec{x}^*)$ and $\nabla c(\vec{x}^*)$.

Lagrangian function

The Lagrangian function

$$\mathcal{L}(\vec{x}, \lambda) = f(\vec{x}) - \lambda c(\vec{x}) \tag{2}$$

- $\nabla_{\vec{x}}\mathcal{L} = \dfrac{\partial \mathcal{L}}{\partial \vec{x}} = \nabla f(\vec{x}) - \lambda \nabla c(\vec{x})$.
- $\nabla_{\lambda}\mathcal{L} = \dfrac{\partial \mathcal{L}}{\partial \lambda} = -c(\vec{x})$.
- Therefore, at the optimal solution , $\nabla \mathcal{L} = \begin{pmatrix} \nabla_{\vec{x}}\mathcal{L}(\vec{x}^*) \\ \nabla_{\lambda}\mathcal{L}(\vec{x}^*) \end{pmatrix} = 0$.
- If $c(\vec{x}^*)$ is inactive , $\lambda^* = 0$. $\Rightarrow$ The complementarity condition $\lambda^* c(\vec{x}^*) = 0$.
- The scalar λ is called *Lagrange multiplier*.

Example 3

Example

$$
\begin{aligned}
\min_{x_1, x_2} \quad & f(x_1, x_2) = x_1 + x_2 \\
\text{s.t.} \quad & c_1(\vec{x}) = 2 - x_1^2 - x_2^2 \geq 0 \\
& c_2(\vec{x}) = x_2 \geq 0
\end{aligned}
$$

- $\nabla c_1 = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}$, $\nabla c_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
- The optimal solution $\vec{x}^* = (-\sqrt{2}, 0)^T$, at which $\nabla c_1(\vec{x}^*) = \begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix}$.
- $\nabla f(\vec{x}^*)$ is a linear combination of $\nabla c_1(\vec{x}^*)$ and $\nabla c_2(\vec{x}^*)$.

Example 3

- For this example, the Lagrangian $\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \lambda_1 c_1(\vec{x}) - \lambda_2 c_2(\vec{x})$, and

$$\nabla \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) = \begin{pmatrix} \nabla_{\vec{x}} \mathcal{L} \\ \nabla_{\lambda_1} \mathcal{L} \\ \nabla_{\lambda_2} \mathcal{L} \end{pmatrix} = \begin{pmatrix} \nabla f(\vec{x}^*) - c_1(\vec{x})/2\sqrt{2} - c_2(\vec{x}) \\ -c_1(\vec{x}^*) \\ -c_2(\vec{x}^*) \end{pmatrix} = \vec{0}.$$

  - What is $\vec{\lambda}^*$?
  - The examples suggests the first order necessity condition for constrained optimizations is the gradient of the Lagrangian is zero. But is it true?

Example 4

Example

$$\min_{x_1,x_2} \quad f(x_1, x_2) = x_1 + x_2$$
$$\text{s.t.} \quad c_1(\vec{x}) = (x_1^2 + x_2^2 - 2)^2 = 0$$

- $\nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\nabla c(\vec{x}) = \begin{pmatrix} 4(x_1^2 + x_2^2 - 2)x_1 \\ 4(x_1^2 + x_2^2 - 2)x_2 \end{pmatrix}$.
- Optimal solution is $(-1, -1)$, but $\nabla c(-1, -1) = (0, 0)^T$ is not parallel to ∇f.

Example 5

Example

$$\begin{aligned} \min_{x_1, x_2} \quad & f(x_1, x_2) = x_1 + x_2 \\ \text{s.t.} \quad & c_1(\vec{x}) = 1 - x_1^2 - (x_2 - 1)^2 \geq 0 \\ & c_2(\vec{x}) = -x_2 \geq 0 \end{aligned}$$

- $\nabla c_1 = \begin{pmatrix} -2x_1 \\ -2(x_2 - 1) \end{pmatrix}$, $\nabla c_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$, and $\nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
- The only solution is $(0, 0)$. $\nabla c_1(0, 0) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$,
  $\nabla c_2(0, 0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$.
- At the optimal solution, ∇f is not a linear combination of ∇c_1 and ∇c_2.

Regularity conditions

Regularity conditions: conditions of the constraints

Linear independence constraint qualifications (LICQ)

Given a point $\vec{x}$ and its active set $\mathcal{A}(\vec{x})$, LICQ holds if the gradients of the constraints in $\mathcal{A}(\vec{x})$ are linearly independent.

