New Problem Today: Nonlinear Optimization

Given a function $f: \mathbb{R}^n \to \mathbb{R}$ — objective function

and a set $S \subseteq \mathbb{R}^n$ — feasible set

find $\underline{x}^* \in S$ such that f attains a minimum on S at $\underline{x}^*$.

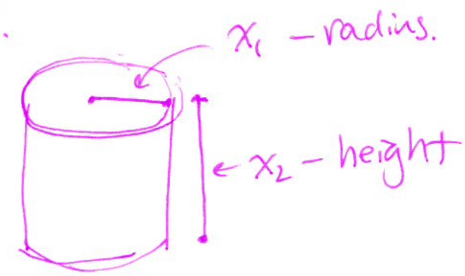
i.e. $f(\underline{x}^*) \leq f(x)$ for all $\underline{x} \in S$

Normally the set S is defined using constraints.

$$S = \{ \underline{x} \in \mathbb{R}^n \mid \underbrace{g(\underline{x}) = \underline{0}}_{\text{equality constraints}}, \underbrace{h(\underline{x}) \leq \underline{0}}_{\text{inequality constraints}} \}$$

If $S = \mathbb{R}^n$, the optimization problem is said to be unconstrained.

eg//. Minimize the surface are of a cylinder given a ~~to~~ volume constraint.



Surface area $f(x_1, x_2) = 2\pi x_1 \cdot x_2 + 2\pi x_1^2$

Subject to $g(x_1, x_2) = \pi x_1^2 x_2 = 355$

Local vs. Global Minima

Def The point $\underline{x}^*$ is a global minimum of f if

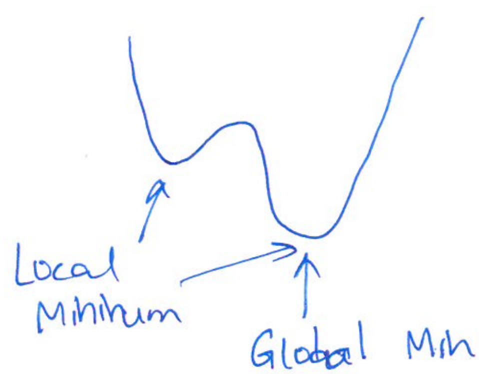
$f(\underline{x}^*) \leq f(\underline{x})$ for any $\underline{x} \in S$

Def The point $\underline{x}^*$ is a local minimum of f if

$f(\underline{x}^*) \leq f(x)$ ~~∀~~ within some neighbourhood of x^*

(there exists $\epsilon > 0$ such that $\forall \|\underline{x} - \underline{x}^*\| < \varepsilon$
then $f(\underline{x}^*) \leq f(\underline{x})$)

The techniques we discuss today will find local minima.
It's generally hard to find global minima unless
f has special properties



Unconstrained Optimization — optimality condition.

In 1D. $f: \mathbb{R} \to \mathbb{R}$ a local minima always has:

- $f'(x) = 0$ —— critical point.
- $f''(x) > 0$.

<u>Def</u> A <u>critical point</u> of $f: \mathbb{R}^n \to \mathbb{R}$ is a point $\underline{x} \in \mathbb{R}^n$ such that the <u>gradient</u> of f is $\underline{0}$ at $\underline{x}$

$$\nabla f(\underline{x}) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} = \underline{0}$$

Not all critical points are minima, can also be maxima <u>or</u> saddle points.

eg// $f(\underline{x}) = x_1^2 + x_2^2$. $\quad f: \mathbb{R}^2 \to \mathbb{R}$

$$\nabla f(\underline{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1}(x_1^2 + x_2^2) \\ \frac{\partial}{\partial x_2}(x_1^2 + x_2^2) \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$\Rightarrow$ critical point at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

eg// $f(\underline{x}) = x_1^2 - x_2^2$

$$\nabla f(\underline{x}) = \begin{bmatrix} 2x_1 \\ -2x_2 \end{bmatrix} \quad \Rightarrow \text{critical point at} \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We can distinguish among critical points of $f: \mathbb{R}^n \to \mathbb{R}$ by considering the <u>Hessian</u> matrix

$$H_f(\underline{x}) = \begin{bmatrix} \frac{\partial^2 f(\underline{x})}{\partial x_1^2} & \frac{\partial^2 f(\underline{x})}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\underline{x})}{\partial x_2^2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

If the second partial derivatives of f are continuous

then $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ $H_f(x)$ is symmetric

At a critical point $\underline{x}^*$, if $H_f(x^*)$ is

singular $\Rightarrow$ inconclusive

positive definite $\Rightarrow$ $\underline{x}^*$ minimum

negative definite $\Rightarrow$ $\underline{x}^*$ maximum

indefinite. $\Rightarrow$ x^* saddle point.

