Today

- nonlin. eq
- optimization

Announcements

- HW8
- 4CH

Fixed Point Iteration

In 1D:

lin. $|g'(x^*)| < 1$

quad. $g'(x^*) = 0$

$f: \mathbb{R}^n \to \mathbb{R}^n$

$$\begin{aligned} x_0 &= \langle \text{starting guess} \rangle \\ x_{k+1} &= g(x_k) \end{aligned}$$

$g: \mathbb{R}^n \to \mathbb{R}^n$

When does this converge?

inputs

$$\partial g(x) = \begin{pmatrix} \partial_{x_1} g_1 & \\ \partial_{x_1} g_2 & \\ & \end{pmatrix}$$

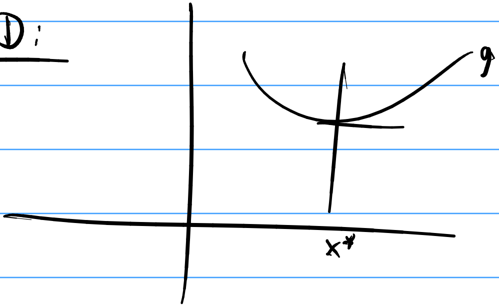
outputs

$\|\partial g\|_p < 1 \Rightarrow \rho(\partial g(x^*)) < 1 \Rightarrow$ lin. conv.

$$J_g(x^*) = 0 \Rightarrow \text{quad. conv.}$$

$$\vec{g}(\vec{x}+\vec{h}) = \vec{g}(\vec{x}) + J_g(\vec{x}) \cdot \vec{h}$$

In 1D:



$$f_1(x_1, x_2) = 0$$
$$f_2(x_1, x_2) = 0$$

Newton's Method

What does Newton's method look like in n dimensions?

$$0 = f(\vec{x}_k + \vec{h}) = f(\vec{x}_k) + J_f(\vec{x}_k)\cdot h + \cancel{O(h^2)}$$

$$-f(x_k) = J_f(\vec{x}_k)\cdot\vec{h}$$

$$-J_f(x_k)^{-1} f(x_k) = \vec{h}$$

$$\vec{x}_{k+1} = \vec{x}_k - J_f^{-1}(x_k)\cdot f(x_k) \quad nD$$

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

Downsides of n-dim. Newton?

- local conv.
- J_f

Demo: Newton's method in n dimensions

Secant in n dimensions?

$x_k = []$ $x_{k-1} = []$

$f(x_k) = []$ $f(x_{k-1}) = []$

What would the secant method look like in n dimensions?

$$\frac{f(x + h\hat{e}_1) - f(x)}{h} \to \text{first column of } J_f$$

$\tilde{J}_0 = I$ $\rightsquigarrow$ with eval $\hat{f}(x_1), f(x_0)$,

update J_1

Broyden's method

Outline

Optimization: Problem Statement

Have: Objective function $f : \mathbb{R}^n \to \mathbb{R}$
Want: Minimizer $\mathbf{x}^* \in \mathbb{R}^n$ so that

$$f(\mathbf{x}^*) = \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{g}(\mathbf{x}) = \mathbf{0} \quad \text{and} \quad \mathbf{h}(\mathbf{x}) \leqslant \mathbf{0}.$$

- $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) \leqslant \mathbf{0}$ are called constraints.
  They define the set of feasible points $\mathbf{x} \in S \subseteq \mathbb{R}^n$.
- If $\mathbf{g}$ or $\mathbf{h}$ are present, this is constrained optimization.
  Otherwise unconstrained optimization.
- If $\mathbf{f}$, $\mathbf{g}$, $\mathbf{h}$ are *linear*, this is called linear programming.
  Otherwise nonlinear programming.

Optimization: Observations

Q: What if we are looking for a *maximizer* not a minimizer? consider $-f$

Give some examples:

- training ANNs
- robot path planning

What about multiple objectives?

