Numerical Optimization
Unit 7: Constrained Optimization Problems

Che-Rung Lee

Scribe: 周宗毅

March 28, 2011
Problem formulation

General formulation

\[
\min_{\vec{x}} \quad f(\vec{x}) \\
\text{s.t.} \quad c_i(\vec{x}) = 0, \quad i \in \mathcal{E} \\
\quad c_i(\vec{x}) \geq 0, \quad i \in \mathcal{I}.
\] (1)

- $\mathcal{E}$ is the index set for equality constraints; $\mathcal{I}$ is the index set for inequality constraints.
- $\Omega = \{\vec{x} | c_i(\vec{x}) = 0, i \in \mathcal{E} \text{ and } c_j(\vec{x}) \geq 0, j \in \mathcal{I}\}$ is the set of feasible solutions.
- The function $f(\vec{x})$ and $c_i(\vec{x})$ can be linear or nonlinear.
Example 1

\[ \min_{x_1, x_2} f(x_1, x_2) = x_1 + x_2 \]
\[ \text{s.t. } c(x_1, x_2) = x_1^2 + x_2^2 - 2 = 0. \]

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

1. \( f(\vec{x}^* + \vec{s}) \geq f(\vec{x}^*) \) for small enough \( \vec{s} \). (why?)

\[
f(\vec{x}^* + \vec{s}) = f(\vec{x}^*) + \nabla f(\vec{x}^*)^T \vec{s} + O(\|\vec{s}\|^2) \Rightarrow \nabla f(\vec{x}^*)^T \vec{s} \geq 0, \quad \forall \vec{s}, \|\vec{s}\| \leq \epsilon
\]

2. \( \vec{c}(\vec{x}^*) = \vec{c}(\vec{x}^* + \vec{s}) = 0 \) for small enough \( \vec{s} \). (why?)

\[
\vec{c}(\vec{x}^* + \vec{s}) \approx c(\vec{x}^*) + \nabla c(\vec{x}^*)^T \vec{s} = 0 \Rightarrow \nabla c(\vec{x}^*)^T \vec{s} = 0, \quad \forall \vec{s}, \|\vec{s}\| \leq \epsilon
\]

3. From 1. and 2., we can infer that \( \nabla f \) must be parallel to \( \nabla c \). (why?)

If \( \nabla f \) is not parallel to \( \nabla c \), there will be an \( \vec{s} \) that makes \( \nabla f^T \vec{s} < 0 \) and \( \nabla c^T \vec{s} = 0 \), as shown in the figure.
Example 2

\[
\min_{x_1, x_2} f(x_1, x_2) = x_1 + x_2 \\
\text{s.t. } c(\vec{x}) = 2 - x_1^2 - x_2^2 \geq 0
\]

What are the properties of the optimal solution in Example 2?

1. If \( f(\vec{x}^*) \) is inside the circle, then \( \nabla f(\vec{x}^*) = 0 \). (why?)

2. If \( f(\vec{x}^*) \) 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 \( \lambda \).
   
   - In the first case, \( \lambda = 0 \).
   - In the second case, \( \lambda \) is the scaling factor of \( \nabla f(\vec{x}^*) \) and \( \nabla c(\vec{x}^*) \).
The Lagrangian function

\[ \mathcal{L}(\vec{x}, \lambda) = f(\vec{x}) - \lambda c(\vec{x}) \]  \hspace{1cm} (2)

- \( \nabla_{\vec{x}} \mathcal{L} = \frac{\partial \mathcal{L}}{\partial \vec{x}} = \nabla f(\vec{x}) - \lambda \nabla c(\vec{x}). \)
- \( \nabla_{\lambda} \mathcal{L} = \frac{\partial \mathcal{L}}{\partial \lambda} = -c(\vec{x}). \)

Therefore, at the optimal solution, \( \nabla \mathcal{L} = \left( \begin{array}{c} \nabla_{\vec{x}} \mathcal{L}(\vec{x}^*) \\ \nabla_{\lambda} \mathcal{L}(\vec{x}^*) \end{array} \right) = 0. \)

- If \( c(\vec{x}^*) \) is inactive, \( \lambda^* = 0. \) \( \Rightarrow \) The complementarity condition \( \lambda^* c(\vec{x}^*) = 0. \)
- The scalar \( \lambda \) is called \textit{Lagrange multiplier}. 

Example 3

**Example**

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

- \[\nabla c_1 = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}, \quad \nabla c_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \nabla f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.\]
- 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

\[ \begin{align*}
\min_{x_1, x_2} & \quad f(x_1, x_2) = x_1 + x_2 \\
\text{s.t.} & \quad c_1(\vec{x}) = (x_1^2 + x_2^2 - 2)^2 = 0
\end{align*} \]

- \( \nabla f = \begin{pmatrix} 1 \\
1 \end{pmatrix} \) and \( \nabla \vec{c}(\vec{x}) = \begin{pmatrix} 4(x_1^2 + x_2^2 - 2)x_1 \\
4(x_1^2 + x_2^2 - 2)x_2 \end{pmatrix} \).