KKT conditions

KKT conditions: the first order necessary condition for the COP

The KKT conditions(Karush-Kuhn-Tucker)

Suppose $\vec{x}^*$ is a solution to the problem defined in (1), where f and c_i are continuously differentiable and the LICQ holds at $\vec{x}^*$. Then there exist a lagrangian multiplier vector $\vec{\lambda}^*$ s.t. the following conditions are satisfied at $(\vec{x}^*, \vec{\lambda}^*)$

1. $\nabla_{\vec{x}^*} \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) = 0$
2. $c_i(\vec{x}^*) = 0 \quad \forall i \in \mathcal{E}$
3. $c_i(\vec{x}^*) \geq 0 \quad \forall i \in \mathcal{I}$
4. $\lambda_i^* c_i(\vec{x}^*) = 0$ (Strict complementarity condition: either $\lambda_i^* = 0$ or $c_i(\vec{x}^*) = 0$.)
5. $\lambda_i^* \geq 0, \forall i \in \mathcal{I}$ ($\lambda_i^* > 0, \forall i \in \mathcal{I} \cap \mathcal{A}^*$ if the strict complementarity condition holds.)

Two definitions for the proof of KKT

Tangent and tangent cone

A vector $\vec{d}$ is said to be a *tangent* to a point set Ω at point $\vec{x}$ if there are a feasible sequence $\{\vec{z}_k\}$ approaching to $\vec{x}$ and a positive sequence $\{t_k\}$ converging to 0, such that

$$\lim_{k \to \infty} \frac{\vec{z}_k - \vec{x}}{t_k} = \vec{d}.$$

The set of all tangents to Ω at $\vec{x}^*$ is called the *tangent cone*.

The set of linearized feasible directions

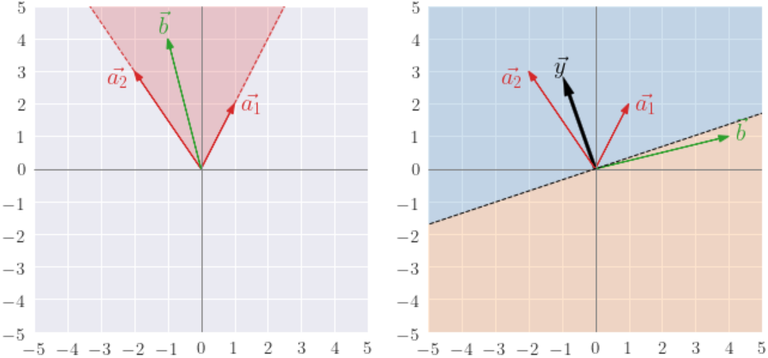
Given a feasible point $\vec{x}$ and the active constraint set $\mathcal{A}(\vec{x})$, the set of linearized feasible directions is defined as

$$\mathcal{F}(\vec{x}) = \left\{ \vec{d} \;\middle|\; \begin{array}{ll} \vec{d}^T \nabla c_i(\vec{x}) = 0 & \forall i \in \mathcal{E}, \\ \vec{d}^T \nabla c_i(\vec{x}) \geq 0 & \forall i \in \mathcal{A}(\vec{x}) \cap \mathcal{I} \end{array} \right\}.$$

Theorem (Farkas lemma)

Let A be an $m \times n$ matrix and $\vec{b}$ be an m vector. Exact one of the following two statements is true:

1. *There exists a $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{b}$ and $\vec{x} \geq 0$.*
2. *There exists a $\vec{y} \in \mathbb{R}^m$ such that $A^T\vec{y} \geq 0$ and $\vec{b}^T\vec{y} < 0$.*



Proof of Farkas lemma

- First, we show that both cannot happen simultaneously by contradiction.
- Suppose (1) and (2) are both held. Since $A^T\vec{y} \geq 0$ and $\vec{x} \geq 0$, $\vec{y}^T A\vec{x} \geq 0$. But $\vec{y}^T A\vec{x} = \vec{y}^T\vec{b} < 0$, which causes a contradiction.
- Second, we show that if (1) does not hold, then (2) must be true.
- Since the cone

$$Q = \{\vec{q} | \vec{q} = A\vec{x}, \vec{x} > 0\}$$

  is a convex set, we can find a separation hyper-plane to separate it from any other vectors outside the cone. So we can find a hyperplane whose normal vector is $\vec{y}$, and $A\vec{y} > 0$ and $\vec{b}^T\vec{y} < 0$. This means Q and $\vec{b}$ are on different side of the hyperplane.

