Math 273a: Optimization
Nonlinear optimization with equality constraints

Instructor: Wotao Yin
Department of Mathematics, UCLA
Fall 2015

material taken from the textbook Chong-Zak, 4th Ed.
About this lecture

- recognize a solution to the problem

\[
\begin{align*}
\text{minimize } & f(x) \\
\text{subject to } & h(x) = 0.
\end{align*}
\]

where \( f : \mathbb{R}^n \rightarrow \mathbb{R}, \ h : \mathbb{R}^n \rightarrow \mathbb{R}^m \)

\[
h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix}
\]

- assume continuous differentiability: \( f, h \in C^1 \).

- goals of this lecture
  1. 1st and 2nd order optimality conditions
  2. solutions to certain simple/important NLP with equality constraints
- **Jacobian of** $h$:

$$Dh(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \vdots \\ \nabla h_m(x)^T \end{bmatrix}$$

- **definition**: a point $x^* \in \mathbb{R}^n$ satisfying $h(x^*) = 0$ is a *regular point* of the constraints if $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ linearly independent, or equivalently, $\text{rank}(Dh(x^*)) = 0$. 
Example 20.2

$S = \{ [x_1, x_2, x_3]^T : x_2 - x_3^2 = 0 \}$

$h_1(x) = x_2 - x_3^2 = 0$
Example 20.3

\[ h_1(x) = x_1 = 0 \]

\[ h_2(x) = x_2 - x_3^2 = 0 \]
Tangent space

- **tangent space** at \( x^* \) on \( S = \{ x \in \mathbb{R}^n : h(x) = 0 \} \) is the set

\[
T(x^*) := \{ y \in \mathbb{R}^n : Dh(x^*)y = 0 \} = \mathcal{N}(Dh(x^*)).
\]

- if \( x^* \) is regular, then \( \dim(T(x^*)) = n - m \).

- **tangent plane** at \( x^* \) is the set

\[
TP(x^*) := T(x^*) + x^*.
\]
Tangent space
Normal space

- **normal space** $N(x^*)$ at $x^*$ on $S = \{x \in \mathbb{R}^n : h(x) = 0\}$ is the set
  \[ N(x^*) := \mathcal{R}(Dh(x^*)^T) = \text{span}\{\nabla h_1(x^*), \ldots, \nabla h_m(x^*)\} \]

- **normal plane** is $NP(x^*) := N(x^*) + x^*$

- **lemma:** $T(x^*) = N(x^*)^\perp$, $N(x^*) = T(x^*)^\perp$, $\mathbb{R}^n = N(x^*) \oplus T(x^*)$. 
Lagrange’s theorem

Theorem (Lagrange’s theorem, \( n = 2, m = 1 \))

Let \( x^* \) be a local minimizer of \( \{f(x) : h(x) = 0\} \). Then \( \nabla f(x^*) \) and \( \nabla h(x^*) \) are parallel, that is, if \( \nabla h(x^*) \neq 0 \), then there exists \( \lambda^* \in \mathbb{R} \) (known as the Lagrange multiplier) such that

\[
\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0.
\]
Theorem (Lagrange’s theorem, 1st-order optimality condition)

Let \( x^* \) be a local minimizer of \( \{ f(x) : h(x) = 0 \} \), where \( h : \mathbb{R}^n \to \mathbb{R}^m, m < n \).
Assume that \( x^* \) is regular. Then, there exists \( \lambda^* \in \mathbb{R}^m \) (the Lagrange multipliers) such that

\[
\nabla f(x^*) + D h(x^*)^T \lambda^* = 0.
\]
Example 20.7

minimize $f(x) = x_1^2 + x_2^2$

subject to $h(x) = x_1^2 + 2x_2^2 - 1 = 0$. 
An example where the Lagrange condition is not satisfied
Different cases where the Lagrange condition satisfied

(a) maximizer; (b), (c) minimizer; (d) neither. Courtesy of Seeley'70
Second-order conditions

