Optimality conditions for unconstrained optimization

$$\boxed{\min_x f(x) \quad x \in \mathbb{R}^n}$$

Def Descent direction

$f: \mathbb{R}^n \to \mathbb{R}$, $d \in \mathbb{R}^n$ is a descent direction for f at point $x \in \mathbb{R}^n$ if $\exists \bar{\alpha} > 0$

$$f(x + d\alpha) < f(x) \text{ for all } \alpha: 0 \le \alpha \le \bar{\alpha}$$

Lemma For a continuously differentiable function f, if a direction d such that $\langle \nabla f(x), d \rangle < 0$ then d is a descent direction at x.

$\triangledown$ $g(\alpha) := f(x + \alpha d)$ $\quad g: \mathbb{R} \to \mathbb{R}$ as a function of α

$g'(\alpha) = d^T \nabla f(x + \alpha d)$ $\quad$ (chain rule in multivariate case)

$$g(\alpha) = g(0) + g'(0)\alpha + \bar{o}(\alpha)$$

$$f(x + \alpha d) = f(x) + \alpha \nabla^T f(x) d + \bar{o}(\alpha)$$

$$\frac{f(x + \alpha d) - f(x)}{\alpha} = \nabla^T f(x) \cdot d + \frac{\bar{o}(\alpha)}{\alpha} \xrightarrow[\alpha \to 0]{} \nabla^T f(x) d$$

So, the sign of derivative is the same as $\nabla^T f(x) \cdot d$ $\blacktriangle$

Note:

Some descent directions might not satisfy $\nabla f^T(x) \cdot d < 0$:

$d = (0, 1)$

$f(x) = x_1^2 - x_2^2$

$\nabla f(x) = (2x_1, -2x_2)$

$\nabla f(1, 0) = (2, 0)$

$\langle \nabla f, d \rangle = 0$, but

$f(x + \alpha d)\big|_{x = (1,0)} = 1^2 - \alpha^2 < f(x) = 1$

FONC (first order necessary condition)

Thm If $\bar{x}$ is an unconstrained local min of f
If f is differentiable $f: \mathbb{R}^n \to \mathbb{R}$
Then $\nabla f(\bar{x}) = 0$.

$\triangledown$ Otherwise $-\nabla f(\bar{x})$ would be a descent direction $\blacktriangle$

Note $\circledast$:

Some points with no descent directions might not be local minima:

$$f(x_1, x_2) = (x_1^2 - 2x_2)(2x_1^2 - x_2)$$

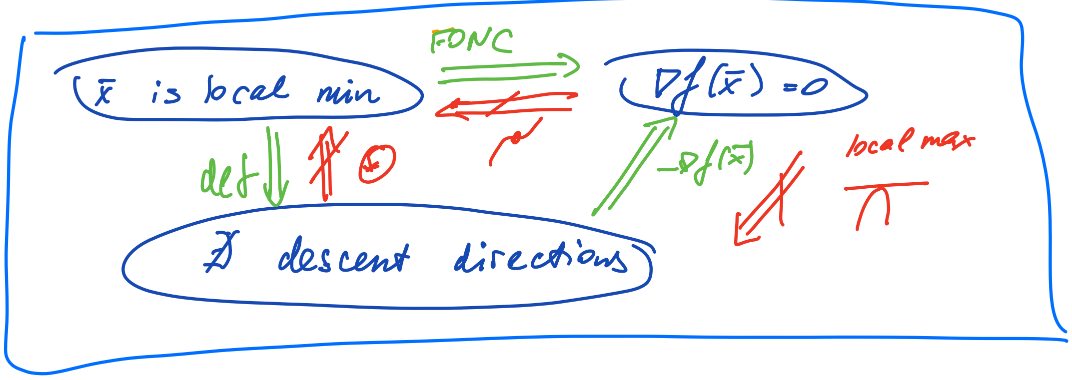
Proof of Note $\circledast$: Not a local min: $f(0,0) = 0$, $\quad f(t, t^2) = (t^2 - 2t^2)(2t^2 - t^2) = -t^4 < 0 \quad \forall t > 0$

Take a direction (d_1, d_2), $\quad g(\alpha) = f(x + \alpha d)\big|_{x=0}$, $\quad g(0) = 0$

$$g(\alpha) = (\alpha^2 d_1^2 - 2\alpha d_2)(2\alpha^2 d_1^2 - \alpha d_2) = 2\alpha^4 d_1^4 - \alpha^3 5 d_1^2 d_2 + 2\alpha^2 d_2^2 = 2\alpha^2\left(d_1^4 \alpha^2 - \frac{5}{2}\alpha\, d_1^2 d_2 + d_2^2\right) > 0$$

for small α

$$D = \frac{25}{4} d_1^4 d_2^2 - 4 d_1^4 d_2^2 = \frac{9}{4} d_1^4 d_2^2$$



SONC (Second order necessary condition)

Thm If f is twice CTS differentiable
If $\bar{x}$ is a strict local minimizer
Then $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x}) \succeq 0$.

