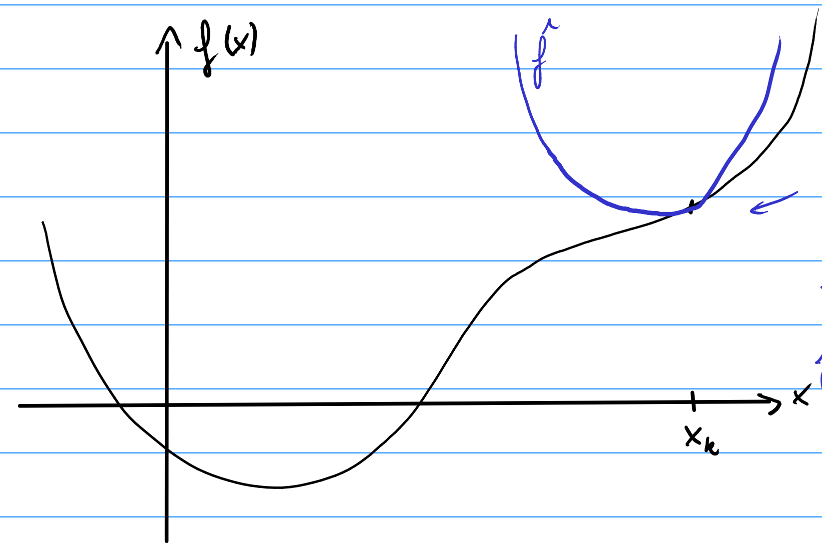
Let's steal the idea from Newton's method for equation solving.

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



$\tilde{f}(x_k) = f(x_k)$

$\tilde{f}'(x_k) = f'(x_k)$

$\tilde{f}''(x_k) = f''(x_k)$

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

from Taylor's theorem

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

Now minimize that.

$$\text{minimize } \tilde{f} \rightarrow \tilde{f}'(x_k + h) = f'(x_k) + f''(x_k) \cdot h = 0$$

w.r.t. h

$$h = -\frac{f'(x_k)}{f''(x_k)}$$

eqn solv.

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

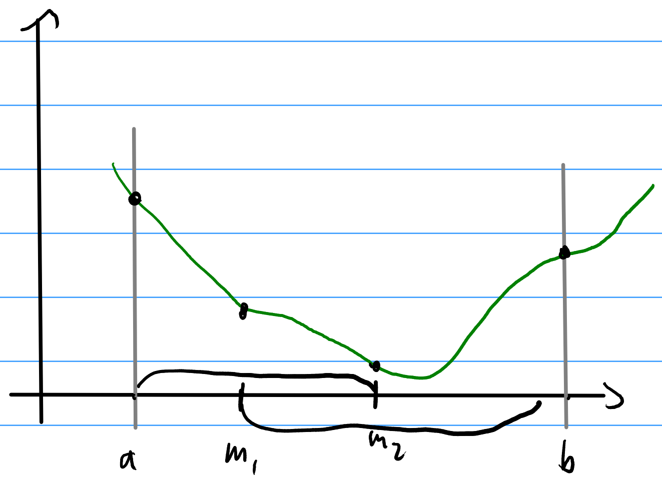
opt

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

Newton for opt is solving f'(x) = 0 → quadratically con

Golden Section Search

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



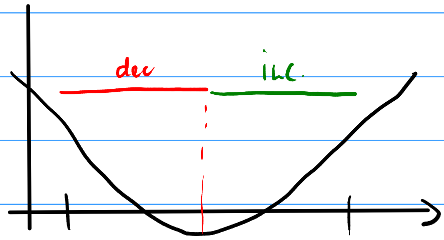
Is one middle point in the bracket good enough?

Next: what condition are we going to maintain throughout?

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

No, so let's promise less and assume more.

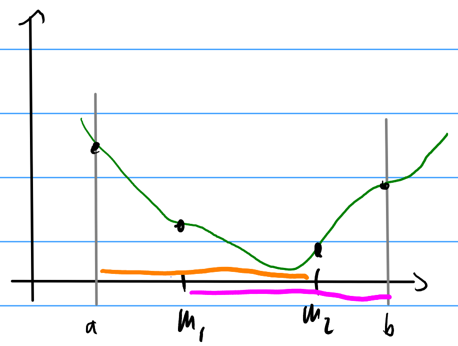
What does it mean for f to be 'unimodal'?



