CSC 338 Lecture 10

Computational Problem: Unconstrained optimization:

find a local minimum of a function $f: \mathbb{R}^n \to \mathbb{R}$.

(x^* is a local min if there exists $\epsilon > 0$ s.t. $\|\underline{x} - \underline{x}^*\| < \epsilon \Rightarrow f(\underline{x}^*) \le f(\underline{x})$)

A local min is always a critical point. $\nabla f(\underline{x}) = 0$.

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

~~eg// $f(x, y) = xy + y^2$~~

eg// $f(x_1, x_2) = x_1 x_2 + x_2^2 - x_1$.

$\nabla f(\underline{x}) = \begin{bmatrix} x_2 - 1 \\ x_1 + 2x_2 \end{bmatrix}$ $\nabla f = \underline{0}$ at $\underline{x} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$

Critical points can be max, min, or saddle pt.

To distinguish, we consider the ~~the~~ second derivatives

$\Rightarrow$ Hessian Matrix.

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

eg// $H_f(\underline{x}) = \begin{bmatrix} \frac{\partial (x_2 - 1)}{\partial x_1} & \frac{\partial}{\partial x_1}(x_1 + 2x_2) \\ \frac{\partial}{\partial x_2}(x_2 - 1) & \frac{\partial}{\partial x_2}(x_1 + 2x_2) \end{bmatrix}$

$= \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$

If the second partial derivatives of f are continuous, then $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$

In other words, $H_f(\underline{x})$ is symmetric.

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

positive definite	$\Rightarrow \underline{x}^*$ is a	min	like $f''(x) > 0$
negative def	$\Rightarrow \underline{x}^*$	max	$f''(x) < 0$
indefinite	$\Rightarrow \underline{x}^*$	saddle pt	- pos in some dir, neg in others
singular	$\Rightarrow$	inconclusive	$f''(x) = 0$

Def A matrix M is pos def if $\underline{x}^T M \underline{x} > 0$ $\forall x \in \mathbb{R}^n$, $\underline{x} \neq 0$.

M neg def if $-M$ is pos def.

indefinite if it is neither pos def or neg def, but not singular.

Q: How to test if M is pos def?

A: Try to compute its Cholesky factorization, see if we succeed!

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

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

$H_f(\underline{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

eg// from earlier

$\Rightarrow$ indefinite

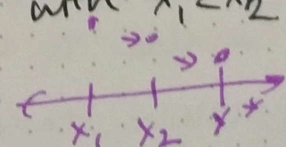
$\Rightarrow$ saddle pt.

we'll focus first on $f: \mathbb{R} \to \mathbb{R}$.

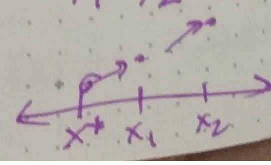
<u>Golden Section Search</u> ~ like bisection search; won't need derivative info

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

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

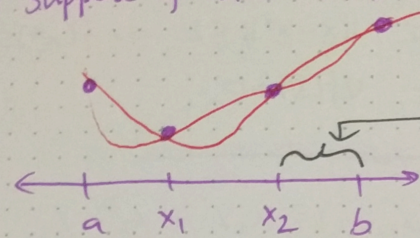


$x_1 > x^*$ implies $f(x_1) < f(x_2)$

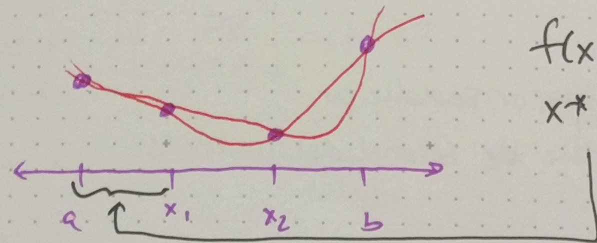


Idea: Compute the function values, then use the unimodal property to "shrink" the interval in which x^* lies:

e.g.// suppose f unimodal on $[a,b]$.



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



$f(x_1) > f(x_2)$ so

x^* can't be in the interval $[a, x_1)$

Idea: Choose x_1, x_2, then rule out either $(x_2, b]$ or $[a, x_1)$

But how do we choose x_1, x_2?

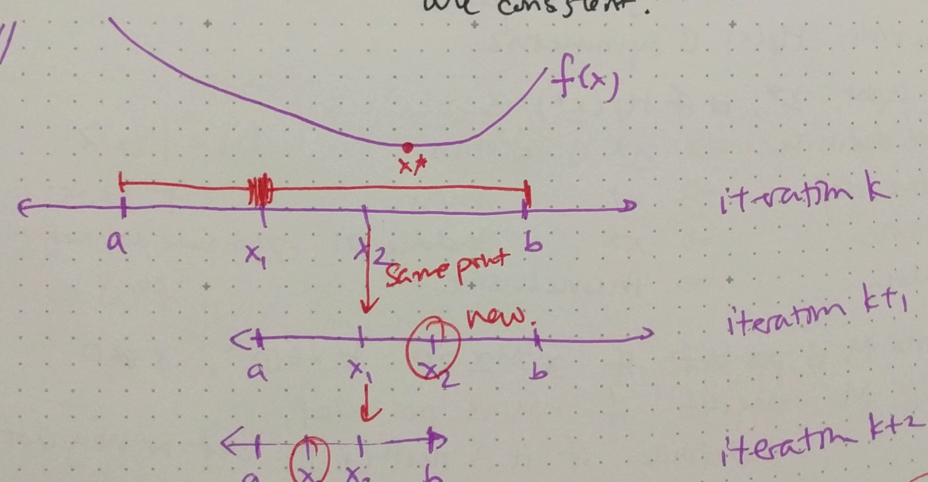
Consideration: ① we need to evaluate $f(x_1), f(x_2)$, so it would be nice if the x_i not ruled out is reused in the next iteration.