Recall Def: A matrix M is positive definite if

$\underline{x}^T M \underline{x} > 0 \quad \forall\, \underline{x} \in \mathbb{R}^n, \ \underline{x} \neq \underline{0}$.

Def: A matrix M is negative definite if $-M$ is positive definite.

Def. A matrix M is indefinite if it is neither pos def. nor negative definite, but is nonsingular

Q: How do we test if a symmetric matrix is pos. def?

A: Try to compute its Cholesky factorization.

Cholesky factorization will succeed iff the matrix is positive definite.

eg// $f(\underline{x}) = x_1^2 + x_2^2$ $\quad \nabla f(\underline{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$

$H_f(\underline{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ which is pos. def. so $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is a minimum.

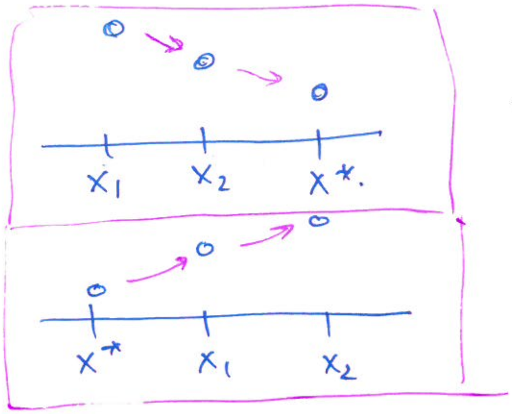
eg// $f(\underline{x}) = x_1^2 - x_2^2$ $\quad \nabla f(\underline{x}) = \begin{bmatrix} 2x_1 \\ -2x_2 \end{bmatrix}$

$H_f(\underline{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$ which is indefinite so $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is a saddle point.

We'll focus on finding a minimum of $f : \mathbb{R} \to \mathbb{R}$ (one dimensional case).

Golden Section Search ~ like bisection search
~ won't use derivative information

Def. A function $f : \mathbb{R} \to \mathbb{R}$ is unimodal on an interval $[a, b]$ if there is a unique value $x^* \in [a, b]$ such that $f(x^*)$ is the (local) minimum of f on $[a, b]$ and for any $x_1, x_2 \in [a, b]$ with $x_1 < x_2$.



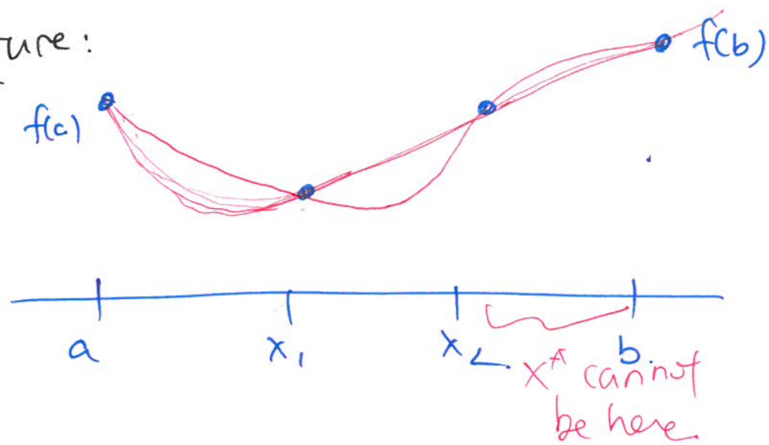
$x_2 < x^*$ implies. $f(x_1) > f(x_2)$

$x_1 > x^*$ implies. $f(x_1) < f(x_2)$

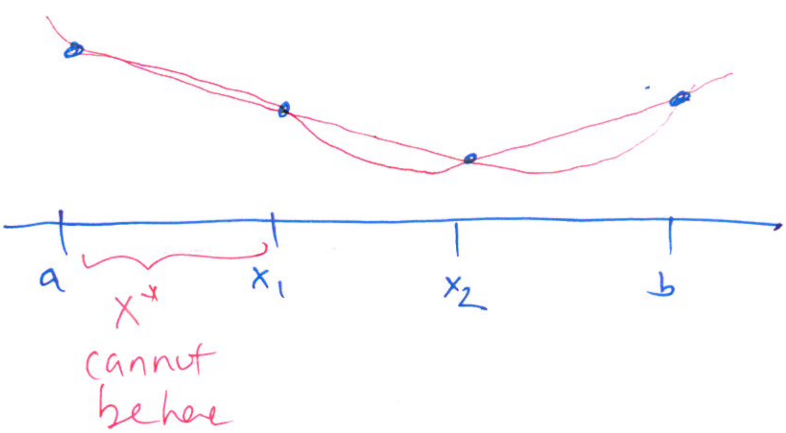
Idea behind golden section search:

- compute function values $f(.)$
- use the unimodal property to "shrink" the interval in which x^* lies.

