MATH 4501558
FALL 2015

LECTURE 1

INTRODUCTION

Unconstrained Optimization

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

$f : \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable

Local minimizer

A point $x_* \in \mathbb{R}^n$ such that

$$f(x_*) \leqslant f(x)$$

for all $x \in B(x_*, \delta) := \{ x \in \mathbb{R}^n \mid \|x - x_*\|_2 \leqslant \delta \}$ for some $\delta \in \mathbb{R}^+$.

(i) Optimality conditions
conditions in terms of derivatives of $f(x)$ that distinguish local minimizers.

(ii) Numerical algorithms
generates a convergent sequence $\{x^{(k)}\}$ such that $\lim_{k \to \infty} x^{(k)}$ is a local minimizer.

Generalizations

Constraints – (smooth) nonlinear programs
(PART I)

$f(x)$ is not differentiable everywhere – nonsmooth optimization
(PART II)

$f : \mathbb{R} \to \mathbb{R}$

optimality conditions

(i) $f'(x_*) = 0$

(ii) $f''(x_*) > 0$

$\Longrightarrow$

x_* is a local minimizer

SUFFICIENT CONDITIONS

x_* is a local minimizer

$\Longrightarrow$

(i) $f'(x_*) = 0$

(ii) $f''(x_*) \geq 0$

NECESSARY CONDITIONS

Background

$\nabla f(x) : \mathbb{R}^n \to \mathbb{R}^n$

$$\nabla f(x) := \begin{bmatrix} \partial f(x)/\partial x_1 \\ \vdots \\ \partial f(x)/\partial x_n \end{bmatrix}$$

Gradient Vector

$\nabla^2 f(x) : \mathbb{R}^n \to \mathbb{R}^{n \times n}$

$$\nabla^2 f(x) := \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & & \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

Hessian matrix

Hessian matrix is symmetric. $\left(\text{i.e. } \frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i} \right)$

Let $A \in \mathbb{R}^{n \times n}$ be symmetric.

(i) Eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are real.

(ii) Associated eigenvectors $v_1, v_2, \ldots, v_n$ can be chosen such that $v_j^T v_i = 0 \quad j \neq i$.

THM (Taylor)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice-continuously differentiable.

(i) For each $x, \underline{x} \in \mathbb{R}^n$

$$f(x) = f(\underline{x}) + \nabla f(\underline{x})^T p + \frac{1}{2} p^T \nabla^2 f(\underline{x} + t p) \, p$$

for some $t \in (0, 1)$ where $p = x - \underline{x}$.

(ii) For each $x, \underline{x} \in \mathbb{R}^n$

$$\nabla f(x) = \nabla f(\underline{x}) + \int_0^1 \nabla^2 f(\underline{x} + t p) \, p \, dt$$

where $p = x - \underline{x}$.

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is

(i) positive semi-definite (PSD) if

$$p^T A p \geq 0 \qquad \forall p \in \mathbb{R}^n$$

(ii) positive definite (PD) if

$$p^T A p > 0 \quad \forall p \in \mathbb{R}^n \setminus \{0\}.$$

THM

$$(*)\quad A \text{ is PSD} \iff \lambda_1, \lambda_2, \dots, \lambda_n \geq 0$$

$$A \text{ is PD} \iff \lambda_1, \lambda_2, \dots, \lambda_n > 0$$

PROOF OF (*)

Consider the decomposition

$$A = V \Lambda V^T$$

$$= \underbrace{\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}}_{\text{eigenvectors of } A} \overbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_n \end{bmatrix}}^{\text{eigenvalues of } A} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}$$

where $V \in \mathbb{R}^{n \times n}$ is orthogonal, $\Lambda \in \mathbb{R}^{n \times n}$ is diagonal.

Suppose $\lambda_1, \dots, \lambda_n \geq 0$. For each $p \in \mathbb{R}^n$

$$p^T A p = \underbrace{p^T V}_{y^T} \Lambda \underbrace{V^T p}_{y}$$

$$= \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 \geq 0.$$

Thus A is PSD.

Suppose $\lambda_j < 0 \ \exists j$ and ~~let~~ recall v_j ~~be~~ is the corresponding eigenvector.

$$v_j^T A v_j = v_j^T V \Lambda V^T v_j$$

$$= e_j^T \Lambda e_j = \lambda_j < 0.$$

Thus A is not PSD.

ptimality Conditions

THM (Sufficient Conditions)

Let $f:\mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Then x_* is a local minimizer if

$$(i)\ \nabla f(x_*) = 0 \quad \text{and} \quad (ii)\ \nabla^2 f(x_*) \text{ is PD.}$$

THM (Necessary Conditions)

Let $f:\mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. If x_* is a local minimizer, then

$$(i)\ \nabla f(x_*) = 0 \quad \text{and} \quad (ii)\ \nabla^2 f(x_*) \text{ is PSD.}$$

Ex

$$f(x) = 6x_1^2 + 4x_1x_2 + 6x_2^2 + 3x_1 + x_2$$

$$= \frac{1}{2}x^T \begin{bmatrix} 12 & 4 \\ 4 & 12 \end{bmatrix} x + \begin{bmatrix} 3 & 1 \end{bmatrix} x$$

$$\nabla f(x) = \begin{bmatrix} \partial f(x)/\partial x_1 \\ \partial f(x)/\partial x_2 \end{bmatrix} = \begin{bmatrix} 12 & 4 \\ 4 & 12 \end{bmatrix} x + \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 \end{bmatrix} = \begin{bmatrix} 12 & 4 \\ 4 & 12 \end{bmatrix}$$

i) $\nabla f(x_*) = 0 \iff x_* = \begin{bmatrix} -1/3 \\ 0 \end{bmatrix}$

i) $\nabla^2 f(x_*) = \begin{bmatrix} 12 & 4 \\ 4 & 12 \end{bmatrix}$ is PD, indeed $\lambda_1 = 8$, $\lambda_2 = 16$.

