6 Optimization

Relax $f(x) = 0$?

$\min_x \|f(x)\|$ Do we need the norm?
Do we need the output to be n-dimensional?

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

Want: minimizer

$$\min_x f(x) \quad \text{subject to} \quad \underbrace{g(x) = 0 \quad \text{and} \quad h(x) \leq 0}_{\text{constraints}},$$

define feasible points $x \in S \subseteq \mathbb{R}^n$

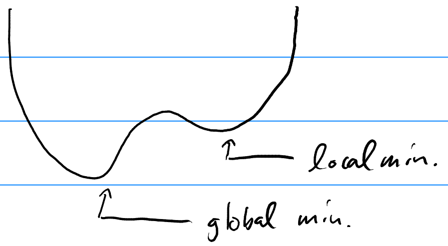
"Optimization" — much more general than our starting point
"What's the cheapest/fastest/shortest way to ...?"

constrained/unconstrained optimization

linear/nonlinear programming (f, g, h)

6.2 Existence / Uniqueness

Need assumptions on f



- $f: S \to \mathbb{R}$ continuous on $S \subseteq \mathbb{R}^n$ closed and bounded

  $\Rightarrow$ has minimum

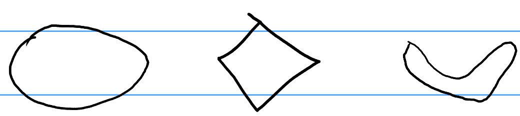
- $f: S \to \mathbb{R}^n$ called *coercive* on $S \subseteq \mathbb{R}^n$ unbounded

  if $\lim_{\|x\| \to \infty} f(x) = +\infty$ What does this mean?

  $\Rightarrow$ f has a global minimum (possibly non-unique — why?)

- $S \subseteq \mathbb{R}^n$ *convex* if for all $0 \le \alpha \le 1$, $x, y \in S$

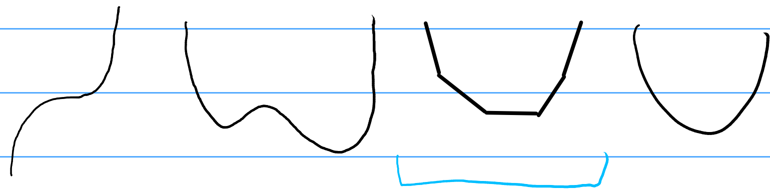
  $\alpha x + (1-\alpha) y \in S$



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

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$ ← convex

$<$ ← strictly convex



O name an example of convex, but not strictly convex

f convex ⇒ f continuous at interior points

f convex ⇒ local minimum is global min

f strictly convex ⇒ local min is unique global min

6.3 Optimality conditions

assume f smooth

1D: necessary $f'(x^*) = 0$ ⇐ x^* extremal pt.

sufficient $f'(x^*) = 0$ $f''(x^*) > 0$ ⇒ x^* loc. minimum

nD: necessary $\nabla f(x^*) = 0$ ⇐ x^* extremal pt.

sufficient $\nabla f(x^*) = 0$ $H_f(x^*)$ pos. def. ⇒ x^* loc. minimum

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

O Hypothetical approach for finding minima?
Symmetric? Test for PD'ness?

WS19 p1

6.4 Sensitivity and Conditioning

1D: $f(x^* + h) = f(x^*) + \underbrace{f'(x^*)}_{=0} h + f''(x^*) \frac{h^2}{2} + O(h^3)$

Now suppose we have approx. min $\tilde{x}$ with

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

Solve above for h:

$$|x - \tilde{x}| = |h| \leq \sqrt{2\, tol / f''(x^*)}$$

IOW: Half as many digits in $\tilde{x}$ as in $f(\tilde{x})$

⚠ Important to keep in mind when setting tolerances

[Can do better only when derivatives known – solve $\nabla f = 0$

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

with $\|s\| = 1$

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

conditioning of H_f determines sensitivity of x^*.

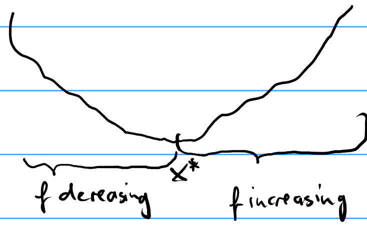
Lec 22

Announcements:

- exam grades posted
- no class Friday
- HW3 due tonight
- HW4/HW3 sol out some time tomorrow

6.5 Methods in 1D (unconstrained)

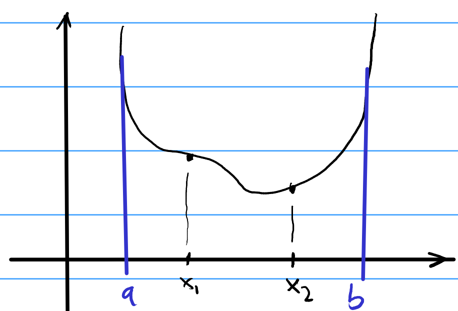
Make new assumption : "unimodal"



f unimodal if for all $x_1 < x_2$

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

Golden Section Search



Suppose we have interval with f unimodal.
Want to maintain this.