- Optimal solution is \((-1, -1)\), but \( \nabla c(-1, -1) = (0, 0)^T \) is not parallel to \( \nabla f \).
Example 5

Example

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

- \( \nabla c_1 = \begin{pmatrix} -2x_1 \\ -2(x_2 - 1) \end{pmatrix} \), \( \nabla c_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \), 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, \( \nabla f \) is not a linear combination of \( \nabla c_1 \) and \( \nabla c_2 \).
Regularity conditions: conditions of the constraints

**Linear independence constraint qualifications (LICQ)**

Given a point $\bar{x}$ and its active set $A(\bar{x})$, LICQ holds if the gradients of the constraints in $A(\bar{x})$ 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 $\lambda^*$ such that the following conditions are satisfied at $(\vec{x}^*, \lambda^*)$

1. $\nabla_{\vec{x}^*} \mathcal{L} (\vec{x}^*, \lambda^*) = 0$
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. $\lambda^*_i c_i(\vec{x}^*) \geq 0$ (Strict complementarity condition: either $\lambda^*_i = 0$ or $c_i(\vec{x}^*) = 0$.)
5. $\lambda^*_i \geq 0, \forall i \in \mathcal{I}$ (If $\lambda^*_i > 0$, $\forall i \in \mathcal{I} \cup \mathcal{A}^*$ if the strict complementarity condition holds.)
Two definitions for the proof of KKT

**Tangent cone**

A vector $\vec{d}$ is said to be a *tangent* to a point set $\Omega$ at point $\vec{x}$ if there are a sequence $\{\vec{z}_k\}$ and a sequence $\{t_k\}$, in which $t_k > 0$ and $\{t_k\}$ converges to 0, such that

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

The set of all tangents to $\Omega$ 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

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

It can be shown that $\mathcal{F}(\vec{x})$ is a cone.
Outline of the proof of the KKT conditions

1. \( \forall \vec{d} \in \text{tangent cone at } \vec{x}^* \quad d^T \nabla f \geq 0. \) (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, \( d^T \nabla f \geq 0 \) for \( \forall d \in F(\vec{x}^*) \)

4. By Farkas lemma, either one need be true.\(^1\)
   (a) \( \exists d \in \mathbb{R}^n, d^T \nabla f < 0, B^T d \geq 0 \quad c^T d = 0 \)
   (b) \( \nabla f \in \{By + Cw | y \geq 0\} \)

5. Since (a) is not true (Because of 3), (b) must be true.

\(^1\)The proof of Farkas lemma can be found in last year’s homework 4.
Example 6

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

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

\[
\vec{x}^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad \nabla f(\vec{x}^*) = \begin{pmatrix} \frac{2(x_1^* - \frac{3}{2})}{4(x_2^* - \frac{1}{2})^3} \end{pmatrix} = 11 \begin{pmatrix} -1 \\ -1/2 \end{pmatrix}.
\]

\[
\vec{\lambda}^* = \begin{pmatrix} \frac{3}{4} \\ \frac{1}{4} \\ 0 \\ 0 \end{pmatrix}^T
\]
The second order condition

- With constraints, we don’t need to consider all the directions. The directions we only need to worried about are the ”feasible directions”.

- The critical cone $\mathcal{C}(\bar{x}^*, \bar{\lambda}^*)$ is a set of directions defined at the optimal solution $(\bar{x}^*, \bar{\lambda}^*)$

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

The second order necessary condition

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

Proof

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

\[
\mathcal{L}(\vec{z}, \vec{\lambda}^*) = \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) + (\vec{z} - \vec{x}^*)^T \nabla_x \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) + \frac{1}{2}(\vec{z} - \vec{x}^*)^T \nabla^2_{xx} \mathcal{L}(\vec{x}^*, \vec{\lambda}^*)(\vec{z} - \vec{x}^*) + O(\|\vec{z} - \vec{x}^*\|^3)
\]

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.

\[
\mathcal{L}(\vec{z}, \vec{\lambda}^*) = f(\vec{z}) - \sum \lambda^*_i c_i(\vec{z}) = f(\vec{z}) - \sum \vec{\lambda}^*_i (c_i(\vec{x}^*) + \nabla c_i(\vec{x}^*)^T \vec{w}) = f(\vec{z})
\]

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

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

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

\[
\nabla_x \mathcal{L} = \begin{pmatrix}
-0.2(x_1 - 4) + 2\lambda x_1 \\
2x_2 - 2\lambda x_2
\end{pmatrix},
\nabla_{xx} \mathcal{L} = \begin{pmatrix}
-0.2 - 2\lambda & 0 \\
0 & 2 - 2\lambda
\end{pmatrix}
\]

at \( \vec{x}^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \), \( \lambda^* = 0.3 \)

\[\nabla C(\vec{x}^*) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}\]

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

\[\nabla_{xx} \mathcal{L}(\vec{x}^*, \lambda^*) = \begin{pmatrix}
-0.4 & 0 \\
0 & 1.4
\end{pmatrix}
\]

\[
\begin{pmatrix} 0 & w_2 \end{pmatrix} \begin{pmatrix}
-0.4 & 0 \\
0 & 1.4
\end{pmatrix} \begin{pmatrix} 0 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 & 1.4w_2 \end{pmatrix} \begin{pmatrix} 0 \\ w_2 \end{pmatrix} = 1.4w_2^2 > 0
\]

(UNIT 7) Numerical Optimization March 28, 2011 18 / 29
Is there any easy way to check the condition?

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

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

- To check $\vec{w}^T \nabla_{xx} \mathcal{L}^* \vec{w} \geq 0$, $\iff \vec{u}^T Z^T \nabla_{xx} \mathcal{L}^* Z \vec{u} \geq 0$ for all $\vec{u}$
  $\iff Z^T \nabla_{xx} \mathcal{L}^* Z$ is positive semidefinite.

- The matrix $Z^T \nabla_{xx} \mathcal{L}^* Z$ 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 $\vec{x}^*$.
  
  $$A(\vec{x}^*)^T = [\nabla c_i(\vec{x}^*)]_{i \in A(\vec{x}^*)}$$

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

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

- 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

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

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

- 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.

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

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

\[
\begin{align*}
\max \mathcal{L}(\vec{x}, \vec{\lambda}) \\
\text{s.t. } \nabla_x \mathcal{L}(\vec{x}, \vec{\lambda}) = 0, \quad \lambda \geq 0
\end{align*}
\]
Example

\[
\begin{align*}
\min_{x_1, x_2} & \quad 0.5(x_1^2 + x_2^2) \\
s.t. & \quad x_1 - 1 \geq 0
\end{align*}
\]

- \( \mathcal{L}(X_1, X_2, \lambda) = 0.5(x_1^2 + x_2^2) - \lambda_1(x_1 - 1), \)
- \( \nabla_x \mathcal{L} = \begin{pmatrix} x_1 - \lambda_1 \\ x_2 \end{pmatrix} = 0, \) 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 $\bar{x}$ and $\bar{\lambda}$ feasible, $q(\bar{\lambda}) \leq f(\bar{x})$

$q(\lambda) = \inf_{\bar{x}} (f(\bar{x}) - \bar{\lambda}^T c(\bar{x})) \leq f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) \leq f(\bar{x})$

**Example**

\[
\begin{align*}
\min_{\bar{x}} & \bar{c}^T \bar{x} \quad \text{s.t.} \quad A\bar{x} - \bar{b} \geq 0, \quad \bar{x} \geq 0 \\
\mathcal{L}(\bar{x}, \bar{\lambda}) &= \bar{c}^T \bar{x} - \bar{\lambda}^T (A\bar{x} - \bar{b}) = (\bar{c}^T - \bar{\lambda}^T A)\bar{x} + \bar{b}^T \bar{\lambda}
\end{align*}
\]

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

\[q(\bar{\lambda}) = \inf_{\bar{x}} \mathcal{L}(\bar{x}, \bar{\lambda}) = \bar{b}^T \bar{\lambda}\]

The dual problem becomes

\[
\max_{\bar{\lambda}} \bar{b}^T \bar{\lambda} \quad \text{s.t.} \quad A^T \bar{\lambda} \leq 0 \text{ and } \bar{\lambda} \geq 0.
\]
The rock-paper-scissors game (two person zero sum game)

The payoff matrix $A =$

<table>
<thead>
<tr>
<th>opp</th>
<th>you</th>
<th>Rock</th>
<th>Paper</th>
<th>Scissors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rock</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>Paper</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Scissors</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

- Suppose the opponent’s strategy is $\vec{x} = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$.

- What should your strategy be to maximize the payoff?
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 winning 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 \( \vec{x}, \vec{y} \in \{0, 1\} \).

\[
\begin{align*}
\min_i \max_j a_{ij} &= \min_i \left\{ \max_j a_{1,j}, \max_j a_{2,j} \right\} = \min\{2, 4\} = 2. \\
\max_j \min_i a_{ij} &= \max_j \left\{ \min_i a_{i,1}, \min_i a_{i,2} \right\} = \max\{-1, 2\} = 2.
\end{align*}
\]

Example

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

\[
\begin{align*}
\min_i \max_j a_{ij} &= \min_i \left\{ \max_j a_{1,j}, \max_j a_{2,j} \right\} = \min\{2, 4\} = 2. \\
\max_j \min_i a_{ij} &= \max_j \left\{ \min_i a_{i,1}, \min_i a_{i,2} \right\} = \max\{-1, 1\} = 1.
\end{align*}
\]
Strong duality theorem

\[
\max_{\bar{y}} \min_{\bar{x}} F(\bar{x}, \bar{y}) = \min_{\bar{x}} \max_{\bar{y}} F(\bar{x}, \bar{y}) \quad \text{if and only if there exists a point } (\bar{x}^*, \bar{y}^*) \\
\text{such that } F(\bar{x}^*, \bar{y}) \leq F(\bar{x}^*, \bar{y}^*) \leq F(\bar{x}, \bar{y}^*).
\]

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