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

No, but the method we derive we'll use anyway.

So how do we maintain unimodality in each bracket?

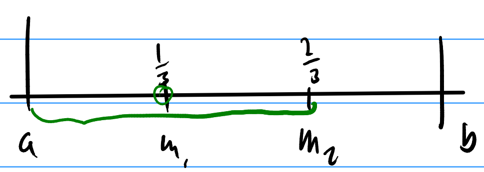


$f(m_1) > f(m_2)$ reduce to $[m_1, b]$

$f(m_1) < f(m_1)$ reduce to $[a, m_2]$

$f(m_1) = f(m_2)$ choose either

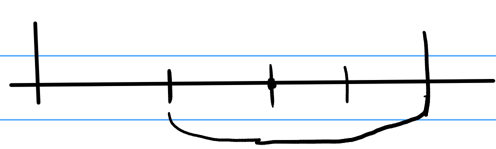
Where do we put the midpoints?



Golden ratio/section

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

$$m_1 = a + \left(1 - \frac{\sqrt{5}-1}{2}\right)(b-a)$$



Golden section search

What's the convergence order of Golden Section Search?

linear

Steepest Descent

What do we do in n dimensions?

Maybe go in the direction of the steepest descent.

What does that mean mathematically?

$$\vec{d} = -\nabla f(\vec{x}) \qquad \vec{x}_{k+1} = \vec{x}_k + \alpha \vec{d}$$

And how far do we go?

Good question -> Leave that to a one-dimensional opt. method

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

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

Search direction $\vec{d} = -\begin{pmatrix} x_0 \\ 5x_1 \end{pmatrix}$

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

linear

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

Newton's method in n dimensions

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

1D: $\tilde{f}(x_k+h) = f(x_k) + f'(x_k)\cdot h + \frac{f''(x_k)}{2}\cdot h^2$

nD: $\tilde{f}(\vec{x}_k+\vec{h}) = f(\vec{x}_k) + \vec{\nabla} f(x_k)\cdot\vec{h} + h^T H_f(x_k)\, h/2$

$$H_f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 x_1} & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f \\ \vdots & & \\ \frac{\partial^2}{\partial x_n \partial x_1} f & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f \end{pmatrix}$$

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

$$\tilde{f}(x_k+\hat{h}) = f(\vec{x}_k) + \vec{\nabla} f(x_k)\cdot\vec{h} + \vec{h}^T H_f(x_k)\, \vec{h}/2$$

$$0 = \nabla_h \tilde{f}(x_k+h) = \dot{\nabla} f(\vec{x}_k) + H_f(x_k)\cdot\vec{h}$$

$$\rightarrow \quad H_f(\vec{x}_k)\cdot\vec{h} = -\nabla f(x_k)$$

$$\vec{h} = -H_f(x_k)^{-1}\cdot\nabla f(x_k)$$

$$\vec{x}_{n+1} = \vec{x}_n - H_f(x_k)^{-1}\cdot\nabla f(x_k)$$

1D: $\frac{f'}{f''}$

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

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

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

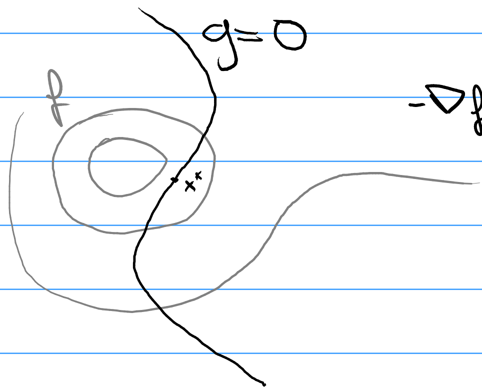
What if we don't even have one derivative, let alone two?!

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 <u>and</u> 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:

Can you do an example?

Minimize $(x-2)^4 + 2(y-1)^2$ subject to $x+4y=3$