- Pick x_1, x_2
- $f(x_1) > f(x_2)$
  reduce to (x_1, b)
- $f(x_1) \leq f(x_2)$
  reduce to (a, x_2)

Where to put x_1, x_2?

- Symmetric $x_1 = a + \tau(b-a)$
  $x_2 = a + (1-\tau)(b-a)$
- Want to reuse function values : $\tau^2 = 1 - \tau$
  $\Rightarrow \tau = \frac{\sqrt{5}-1}{2} \approx .618$

linearly convergent.

Want better convergence!

Idea: Use function approximation like Newton

Newton's method

Is approximation by a line any good?

$$f(x+h) \approx f(x) + f'(x)h + f''(x)\frac{h^2}{2} \quad =: \hat{f}(h)$$

Minimize! $\rightsquigarrow$ $0 = \hat{f}'(h) = f'(x) + f''(x)h$

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

$$\begin{cases} x_0 = \langle \text{starting guess} \rangle \\ x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \end{cases}$$

Notice something?

Identical to Newton for solving $f'(x)=0$!

Locally quadratically convergent (but may fail)

because

Good idea: Combine $\underbrace{\text{slow-and-safe}}_{\text{bracketing}}$ with $\underbrace{\text{fast-and-risky}}_{\text{Newton}}$.

<lec 23>

6.6 Methods in nD (unconstrained)

Steepest descent

- Want to decrease f? What's the direction of steepest descent?

  $-\nabla f$

- How far do you go?

  ? → line search!

$$
\left[\begin{array}{l}
x_0 = \langle \text{starting guess} \rangle \\
\\
s_k = -\nabla f(x_k) \\
\text{Choose } \alpha_k \text{ to minimize } f(x_k + \alpha_k s_k) \\
x_{k+1} = x_k + \alpha_k s_k
\end{array}\right.
$$

Demo

WS 21p1

Newton's method

$$f(x+s) \approx f(x) + \nabla f(x)^T s + \frac{1}{2} s^T H_f(x) s \qquad =: \hat{f}(s)$$

$$\nabla \hat{f}(s) = 0 \qquad \rightsquigarrow \qquad H_f(x) s = -\nabla f(x)$$

$$
\left[\begin{array}{l}
x_0 = \langle \text{starting guess} \rangle \\
\\
\text{Solve } H_f(x_k) s_k = -\nabla f(x_k) \quad \text{for } s_k \\
x_{k+1} = x_k + s_k
\end{array}\right.
$$

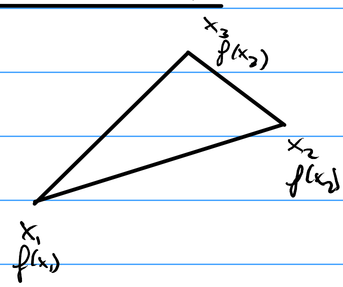
Demo

Drawbacks: → second derivatives (!) (fixed by CG)

→ local conv.

→ works poorly when H_f is near-indefinite

Nelder-Mead



Demo (gif)

$f(x_1) \leq f(x_2) \leq f(x_3)$

<lec 24>

- project posted

6.7 Nonlinear least squares

What if f to be minimized is actually a 2-norm?

$$f(x) = \| r(x) \| \qquad r(x) = y - f(x)$$

Equivalent:

$$\phi(x) = \frac{1}{2} r(x)^T r(x)$$

$$\frac{\partial}{\partial x_i} \phi = \frac{1}{2} \sum_{j=1}^{n} \frac{\partial}{\partial x_i} r_j(x)^2 = \sum_j \left(\frac{\partial}{\partial x_i} r_j \right) r_i$$

$$\nabla \phi = J_r^T(x)\, r(x)$$

just "J" for this section

$$H_\phi = J(x)^T J(x) + \sum_{i=1}^{n} r_i H_{r_i}(x)$$

Newton step: $H_\phi \underbrace{s_k}_{\text{increment}} = -\nabla \phi$

$$\left(J^T J + \sum_i H_{r_i} r_i \right) s_k = -J^T r$$

$\sum_i H_{r_i} r_i$ inconvenient to compute

and: multiplied by $\underbrace{\text{ith residual component}}_{\text{hopefully small}}$

→ forget about it.

Gauss-Newton method

$$J(x_k)^T J(x_k) s_k = -J(x_k)^T r(x_k)$$

o [remind you of something?

$$J(x_k) s_k \cong -r(x_k)$$

o [How do we solve that?

o [What is this method good for?