Relation of Farkas lemma and KKT conditions

Farkas's Lemma (1902) plays an important role in the proof of the KKT condition. The most critical part in the proof of the KKT condition is to show that the Lagrange multiplier $\vec{\lambda}^* \geq 0$ for inequality constraints. We can say if the LICQ condition is satisfied at $\vec{x}^*$, then any feasible direction $\vec{u}$ at $\vec{x}^*$ must have the following properties:

1. $\vec{u}^T \nabla f(\vec{x}^*) \geq 0$ since $\vec{x}^*$ is a local minimizer. (Otherwise, we find a feasible descent direction that decreases f.)
2. $\vec{u}^T \nabla c_i(\vec{x}^*) = 0$ for equality constraints, $c_i = 0$.
3. $\vec{u}^T \nabla c_i(\vec{x}^*) \geq 0$ for inequality constraints, $c_i \geq 0$.

Here is how Farkas Lemma enters the theme. Let $\vec{b}$ be $\nabla f(\vec{x}^*)$, $\vec{y}$ be $\vec{u}$ (any feasible direction at $\vec{x}^*$), the columns of A be $\nabla c_i(\vec{x}^*)$). Since no such $\vec{u}$ exists, according to the properties of $\vec{y}$, statement (1) must hold. The vector $\vec{x}$ in (1) corresponds to $\vec{\lambda}^*$, which just gives us the desired result of the KKT condition.

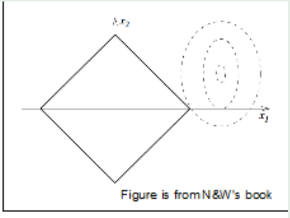
Outline of the proof of the KKT conditions

1. $\forall \vec{d} \in$ tangent cone at $\vec{x}^*$ $\vec{d}^T \nabla f \geq 0$. (Using the idea of tangent cone to prove it)
2. Tangent cone at $\vec{x}^*$ = feasible directions at $\vec{x}^*$
3. By 1 and 2 , $\vec{d}^T \nabla f \geq 0$ for $\forall \vec{d} \in F(\vec{x}^*)$
4. By Farkas lemma , either one need be true.
   1. There exists a $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{b}$ and $\vec{x} \geq 0$.
   2. There exists a $\vec{y} \in \mathbb{R}^m$ such that $A^T \vec{y} \geq 0$ and $\vec{b}^T \vec{y} < 0$.
5. Since (2) is not true (Because of 3) , (1) must be true.

Example 6

Example

$$\begin{array}{ll} \min_{x_1,x_2} & (x_1 - \frac{3}{2})^2 + (x_2 - \frac{1}{2})^4 \\ \text{s.t.} & c_1(\vec{x}) = 1 - x_1 - x_2 \geq 0 \\ & c_2(\vec{x}) = 1 - x_1 + x_2 \geq 0 \\ & c_3(\vec{x}) = 1 + x_1 - x_2 \geq 0 \\ & c_4(\vec{x}) = 1 + x_1 + x_2 \geq 0 \end{array}$$



$$\nabla c_1 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \nabla c_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \nabla c_3 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \nabla c_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$\vec{x}^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \text{ and } \nabla f(\vec{x}^*) = \begin{pmatrix} 2(x_1^* - \frac{3}{2}) \\ 4(x_2^* - \frac{1}{2})^3 \end{pmatrix} = \begin{pmatrix} -1 \\ -1/2 \end{pmatrix}.$$

$$\vec{\lambda}^* = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 \end{pmatrix}^T$$

The second order condition

- With constraints, we don't need to consider all the directions. The directions we only need to worry about are the "feasible directions".
- The critical cone $\mathcal{C}(\vec{x}^*, \vec{\lambda}^*)$ is a set of directions defined at the optimal solution $(\vec{x}^*, \vec{\lambda}^*)$

$$\vec{w} \in \mathcal{C}(\vec{x}^*, \vec{\lambda}^*) \Leftrightarrow \begin{cases} \nabla c_i(\vec{x}^*)^T \vec{w} = 0 & \forall i \in \mathcal{E} \\ \nabla c_i(\vec{x}^*)^T \vec{w} = 0 & \forall i \in \mathcal{A}(\vec{x}^*) \cap \mathcal{I},\ \lambda_i^* > 0 \\ \nabla c_i(\vec{x}^*)^T \vec{w} \geq 0 & \forall i \in \mathcal{A}(\vec{x}^*) \cap \mathcal{I},\ \lambda_i^* = 0 \end{cases}$$

