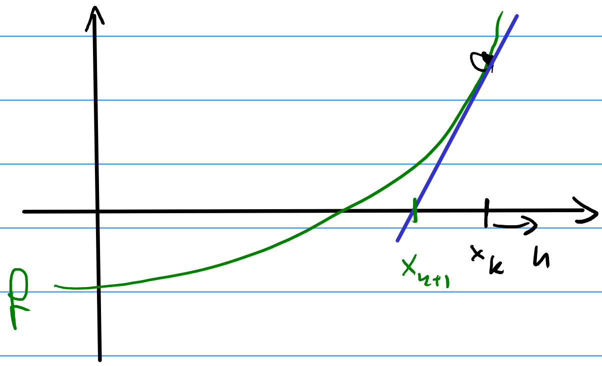
Newton's method

Suppose x_k is our current guess of the zero.



Idea: Build a solvable approximate version of f using $f(x_k)$ and $f'(x_k)$

$$f(x_k + h) = f(x_k) + h \cdot f'(x_k)$$

Find the zero of the approximate version.

Name some downsides of Newton's method.

Does not converge globally (only converges locally, i.e. if starting guess is already near the solution)

Name an upside:

Seems to converge quickly as soon as we're close.

Recap: Rate of convergence--what's that?

$$e_k = \| x_{true} - x_k \|$$

"Linearly" (or order 1 convergent)

$$e_{k+1} \le C \cdot e_k \quad (C < 1)$$

↳ e.g. bisection

Example:

$e_4 = 0.01 \qquad C = 0.1$

$e_5 \le 0.1 \cdot 0.01 = 0.001$

$e_6 \le 0.0001$

Lesson: Gains a constant number of digits per iteration

Quadratically (or order 2 convergent)

$$e_{k+1} \le C \cdot e_k^2$$

Lesson: Doubles number of digits each step

$e_4 = 0.01 \qquad C = 1$

$e_5 = e_4^2 = 0.01 \cdot 0.01 = 0.0001$

$e_6 = 10^{-8}$

Solving systems of nonlinear equations

Want to solve $\vec{f}(\vec{x}) = \vec{0}$. $\qquad f: \mathbb{R}^n \to \mathbb{R}^n$ why? !

Let's try to carry over our 1-dimensional ideas.

Let's first get an idea of what behavior can occur.

Based on the demo: Does bisection stand a chance?

Let's try Newton's method then. What's the linear approximation of $\vec{f}$?

$$\vec{\tilde{f}}(\vec{x}_n + \vec{h}) = \vec{f}(\vec{x}_n) + \underbrace{J_{\vec{f}}(\vec{x}_n)}\vec{h}$$

$$\vec{f} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$$

$$J_f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix}$$

OK, now solve that for h. $\qquad \vec{f}(\vec{x}) = \vec{0} \leftarrow$

$$\vec{0} = \vec{f}(\vec{x}_n) + J_{\vec{f}}(\vec{x}_n)\vec{h} \quad \rightsquigarrow \quad J_f(\vec{x}_n)\vec{h} = -\vec{f}(\vec{x}_n)$$

$$\vec{h} = -J_{\vec{f}}^{-1}(\vec{x}_n)\vec{f}(\vec{x}_n)$$

$$\vec{x}_{n+1} = \vec{x}_n + \vec{h}$$

Let's do an example of that:

$$f(x,y) = \begin{pmatrix} x + 2y - 2 \\ x^2 + 4y^2 - 4 \end{pmatrix} \begin{matrix} - f_1 \\ - f_2 \end{matrix} \qquad \vec{f}(\vec{x}) = \vec{0}$$

$$J_f = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2x & 8y \end{pmatrix}$$

What are the downsides of this method?

So how about (an n-dimensional analog of) the secant method?

12

Optimization

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

Instead minimize $\|g(x)\|_2 = f(\vec{x})$

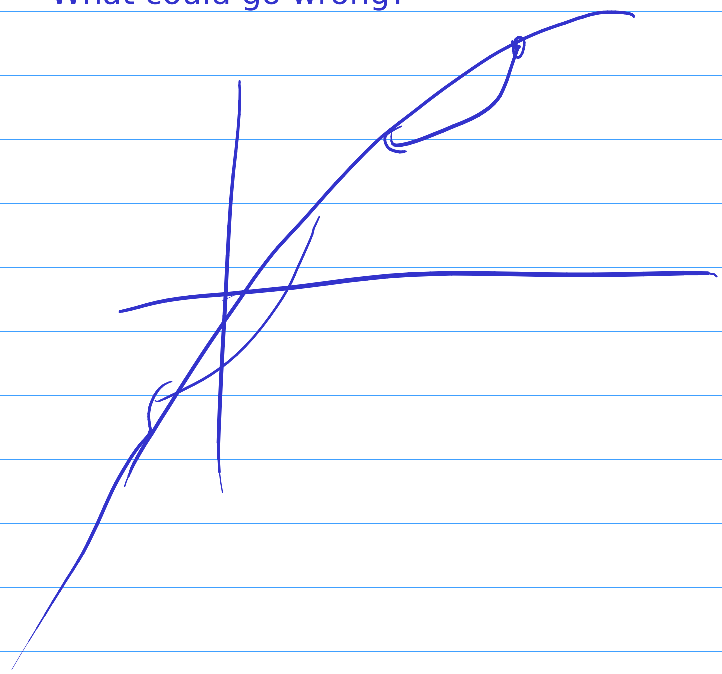
Create a problem statement for "optimization".