$\triangledown$

$$f(\bar{x}+\alpha d) = f(\bar{x}) + \alpha d^T \underbrace{\nabla f(\bar{x})}_{=0} + \frac{\alpha^2}{2} d^T \nabla^2 f(\bar{x}) d + \bar{o}(\alpha^2)$$

$$\frac{f(\bar{x}+\alpha d) - f(\bar{x})}{\alpha^2} = \frac{1}{2} d^T \nabla^2 f(\bar{x}) d + \frac{\bar{o}(\alpha^2)}{\alpha^2} > 0 \text{ for small } \alpha > 0$$

So, $d^T \nabla^2 f(\bar{x}) d \geq 0$
for any direction d, so $\nabla^2 f(\bar{x}) \succeq 0$ ▲

Note: even if $\nabla f(x^*) = 0$ and $\nabla^2 f(x) \succeq 0$ the point might be not a strict minimizer ($=0$).

SOSC (Second order sufficient condition)

Thm If $f: \mathbb{R}^n \to \mathbb{R}$ is twice CTS differentiable
$\exists\ \bar{x}: \nabla f(\bar{x}) = 0$, $\nabla^2 f(\bar{x}) \succ 0$
Then $\bar{x}$ is a strict local minimiz

$\triangledown$ Let λ be the min eigenvalue of $\nabla^2 f(x)$

$$\Rightarrow y^T \nabla^2 f(\bar{x}) y \geq \lambda \|y\|^2 \quad \forall y \in \mathbb{R}^n \quad \left(\text{Let } y = \alpha d,\ \|d\| = 1\right)$$

$$f(\bar{x}+y) - f(\bar{x}) = \frac{1}{2} y^T \nabla^2 f(\bar{x}) y + \bar{o}(\|y\|^2) \geq$$

$$\geq \frac{1}{2}\|y\|^2 \lambda + \bar{o}(\|y\|^2)$$

$$\frac{f(\bar{x}+y) - f(\bar{x})}{\|y\|^2} \geq \frac{1}{2}\lambda + \frac{\bar{o}(\|y\|^2)}{\|y\|^2} \xrightarrow[\alpha \to 0]{} \frac{\lambda}{2} > 0$$

So, $f(\bar{x}+y) > f(\bar{x})$ for $\|y\|$ small enough ▲

Note: Strict local minimizer might not have $\nabla^2 f(x) \succ 0$

Two problems: ① We got no info about global minima

② FONC + SONC + SOSC can be together inconclusive.

Example $f(x) = 4x_1^2 - x_2^3$ at $(0,0)$

$\nabla f(0,0) = (8x_1, -3x_2^2) = (0,0)$ FONC ✓

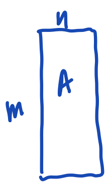
$\nabla^2 f(0,0) = \begin{pmatrix} 8 & 0 \\ 0 & -6x_2 \end{pmatrix} = \begin{pmatrix} 8 & 0 \\ 0 & 0 \end{pmatrix}$ SONC ✓

SOCS ✗ since 0 eigenvalue

② Not a local min $d = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is problematic $f(\alpha d) = \begin{pmatrix} 0 \\ -\alpha^3 \end{pmatrix}$

Application 1 Least squares problem

$\min_x \|Ax - b\|_2^2$

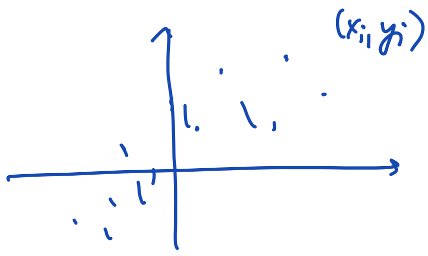


$m \gg n$

(cannot solve system directly)

That is, we have more data than dimensions

Note: Curve fitting is also a least squares problem!



$p(x) = c_3x^3 + c_2x^2 + \dots + c_0$

Suppose A has linearly independent columns

$f(x) = \|Ax - b\|^2 = x^T A^T A x - 2x^T A^T b + b^T b$

$\nabla f(x) = 2A^T A x - 2A^T b$ candidates for solutions

$\nabla f(x) = 0$ $x = (A^T A)^{-1} A^T b$

Note: nullspace of $A^T A$:

$x^T A^T A x = 0$

$\|Ax\|^2 = 0 \Rightarrow Ax = 0 \Rightarrow x = 0 \Rightarrow A^T A$ is invertible (full rank)

Hence, $\nabla^2 f(x) \succ 0 \Rightarrow x$ is a strict local min

Uniqueness: objective is radially unbounded

Application 2 Fermat – Weber facility location problem

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{m} \|x - z_i\|$$

place new grocery store to minimize travel to the houses

Application 3 More general function fitting

x_{ij}

$$\min_{a_k,u,v} \sum_{ij} \left(x_{ij} - \sum_{k=1}^{r} a_k u_i v_j\right)^2$$

$$\min_{a,u,v} \sum_{ij} \left(x_{ij} - \sum_{k} a_k \exp(-|u_i - v_j|^2)\right)$$