The second order necessary condition

Suppose $\vec{x}^*$ is a local minimizer at which the LICQ holds, and $\vec{\lambda}^*$ is the Lagrange multiplier. Then $\vec{w}^T \nabla^2_{xx} \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) \vec{w} \geq 0, \quad \forall \vec{w} \in \mathcal{C}(\vec{x}^*, \vec{\lambda}^*)$.

Proof

We perform Taylor expansion at $\vec{x}^*$ and evaluate its neighbor $\vec{z}$,

$$\begin{aligned}\mathcal{L}(\vec{z}, \vec{\lambda}^*) =& \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) + (\vec{z} - \vec{x}^*)^T \nabla_x \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) \\ &+ \frac{1}{2}(\vec{z} - \vec{x}^*)^T \nabla_{xx}^2 \mathcal{L}(\vec{x}^*, \vec{\lambda}^*)(\vec{z} - \vec{x}^*) + O(\|\vec{z} - \vec{x}^*\|^3)\end{aligned}$$

Since $\mathcal{L}(\vec{x}^*, \vec{\lambda}^*) = f(\vec{x}^*)$ (why?) and $\nabla_x \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) = 0$. Let $\vec{w} = \vec{z} - \vec{x}^*$, which is in the critical cone.

$$\begin{aligned}\mathcal{L}(\vec{z}, \vec{\lambda}^*) &= f(\vec{z}) - \sum_{\forall\, i} \lambda_i^* c_i(\vec{z}) \\ &= f(\vec{z}) - \sum_{\forall\, i} \vec{\lambda}_i^* (c_i(\vec{x}^*) + \nabla c_i(\vec{x}^*)^T \vec{w}) = f(\vec{z})\end{aligned}$$

Thus, $f(\vec{z}) = \mathcal{L}(\vec{z}, \vec{\lambda}^*) = f(\vec{x}^*) + \frac{1}{2}\vec{w}^T \nabla_{xx}^2 \mathcal{L}(\vec{x}^*, \vec{\lambda}^*)\vec{w} + O(\|\vec{z} - \vec{x}^*\|^3)$, which is larger than $f(\vec{x}^*)$ if $\vec{w}^T \nabla_{xx}^2 \mathcal{L}(\vec{x}^*, \vec{\lambda}^*)\vec{w} \geq 0$.

Example

$$\min_{x_1,x_2} -0.1(x_1-4)^2 + x_2^2 \text{ s.t. } x_1^2 + x_2^2 - 1 \geq 0.$$

$\mathcal{L}(\vec{x},\lambda) = -0.1(x_1-4)^2 + x_2^2 - \lambda(x_1^2+x_2^2-1)$

$\nabla_x \mathcal{L} = \begin{pmatrix} -0.2(x_1-4) - 2\lambda x_1 \\ 2x_2 - 2\lambda x_2 \end{pmatrix}, \nabla^2_{xx}\mathcal{L} = \begin{pmatrix} -0.2-2\lambda & 0 \\ 0 & 2-2\lambda \end{pmatrix}$

The problem is unbounded, but we want to show that $\vec{x}^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is a local minimum, at which $\lambda^* = 0.3$ and $\nabla C(\vec{x}^*) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$.

The critical cone $\mathcal{C}(\vec{x}^*) = \left\{ \begin{pmatrix} 0 \\ w_2 \end{pmatrix} \middle| w_2 \in \mathbb{R} \right\}$

The Hessian $\nabla^2_{xx}\mathcal{L}(\vec{x}^*,\vec{\lambda}^*) = \begin{pmatrix} -0.4 & 0 \\ 0 & 1.4 \end{pmatrix}$ is not positive definite, but the projected Hessian $\begin{pmatrix} 0 & w_2 \end{pmatrix} \begin{pmatrix} -0.4 & 0 \\ 0 & 1.4 \end{pmatrix} \begin{pmatrix} 0 \\ w_2 \end{pmatrix} = 1.4w_2^2 \geq 0$ is.

Some easy way to check the condition

Is there any easy way to check the condition?

- Let Z be a matrix whose column vectors span the subspace of $\mathcal{C}(\vec{x}^*, \vec{\lambda}^*)$

$$\Rightarrow \begin{cases} \forall \vec{w} \in \mathcal{C}(\vec{x}^*, \vec{\lambda}^*), & \exists \vec{u} \in \mathbb{R}^m \quad s.t.\ \vec{w} = Z\vec{u} \\ \forall \vec{u} \in \mathbb{R}^m, & Z\vec{u} \in \mathcal{C}(\vec{x}^*, \vec{\lambda}^*) \end{cases}$$

- To check $\vec{w}^T \nabla_{xx} \mathcal{L}^* \vec{w} \geq 0$, $\Leftrightarrow \vec{u}^T Z^T \nabla^2_{xx} \mathcal{L}^* Z \vec{u} \geq 0$ for all $\vec{u}$ $\Leftrightarrow Z^T \nabla^2_{xx} \mathcal{L}^* Z$ is positive semidefinite.
- The matrix $Z^T \nabla^2_{xx} \mathcal{L}^* Z$ is called the *projected Hessian*.

Active constraint matrix

- Let $A(\vec{x}^*)$ be the matrix whose rows are the gradient of the active constraints at the optimal solution $\vec{x}^*$.

$$A(\vec{x}^*)^T = [\nabla c_i(\vec{x}^*)]_{i\in\mathcal{A}(\vec{x}^*)}$$

- The critical cone $\mathcal{C}(\vec{x}^*, \vec{\lambda}^*)$ is the null space of $A(\vec{x}^*)$

$$\vec{w} \in \mathcal{C}(\vec{x}^*, \vec{\lambda}^*) \Leftrightarrow A(\vec{x}^*)\vec{w} = 0$$

- We don't consider the case that $\lambda^* = 0$ for active c_i. (Strict complementarity condition.)

Compute the null space of $A(\vec{x}^*)$

- Using QR factorization

$$A(\vec{x}^*)^T = Q\begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}\begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1$$

$A \in \mathbb{R}^{m\times n},\ Q \in \mathbb{R}^{n\times n},\ R \in \mathbb{R}^{m\times m},\ Q_1 \in \mathbb{R}^{n\times m},\ Q_2 \in \mathbb{R}^{n\times (n-m)}$

- The null space of A is spanned by Q_2, which means any vectors in the null space of A is a unique linearly combination of Q_2's column vectors.

$$\vec{z} = Q_2\vec{v} \qquad A\vec{z} = R^T Q_1^T Q_2 \vec{v} = 0$$

To check the second order condition is to check if $Q_2^T \nabla^2 \mathcal{L}^* Q_2$ is positive definite.

Duality

Consider the problem: $\min_{\vec{x}\in\mathbb{R}^n} f(\vec{x})$ subject to $c(\vec{x}) = \begin{pmatrix} c_1(\vec{x}) \\ c_2(\vec{x}) \\ \vdots \\ \\ c_m(\vec{x}) \end{pmatrix} \geq 0$

Its Lagrangian function is

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \vec{\lambda}^T c(\vec{x})$$

The dual problem is defined as

$$\max_{\vec{\lambda}\in\mathbb{R}^n} q(\vec{\lambda}) \quad \text{s.t.} \quad \vec{\lambda} \geq 0$$

where $q(\vec{\lambda}) = \inf_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda})$.

Duality

- Infimum is the global minimum of $\mathcal{L}(\cdot, \vec{\lambda})$, which may not be defined or difficult to compute.
- For f and $-c_i$ are convex, $\mathcal{L}$ is also convex $\Rightarrow$ the local minimizer is the global minimize.
- Wolfe's duality: another formulation of duality when function is differentiable.

$$\max_{\vec{x}.\vec{\lambda}} \mathcal{L}(\vec{x}, \vec{\lambda})$$

$$s.t.\ \nabla_x \mathcal{L}(\vec{x}, \vec{\lambda}) = 0,\ \lambda \geq 0$$

Example

Example

$\min_{(x_1\ x_2)} 0.5(x_1^2 + x_2^2) \quad s.t.\ x_1 - 1 \geq 0$

- $\mathcal{L}(X_1\ X_2, \vec{\lambda}) = 0.5(x_1^2 + x_2^2) - \lambda_1(x_1 - 1)$,
- $\nabla_x \mathcal{L} = \begin{pmatrix} x_1 - \lambda_1 \\ x_2 \end{pmatrix} = 0$, which implies $x_1 = \lambda_1$ and $x_2 = 0$.
- $q(\lambda) = \mathcal{L}(\lambda_1, 0, \lambda_1) = -0.5\lambda_1^2 + \lambda_1$.
- The dual problem is

