First-Order Optimality Condition

- For function of one variable, one can find extremum by differentiating function and setting derivative to zero.

- Generalization to function of $n$ variables is to find critical point, i.e., solution of nonlinear system

$$\nabla f(x) = 0$$

where $\nabla f(x)$ is gradient vector of $f$, whose $i$th component is $\partial f(x)/\partial x_i$.

- For continuously differentiable $f: S \subseteq \mathbb{R}^n \to \mathbb{R}$, any interior point $x^*$ of $S$ at which $f$ has local minimum must be critical point of $f$.

- But not all critical points are minima: they can also be maxima or saddle points.
\[ f(\mathbf{x}) = f(x_0) + \nabla f(x_0) \cdot (x-x_0) + \frac{1}{2} (x-x_0)^T H_f (x-x_0) + \ldots \]

\[ f(\mathbf{x}) \geq f(x_0) \quad \forall x \quad \Rightarrow \nabla f(x_0) = 0 \quad \text{critical pt} \]

\[ \text{min } f(x) \quad \text{s.t. } g(x) = 0 \quad \text{critical point} \quad \leftrightarrow \quad \text{candidate pt for min} \]
Constrained Optimality

- If problem is constrained, only feasible directions are relevant.
- For equality-constrained problem

$$\min f(x) \text{ subject to } g(x) = 0$$

where $f: \mathbb{R}^n \to \mathbb{R}$ and $g: \mathbb{R}^n \to \mathbb{R}^m$, with $m \leq n$, necessary condition for feasible point $x^*$ to be solution is that negative gradient of $f$ lie in space spanned by constraint normals,

$$-\nabla f(x^*) = J_g^T(x^*)\lambda$$

where $J_g$ is Jacobian matrix of $g$, and $\lambda$ is vector of Lagrange multipliers.

- This condition says we cannot reduce objective function without violating constraints.
1st order Nec Cond

\[-\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) + \cdots = [\nabla g_1(x^*) \nabla g_2(x^*) \nabla g_3(x^*) \cdots] \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \end{bmatrix} \]

0 = g_i(x) = g_1(x) = 0, g_2(x) = 0

\[g: \mathbb{R}^n \rightarrow \mathbb{R}^m\]

\[J_g(x^*) = \left[ \begin{array}{c|c} \nabla g_1 & \vdots \\ \vdots & \nabla g_m \end{array} \right] \]

JACOBIAN

LAGRANGE MULTIPLIERS
Example: Constrained Optimization

- Consider quadratic programming problem
  \[ \min_x f(x) = 0.5x_1^2 + 2.5x_2^2 \quad \text{Quad} \]

subject to
  \[ g(x) = x_1 - x_2 - 1 = 0 \quad \text{Lin} \]

- Lagrangian function is given by
  \[ \mathcal{L}(x, \lambda) = f(x) + \lambda g(x) = 0.5x_1^2 + 2.5x_2^2 + \lambda(x_1 - x_2 - 1) \]

- Since
  \[ -\nabla f = \nabla g \lambda_1 = \nabla f + \lambda \nabla g = 0 \quad \nabla g = [1 \quad -1] \]

we have
  \[ \nabla_x \mathcal{L}(x, \lambda) = \nabla f(x) + J_g^T(x) \lambda = \begin{bmatrix} x_1 \\ 5x_2 \end{bmatrix} + \lambda \begin{bmatrix} -1 \\ -1 \end{bmatrix} = 0 \]
Example, continued

\[ \min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad g(\mathbf{x}) = 0 \]

contours of \(0.5x_1^2 + 2.5x_2^2\)

line \(x_1 - x_2 = 1\)

feasible space
Example, continued

- So system to be solved for critical point of Lagrangian is
  \[ \nabla f + \lambda \nabla g = 0 \]
  \[ g(x) = 0 \]
  \[ x_1 + \lambda = 0 \]
  \[ 5x_2 - \lambda = 0 \]
  \[ x_1 - x_2 = 1 \]

which in this case is linear system

\[
\begin{bmatrix}
1 & 0 & 1 \\
0 & 5 & -1 \\
1 & -1 & 0
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\lambda
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix}
\]

- Solving this system, we obtain solution
  \[ x_1 = 0.833, \quad x_2 = -0.167, \quad \lambda = -0.833 \]
Examples: Optimization Problems

