Optimization Problems

• A set of decision variables, $x$, in vector or matrix form with dimension $n$
• A continuous and sometime differentiable objective function $f(x)$
• A feasible region where $x$ can be in

• One can smooth them by reformulation as constrained optimization:

  $$\max \min_i \{f_i(x), i=1,\ldots,n\} \rightarrow$$

  $$\max \alpha \quad \text{s.t.} \quad \alpha - f_i(x) \leq 0, \text{ for } i=1,\ldots,n$$
Function, Gradient Vector and Hessian Matrix

• A function $f$ of $x$ in $\mathbb{R}^n$

• The Gradient Vector of $f$ at $x$

$$\nabla f(x) = \left( \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

• The Hessian Matrix of $f$ at $x$

$$\nabla^2 f(x) = \begin{pmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \ldots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_n}
\end{pmatrix}$$

• Taylor’s Expansion Theorem
Convex and Concave Functions

\( f(x) \) is a \textit{convex function} if and only if for any given two points \( x_1 \) and \( x_2 \) in the function domain and for any constant \( 0 \leq \alpha \leq 1 \)

\[
f(\alpha x_1 + (1- \alpha)x_2) \leq \alpha f(x_1) + (1- \alpha)f(x_2)
\]

Strictly convex if \( x_1 \neq x_2 \),

\[
f(0.5x_1 + 0.5x_2) < 0.5f(x_1) + 0.5f(x_2)
\]
Convex Quadratic Functions

\[ f(x) = x^T Q x + c^T x \] is a convex function if and only if Hessian matrix $Q$ is positive semi-definite (PSD).

\[ f(x) = x^T Q x + c^T x \] is a strictly convex function if and only if $Q$ is positive definite (PD).

$Q$ is PSD if and only if $x^T Q x \geq 0$ for all $x$.

A 2x2 matrix is PSD (or PD) if and only if two diagonal entries and the determinant are nonnegative (or positive)
Convex Sets

• A set is **convex** if every line segment connecting any two points in the set is contained entirely within the set
  - Ex - polyhedron
  - Ex - ball

• An **extreme point** of a convex set is any point that is not on any line segment connecting any other two distinct points of the set

• The intersection of convex sets is a convex set

• A set is closed if the limit of any convergent sequence of the set belongs to the set
If $f(x)$ is a convex function, then the lower level set \( \{x: f(x) \leq b\} \) is a convex set for any constant $b$.

The graph of a convex function lies above its tangent line (planes). The Hessian matrix of a convex function is positive semi-definite.
Optimization Problem Classes

- **Unconstrained Optimization**
  - Convex or Nonconvex

- **Constrained Optimization**
  - Conic Linear Optimization/Programming (CLO/CLP)
  - Convex Constrained Optimization (CCO)
    - Feasible region/set convex; objective general
  - Generally Constrained Optimization (GCO)
  - Convex Optimization (CO)
    - Minimize a convex function over a convex feasible set
    - Maximize a concave function over a convex feasible set

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad x \in X
\end{align*}
\]
### Optimization Problem Forms

<table>
<thead>
<tr>
<th>Min $c^T x$</th>
<th>Min $f(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x \in K$</td>
<td></td>
</tr>
<tr>
<td>$Ax - b = 0$</td>
<td>s.t. $h_i(x) = 0, i=1,...,m$</td>
</tr>
<tr>
<td></td>
<td>$c_i(x) \geq 0, i=1,...,p$</td>
</tr>
</tbody>
</table>

**Conic Linear Optimization (CLO)**

- $A$: an $m \times n$ matrix
- $c$: objective coefficient
- $K$: a closed convex cone

This is convex optimization

**Generally Constrained Optimization (GCO)**

Each function can be continuous, continuously differentiable ($C^1$), or twice continuously differentiable ($C^2$)

It is CCO if $c_i$ are all concave, and $h_i$ are all linear/affine functions.

In addition, if $f$ is convex, it is CO.
Why do we care about convex optimization?

