- poweriter.m
- learn names:
- steer paper choices

M116 lec 2 ① 3/27/14.

Today: iter. method for EVP, conditioning, stability.

Last time: direct vs iterative
complexity
convergence rate of error to zero vs k (# iters, effort):
types: k^{-p} (ask) algebraic order p
r^k rate $r \in (0,1)$ (ask) geometric / exp. / linear. } (ask which decays faster?)
EVP: Power iteration for dominant eigvec v_1

$r^k = o(k^{-p})$
little-oh meaning?
$f(k) = o(g(k))$
if $\lim_{k\to\infty} \frac{f(k)}{g(k)} = 0$
HW: pf.

recall Power iteration: $x^{(0)}$ random, $x^{(k+1)} = Ax^{(k)}$ then normalize each step.

Showed $x^{(k)} \to v_1$ w/ exp. rate $r = \left|\frac{\lambda_2}{\lambda_1}\right|$

code: poweriter.m, can be terribly slow. (rerun. w/ random A's)

how incr. rate? get other evecs v_m?

$A = V\Lambda V^T$ let $\mu \in \mathbb{R}$,

$(A-\mu I)^{-1} = V \operatorname{diag}\{(\lambda_j - \mu)^{-1}\} V^T$ ← i) why?

ii) is a spectral decomp. means $(A-\mu I)^{-1}$ has a huge eigval if μ near a λ_j.

[gives I when pre- or post-mult. by $A - \mu I = V(\Lambda - \mu I)V^T$ since $V^TV = VV^T = I$.

"Inverse iteration"

Let μ be an estimate for λ_m, $x^{(0)}$ rand. $\in \mathbb{R}^n$
for $k = 1, 2, \ldots$
  solve $(A-\mu I)w = x^{(k-1)}$ for $w \in \mathbb{R}^n$.
  $x^{(k)} = w/\|w\|$
end.

- applies power iteration to $(A-\mu I)^{-1}$
- convergence: let λ_m be closest to μ, λ_s 2nd-closest, then reuse thm on power iter,

Thm: $\|x^{(k)} - (\pm)v_m\| = O\left(\left|\frac{\lambda_m - \mu}{\lambda_s - \mu}\right|^k\right)$

code: poweriter.m pt II, fast.

✱ Estimating λ_m from x, an estimate for v_m? (ask)

Rayleigh quotient $R(x) := \frac{x^TAx}{x^Tx}$

Lord Rayleigh, 1880.

why good?

$x = \sum \alpha_j v_j$ $x^Tx = \sum_{ij} \alpha_i \alpha_j \underbrace{v_i^T v_j}_{\delta_{ij}} = \sum \alpha_j^2$

$x^TAx = \sum_{ij} \alpha_i \alpha_j \underbrace{v_i^TAv_j}_{\lambda_j \delta_{ij}} = \sum \lambda_j \alpha_j^2$

so $R(x) = \frac{\sum \lambda_j \alpha_j^2}{\sum \alpha_j^2} \approx \lambda_m$ if $\begin{cases} \alpha_m \text{ large} \\ \alpha_j \text{ small}, j \neq m \end{cases}$

Thm: say $\|x - v_m\| = \varepsilon$ then $R(x) - \lambda_m = O(\varepsilon^2)$ as $\varepsilon \to 0$.

first:

pf 2: $R(x)$ stationary at $x = v_m$ ie $\vec{\nabla} R = \vec{0}$ (HW: check), and $R(v_m) = \lambda_m$

since $R(x)$ smooth for $x \neq 0$, holds by calculus.

pf 1: $\underbrace{\| \sum \alpha_j v_j - v_m \|^2}_{\text{since o.n.b.: } (\alpha_m - 1)^2 + \sum_{j \neq m} \alpha_j^2} = \varepsilon^2$ so $|\alpha_m - 1| \le \varepsilon$ & $\sum_{j \neq m} \alpha_j^2 \le \varepsilon^2$.

R(x) scale-invariant:
rescale x so $\alpha_m = 1$, then $\sum_{j \neq m} \alpha_j^2 \le \frac{\varepsilon^2}{(1-\varepsilon)^2} = \delta$

$$R(x) = \frac{\lambda_m + \sum_{j \neq m} \lambda_j \alpha_j^2}{1 + \sum_{j \neq m} \alpha_j^2} \le \frac{\lambda_m + \max_j |\lambda_j| \delta}{1 - \delta} \leftarrow \text{worst case}$$

$= \lambda_m (1 + O(\delta))$ using eg $\frac{1}{1-\delta} = 1 + O(\delta)$ etc.

asymp rules

sim for lower bnd.

(show) $\hat{\lambda}^{(k)} = R(x^{(k)})$ in inverse iter code. how many more digits $\hat{\lambda}^{(k)}$ converge than $x^{(k)}$? twice.

Any ideas how to improve inverse iter? use $\mu = R(x^{(k-1)}) = \hat{\lambda}^{(k-1)}$ best avail. eigval.

