Optimization Problems

• A set of decision variables, \( x \), in vector or matrix form with dimension \( n \) or \( n \times n \)

• 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:

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

\[
\begin{align*}
\max & \quad \min_{i} \{ f_{i}(x), i=1,...,n \} \\
\text{s.t.} & \quad \alpha - f_{i}(x) \leq 0, \text{ for } i=1,...,n
\end{align*}
\]
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 \( x \)

\[
\nabla^2 f(x) = \begin{pmatrix}
\frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}
\end{pmatrix}
\]

- **Taylor’s Expansion Theorem**
Convex and Concave Functions

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

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

Strictly convex if \( x_1 \neq x_2 \), \( f(0.5x_1 + 0.5x_2) < 0.5f(x_1) + 0.5f(x_2) \)
More on Convex Functions

$f(x)$ is a (strictly) convex function if and only if its Hessian matrix is (positive definite PD) positive semi-definite (PSD) in the domain of the function.

A symmetric matrix $Q$ is PSD (or PD) if and only if $x^TQx \geq (or >) 0$ for all $x \neq 0$.

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

$f(x)$ is a (strictly) concave function if $-f(x)$ is a (strictly) convex function
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

• A set is compact if it is bounded and closed.
Convexity of Function and Level 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 is 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

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

**Conic Linear Optimization (CLO)**

\[
\begin{align*}
\text{min} & \quad c^T x \\
\text{s.t.} & \quad A x - b = 0, \\
& \quad x \in K
\end{align*}
\]

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

This is convex optimization

**Generally Constrained Optimization (GCO)**

\[
\begin{align*}
\text{min} & \quad f(x) \\
\text{s.t.} & \quad h_i(x) = 0, \; i=1,\ldots,m \\
& \quad c_i(x) \geq 0, \; i=1,\ldots,p
\end{align*}
\]

- 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”:
  \[
  \begin{align*}
  \min & \quad c^T x, \quad \text{s.t.} \quad ||x||^2 = 1 \\
  \quad & \quad \uparrow \\
  \min & \quad c^T x, \quad \text{s.t.} \quad ||x||^2 \leq 1
  \end{align*}
  \]
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

CME307/MS&E311 Optimization Lecture Notes #09
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**

- \( Ax - b = 0, \)
- \( x \in K \)
- \( A: \text{an } m \times n \text{ matrix} \)
- \( b: \text{m-dimensional vector} \)
- \( K: \text{a closed convex cone} \)

**Objective**

\[ p^* = \min 0^T x \]

\[ \text{s.t. } Ax - b = 0, \]

\[ x \in K \]

**System Y**

- \( b^T y = 1 (> 0) \)
- \( A^T y + s = 0, \)
- \( s \in K^* \)
- \( K^* \text{ is the dual cone} \)

**Objective**

\[ d^* = \max b^T y \]

\[ \text{s.t. } A^T y + s = 0, \]

\[ s \in K^* \]
Alternative Systems and CLO Pairs II

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

\[
\begin{align*}
\mathbf{c}^T \mathbf{x} &= -1 (< 0) \\
\mathbf{A} \mathbf{x} &= 0, \\
\mathbf{x} &\in K
\end{align*}
\]

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

\[
\begin{align*}
\mathbf{A}^T \mathbf{y} + \mathbf{s} - \mathbf{c} &= 0, \\
\mathbf{s} &\in K^*
\end{align*}
\]

\[
\begin{align*}
p^* &= \min_{\mathbf{x}} c^T \mathbf{x} \\
&\text{s.t. } \mathbf{A} \mathbf{x} = 0, \\
&\quad \mathbf{x} \in K^*
\end{align*}
\]

\[
\begin{align*}
d^* &= \max_{\mathbf{y}} 0^T \mathbf{y} \\
&\text{s.t. } \mathbf{A}^T \mathbf{y} + \mathbf{s} - \mathbf{c} = 0, \\
&\quad \mathbf{s} \in K
\end{align*}
\]
Feasibility Test Machine

