WI rew

Review
Critical points:

Assume $f: \mathbb{R}^n \to \mathbb{R}$

Gradient $\nabla f(\vec{x}_0) = \left(\frac{\partial f}{\partial x_1}, \dots \frac{\partial f}{\partial x_n}\right)$ $\quad \vec{x} = (x_1, x_2, \dots x_n)$

necessary condition

Thm Let $\vec{x}^*$ be a local minimiser of f and let $f \in C^1(\mathcal{U})$ (neighborhood of $\mathcal{U}$) $\Rightarrow \nabla f(\vec{x}^*) = \vec{0}$

Not sufficient $\Rightarrow$ max and saddle points also have vanishing gradients

Hessian of f: $\nabla^2 f(\vec{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{1 \le i,j \le n}$ $\Leftarrow$ symmetric $n \times n$ matrix

Recall in 1D: x^* is a local min of $f \in C^2([a,b])$ then $f'(x^*) = 0$ and $f''(x^*) \geq 0$ guarantee local minimiser

Higher space dimension $\Rightarrow$ Hessian has to be pos. semi def (necessary cond)

$\vec{A}$ is pos. semi def $\Rightarrow \vec{x}^T \vec{A} \vec{x} \geq 0$

→ sufficient condition

Thm Let $f \in C^2(\mathcal{U})$ for some open set $\mathcal{U}$ and $\vec{x}^* \in \mathcal{U}$. If
$x^* \in \mathcal{U}$ is local min $\Rightarrow \nabla f(\vec{x}^*) = \vec{0}$ and $\nabla^2 f(\vec{x}^*)$ is pos. semidef.
If $\nabla f(\vec{x}^*) = \vec{0}$ and $\nabla^2 f(\vec{x}^*)$ pos. def $\Rightarrow \vec{x}^*$ is a strict local min.

Only local minimisers, no global. If f is convex then the above thm ensures global minimiser

(2) $\Rightarrow 2^{nd}$ partial derivative test

Ex Quadratic function

$$f(\vec{x}) = \frac{1}{2} \vec{x} \vec{A} \vec{x} + \vec{b}^T \vec{x} + c \qquad \begin{matrix} \vec{A} \in \mathbb{R}^{n \times n} \\ \vec{b} \in \mathbb{R}^n \end{matrix} \text{ symmetric}$$

Write inner product:

$$\vec{x}^T \vec{A} \vec{x} = (x_1 \dots x_n) \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (x_1 \dots x_n) \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots a_{1n}x_n \\ \vdots \\ a_{n1}x_1 + \dots + a_{nn}x_n \end{pmatrix}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

$$= \cancel{\sum_{i=1}^{n} a_{ii} x_i^2 + 2 \sum_{1 \le i,j \le n} a_{ij} x_i x_j} \Leftarrow \text{skip}$$

Calculate partial derivative

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^{n} a_{ij} x_j + b_i \qquad \frac{\partial^2 f}{\partial x_i^2} = a_{ij}$$

$$\Rightarrow \nabla f(\vec{x}) = \vec{A}\vec{x} + b \qquad \nabla^2 f(\vec{x}) = \vec{A} \leftarrow \text{pos.}$$

If $\vec{A}$ is pos. semidef $\Rightarrow f$ is convex

Typical example

$$f(\vec{x}) = \|\vec{A}\vec{x} - \vec{b}\|^2 = (\vec{A}\vec{x} - b)^T(\vec{A}\vec{x} - \vec{b}) = \vec{x}^T \underbrace{\vec{A}^T\vec{A}}\vec{x} - 2\vec{b}^T\vec{A}\vec{x} + \vec{b}^T\vec{b}$$

is always symmetric & pos. semidefinite

Directional derivative: measures how much a function changes in a given direction $\vec{u} \in \mathbb{R}^n$

$$\nabla_{\vec{u}} f(\vec{x}) = \lim_{\varepsilon \to 0} \frac{f(\vec{x} + \varepsilon\vec{u}) - f(\vec{x})}{\varepsilon}$$

In particular $\nabla_{\vec{u}} f(\vec{x}) = \vec{u} \cdot \nabla f(\vec{x})$

Gradient always goes in direction of maximal ascent because

$$\nabla_{\vec{u}} f(\vec{x}) = \nabla f(\vec{x}) \cdot \vec{u} = \|\nabla f(\vec{x})\| \, \|\vec{u}\| \cos\theta$$

angle between $\nabla f(\vec{x})$ and $\vec{u}$

If we max $\nabla_{\vec{u}} f(\vec{x})$ over all directions $\vec{u}$ (assuming $\|\vec{u}\| = 1$) we see that max occurs when $\theta = 0 \Rightarrow \nabla_{\vec{u}} f(\vec{x}) = \|\nabla f(\vec{x})\| \Rightarrow \vec{u} = \dfrac{\nabla f(\vec{x})}{\|\nabla f(\vec{x})\|}$