• It guarantees that every local optimizer is a global optimizer
• It guarantees that every (first-order) KKT (or stationary) point/solution is a global optimizer
• This is significant because all of our numerical optimization algorithms search/generate a KKT point/solution
• Sometime the problem can be “convexfied”:
  \[
  \min c^T x, \quad \text{s.t.} \quad ||x||^2 = 1 \\
  \]
  \[
  \uparrow \\
  \min c^T x, \quad \text{s.t.} \quad ||x||^2 \leq 1 \\
  \]
Optimization **Theory**: Mathematical Foundations

- Taylor’s Expansion Theorem
- Implicit Function Theorem
- Separating Hyperplane Theorem
- Supporting Hyperplane Theorem
- Caratheodory’s Theorem
- Duality and KKT Optimality Conditions
- Alternative Linear System/Farkas’ Lemma
- Implicit Function Theorem
- Separating Hyperplane Theorem
- Supporting Hyperplane Theorem
- Caratheodory’s Theorem
- Duality and KKT Optimality Conditions
- Alternative Linear System/Farkas’ Lemma
Theory: Feasibility Conditions

• Feasibility Conditions or Farkas’ Lemmas are developed to characterize and certify feasibility or infeasibility of a feasible region

• Alternative Systems X and Y: X has a feasible solution if and only if Y has no feasible solution
  • X and Y cannot both have feasible solution
  • Exactly one of them has a feasible solution

• They can be viewed as special cases of Linear Programming primal and dual pairs
Alternative Systems and CLO Pairs I

System X
- \( A: \text{an } m \times n \text{ matrix} \)
- \( b: \text{m-dimension vector} \)
- \( K: \text{a closed convex cone} \)

\[
Ax - b = 0, \quad x \in K
\]

\[
p^* = \text{min } 0^T x \\
\text{s.t. } Ax - b = 0, \quad x \in K
\]

System Y
- \( K^* \text{ is the dual cone} \)

\[
b^Ty = 1(>0) \\
A^Ty + s = 0, \quad s \in K^*
\]

\[
d^* = \text{max } b^Ty \\
\text{s.t. } A^Ty + s = 0, \quad s \in K^*
\]
Alternative Systems and CLO Pairs II

**System X**

- $c^T x = -1(<0)$
- $Ax = 0,$
- $x \in K$

**System Y**

- $A^T y + s - c = 0,$
- $s \in K^*$

**System X**

- $A$: an $m \times n$ matrix
- $c$: $n$-dimensional vector
- $K$: a closed convex cone

**System Y**

- $K^*$ is the dual cone

**CME307/MS&E311 Optimization Lecture Notes #10**
Feasibility Test Machine

Is system X feasible?

Yes

Is system Y feasible?

“No” under any circumstances

No

Is system Y feasible?

“Yes” under certain conditions of cone $K$ and data matrix $A$:

a) $K$ is a polyhedron cone, or

b) $Ax$ or $A^Ty$ has an interior solution
## General Rules to Construct the CLO Dual

<table>
<thead>
<tr>
<th>OBJ Vector/Matrix</th>
<th>RHS Vector/Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHS Vector/Matrix</td>
<td>OBJ Vector/matrix</td>
</tr>
<tr>
<td>$A$</td>
<td>$A^T$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Max model</th>
<th>Min model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_j \geq_k 0$</td>
<td>$j$th constraint $\geq_{K^*}$</td>
</tr>
<tr>
<td>$x_j \leq_k 0$</td>
<td>$j$th constraint $\leq_{K^*}$</td>
</tr>
<tr>
<td>$x_j$ free</td>
<td>$j$th constraint $=$</td>
</tr>
<tr>
<td>$i$th constraint $\leq_k$</td>
<td>$y_i \geq_{K^*} 0$</td>
</tr>
<tr>
<td>$i$th constraint $\geq_k$</td>
<td>$y_i \leq_{K^*} 0$</td>
</tr>
<tr>
<td>$i$th constraint $=$</td>
<td>$y_i$ free</td>
</tr>
</tbody>
</table>

The dual of the dual is the primal
Theory: Optimality Conditions

- Optimality (KKT) Conditions are developed to characterize and certify possible minimizers
  - Feasibility of original variables
  - Optimality conditions consist of original variables and Lagrange multipliers
  - Zero-order, First-order, Second-order, necessary, sufficient

- They may not lead directly to a very efficient algorithm for solving problems, but they do have a number of benefits:
  - They give insight into what optimal solutions look like
  - They provide a way to set up and solve small problems
  - They provide a method to check solutions to large problems
  - The Lagrange multipliers can be seen as sensitivities of the constraints

- A minimizers may not satisfy optimality conditions unless certain constraint qualifications hold.
KKT Optimality Condition Test Machine

Is \( \mathbf{x} \) a (local) optimizer?

