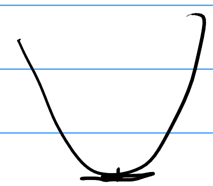
12 Optimization

Let's try to weaken the requirement $f(\vec{x}) = \vec{0}$. $(f: \mathbb{R}^n \to \mathbb{R}^n)$

$$\min \|f(x)\|_2$$

Create a problem statement for "optimization".

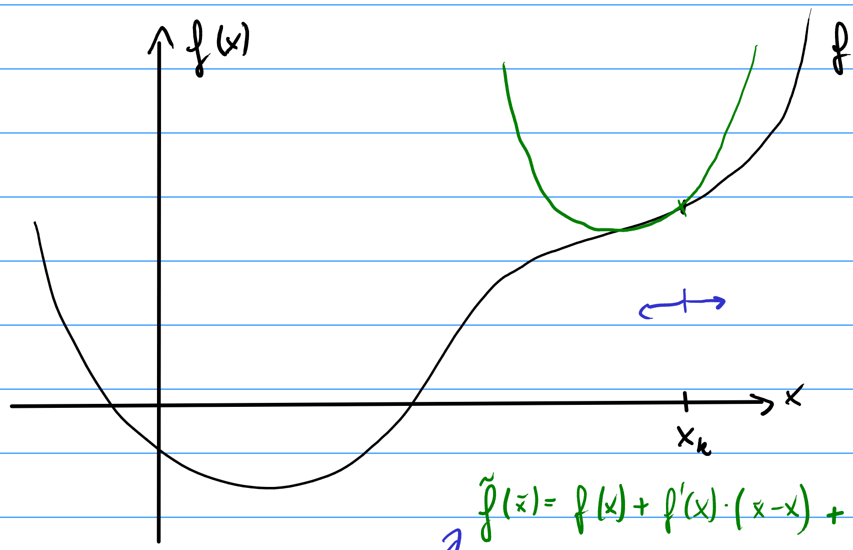
$$\min g(x)$$



$g'(x)=0$

Let's steal the idea from Newton's method for equation solving.

Build a simple version of f and minimize that. Let's try in 1D first.



$f(x) = x^3 + 2x^2 + 3x - 4$

$$\tilde{f}(\bar{x}) = f(x) + f'(x)\cdot(\bar{x} - x) + f''(x)\cdot\frac{(\bar{x}-x)^2}{2}$$

→ Taylor $\left(\cdots + f^{(n)}(x)\cdot\frac{(\bar{x}-x)^n}{n!}\right)$

Does a linear approximation (a line) help at all?

$h = \bar{x} - x$

$$\tilde{f}(x+h) = f(x) + f'(x)\cdot h + f''(x)\cdot\frac{h^2}{2}$$

$f(x) = x^3 + 2x^2 + 3x - 4$ $\quad f'(x) = 3x^2 + 4x + 3$ $\quad f''(x) = 6x + 4$

$$\hat{f}(\bar{x}) = -4 + \overset{f'(0)\cdot\bar{x}}{3\bar{x}} + \overset{f''(0)\cdot\bar{x}^2/2}{4\cdot\bar{x}^2/2}$$

$\tilde{f}(x+h) =$

① $x_{k+1} = x_k + h$

Now minimize that.

$\left[x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \right.$ (eqn solve) $\left.\right]$ ② $\rightsquigarrow h = -\frac{f'(x)}{f''(x)}$

$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$ (opt) ③

$f(x) = \cos(x)$

$f'(x) = -\sin(x)$

$f''(x) = -\cos(x)$

$$x_{k+1} = x_k - \frac{-\sin(x_k)}{-\cos(x_k)}$$

Does that look at all familiar?