Picture:

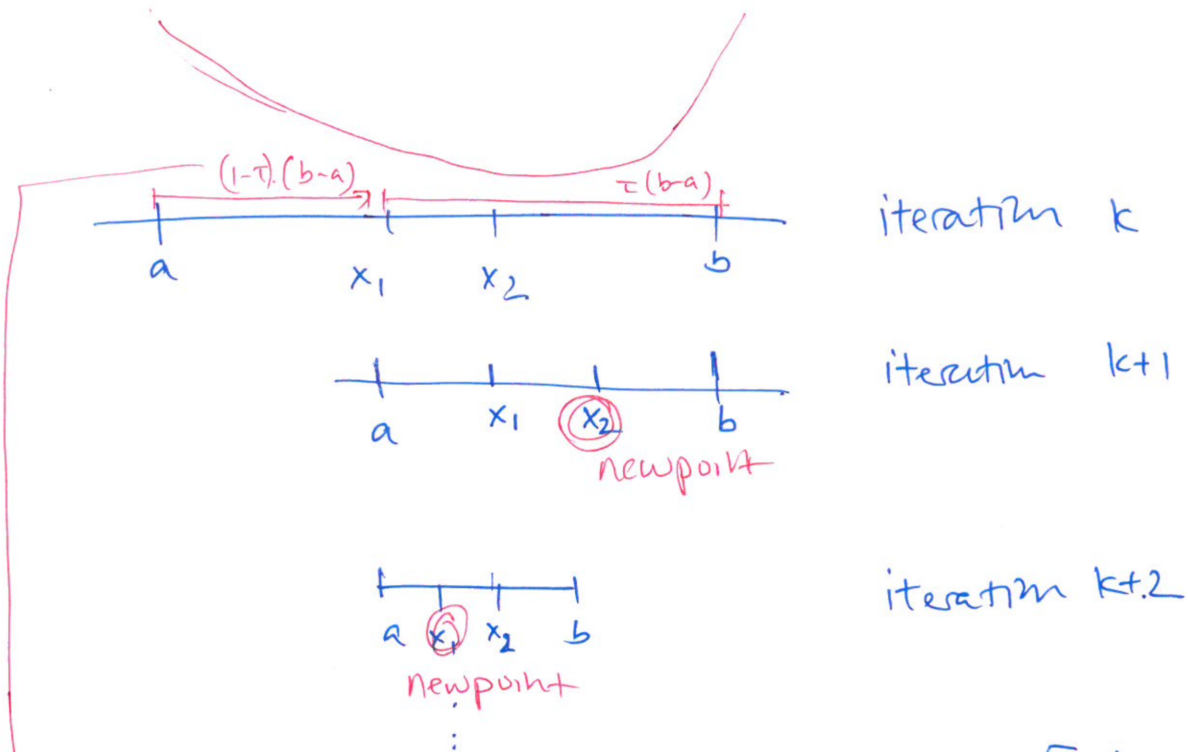


In this picture, $f(x_1) < f(x_2)$, so x^* cannot be in the interval $(x_2, b]$.



$f(x_1) > f(x_2)$ So x^* cannot be in the interval $[a, x_1)$

How do we choose x_1, x_2 ?



If we solve for τ, we get $\tau = \frac{\sqrt{5}-1}{2} = 0.618$

$\Rightarrow$ See. Algo 6.1 or ~~the~~ slides 19

Convergence of the golden section search is linear

with $C = 0.618 = \tau$

Golden Section Search.

Good: Guaranteed convergence, easy to implement.
Does not require derivative

Bad: Finding an initial bracket $[a, b]$ can be difficult

Golden Section search conv. slowly.

finds a local optima.

Newton's Method

Idea: use a quadratic approximation of f near x, and find the minimum of the quadratic approx.

Taylor series Expansion

$$f_{approx}(x+h) = f(x) + f'(x)h + \frac{1}{2} f''(x) h^2$$

To find the minimum of this quadratic, set...

$$\frac{\partial f_{approx}(x+h)}{\partial h} = 0. = f'(x) + f''(x)\, h.$$

$$\Rightarrow \boxed{h = -\frac{f'(x)}{f''(x)}}$$

Algo: Start with x_0 close to x^* ← otherwise, we won't converge.

make updates

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}.$$

The convergence rate of Newton's method is quadratic. But the convergence is not guaranteed.

Even if Newton's method converges, we might not find a minimum → can converge to maximum of inflection point.