- Is system X feasible?
  - Yes
  - If system Y is feasible, then...
    - "Not" under any circumstances
  - No
    - If system Y is feasible, then...
      - "Yes" under certain conditions of cone $K$ and data matrix $A$:
        a) $K$ is a polyhedron cone, or
        b) $Ax$ or $A^T y$ 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>

- **Max model**
  - \( x_j \geq_k 0 \)
  - \( x_j \leq_k 0 \)
  - \( x_j \) free
  - \( i \)th constraint \( \leq_k \)
  - \( i \)th constraint \( \geq_k \)
  - \( i \)th constraint =

- **Min model**
  - \( j \)th constraint \( \geq_{k^*} \)
  - \( j \)th constraint \( \leq_{k^*} \)
  - \( j \)th constraint =
  - \( y_i \geq_{k^*} 0 \)
  - \( y_i \leq_{k^*} 0 \)
  - \( y_i \) free

---

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 \( x \) a (local) optimizer?

“Yes” only under certain circumstances

Higher Order Test

Passed

KKT Optimality Condition Test

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

“Not” under certain constraint qualifications:

a) Feasible region has an interior, or
b) \( x \) is a regular point on the hypersurface of active constraints

Failed
0-Order Condition: Duality Theorems for CLO

**Primal Problem**
- \( p^* = \min_c \ c^T x \)
- \[ A x - b = 0, \]
- \( x \in K \)

**Dual Problem**
- \( d^* = \max_{b^T y} \)
- \[ A^T y + s - c = 0, \]
- \( s \in K^* \)

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

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

- “Yes” under certain conditions of cone \( K \) and data matrix \( A, b, c \):
  - **a)** \( K \) is a polyhedron cone, or
  - **b)** either one has an interior feasible solution
The Lagrange Function of GCO

\[
\begin{align*}
\min & \quad f(\mathbf{x}) \\
\text{s.t.} & \quad c_i(\mathbf{x}) (\leq, =, \geq) 0, \ i=1,...,m \\
\end{align*}
\]

Restriction on multipliers \( y_i, \)

\[
\begin{align*}
y_i (\leq,"\text{free"},\geq) 0, \ i=1,...,m \\
\end{align*}
\]

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 \( c_i(\mathbf{x}) \geq 0 \)
\( y_i = 0 \) : no penalty if \( c_i(\mathbf{x}) \) is strictly satisfied (complementarity)
The Lagrangian Duality for GCO

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

Let \( \phi(y) = \inf_x L(x,y) \)

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

0-Order Condition: \( p^* = d^* \)

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

**Strong Duality Theorem**
They must equal? **Not necessarily!**

Sufficient!
Zero-Order Optimality Test for CLO and GCO

- **Is \(x\) an optimizer?**
  - “Yes” under any circumstances
  - **Higher order test**

- **Is \(x\) not a (local) optimizer?**
  - a) “Not” for sure if \(K\) is a polyhedral cone in CLO; or
  - b) “Not” for sure 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

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

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

The curve (surface) of the objective function is tangential to the constraint curve (surface) at the optimal point.
Optimality Test for CCO

Is $x$ a (local) optimizer?

- "Yes" if $f$ is also (locally) convex

2-order test

Passed

1-order KKT Optimality Test

Failed

Is $x$ not a (local) optimizer?

"Not" for sure when the feasible region has an interior or it is a polyhedral.
Optimality Test for GCO

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

- "Yes" if it is a (locally) convex problem

1-order KKT Optimality Test

Passed

Failed

2-order test

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

- "Not" when \( x \) is a regular point on the hypersurface of active constraints
2-Order KKT Condition for GCO

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

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

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

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

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

• Inverse learning such as sensor localization
  • SOCP: KKT conditions explain observations
  • SDP: Duality explains localizability

• Distributionally robust optimization/learning
  – A model to deal with inaccurate sample-distributions in stochastic optimization and prediction

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