$$f'(x) = h'(x) \cdot g'(h(x))$$

$$f(x) = -\exp(-x^2) \quad \rightsquigarrow \quad f(x) = g(h(x))$$

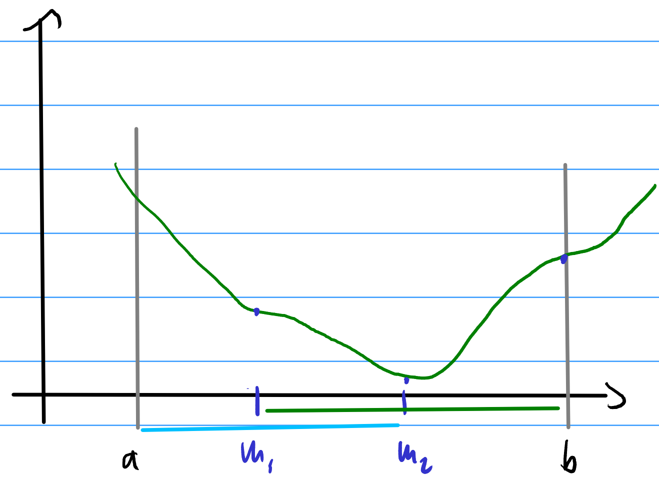
$$f'(x) = (-2x) \cdot (-\exp(-x^2)) \qquad g(x) = -\exp(x)$$

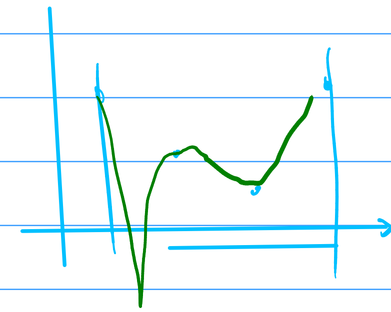
$$f''(x) = (-2)(-\exp(-x^2)) \qquad h(x) = -x^2$$

$$+ \; -4x^2 \exp(-x^2)$$

Golden Section Search

Let's try to create an analog to 'bisection', with a type of bracket.





Is one middle point in the bracket good enough?

No.

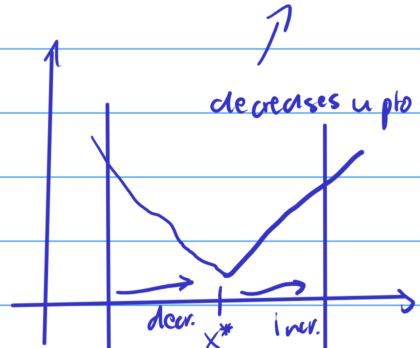
Next: what condition are we going to maintain throughout?

In particular: Is "the minimum is in the bracket" feasible?

No.

What does it mean for f to be 'unimodal'?

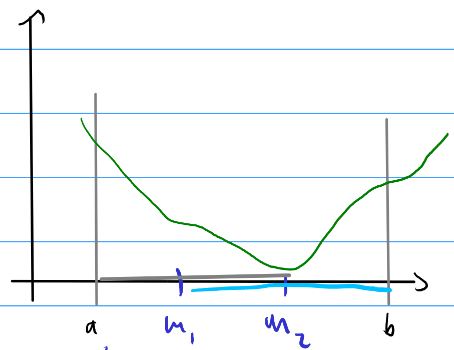
decreases up to some x^*, increases after that



Reality check: Do we typically know that a function is unimodal in a bracket?

No.

So how do we maintain unimodality in each bracket?

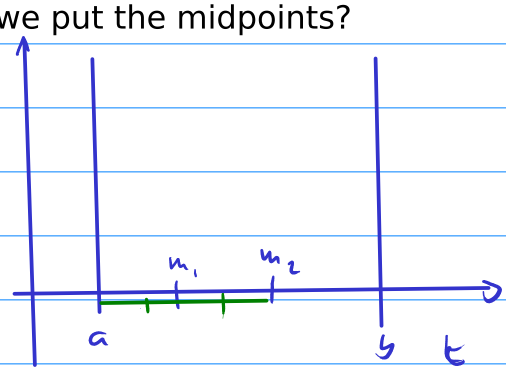


$f(m_1) < f(m_2) \rightsquigarrow$ zoom into $[a, m_2]$

$f(m_1) > f(m_2) \rightsquigarrow$ zoom into $[m_1, b]$

$f(m_1) = f(m_2) \rightsquigarrow$ do $\left\{ \begin{array}{l} \text{choose either} \\ \text{zoom into } [m_1, m_2] \end{array} \right.$

Where do we put the midpoints?



$$m_1 = a + \left(\underbrace{\frac{\sqrt{5}-1}{2}}_{.618}\right) \cdot (b-a)$$

$$m_2 = a + \left(1 - \frac{\sqrt{5}-1}{2}\right) \cdot (b-a)$$

→ Wikipedia "Golden section search"

What's the convergence order of Golden Section Search?

linear

Steepest Descent

What do we do in n dimensions?