By sufficient conditions x_* is a local min.
By necessary conditions f has no other local min.

PROOF OF SUFFICIENT CONDITIONS

There exists a ball $B(x_*, \delta)$ such that

$$\nabla^2 f(x) \text{ is PD} \quad \forall x \in B(x_*, \delta).$$

By Taylor's thm for each $x \in B(x_*, \delta) \setminus \{x_*\}$

$$f(x) = f(x_*) + \nabla f(x_*)^T p + \frac{1}{2} p^T \nabla^2 f(x_* + tp) p$$

$$= f(x_*) + \frac{1}{2} p^T \nabla^2 f(x_* + tp) p$$

for some $t \in (0,1)$ where $p = x - x_*$. Since $x_* + tp \in B(x_*, \delta)$, we have $\nabla^2 f(x_* + tp)$ is PD implying

$$f(x) > f(x_*)$$

□

Convexity

$f : \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(\alpha x + (1-\alpha) y) \leqslant \alpha f(x) + (1-\alpha) f(y)$$

$$\forall \alpha \in [0,1] \quad \text{and} \quad \forall x, y \in \mathbb{R}^n$$

THM

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a convex diffentrable function. Then $x_* \in \mathbb{R}^n$ such that $\nabla f(x_*) = 0$ is a ~~local~~ global minimizer of f.

F

Suppose x is not a global minimizer of f. Thus there exists $y \in \mathbb{R}^n$ such that

$$f(y) < f(x).$$

But then

$$\nabla f(x)^T (y-x) = \frac{d}{d\alpha}\left(f(x+\alpha(y-x))\right)\Big|_{\alpha=0}$$

$$= \lim_{h\to 0} \frac{f(x+h(y-x)) - f(x)}{h}$$

$$\leqslant \lim_{h\to 0} \frac{h f(y) + (1-h) f(x) - f(x)}{h}$$

$$= f(y) - f(x) < 0.$$

Consequently, $\nabla f(x) \neq 0$. □

Numerical Algorithms

① Newton's Method (Pure)

Define a sequence $\{x^{(k)}\}$ such that $x^{(k+1)}$ is a local minimizer of

$$q^{(k)}(x) := \frac{1}{2}(x-x^{(k)})^T \nabla^2 f(x^{(k)})(x-x^{(k)})$$
$$+ \nabla f(x^{(k)})^T (x-x^{(k)}) + f(x^{(k)})$$

us

$$\nabla q^{(k)}(x^{(k+1)}) = 0$$

$$\nabla^2 f(x^{(k)})(x^{(k+1)} - x^{(k)}) + \nabla f(x^{(k)}) = 0$$

$$x^{(k+1)} = x^{(k)} - \left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^{(k)})$$

Equivalently

$$x^{(k+1)} = x^{(k)} + p^{(k)}$$

where $p^{(k)}$ satisfies

$$\nabla^2 f(x^{(k)})\, p^{(k)} = -\nabla f(x^{(k)})$$

② BFGS - A Quasi Newton's Method

Based on

$$q^{(k)}(x) := \frac{1}{2}(x-x_k)^T B_k (x-x_k)$$
$$+ \nabla f(x_k)^T (x-x_k) + f(x_k)$$

where $B_k \approx \nabla^2 f(x_k)$.

$\nabla q^{(k)}(x^{(k+1)}) = 0$ yields

$$x^{(k+1)} = x^{(k)} + p^{(k)}$$
$$B_k p^{(k)} = -\nabla f(x^{(k)})$$

Update B_k using $x^{(k)}$, $x^{(k+1)}$, $\nabla f(x^{(k)})$, $\nabla f(x^{(k+1)})$.

We would like B_{k+1} to satisfy

$$B_{k+1}^T = B_{k+1}$$

secant equation $\Big($ $B_{k+1} s_k = y_k$ where $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$

$s_k = x_{k+1} - x_k$

Secant Equation

By Taylor's thm

$$\nabla f(x_{k+1}) = \nabla f(x_k) + \int_0^1 \nabla^2 f(x_k + t s_k)\, s_k \, dt$$

Assume $\|s_k\| \approx 0$ so that

$$B_{k+1} \approx \nabla^2 f(x_{k+1}) \approx \nabla^2 f(x_k + t s_k) \quad \forall t \in [0,1]$$

This leads to $B_{k+1} s_k = y_k$.

Rank 2 update rule

$$B_{k+1} = B_k + \alpha \underbrace{u u^T}_{\text{rank } 1} + \beta v v^T$$

$\alpha, \beta \in \mathbb{R}$ and $u, v \in \mathbb{R}^n$

$$\underbrace{B_{k+1} s_k}_{y_k} = B_k s_k + \alpha u (u^T s_k) + \beta v (v^T s_k)$$

① Choose α and u as follows:

$$u = y_k \qquad \alpha = 1/(u^T s_k)$$

② Choose β and v as follows:

$$v = B_k s_k \qquad \beta = -1/(v^T s_k)$$

Consequently,

$$B_{k+1} = B_k + \frac{y_k y_k^T}{(u^T s_k)} - \frac{B_k s_k s_k^T B_k}{v^T s_k}$$