$$\max_{\lambda_1 \geq 0} -0.5\lambda_1^2 + \lambda_1$$

Weak duality

Weak duality: For any $\vec{x}$ and $\vec{\lambda}$ feasible, $q(\vec{\lambda}) \le f(\vec{x})$
$q(\vec{\lambda}) = \inf_{\vec{x}}(f(\vec{x}) - \vec{\lambda}^T c(\vec{x})) \le f(\vec{x}) - \vec{\lambda}^T c(\vec{x}) \le f(\vec{x})$

Example

$$\min_{\vec{x}} \vec{c}^T\vec{x} \quad s.t.\ A\vec{x} - \vec{b} \ge 0,\ \vec{x} \ge 0$$

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = \vec{c}^T\vec{x} - \vec{\lambda}^T(A\vec{x} - \vec{b}) = (\vec{c}^T - \vec{\lambda}^T A)\vec{x} + \vec{b}^T\vec{\lambda}$$

Since $\vec{x} \ge 0$, if $(\vec{c} - A^T\vec{\lambda})^T < 0$, $\inf_{\vec{x}} \mathcal{L} \to -\infty$. We require $\vec{c}^T - A^T\vec{\lambda} > 0$.

$$q(\vec{\lambda}) = \inf_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}) = \vec{b}^T\vec{\lambda}$$

The dual problem becomes

$$\max_{\lambda} \vec{b}^T\vec{\lambda} \qquad s.t. \quad A^T\vec{\lambda} \le \vec{c} \ \text{ and } \vec{\lambda} \ge 0.$$

The rock-paper-scissors game (two person zero sum game)

The payoff matrix $A =$

	Rock	Paper	Scissors
Rock	0	1	-1
Paper	-1	0	1
Scissors	1	-1	0

- Suppose the opponent's strategy is $\vec{x} = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$.
- What should your strategy be to maximize the payoff?

Problem formulation

- Let $\vec{y} = (y_1, y_2)^T$. We can express this problem as

$$\max_{\vec{y}} \vec{x}^T A\vec{y} = \max_{\vec{y}} \frac{-1}{2} y_1 + \frac{1}{2} y_2$$

  Therefore, to maximize your wining chance, you should throw paper.
- On the other hand, the problem of your opponent is

$$\min_{\vec{x}} \vec{x}^T A\vec{y}$$

- What if you do not know your opponent's strategy? It becomes a min-max or max-min problem.

$$\max_{\vec{y}} \min_{\vec{x}} \vec{x}^T A\vec{y}$$

Two examples

Example

Consider the payoff matrix $A = \begin{pmatrix} -1 & 2 \\ 4 & 3 \end{pmatrix}$, and $\vec{x}, \vec{y} \in \{0, 1\}$.

- $\min_i \max_j a_{ij} = \min_i \left\{ \max_j a_{1,j}, \max_j a_{2,j} \right\} = \min\{2, 4\} = 2.$
- $\max_j \min_i a_{ij} = \max_j \left\{ \min_i a_{i,1}, \min_i a_{i,2} \right\} = \max\{-1, 2\} = 2.$

Example

Consider the payoff matrix $A = \begin{pmatrix} -1 & 2 \\ 4 & 1 \end{pmatrix}$

- $\min_i \max_j a_{ij} = \min_i \left\{ \max_j a_{1,j}, \max_j a_{2,j} \right\} = \min\{2, 4\} = 2.$
- $\max_j \min_i a_{ij} = \max_j \left\{ \min_i a_{i,1}, \min_i a_{i,2} \right\} = \max\{-1, 1\} = 1.$

Strong duality theorem

Strong duality theorem

$\max_{\vec{y}} \min_{\vec{x}} F(\vec{x}, \vec{y}) = \min_{\vec{x}} \max_{\vec{y}} F(\vec{x}, \vec{y})$ if and only if there exists a point $(\vec{x}^*, \vec{y}^*)$ such that $F(\vec{x}^*, \vec{y}) \leq F(\vec{x}^*, \vec{y}^*) \leq F(\vec{x}, \vec{y}^*)$.

- Point $(\vec{x}^*, \vec{y}^*)$ is called a saddle point.