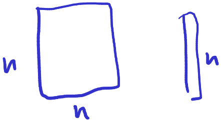
- HW 9
- Examlet 4

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

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



Secant in n dimensions?

What would the secant method look like in n dimensions?

Want: approximate Jacobian

$$\otimes \qquad \tilde{J}_{k+1}(\vec{x}_{k+1} - \vec{x}_k) = \vec{f}(\vec{x}_{k+1}) - \vec{f}(\vec{x}_k)$$

n^2 unknowns

n equations

natural idea: update Jacobian guess $\tilde{J}_{k+1}$ according to $\otimes$

Broyden's method

- minimize $\|\tilde{J}_{k+1} - \tilde{J}_k\|_F$ subject to $\otimes$ — bad Broyden
- minimize $\|\tilde{J}_{k+1}^{-1} - \tilde{J}_k^{-1}\|_F$ subject to $\otimes$ — good Broyden

Numerically Testing Derivatives

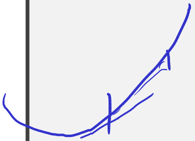
Getting derivatives right is important. How can I test/debug them?

$$T(h) = \left\| \frac{\vec{f}(\vec{x}+h\vec{s}) - \vec{f}(\vec{x})}{h} - J_f(\vec{x})\,\vec{s} \right\| \to 0$$

$\|\vec{s}\|_2 = 1$

have f

trying to come up with



as you vary $h \to 0$, $T = O(h^1)$.

Ideas: • do ↑ term-by-term

• automatic differentiation (AD) → "forward mode", "reverse mode"

Outline

Optimization: Problem Statement

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

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = 0 \quad \text{and} \quad \boldsymbol{h}(\boldsymbol{x}) \leq 0.$$

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

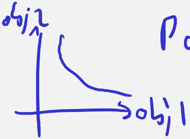
Optimization: Observations

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

Give some examples:

- ML
- Molecular Dynamics
- Solving dual problems (will see this)
- "build a bridge"

What about multiple objectives?



Pareto optimality

two objectives $f_1, f_2 \to f = \alpha f_1 + \beta f_2$

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

there exists a min

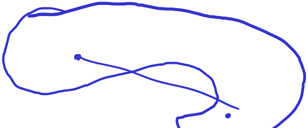
$f : S \to \mathbb{R}$ is called *coercive* on $S \subseteq \mathbb{R}^n$ (~~which must be unbounded~~) if

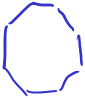
$$\lim_{\|x\| \to +\infty} f(x) \to +\infty$$

If f is coercive, and continuous, and S is closed, $(x, y) : x \geq 0$

a global minimum exists

Convexity





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

$$\alpha \vec{x} + (1-\alpha)\vec{y} \in S$$

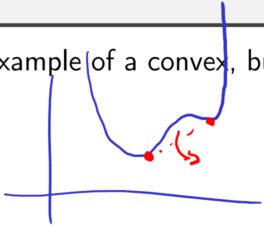
$f : S \to \mathbb{R}$ is called convex on $S \subseteq \mathbb{R}^n$ if for $\boldsymbol{x}, \boldsymbol{y} \in S$ and all $0 \le \alpha \le 1$

$$f(\alpha\vec{x} + (1-\alpha)\vec{y}) \le \alpha f(\vec{x}) + (1-\alpha) f(\vec{y})$$

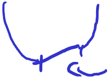
"convex"

$<$ "strictly convex"

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

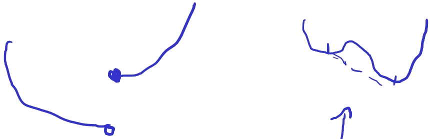


- linear function
-



linear piece

Convexity: Consequences



If f is convex, ...

- f is continuous at interior points.
- any local minimum is a global min

If f is strictly convex, ...

a local min is a unique global min.

Optimality Conditions

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

1D:

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

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

$x \neq 0$: $x^T H x \geq 0$ → PSD

> 0 → PD

nD:

necessary: $\nabla f(x^*) = 0$

sufficient: $\nabla f(x^*) = 0$

Hessian pos def.

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

$$H_f = \begin{bmatrix} \partial_{x_1 x_1} f & \cdots & \partial_{x_1 x_n} f \\ \vdots & & \\ \partial_{x_n x_1} f & \cdots & \end{bmatrix}$$

Optimization: Observations

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

solve $\nabla f = 0$.

Q: Is the Hessian symmetric?

yes, Schwarz's theorem says so.

Q: How can we practically test for positive definiteness?

Cholesky factorize it $H = LL^T$

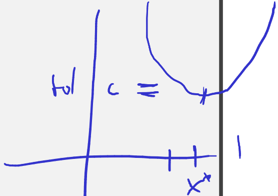
Sensitivity and Conditioning (1D)

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

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

Assume $|f(\tilde{x}) - f(x^*)| < tol.$

$$|x-x^*| < \sqrt{2\,tol / f''(x^*)}$$



common tolerance for $|x-x^*|$ in double prec. 10^{-8} $\sqrt{\varepsilon_{mach}}$

Want to dodge this? Solve $\nabla f = 0$ instead

Sensitivity and Conditioning (nD)

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

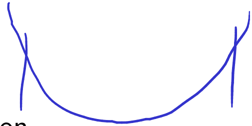
$$\|\vec{s}\|_2 = 1 \qquad \vec{x} = \vec{x}^* + h\vec{s}$$

Assume: $|f(\vec{x}) - f(\vec{x}^*)| < tol.$

$$f(\vec{x}^* + h\vec{s}) = f(\vec{x}^*) + h \cancel{\underbrace{\nabla f(\vec{x}^*)}_{=0}{}^T s} + \frac{h^2}{2}\vec{s}^T H_f(x^*)\vec{s} + O(h^3)$$

$$\|\vec{x} - \vec{x}^*\| \le \sqrt{\frac{2\,tol}{\lambda_{min}(H_f(x^*))}}$$

Unimodality

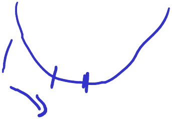


Would like a method like bisection, but for optimization.
In general: No invariant that can be preserved.
Need *extra assumption.*

f is called unimodal on an open interval if there exists an x^* in the interval so that for all $x_1 < x_2$

$$x_2 < x^* \Rightarrow f(x_1) > f(x_2)$$

$$x^* < x_1 \Rightarrow f(x_1) < f(x_2)$$



In-Class Activity: Optimization Theory

In-class activity: Optimization Theory