November 5, 2024

Announcements

Goals

Review

"Jax"

- Automatic differentiation (AD)
- Testing derivatives

$\vec{f}$

∂f

Numerically Testing Derivatives

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

$$\left\| \frac{\vec{f}(\vec{x}+h\vec{s}) - \vec{f}(\vec{x})}{h} - J_f(\vec{x})\,\vec{s} \right\| = O(h) \qquad \|\vec{s}\| = 1$$

$$f(\vec{x}+h\vec{s}) = f(\vec{x}) + h\,J_f(\vec{x})\,\vec{s} + O(h^2)$$

$$\vec{\alpha} = (2,3,1) \leftarrow \text{multi-index}$$

$$\partial^{\vec{\alpha}} f = \partial_x^{\alpha_x} \partial_y^{\alpha_y} \partial_z^{\alpha_z} f$$

$$\vec{h}^{\vec{\alpha}} = h_x^{\alpha_x} h_y^{\alpha_y} h_z^{\alpha_z}$$

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

$$\vec{\alpha}! = \alpha_x! \, \alpha_y! \, \alpha_z!$$

$$|\vec{\alpha}| = \alpha_x + \alpha_y + \alpha_z$$

$$f(\vec{x} + \vec{h}) = f(\vec{x}) + \dots$$

$$= \sum_{|\vec{\alpha}| \leq p} \frac{\partial^{\alpha} f}{\alpha!} \vec{h}^{\alpha} + \dots$$

Outline

Optimization: Problem Statement

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

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

- $\boldsymbol{g}(\boldsymbol{x}) = 0$ and $\boldsymbol{h}(\boldsymbol{x}) \le 0$ 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?

Give some examples:

$f(x) = ?$ $\quad \rightsquigarrow \quad \|f(x) - z\|_2 \to \min$

What about multiple objectives?

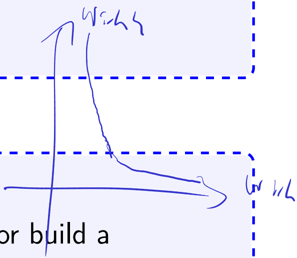
Optimization: Observations

Q: What if we are looking for a *maximizer* not a minimizer?
Give some examples:

- What is the fastest/cheapest/shortest... way to do...?
- Solve a system of equations 'as well as you can' (if no exact solution exists)–similar to what least squares does for linear systems:

$$\min \underbrace{\|F(\boldsymbol{x})\|}_{f(\vec{x})}$$

What about multiple objectives?



- In general: Look up Pareto optimality.
- For 450: Make up your mind–decide on one (or build a combined objective). Then we'll talk.

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

it has a minimum.

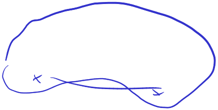
$f : S \to \mathbb{R}$ is called *coercive* on $S \subseteq \mathbb{R}^n$ if

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

If f is coercive and continuous and S is closed, ...

global min im exists.

Convexity



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

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

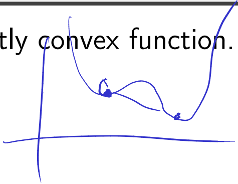
$f : S \to \mathbb{R}$ is called convex on $S \subseteq \mathbb{R}^n$ if for $\boldsymbol{x}, \boldsymbol{y} \in S$ and all $0 \le \alpha \le 1$

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

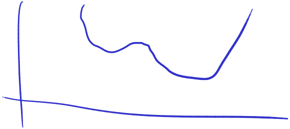
$<$ strictly

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

$<$, not s.c.



Convexity: Consequences



If f is convex, . . .

all local minima
are global minima

If f is strictly convex, . . .

all local minima
are unique global minima

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: $f''(x) > 0 \quad (x \in B(x^*, \varepsilon))$

for local minima

nD:

- necessary: $\nabla f(x^*) = 0$
- sufficient $H_f(x)$ positive semidef $x \in B(x^*, \varepsilon)$

(same)

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

last chapter

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

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

$$H_f = \begin{pmatrix} \partial_{xx} f & \partial_{xy} \\ \partial_{yx} f & \ddots \end{pmatrix}$$

Schwarz's thm: H_f symm.

eigenvector

$$H_f = X \; D \; X^T$$

$$\vec{h}^T H_f \vec{h} = h^T X D X^T \vec{h} = \vec{z}^T D \vec{z}$$

$$\vec{z} = X \vec{h} \qquad = \sum_i d_{ii} \, z_i^2$$

Optimization: Observations

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

Solve $\nabla f \stackrel{?}{=} 0$.

Q: Is the Hessian symmetric?

✓

Q: How can we practically test for positive definiteness?

Cholesky!

$\vec{x}^T A \vec{x} \geq 0$

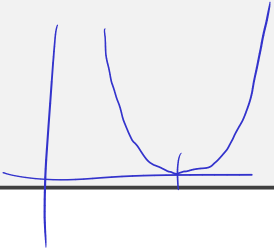
Sensitivity and Conditioning (1D)

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

Suppose $|f(\tilde{x}) - f(x^*)| < \text{tol}$. ($x^*$ is true min)

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

Ignore $O(h^3)$ term and solve for h:



$$|\tilde{x} - x^*| \le \sqrt{\frac{2\,\text{tol}}{f''(x^*)}}$$

wide bowls bad

Sensitivity and Conditioning (nD)

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

$$f(\vec{x}+\vec{h}) = f(\vec{x}) + \cancel{\underbrace{\nabla f}_{=0} \cdot \vec{h}} + \frac{1}{2}\vec{h}^T H_f \vec{h} + O(\|h\|^3)$$

$$\|\vec{h}\|^2 \leq \frac{2tol}{\lambda_{min}(H_f(x^*))}$$

" cond. of Hessian "

Unimodality

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