Go in dir of steepest desc.

What does that mean mathematically?

$$x_{k+1} = x_k + \alpha \cdot (-\nabla f(x_k))$$

And how far do we go?

Find α by 1D optimization

Do an example: $f(x) = \frac{1}{2} x_0^2 + 2.5\, x_1^2$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_0} \\ \frac{\partial f}{\partial x_1} \end{pmatrix} = \begin{pmatrix} x_0 \\ 5 x_1 \end{pmatrix}$$

What's the convergence order in the example in the demo?

~ linear

Can we do better by using information from the second derivative?

$\rightarrow$ Newton :)

Newton's method in n dimensions

Step 1: Write down a quadratic approximation $\tilde{f}$ to f at x_k .

$$\tilde{f}(\vec{x}+\vec{h}) = f(\vec{x}) + \vec{\nabla} f(\vec{x}) \cdot \vec{h} + \frac{1}{2} \vec{h}^T \underbrace{H_f}_{\text{matrix}}(x) \cdot \vec{h}$$

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n} \end{pmatrix}$$

Step 2: Find minimum of $\tilde{f}$. To do so, take derivative and set to zero.

$$\nabla_h \tilde{f}(x) = \nabla f(x) + H_f(x) \cdot h \overset{!}{=} 0$$

$$h = -H_f(x) \nabla f(x)$$

$$x_{k+1} = x_k + h = x_k - H_f(x) \cdot \nabla f(x)$$

Do an example: $f(x) = \frac{1}{2} x_0^2 + 2.5\, x_1^2$

$$\nabla f(x) = \begin{pmatrix} x_0 \\ 5x_1 \end{pmatrix}$$

$$H_f(x) = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$$

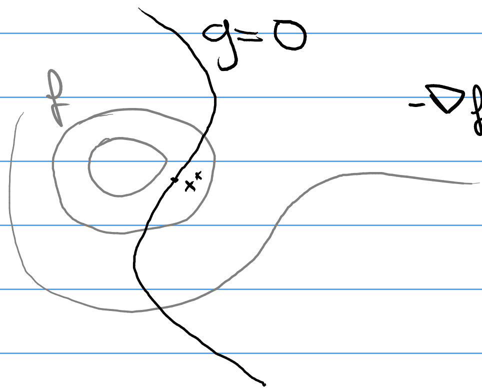
What if we don't even have one derivative, let alone two?!

Constrained Optimization

Modify the problem statement of optimization to accommodate a constraint.

What does a solution/minimum x^* of this problem look like?

I.e. what are some necessary conditions on x^* ?



$$-\nabla f = \nabla g \cdot \lambda = \nabla g^T \lambda$$
$$= J_g^T \lambda \quad \text{for some } \lambda$$

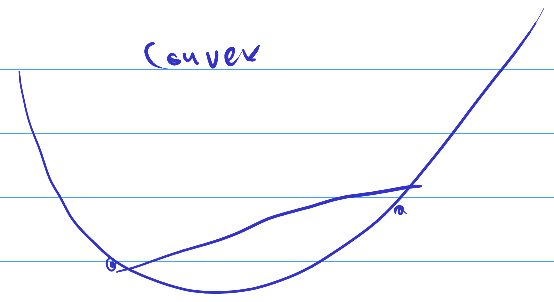
Miracle: Reduce constrained to un-constrained optimization.

Define a new function of more unknowns: x and λ , $\lambda \in \mathbb{R}^m$

$\mathcal{L}(x, \lambda) :=$

What are the necessary conditions for an un-constrained minimum of $\mathcal{L}$?

Using Newton's method on $\mathcal{L}$ gets a new name:


Convex