"Rayl. Quot Iter". (inv iter. update x) $\rightleftarrows$ (Rayl. update $\hat{\lambda}$)

Thm: $\forall x^{(0)}$ except set meas. zero, conv. obeys $\varepsilon_{k+1} = O(\varepsilon_k^3)$

where $\varepsilon_k := \| x^{(k)} - (\pm) v_m \|_2$

or $:= |\hat{\lambda}^{(k)} - \lambda_m|$ w/ const in $O(\cdot)$ uniform for k suff. large.

each iter triples # correct digits!

"cubic" convergence: means $\varepsilon_k \le C \varepsilon_{k-1}^3 \le C' \varepsilon_{k-2}^9$ ---

(ask: faster than exp.?) ie $\varepsilon_k = O(r^{(3^k)})$ $r \in (0,1)$.

- $\exists$ algs that are 'quadratic' conv. $\varepsilon_{k+1} = O(\varepsilon_k^2)$ eg Newton's method finding roots.

Pf sketch: say $\| x^{(k)} - \pm v_m \| = \varepsilon$

then $|\hat{\lambda}^{(k)} - \lambda_m| = O(\varepsilon^2)$

so $\left| \frac{\lambda_m - \mu}{\lambda_J - \mu} \right| = O(\varepsilon^2) \leftarrow$ so 1 step inv. iter mult. error by this

$\Rightarrow \| x^{(k+1)} - \pm v_m \| = \varepsilon \, O(\varepsilon^2)$.

- $\hat{\lambda}^{(k)}, x^{(k)}$ conv. at same rate.

(code) show pt III note: $O(n^3)$ per step but only <5 needed.

BREAK —— Damian v. excited by exploring which λ_j μ settles on (chaotic)

Much better: $\exists O(n^3)$ $\forall \lambda$'s. eg QR iteration Francis paper, builds on this.

Condition # of problem. (as M126 W12; NLA §12)

numerical problem is map $f: X \to Y$ (input space → space of answers)

eg $f(x) = \tan x$ eval. some func.

eg. $f(\{a_0, \dots a_p\}) = \{z_1, \dots z_p\}$ (poly coeffs → roots of poly.)

f 'well-cond.' if infinitesimal pert. δx in x causes 'small' pert. $\delta f := f(x+\delta x) - f(x)$

Defn Abs. cond. # $\hat{\kappa} = \hat{\kappa}(x) := \lim_{\delta \to 0} \sup_{\|\delta x\| \le \delta} \frac{\|\delta f\|}{\|\delta x\|} \overset{\text{abbrev}}{=} \sup_{\delta x} \frac{\|\delta f\|}{\|\delta x\|}$

"norm of Frechet derivative" "Lipschitz const" at point x. ??? e.g. 2-norms.

if $x \in \mathbb{C}^n$, $f(x) \in \mathbb{C}^m$ vectors, $\frac{\partial f_i}{\partial x_j} = J_{ij}(x)$ matrix $J \in \mathbb{C}^{m\times n}$. Jacobean.

If f smooth then $\delta f \approx J\,\delta x$ as $\|\delta x\| \to 0$. locally linear.

Mult. by given matrix J, what is largest factor by which length of vector can grow?

"max growth factor" $= \|J\|_2$ or $\|J\|$ 2-norm of matrix. $= \sup_{0 \ne x \in \mathbb{C}^n} \frac{\|Jx\|}{\|x\|} = \sup_{\|x\|=1} \|Jx\|$ = smallest C st. $\|Jx\| \le C\|x\|$ $\forall x$

Properties. (give as exer.)

- say $A = \mathrm{diag}\{a_1, \dots a_n\} \in \mathbb{R}^{n\times n}$ what is $\|A\|$? $= \max_j |a_j|$
- say $A = uv^T$ $u \in \mathbb{R}^m, v \in \mathbb{R}^n$ (was tricky) $\in \mathbb{R}^{m\times n}$, a rank-1 matrix. $\|Ax\| = \underbrace{|v^T x|}_{\le \|v\|\|x\| \text{ by what? C-S.}} \|u\|$

  so $\|A\| = \|u\|\|v\|$.
- $\|AB\| \le \|A\|\,\|B\|$ submultiplicative. since $\|ABx\| \le \|A\|\,\|Bx\| \le \|A\|\|B\|\|x\|$ is this C-S? No! Defn of norm.
- square A can have all $\lambda_j = 1$ but huge $\|A\|$, give example.

So for vector func $f: \mathbb{C}^m \to \mathbb{C}^n$, abs cond # $\hat{\kappa}(x) = \|J(x)\|_2$

got here!

more useful: Relative cond # $\kappa := \sup_{\delta x} \frac{\|\delta f\|/\|f\|}{\|\delta x\|/\|x\|} = \frac{\|J(x)\|}{\|f\|/\|x\|}$

rel. change on inputs & output: important since computer introduces relative errors.

say $\kappa \lesssim 10^3$ well-cond.

otherwise ill-cond. $\gg 10^3$.

κ is property of problem not alg. used to solve it.