• “Yes” only under certain circumstances

Is \( \mathbf{x} \) not a (local) optimizer?

• “Not” under certain constraint qualifications:
  a) Feasible region has an interior, or
  b) \( \mathbf{x} \) is a regular point on the hypersurface of active constraints
Duality Theorems for CLO

Primal Problem
A: an m x n matrix
c: objective coefficient
K: a closed convex cone

Weak Duality Theorem
K* is the dual cone

Dual Problem

0-Order Condition:  p* = d*

Strong Duality Theorem: They must equal?

"Yes" under certain conditions of cone K and data matrix A, b, c:
1) K is a polyhedron cone, or
2) either one has an interior feasible solution
The Lagrange Function of GCO

$$\min \quad f(\mathbf{x})$$

s.t.
$$c_i(\mathbf{x}) \ (\leq, =, \geq) 0, \ i=1,\ldots,m$$

Restriction on multipliers $y_i$,
$$y_i \ (\leq,"free",\geq) 0, \ i=1,\ldots,m$$

The Lagrange Function
$$L(\mathbf{x},\mathbf{y}) = f(\mathbf{x}) - \sum_i y_i c_i(\mathbf{x})$$

The Lagrange function can be interpreted as a “penalized” aggregated objective function:

- $y_i$ free: can be penalized either way
- $y_i \geq 0$ : can be penalized when $c_i(\mathbf{x}) \leq 0$
- $y_i \leq 0$ : can be penalized when $c_i(\mathbf{x}) \geq 0$
### The Lagrangian Duality for GCO

- **Weak Duality Theorem**
  - \( p^* \geq d^* \)

- **Strong Duality Theorem**
  - They must equal?

#### \( p^* \)\text{-}Optimal

\[
p^* = \min_{\mathbf{x}} \quad f(\mathbf{x}) \\
\text{s.t.} \quad c_i(\mathbf{x}) \ (\geq, =, \leq) 0, \ i=1,\ldots,m
\]

#### Dual \( d^* \)\text{-}Optimal

\[
d^* = \max_{\mathbf{y}} \quad \phi(\mathbf{y}) \\
\text{s.t.} \quad y_i \ (\leq, ”free”, \geq) 0, \ i=1,\ldots,m
\]

#### Lagrangian Function

Let \( \phi(\mathbf{y}) = \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}) \)

#### 0-Order Condition

- Sufficient!

- **Sufficient!**

0-Order Condition: \( p^* = d^* \)
Zero-Order Optimality Test for CLO and GCO

Is $\mathbf{x}$ an optimizer?

- "Yes" under any circumstances
- Failed

Is $\mathbf{x}$ not a (local) optimizer?

- Passed
- a) "Not" for sure if $K$ is a polyhedral cone in CLO; or
- b) "Not" when Feasible region has an interior in CCO; otherwise
- c) Inconclusive in GCO.

Zero-order condition is sufficient
1 and 2-order Conditions: Unconstrained

• Problem:
  – Minimize \( f(x) \), where \( x \) is a vector that could have any values, positive or negative

• First Order Necessary Condition (min or max):
  – \( \nabla f(x) = 0 \) (\( \partial f / \partial x_i = 0 \) for all \( i \)) is the first order necessary condition for optimization

• Second Order Necessary Condition:
  – \( \nabla^2 f(x) \) is positive semidefinite (PSD)
    • \( [d^T \nabla^2 f(x) d \geq 0 \) for all \( d \)]

• Second Order Sufficient Condition (Given FONC satisfied)
  – \( \nabla^2 f(x) \) is positive definite (PD)
    • \( [d^T \nabla^2 f(x) d > 0 \) for all \( d \neq 0 ] \)
1-Order KKT Condition for GCO

Recall the Lagrange Function

\[ L(x,y) = f(x) - \sum_i c_i(x)y_i \]

\[ \nabla_x L(x,y) = 0, \text{ that is,} \]
\[ \frac{\partial L(x,y)}{\partial x_j} = 0, \text{ for all } j=1,\ldots,n, \text{ and} \]
\[ c_i(x)y_i = 0, \text{ for all } i=1,\ldots,m \]
\[ c_i(x) (\leq, =, \geq) 0, \ y_i (\leq,"free", \geq) 0, \ i=1,\ldots,m \]
Optimality Test for CCO

Is \( x \) a (local) optimizer?