Steepest descent: $\theta = \pi \Rightarrow \vec{u} = -\dfrac{\nabla f(\vec{x})}{\|f(\vec{x})\|}$ ← basis of many algorithms

Frechet differentiability: Let $f: \mathbb{R}^n \to \mathbb{R}^m$. f is called Frechet differentiable if there exists a linear map $Jf: \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\vec{u} \to 0} \frac{\|f(\vec{x} + \vec{u}) - f(\vec{x}) - Jf(\vec{x})\vec{u}\|_2}{\|\vec{u}\|} = 0$$

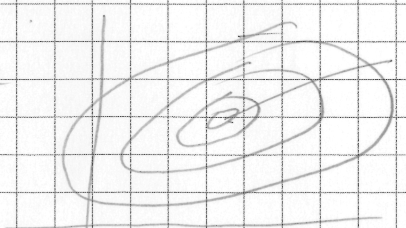
If $f = (f_1(\vec{x}), \ldots, f_m(\vec{x}))^T$ and all partial derivatives exist then Jf can be represented by the Jacobian

$$Df(\vec{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Contour plots: useful for visualisation functions in 2D

Level set is a set of the form $\{\vec{x} : f(\vec{x}) = c\}$ — $c \in \mathbb{R}$, level

Convex function



only one sink

Linear algebra review

Inner product $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$, satisfying (vector space)

i) $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$ symmetry

ii) $\langle a\vec{x} + b\vec{y}, \vec{z} \rangle = a\langle \vec{x}, \vec{z} \rangle + b\langle \vec{y}, \vec{z} \rangle$ linearity

iii) $\langle \vec{x}, \vec{x} \rangle > 0 \quad \forall \vec{x} \in V, \vec{x} \neq \vec{0}$ pos. def.

Every scalar product induces a norm $\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}$

Ex. Euclidean product in $\mathbb{R}^n$

$$\langle \vec{x}, \vec{y} \rangle = \vec{x} \cdot \vec{y} = \vec{x}^T \vec{y} = \sum_{i=1}^{n} x_i y_i \qquad \vec{x} = (x_1 \dots x_n),\ \vec{y} = (y_1 \dots y_n)$$

Norm $\|\ \| : V \to \mathbb{R}$, is a function satisfying

i) $\|\vec{x}\| \geq 0 \quad \forall \vec{x} \in V$ and $\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = \vec{0}$

ii) $\|\lambda \vec{x}\| = |\lambda| \|\vec{x}\| \quad \forall \lambda \in \mathbb{R}, \vec{x} \in V$

iii) $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ triangle inequality

Examples: 1-norm $\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i|$

2-norm $\|\vec{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$ $\Leftarrow$ induced by Euclidean inner product

∞-norm $\|\vec{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$

Cauchy Schwarz $|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\|_2 \|\vec{y}\|_2$

Basis of a linear vector space $\{\vec{b}_1, \dots \vec{b}_k\}$ is a basis of V if for every $\vec{x} \in V$ there exists a unique representation $\vec{x} = \sum_{i=1} \lambda_i \vec{b}_i$

Basis orthogonal if $\langle b_i, b_j \rangle = 0$ for $i \neq j$

orthonormal — " — and $\|b_i\| = 1 \quad \forall i = 1 \dots k$

Classical example: $V = \mathbb{R}^n$, $\{\vec{e}_1, \dots \vec{e}_n\}$ where $\vec{e}_i = (0, \dots \underset{\uparrow \text{ i-th entry}}{1} 0 \dots 0)$

Solving linear systems: $\vec{A}\vec{x} = \vec{b}$ $\quad \vec{A} \in \mathbb{R}^{m \times n}, \vec{x} \in \mathbb{R}^n, b \in \mathbb{R}^m$

$\Rightarrow$ When does a solution exist?
When is it unique?

$\underset{\text{image}}{\operatorname{Im}} (\vec{A}) = \{\vec{A}\vec{x} : \vec{x} \in \mathbb{R}^n\}$

$\underset{\text{kernel}}{\ker} \vec{A} = \{x \in \mathbb{R}^n : \vec{A} \cdot \vec{x} = \vec{0}\}$ $\Leftarrow$ null space

Equivalent conditions:

1. $\vec{A}$ is invertible
2. $\operatorname{rk}(\vec{A}) = n$
3. $\ker \vec{A} = \{\vec{0}\}$
4. $\operatorname{im} \vec{A} = \mathbb{R}^n$
5. rows of $\vec{A}$ are linearly independent
6. Columns of $\vec{A}$ are linearly inde