- **Lagrangian:**
  \[ \mathcal{L}(x, \lambda) := f(x) + \lambda^T h(x) = f(x) + \lambda_1 h_1(x) + \cdots + \lambda_m h_m(x) \]

- **Hessians:** \( F \) of \( f \) and \( H_k \) of \( h_k \)

- **Hessian** of \( \mathcal{L} \):
  \[ \mathcal{L}(x, \lambda) = F(x) + \lambda_1 H_1(x) + \cdots + \lambda_m H_m(x) \]

---

**Theorem (2nd-order necessary condition)**

Let \( x^* \) be a local minimizer of \( \{ f(x) : h(x) = 0 \} \), where \( h : \mathbb{R}^n \rightarrow \mathbb{R}^m, m < n \). Suppose that \( f, h \in C^2 \). Assume that \( x^* \) is regular. Then, there exists \( \lambda^* \) such that

1. \( \nabla f(x^*) + D h(x^*)^T \lambda^* = 0 \);
2. for all \( y \in T(x^*) \), we have \( y^T \mathcal{L}(x^*, \lambda^*) y \geq 0 \).

The role of \( \mathcal{L}(x^*, \lambda^*) \) is similar to the objective Hessian \( F(x^*) \) in the unconstrained case.
Theorem (2nd-order sufficient condition)

Suppose that $f, h \in C^2$ and there exists feasible $x^* \in \mathbb{R}^n$, $\lambda^* \in \mathbb{R}^m$ such that

1. $\nabla f(x^*) + Dh(x^*)^T \lambda^* = 0$;

2. for all $y \in T(x^*)$, $y \neq 0$, we have $y^T L(x^*, \lambda^*) y > 0$.

Then, $x^*$ is a strict local minimizer of $\{f(x) : h(x) = 0\}$. 
Exercise (Example 20.9)

- solve

\[
\text{maximize } \frac{x^T Q x}{x^T P x}
\]

where

\[
P = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}
\]

- reformulate as

\[
\text{maximize } f(x) = x^T Q x
\]

subject to \( h(x) = x^T P x - 1 = 0 \)

- the Lagrange condition gives \( x^* = \pm \left[ \frac{1}{\sqrt{2}}, 0 \right]^T \)

- the second-order sufficient condition is satisfied by \( x^* \)

- any \( tx^* \), \( t \neq 0 \), is the solution to the original problem
Quadratic minimization with linear constraints

- suppose $Q$ invertible and $\text{rank}(A) = m$. consider

  $\min \frac{1}{2} x^T Q x$

  subject to $Ax = b$

- Lagrangian: $L(x, \lambda) = \frac{1}{2} x^T Q x + \lambda^T (b - Ax)$

- first-order condition: $\nabla_x L(x^*, \lambda^*) = Q x^* - A^T \lambda^* = 0$

- hence, $x^* = Q^{-1} A^T \lambda^*$

- since $Ax^* = A Q^{-1} A^T \lambda^* = b$, we get $\lambda^* = (A Q^{-1} A^T)^{-1} b$

- $x^* = Q^{-1} A^T \lambda^* = Q^{-1} A^T (A Q^{-1} A^T)^{-1} b$ is a candidate sol

- if Lagrangian Hessian $L(x^*, \lambda^*) = Q \succ 0$, then $x^*$ is a strict minimizer
Minimal norm solution to linear equations

- Consider Q invertible and
  
  \[ \text{minimize } \frac{1}{2} \| x \|^2 \]
  
  subject to \( Ax = b \)

- Assume \( \text{rank}(A) = m \). Then, the solution is
  
  \[ x^* = A^T (AA^T)^{-1} b. \]

- The solution is the projection of 0 to the affine set \( \{ x \in \mathbb{R}^n : Ax = b \} \)
Summary

- the equality constraints in an NLP define the “constraint surface”
- at a “stationary point” of an NLP, the objective gradient is perpendicular to the tangent of the constraint surface
- the 1st-order conditions equalize $\nabla f$ with a linear combination of all $\nabla h_i$
- like before, 1st-order conditions give candidate solutions, 2nd-order conditions help determine extremal status