Demo

o [Newton on its own is only locally convergent.
Gauss-Newton is clearly similar.
Does it have a reason to be extra-bad? → Yes, what if residuals aren't small?

→ Need good starting guess

WS22p1

o [When can't we find a good Newton step?

Levenberg - Marquardt method

$$\left(J(x_k)^T J(x_k) \underbrace{+ \mu_k I}_{\text{omit at first}} \right) s_k = -J(x_k) r(x_k)$$

Corresponding least squares system?

$$\begin{bmatrix} J(x_k) \\ \sqrt{\mu_k} I \end{bmatrix} s_k \cong \begin{bmatrix} -r(x_k) \\ 0 \end{bmatrix}$$

"Regularization" $\rightarrow$ "make H more positive definite"

6.8 Constrained optimization

$\min_x f(x)$ subject to $g(x)=0$ and ~~$h(x) \leq 0$~~ for now

"equality-constrained optimization"

Necessary condition for unconstrained opt.: $\nabla f(x) = 0$

Why? → Otherwise there's a downhill direction.
Applicable here?
Nope, downhill has to be feasible.

s is a feasible direction at x: $x + \alpha s$ feasible for $\alpha \in [0, r]$ (for some r)

Necessary for minimum:

$\nabla f(x) \cdot s \geq 0$ ("uphill that way")

for any feasible direction s

If not at boundary of feasible set: s and $-s$ f.dir.
$\Rightarrow \nabla f = 0$.
⇝ only the boundary is interesting

Necessary at boundary: $(g(x) = 0)$

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

$-\nabla f(x) \in$ row-span J_g

J_g: $m \times n$

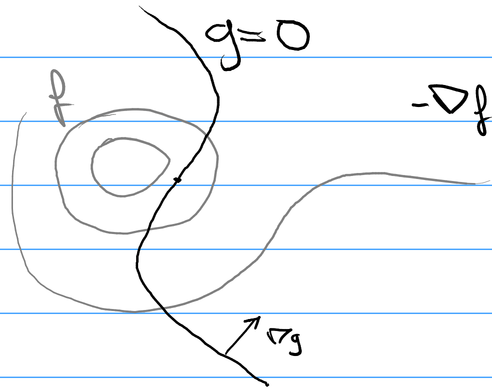
" all descent directions would cause change (→ violation) of constraints"

Also works on interior!

$-\nabla f(x) = J_g^T \lambda$ for some λ

<lec 25>

deadline for hw 4 extended



$$-\nabla f = \nabla g \cdot \lambda = \nabla g^T \lambda$$
$$= J_g^T \lambda \quad \text{for some } \lambda$$

Lagrangian function

$$\mathcal{L}(x, \lambda) := f(x) + \lambda^T g(x)$$

$$\otimes \quad \nabla \mathcal{L} = \begin{pmatrix} \nabla_x \mathcal{L} \\ \nabla_\lambda \mathcal{L} \end{pmatrix} = \begin{pmatrix} \nabla f + J_g(x)^T \lambda \\ g(x) \end{pmatrix} \overset{!}{=} 0$$

necessary condition!

"Lagrange multipliers"

→ use Newton to solve $\otimes$
("sequential quadratic programming")

Inequality-constrained optimization

$\min_x f(x)$ subject to $g(x)=0$ and $h(x)\leq 0$

$g: \mathbb{R}^n \to \mathbb{R}^m$ $h: \mathbb{R}^n \to \mathbb{R}^p$

Idea:

- Define Lagrangian as before.

$$\mathcal{L}(x, \lambda_1, \lambda_2) = f(x) + \lambda_1^T g(x) + \lambda_2^T h(x)$$

- Some ineq. constraints may not be "active"
  called "active" if $h_i(x^*)=0$
  (equality c. is always "active")

- If h_i inactive, must force $\lambda_{2,i}=0$
  ($h_i(x^*)<0$)

  Otherwise: Behavior of h could change location of minimum of $\mathcal{L}$!

- Enforce as $h_i(x^*)\,\lambda_{2,i}=0$ ("complementarity condition")

Assuming J_g and $J_{h,\text{active}}$ has full rank:

$$\left[\begin{array}{ll} \otimes & \nabla_x \mathcal{L}(x^*, \lambda_1^*, \lambda_2^*)=0 \\ \otimes & g(x^*)=0 \\ & h(x^*)\leq 0 \\ & \lambda_2 \geq 0 \\ \otimes & h_i(x^*)\cdot\lambda_{2,i}=0 \quad i=1,\dots,p \end{array}\right]$$

KKT conditions
("Karush-Kuhn-Tucker")

$\leftarrow$ necessary

again: solve $\otimes$ by Newton

WS23p1