02/13/18  Constrained optimization.  Convexity, Lagrangean, duality.

**Def**  A set $S$ is convex if
$x, y \in S \implies \theta x + (1-\theta) y \in S$ for $\theta \in [0, 1]$.

**Def**  $f: \mathbb{R}^n \to \mathbb{R}$ is convex if
$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$
for $x, y \in \mathbb{R}^n$, $\theta \in [0, 1]$.

**Prop.**  A differentiable $f: \mathbb{R}^n \to \mathbb{R}$ is convex iff
$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$ for $x, y$.

$
\Rightarrow$ tangent line is a global underestimator.

**Prop.**  $x$ is a minimizer of a differentiable, convex $f$ iff $\nabla f(x) = 0$.  (= global minimizer)
Constrained optimization

\[ \text{min } \mathbf{w}^{\top} \mathbf{A} \mathbf{x} \quad \text{s.t. } \mathbf{A} \mathbf{x} = \mathbf{b} \]

\[ \Rightarrow \text{min } f_{0}(\mathbf{x}) \quad \text{s.t. } h_{i}(\mathbf{x}) = 0 \quad \text{ex. } h_{i}(\mathbf{x}) = \frac{1}{2} \mathbf{x}^{\top} \mathbf{x} - \mathbf{b}^{\top} \mathbf{x} \]

Assume \( f_{0}, h_{i} \in C^{1} \)

\[ \mathbf{x}_{\text{opt}} \quad h_{i}(\mathbf{x}) = 0 \quad (\mathbf{x}_{\text{feasible}}) \]

\[ f_{0}(\mathbf{x}) = \text{constant.} \]

At \( \mathbf{x} = \mathbf{x}_{\text{opt}} \), normal to \( f_{0}(\mathbf{x}) = f_{0}(\mathbf{x}_{\text{opt}}) \)

is also normal to all \( h_{i}(\mathbf{x}) = 0 \)

\[ \nabla f_{0} + \sum_{i} \nabla h_{i} = 0 \]

\( h_{i} = 0 \)

(1)

(2)

\[ \mathbf{L}(\mathbf{x}, \mathbf{\lambda}) = f_{0}(\mathbf{x}) + \sum_{i} \lambda_{i} h_{i}(\mathbf{x}) \]

\[ \nabla_{\mathbf{x}} \mathbf{L} = 0 \quad \Leftrightarrow \quad (1) \]

\[ \frac{\partial \mathbf{L}}{\partial \mathbf{\lambda}_{i}} = 0 \quad \Leftrightarrow \quad (2). \]

(1) may determine \( \mathbf{\lambda} \) as well.

It is a linear system for them (scratch for singular matrix).
\( \lambda \) is the Lagrange multiplier
\( \nu \) are Lagrange multipliers
(1), (2) are the Karush-Kuhn-Tucker (KKT) conditions, a.k.a. first-order optimality conditions

**Example** ULS: \( d(x, y) = \frac{1}{2} \| x \|_2^2 - \nu^T (Ax - b) \)

\[
\begin{align*}
\nabla_x d &= x^T - \nu^T A = 0 \\
\nabla_{\nu} d &= Ax - b = 0
\end{align*}
\]

then \( AA^T \nu = b \), \( \nu = (AA^T)^{-1} b \)
\( \lambda = \frac{1}{2} \frac{1}{(AA^T)^{-1} b} \)

The KKT conditions are necessary
(\( x^* \) minimizer \( \Rightarrow (1), (2) \)
argue by contradiction)

but they are not in general sufficient
((1), (2) \( \Rightarrow x^* \) minimizer).

**Ex. 1.** (level sets of \( f \) non-convex)

- minimum \( f = c_2 \) local
- maximum \( f = c_3 \) local

\( c_1 < c_2 < c_3 \)
Q. When is KKT sufficient?

Would like to recognize (1) of KKT as a condition of minimization of f over x.

Def. \( g(v) = \min_{x} d(x, v) \)

Note \( g(v) \leq d(x, v) \) & pick feasible: \( h_i(x) = 0 \)

true for all feasible \( x \)

\( \Rightarrow g(v) \leq f_0(x_{opt}) \)

true for all \( v \)

\[ \max g(v) = \frac{f_0(x_{opt})}{g(v_{opt}) \leq f_0(x_{opt})} \]

so \( g(v_{opt}) \) provides a floor on the optimal value of the objective \( f_0(x) \).

\[ \max_{v} g(v) \] is the dual problem
\( v \)  
\( v \) dual variables, \( v_{opt} \) dual optimal
\( g(v) \) is the Lagrangian dual function
\[ \min_{x} f_0(x) = h_i(x) = 0 \] in the primal problem
\( x \) primal variables, \( x_{opt} \) primal optimal
\[ f_0(x^{opt}) - g(y^{opt}) = \text{duality gap} \]
\[ f_0(x^{opt}) \geq g(y^{opt}) = \text{weak duality} \]
\[ f_0(x^{opt}) = g(y^{opt}) = \text{strong duality} \]
(holds / obtained)

**Theorem.** Assume \( f_0(x) \) is differentiable, convex
\( h_i(x) = a_i^T x - b_i \) in affine
Then KKT are sufficient
\((x, y) \) obey (1), (2) \( \Rightarrow \) they are optimal
and strong duality holds.

**Proof.** Let \((x, y) \) obey (1), (2)

(2) \( \Rightarrow \) \( x \) is feasible
Consider \( L(x, y) = f_0(x) + \sum_i y_i h_i(x) \)
It is also a convex & differentiable function
(1) means \( \nabla_x L(x, y) = 0 \) when \( x = \bar{x} \)
\( \Rightarrow \bar{x} \) is a minimizer of \( L(x, y) \).

\[ g(\bar{y}) = \min_x L(x, \bar{y}) \]
\[ = \bar{y}^T \bar{L}(\bar{x}, \bar{y}) \] because \( \bar{x} \) is a minimizer
\[ = f_0(\bar{x}) \] because \( \bar{x} \) is feasible
\( \Rightarrow \) zero dual gap
Lower bound is attained for \( f_0 \)
\( \Rightarrow \bar{x} \) is primal optimal
Upper bound is attained for \( g \)
\( \Rightarrow \bar{y} \) is dual optimal.

**Remark:** We say \( g(\bar{y}) = f_0(\bar{x}) \) is a certificate of optimality of \( \bar{x} \) and \( \bar{y} \).
**Remark:** \( g(\bar{y}) \) is the min of affine functions
\( \Rightarrow \) it is concave