- combine them
- look up Pareto optimal

Existence/Uniqueness

Terminology: global minimum / local minimum

Under what conditions on f can we say something about existence/uniqueness?

If $f : S \to \mathbb{R}$ is continuous on a closed and bounded set $S \subseteq \mathbb{R}^n$, then

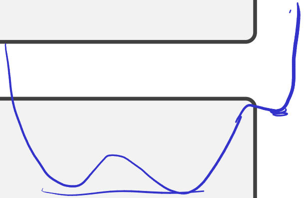
a minimum exists.

$f : S \to \mathbb{R}$ is called *coercive* on $S \subseteq \mathbb{R}^n$ (which must be unbounded) if

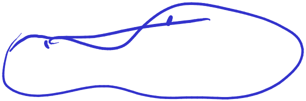
$$\lim_{\|x\| \to \infty} f(x) = +\infty$$

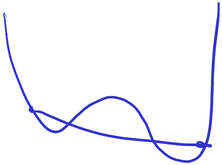
If f is coercive, ...

a global minimum.



Convexity





$S \subseteq \mathbb{R}^n$ is called convex if for all $\mathbf{x}, \mathbf{y} \in S$ and all $0 \leqslant \alpha \leqslant 1$

$$\alpha x + (1-\alpha) y \in S.$$

$f : S \to \mathbb{R}$ is called convex on $S \subseteq \mathbb{R}^n$ if for \ $\mathbf{x}, \mathbf{y} \in S$ and all $0 \leqslant \alpha \leqslant 1$

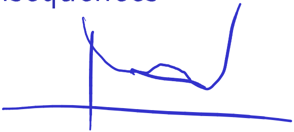
$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

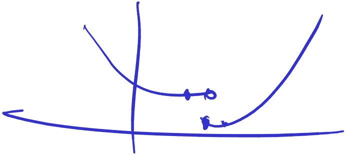
$$<$$

Q: Give an example of a convex, but not strictly convex function.

$\longrightarrow$ likely non-uniqueness of the min.

Convexity: Consequences





If f is convex, . . .

- f is continuous
- any local min is a global min.

If f is strictly convex, . . .

any local min is a unique global min

Optimality Conditions

$f: \mathbb{R}^n \to \mathbb{R} \quad \nabla f = J_f \in \mathbb{R}^{n \times 1}$

If we have found a candidate $\mathbf{x}^*$ for a minimum, how do we know it actually is one? Assume f is smooth, i.e. has all needed derivatives.

In one-d:

Necessary: $f'(x^*) = 0$

Sufficient: $f'(x^*) = 0$ and $f''(x^*) > 0$

$\Rightarrow$ local min

In multiple-d:

necessary: $\nabla f = 0$

sufficient: $\nabla f = \dot{0}$ and $H_f(x^*)$ pos. def.

$\partial_{x_j}\partial_{x_i} f$

symm.

$$H_f = \begin{pmatrix} \partial_{x_1 x_1} f & \partial_{x_1 x_2} f \\ \partial_{x_2 x_1} f & \end{pmatrix}$$

Optimization: Observations

Q: Come up with a hypothetical approach for finding minima.

Solve $\nabla f(x) = 0$ e.g. using Newton

Q: Is the Hessian symmetric?

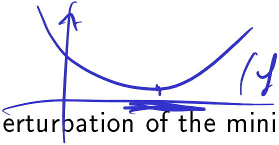
yes

Q: How can we practically test for positive definiteness?

Cholesky.

$x^T A x > 0$

Sensitivity and Conditioning (1D)



$(f(x) - \tilde{f}(x) / < \varepsilon$

How does optimization react to a slight perturbation of the minimum?

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

$$|x - x^*| \le \sqrt{2\varepsilon / f''(x^*)}$$

If $\varepsilon = 10^{-16}$ $\Rightarrow$ tol on solution is 10^{-8}

Sensitivity and Conditioning (nD)

How does optimization react to a slight perturbation of the minimum?