“Yes” if \( f \) is also (locally) convex

1-order KKT Optimality Test

Passed

Failed

Is \( x \) not a (local) optimizer?

“Not” when the feasible region has an interior

2-order test
Optimality Test for GCO

Is $x$ a (local) optimizer?

“Yes” if it is a (locally) convex problem

Is $x$ not a (local) optimizer?

“Not” when $x$ is a regular point on the hypersurface of active constraints

1-order KKT Optimality Test

Passed

Failed

2-order test
2-Order KKT Condition for GCO

Tangent Plane:

\[ T = \{ z : \nabla c_i(x)z = 0, \text{ for all } i, \text{ such that } c_i(x) = 0 \} \]

Necessary Condition:

\[ z^T \nabla_x^2 L(x,y)z \geq 0, \text{ for all } z \text{ in } T \]

Sufficient Condition:

\[ z^T \nabla_x^2 L(x,y)z > 0, \text{ for all non-zero } z \text{ in } T \]

This can be done by checking positive semi-definiteness (or definiteness) of the projected Hessian of the Lagrange function.
Example: Optimality Conditions

\[ \min \quad x_1^2 + x_2^2 \]
\[
\text{s.t.} \quad 1 - 0.25 \cdot (x_1 - 2)^2 - (x_2 - 2)^2 \geq 0
\]

\[
L(x_1, x_2, \lambda) = x_1^2 + x_2^2 - \lambda(1 - 0.25 \cdot (x_1 - 2)^2 - (x_2 - 2)^2)
\]
\[
\left( \frac{\partial L}{\partial x_1} \right) = \left( \begin{array}{c} \frac{\partial L}{\partial x_1} \\ \frac{\partial L}{\partial x_2} \end{array} \right) = \left( \begin{array}{c} 2x_1 \\ 2x_2 \end{array} \right) - \lambda \cdot \left( \begin{array}{c} 0.5(2 - x_1) \\ 2(2 - x_2) \end{array} \right) = 0
\]

\[
1 - 0.25 \cdot (x_1 - 2)^2 - (x_2 - 2)^2 \geq 0, \quad \lambda \geq 0
\]

\[
\lambda(1 - 0.25 \cdot (x_1 - 2)^2 - (x_2 - 2)^2) = 0
\]
Example: KKT Conditions

The curve (surface) of the objective function is tangential to the constraint curve (surface) at the optimal point.
Example: Computation of a KKT Point

\[
\begin{pmatrix}
2x_1 \\
2x_2
\end{pmatrix} - \lambda \cdot \begin{pmatrix}
0.5(2 - x_1) \\
2(2 - x_2)
\end{pmatrix} = 0
\]

\[
x_1 = \frac{2\lambda}{4 + \lambda}; x_2 = \frac{2\lambda}{1 + \lambda}
\]

• If \( \lambda = 0 \), then \( x_1 = 0 \) and \( x_2 = 0 \), and thus the constraint would not hold with equality. Therefore, \( \lambda \) must be positive.

• Plugging the two values of \( x_1(\lambda) \) and \( x_2(\lambda) \) into the constraint with equality gives us \( \lambda = 1.8 \).

• We can then solve for \( x_1 = 0.61 \) and \( x_2 = 1.28 \).
Applications: Optimality Conditions

• The market equilibrium theory
  • Fisher market, Arrow-Debreu market
  • Duality and optimality lead to equilibrium conditions

• Sensor localization
  • SOCP: KKT conditions explain observations
  • SDP: Duality explains localizability

• Offline and Online LP
  – Learning optimal dual solution helps to make primal decisions online

• Non-convex regularization
  • $L_p$ norm regulation function for unconstrained or constrained minimization
  • KKT conditions establish a desired thresh-holding properties at any KKT solution (first or second order)