Lecture 4 ⊛ Handout @ front!

Intro to Optim.jl
- gradient based continuous opt
- uses provided gradients

Example 1: Rosenbrock
Example 2: ||A - UV^T||
Example 3: Flux.jl

Overview of opt. terms & defns
- global min
- strict global min
- local min
- strict local min
- isolated min

1d Taylor

1d Optimiality
- Necessary
- Sufficient

] Prep for multi-d.

Office Hours
3:30-4:30pm on Friday (this week)

TA: Disha Shur

⊛ Adopt Meaningfully attempted pts! ⊛

Optim.jl

A software package for gradient based continuous optimization like we will see in class!

The software needs

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

$$g(\underline{x}) = \frac{\partial f}{\partial x} : \mathbb{R}^n \to \mathbb{R}^n$$

We need to give these as function to the software!

Example:

$$\min \|A - UV^T\|_F^2$$

s.t. U, V are rank k.

$A: m \times n, \quad m \geq n$

$U: m \times k \quad k < n$

$V: n \times k \quad k \leq n$

Actual, "analytic" soln given by the SVD of A.

SVD of A

A ~~B~~ : Real-valued

$A = U \Sigma V^T$

U: $m \times m$, orthog.

V: $n \times n$, orthog.

Σ: $m \times n$, diag. w/ non-neg. values.

By Convention

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & 0 \\ \vdots & & & \\ 0 & 0 & \ddots & \sigma_n \end{bmatrix}$$

$\sigma_i \geq \sigma_{i+1} \quad \forall i$

U

V^T

U^*

$U^T U$

U^*

$^* = \text{opt } U$

$= U[:, 1:K]\, \Sigma[1:K, 1:K]$

$= U_K \Sigma_K$

$= \text{opt } V$

$= V[:, 1:K]$

$= V_K$

$V^{*T} = U_K \Sigma_K V_K^T$

$^* = \sum_i u_i v_i^T \sigma_i$

$= U_K, \quad V^* = V_K \Sigma_K$

$f(U,V) = \|A - UV^T\|_F^2$

$g(U,V) = ??$

$$\underline{x} = \begin{bmatrix} \underline{u}_1 \\ \underline{u}_2 \\ \vdots \\ \underline{u}_K \\ \underline{v}_1 \\ \vdots \\ v_K \end{bmatrix}$$

$U = [\underline{u}_1$

$V = [\underline{v}_1$

$$= \sum (A_{ij} - (UV^T)_{ij})^2$$

$$f(U,V) = \|A - UV^T\|_F^2$$

$$= tr((A - UV^T)^T(A - UV^T))$$

$$= tr(A^TA) - tr(A^TUV^T)$$

$$- tr(VU^TA)$$

$$\Bigg] = \boxed{-2\,tr(VU^TA)}$$

$$+ tr(VU^TUV^T) \rightarrow$$

$$\underline{u}_2 \;-\; \underline{u}_k$$

$$\underline{v}_2 \cdots \underline{v}_k$$

$$tr(X^TX) = \|X\|_F^2$$

$$tr(A) = \sum_i A_{ii}$$

$$tr(X) = tr(X^T)$$

$$tr(XYZ) = tr(ZXY) = tr(YZX)$$

(dims have to work)

$$\frac{\partial f}{\partial U} = -2AV + 2UV^TV$$

$$= -2(A - UV^T)V$$

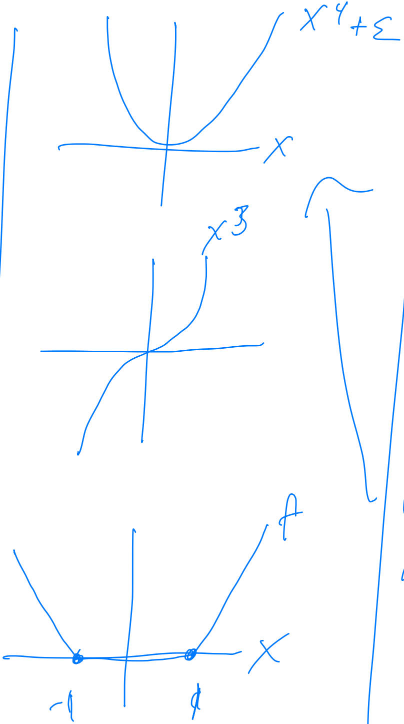
$$\frac{\partial f}{\partial V} = -2(A - UV^T)^T U$$

$$\frac{\partial\left((a - uv)^2\right)}{\partial u} = -2(a - uv)v$$

A basket of definition

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

A global minimizer is a point $\underline{x}^*$ s.t. $f(\underline{x}^*) \le f(\underline{x})$ for all $\underline{x}$ in the domain of f.



A ==Strict global== min is a point $\underline{x}^*$ st

$$f(\underline{x}^*) < f(\underline{x})$$

for all $\underline{x} \neq \underline{x}^*$ in the domain of f.

Only $x^4+\varepsilon$ has a strict global min

Quiz

Does
have a
min?

Ans!

$\|A - UX$

$\|A - C$

any

Idea

$\|A - UV^T\|$

strict global

No!

$X^{-1} V^T\| =$

$uV^T\|$ for

non-sig X.