② It would also be nice if the ratios

$(x_1 - a) : (x_2 - x_1) : (b - x_2)$

are consistent.

e.g.//



Fix the ratio $\frac{b-x_1}{b-a} = \tau$ So $\frac{x_1-a}{b-a} = 1-\tau$, $\frac{x_2-x_1}{b-a} = (1-2\tau)$

$\cancel{\frac{b-x_2}{b-x_1} = \frac{b-x_2}{\tau(b-a)}}$

$\frac{x_2-x_1}{b-x_1} = (1-\tau) = \frac{x_2-x_1}{\tau(b-a)}$

$\to \tau(1-\tau)(b-a) = (1-2\tau)(b-a)$

$\Rightarrow \tau(1-\tau) = (1-2\tau)$

$\Rightarrow \tau^2 + \tau - 1 = 0$

$\Rightarrow$ we have $\tau = \frac{\sqrt{5}-1}{2} = 0.618\ldots$ the golden ratio (see slide 19)

Demo Convergence of Golden See Search liner with $c = 0.618 = \tau$.

Pro: Guaranteed convergence, no derivatives

Bad: Need initial bracket, unimodal, slow conv, local opt.

Newton's Method

Idea: Use a quadratic approximation of f near x, and find min. of the quadratic approx.

$$f(x+h) = \underbrace{f(x) + f'(x)h + \tfrac{1}{2}f''(x)h^2}_{\text{quadratic approx}} + O(h^3)$$

$\Rightarrow$ set $\frac{\partial}{\partial h}\left(f(x) + f'(x)h + \frac{1}{2}f''(x)h^2\right) = 0$

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

Algo: start with x_0 close to x^*.

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

The convergence rate of Newton's method is quadratic. But convergence is not guaranteed. Moreover, even if Newton's method converges, we might not find a min: can converge to a max / inflection point.

Demo

Sensitivity & Conditioning

Recall: The problem of finding a root of $f(x)$ has abs cond. num $\left|\frac{1}{f'(x)}\right|$

Minimization and root finding are related problems, but min problems have worse conditioning ~ it is like finding a multiple root.

$f'(x) \approx 0$ near min

(Root finding if $|f(\hat{x})| \leq \epsilon$ then $|\hat{x} - x^*| \leq \left|\frac{\epsilon}{f'(x^*)}\right|$)

Taylor series expansion

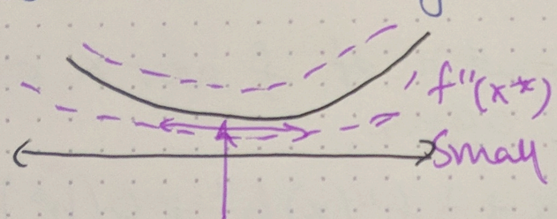
$$f(\hat{x}) = f(x^* + h) = f(x^*) + \underbrace{f'(x^*)}_{=0}h + \tfrac{1}{2}f''(x^*)h^2 + O(h^3)$$

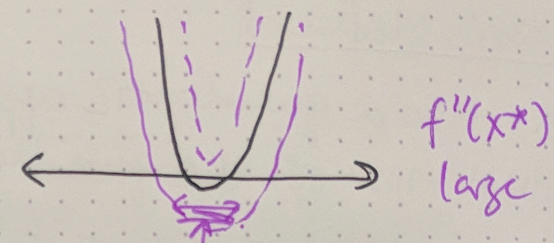
$$\approx f(x^*) + \tfrac{1}{2}f''(x^*)h^2$$

$$h^2 \approx 2 \frac{f(\hat{x}) - f(x^*)}{f''(x^*)} \quad \text{if } f''(x^*) \neq 0$$

So if $|f(\hat{x}) - f(x^*)| \le \epsilon \Rightarrow |\hat{x} - x^*| \le \sqrt{\dfrac{2\epsilon}{f''(x^*)}}$

$\Rightarrow$ smaller second derivative means the min problem has worse conditioning





Notice that there is a square root $\sqrt{\epsilon} > \epsilon$ since ϵ is small.

If $f''(x^*) \approx 1$, error will be $\sqrt{\epsilon}$, so even if ϵ is small (e.g. $\epsilon = \epsilon_{mach}$), the solution $\hat{x}$ can be computed to <u>half</u> as many digits of accuracy as the underlying machine precision

Also, if $f'(x)$ is available, the problem of solving $f'(x)=0$ has condition number $\left|\dfrac{1}{f''(x)}\right|$

This sensitivity / conditioning result generalizes to $f : \mathbb{R}^n \to \mathbb{R}$