- Minimize weight of structure subject to constraint on its strength, or maximize its strength subject to constraint on its weight
- Minimize cost of diet subject to nutritional constraints
- Minimize surface area of cylinder subject to constraint on its volume:

\[ \nabla f + \lambda \nabla g = 0 \]

\[ \min_{x_1, x_2} f(x_1, x_2) = 2\pi x_1 (x_1 + x_2) = \text{area}(r, h) \]

subject to \[ g(x_1, x_2) = \pi x_1^2 x_2 - V = 0 \]

where \( x_1 \) and \( x_2 \) are radius and height of cylinder, and \( V \) is required volume
\[
\begin{align*}
\text{min } f(r, h) &= \text{area } = 2\pi r^2 + 2\pi rh \\
\text{s.t. } g(r, h) &= \text{vol } V - V = \pi r^2 h - V = 0 \\
g(r, h) &= 0 \Rightarrow h = \frac{V}{\pi r^2} \\
\frac{\partial f}{\partial r} &= 2\pi r + 2\pi h \frac{\partial h}{\partial r} = 2\pi r^2 + 2V/r \\
\frac{\partial f}{\partial r} &= 4\pi r - 2\sqrt{r^2} = \frac{1}{r^2}(4\pi r^3 - 2V) = 0 \\
2\pi r^3 &= V \\
r^3 &= \frac{V}{2\pi} \\
\text{if } V &= 1000 \\
r^3 &= \frac{1000}{2\pi} \\
r &= 5.41926
\end{align*}
\]
\[ f(r, h) = 2\pi r^2 + 2\pi rh \]
\[ g(r, h) = \pi r^2 h - V = 0 \]

\[ \nabla f = \begin{bmatrix} 4\pi r + 2\pi h \\ 2\pi r \end{bmatrix} ; \quad \nabla g = \begin{bmatrix} 2\pi rh \\ \pi r^2 \end{bmatrix} \]

\[ x = f + \lambda g \]
\[ \nabla x = \nabla f + \lambda \nabla g \]
\[ \nabla x = g \]
\[ r, h = 0 \]

**MOVE ALONG CONSTRAINT**

WANT move also along level set for \( f = \text{obj} \)

\[ -\nabla f = \lambda \nabla g \]
\[
\begin{align*}
\varphi(r,h) &= 2\pi r^2 + 2\pi rh \\
\psi(r,h) &= \pi r^2 h - V = 0
\end{align*}
\]

\[
\nabla f + \lambda \nabla g = \left( \frac{4\pi r + 2\pi h}{2\pi r} \right) + \lambda \left( \frac{2\pi rh}{\pi r^2} \right) = 0
\]

\[
2\pi r + \lambda \pi r^2 = 0 = \pi r (2 + \lambda r) \quad r \neq 0
\]

\[
\lambda = -\frac{2}{r}
\]

\[
4\pi r + 2\pi h + -\frac{2}{r} 2\pi rh = 4\pi r + 2\pi h - 4\pi h = 4\pi r - 2\pi h = 0 \quad \Rightarrow \quad h = 2r
\]

\[
\varphi(r,2r) = 0 = \pi r^2, 2r - 1000 = 0
\]

\[
2\pi r^3 = 1000 \quad r^3 = 1000/2\pi \quad \Rightarrow \quad r = 5.41...
\]
2-D cylinder demo: level sets of Area, with constraint curve vol=1000

\[ V_f + \lambda V_g = 0 \quad h = 2r \]
Constrained Optimality

- If problem is constrained, only *feasible* directions are relevant.
- For equality-constrained problem

\[
\min f(x) \quad \text{subject to} \quad g(x) = 0
\]

where \( f : \mathbb{R}^n \to \mathbb{R} \) and \( g : \mathbb{R}^n \to \mathbb{R}^m \), with \( m \leq n \), necessary condition for feasible point \( x^* \) to be solution is that negative gradient of \( f \) lie in space spanned by constraint normals,

\[
-\nabla f(x^*) = J_g^T(x^*)\lambda
\]

where \( J_g \) is Jacobian matrix of \( g \), and \( \lambda \) is vector of *Lagrange multipliers*.

- This condition says we cannot reduce objective function without violating constraints.
Lagrangian function $\mathcal{L} : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$, is defined by

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T g(x)$$

Its gradient is given by

$$\nabla \mathcal{L}(x, \lambda) = \begin{bmatrix} \nabla f(x) + J_g^T(x) \lambda \\ g(x) \end{bmatrix}$$

Its Hessian is given by

$$H_{\mathcal{L}}(x, \lambda) = \begin{bmatrix} B(x, \lambda) & J_g^T(x) \\ J_g(x) & O \end{bmatrix}$$

where

$$B(x, \lambda) = H_f(x) + \sum_{i=1}^{m} \lambda_i H_{g_i}(x)$$
Constrained Optimality, continued

- Together, necessary condition and feasibility imply critical point of Lagrangian function,

\[ \nabla L(x, \lambda) = \begin{bmatrix} \nabla f(x) + J_g^T(x)\lambda \\ g(x) \end{bmatrix} = 0 \]

- Hessian of Lagrangian is symmetric, but not positive definite, so critical point of \( L \) is saddle point rather than minimum or maximum

- Critical point \( (x^*, \lambda^*) \) of \( L \) is constrained minimum of \( f \) if \( B(x^*, \lambda^*) \) is positive definite on null space of \( J_g(x^*) \)

- If columns of \( Z \) form basis for null space, then test projected Hessian \( Z^T B Z \) for positive definiteness
If inequalities are present, then KKT optimality conditions also require nonnegativity of Lagrange multipliers corresponding to inequalities, and complementarity condition.