Local min

$\underline{x}^*$ is a local min

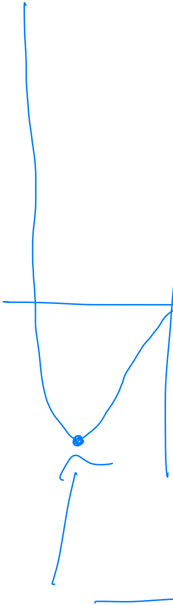
if $f(\underline{x}^*) \le f(\underline{x})$

for all $x \in$ open neighborhood of x^*.

x^* is a strict local min

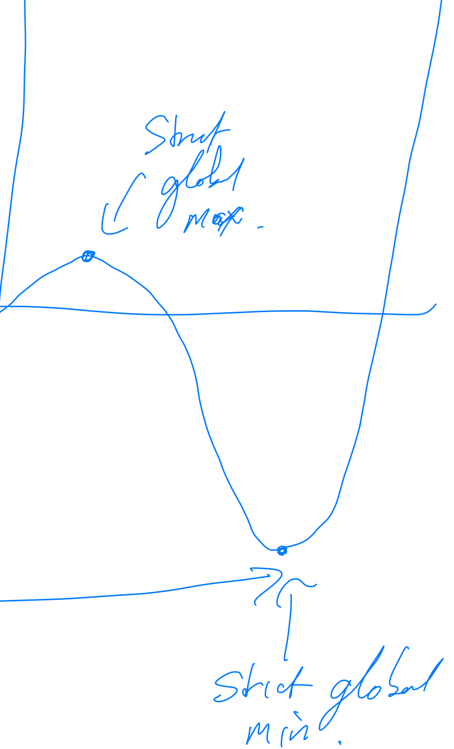
if $f(\underline{x}^*) < f(\underline{x})$

for all $\underline{x} \ne \underline{x}^*$ in a open neighborhood of $\underline{x}^*$.

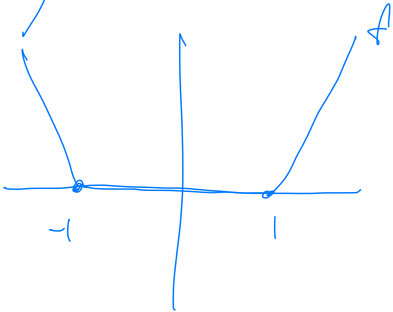


Strict local min

$f(x) = (1-x^2)((x-1)^2-2)$



A point x^* is an isolated local min if its ~~the only local min~~ a local min and there is an open neighborhood where it is the only local min.



Strict $\cancel{\Leftarrow}$ Isolation

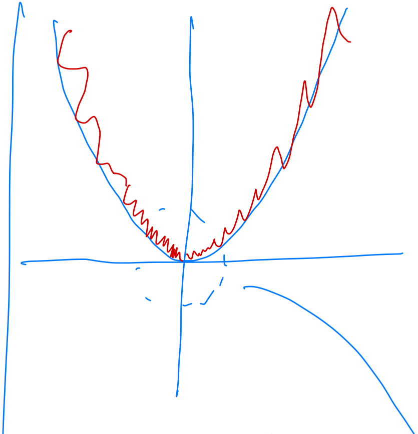
Isolated $\Rightarrow$ Strict!

But Strict $\Rightarrow$ Isolated

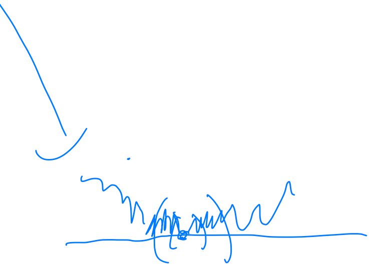
☹

$$f(x) = x^4(\cos(1/x) + 2)$$

$$f(0) = 0$$

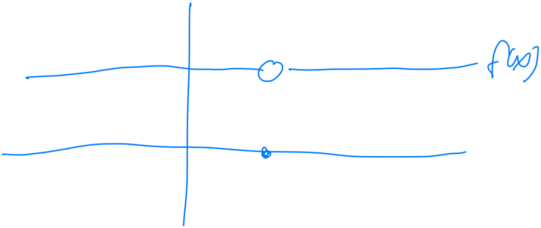


$x = 0$ is a strict local min because $f(x) > 0$ for all $x \neq 0$.



How can we guarantee that $\underline{x}$ is a local min of f?

→ Assume Continuity!



$$f(x) = \begin{cases} 1 & \text{if } x \neq \text{magic} \\ 0 & \text{if } x = \text{magic} \end{cases}$$

Taylor's theorem allows us to

control / determine

behavior of a cont. function

w/ local info!

id.

$$f(x_0+h) = f(x_0) + h f'(x_0+\alpha h)$$

for some $\alpha \in [0,1)$

$$f(x_0+h) = f(x_0) + h f'(x_0) + \frac{h^2}{2} f''(x_0+\alpha h)$$

α depends on h..

$\alpha \in [0,1)$

Necess

Con

Necessar

a loca

true @

$\Rightarrow$ Cr

points

Sufficie

then

loca

(Can

pts

sary vs. Sufficient
ditions

y: those are true at
l min (but may be
other pts too.)
n be used to reject.
as local min.)

: if those are true,
the point is a
l min
be used to accept
as local min.)