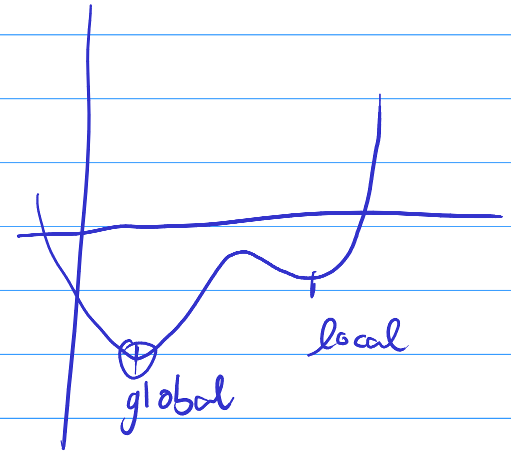
$f: \mathbb{R}^n \to \mathbb{R}$ ← objective function

Find $\vec{x}$ so that $f(\vec{x})$ is minimal.

What if I'm interested in the largest possible value of a function g instead?

What could go wrong?





How can we tell if we've got a (local) minimum in 1D? Remember calculus!

necessary: $f'(x) = 0$ sufficient: $f''(x) > 0$

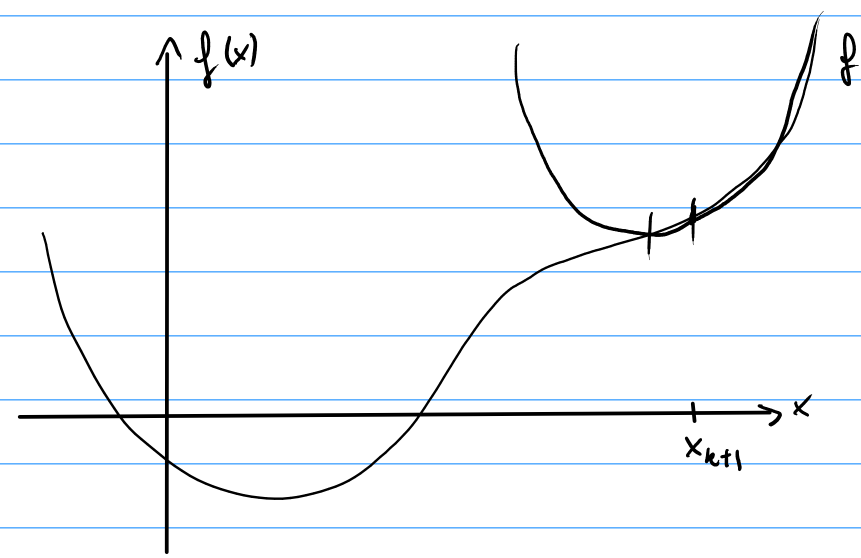
And in n dimensions?

necessary: $\nabla f(x) = 0$ sufficient: $H_f(x)$ is pos. def.

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

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.



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

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

Now minimize that.

$$f(x) = 23x^2 - 4x + 3$$

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

$$\tilde{f}(x_0) = f(x_0)$$

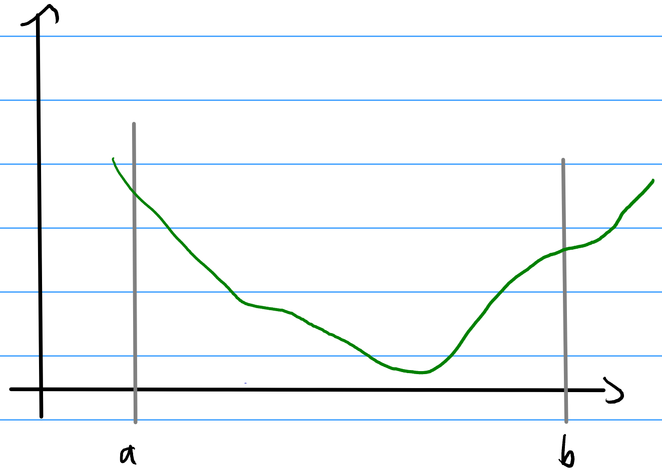
$$\tilde{f}'(x_0) = f'(x_0)$$

$$\tilde{f}''(x_0) = f''(x_0)$$

Does that look at all familiar?

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?

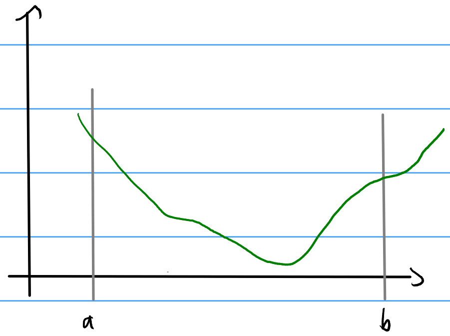
Next: what condition are we going to maintain throughout?

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

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

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

So how do we maintain unimodality in each bracket?



Where do we put the midpoints?

What's the convergence order of Golden Section Search?

Steepest Descent

What do we do in n dimensions?

What does that mean mathematically?

And how far do we go?

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

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

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.

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

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

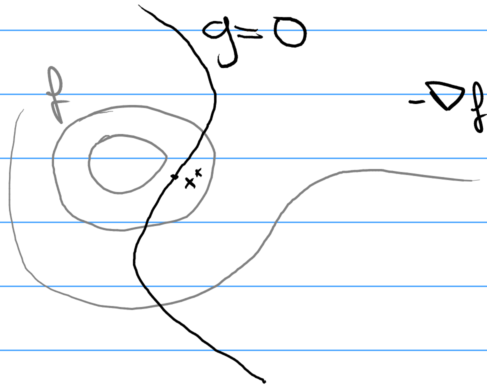
What if we don't even have one derivative, let alone two?!

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: