Legend: Def, **Thm**, Key ideas, Cor

5.1 Power Method

Inner product: $\langle x, y\rangle := \sum_1^n x_i \overline{y_i} = y^* x$.

Induced norm: $\|x\| := \sqrt{\langle x, x\rangle}$.

L_p norm: $\|x\|_p := \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$.

Power Method

Requirement: single λ_1 with maximum modulus and n L.I. eigenvectors.

Process:

(1) Begin with nonzero $x^{(0)} = \sum_{i=1}^n a_i e^{(i)}$

(2) Iterate with $x^{(k)} = Ax^{(k-1)} = A^k x^{(0)}$

Properties:

(1) $x^{(k)} = A^{(k)} x^{(0)} = \lambda_1^k (a_1 e^{(1)} + \epsilon^k)$

(2) *** $x^{(k)}/\lambda_1^k \to a_1 e^{(1)}$

(3) *** $\lim x^{(k+1)}/x^{(k)} = \lambda_1$

Aitken Acceleration

Idea: if $\{r_n\} \to r$ then we can build a new sequence that converges faster.

Aitken Acceleration: the sequence $\{s_n\}$ as defined below converges to r faster:

$$s_n := \frac{r_n r_{n+2} - r_{n+1}^2}{r_{n+2} - 2r_{n+1} + r_n}$$

(i.e., $\lim (s_n - r)/(r_n - r) = 0$).

Warning: Aitken accel. must be stopped once we hit a stationary value (computeres are bad at small number subtraction).

Inverse Power Method

Idea: if A is invertible and has a single λ_n with minimum modulus then we can apply the power method to A^{-1}. (Inverse matrix has reciprocal eigenvalues.) The iteration step is done by $Ax^{(k+1)} = x^{(k)}$ (Gaussian elimination instead of computing A^{-1}).

Usage: computes smallest eigenvalue of A.

Other Variants

(1) Shifted power method: uses $A - \mu I$ and iterates $x^{(k+1)} = (A - \mu I)x^{(k)}$; computes eigenvalue farthest from μ.

(2) Shifted inverse power method: uses $(A - \mu I)x^{(k+1)} = x^{(k)}$; computes eigenvalue of A closest to μ.

(3) Requirement: if for e.g. we want to get λ_3 by prescribing a close enough μ to $|\lambda_3|$, we need $|\lambda_2| > |\lambda_3| > |\lambda_4| \geqslant \ldots$

5.2 Schurs & Gershgorin

Localizing Eigenvalues

Gershgorin: all eigenvalues of $A_{n\times n}$ are contained in the union of D_i, where

$$D_i := \{z \in \mathbb{C} : |z - a_{i,i}| \leqslant \sum_{j\neq i} |a_{i,j}|\}.$$

Generalized Gershgorin: if $P^{-1}AP$ diagonalizes A and B is any matrix, then the eigenvalues of $A + B$ lie in the union of D_i's:

$$D_i := \{z \in \mathbb{C} : |z - \lambda_i| \leqslant \kappa_\infty(P)\|B\|_\infty\}$$

where $\lambda_i \in \Lambda(A)$, $\kappa_\infty(P) = \|P\|_\infty \|P^{-1}\|_\infty$.

(*If A is diagonal then $P = I$. If in addition B has zero diagonal then this special case gives Gershgorin's theorem.*)

Unitary Matrices

Unitary matrix: $UU^* = I$.

Lemmas:

(1) $I - vv^*$ is unitary iff $\|v\|_2^2 = 2$ or 0. *Proof: expand $(I - vv^*)^*(I - vv^*)$ and use that $(vv^*) = vv^*$.*

(2) If $\|x\|_2 = \|y\|_2$ and $\langle x, y\rangle \in \mathbb{R}$ then for some $(I - vv^*)$ we have $(I - vv^*)x = y$. *Proof: let $v = \sqrt{2}(x - y)/\|x - y\|_2$.*

Schur's Factorization

Schur: every square matrix is unitarily similar to a triangular matrix, i.e., any A satisfies $A = U^{-1}BU$ for some unitary U and triangular B.

Corollary: every Hermitian matrix is unitarily similar to a diagonal matrix. *Indeed, if*

$$(UAU^*)^* = UA^*U^*$$

then UAU^* is upper and lower triangular.

5.3 Least-Squares

Orthogonal (set): $\langle v_i, v_j\rangle = 0$ for any $v_i, v_j \in \{v_1, \ldots, v_n\}$.

Orthonormal (set): $\langle v_i, v_j\rangle = \delta_{i,j}$ (1 or 0).

Generalized Pythagorean: if $\langle x, y\rangle = 0$ then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Gram-Schmidt

Requirement: start with a set of L.I. vectors $\{x_1, \ldots, x_n\}$. Goal: get orthonormal vectors.

Process:

(1) Set $u_1 := x_1/\|x_1\|$.

(2) Inductively, $u_k' := x_k - \sum_{i<k} \langle x_k, u_i\rangle u_i$.

(3) Normalization: $u_k := u_k'/\|u_k'\|$.

Corollary: the finite truncation $\{u_1, \ldots, u_k\}$ is an orthonormal basis for $\operatorname{span}\{x_1, \ldots, x_k\}$.

G-S Factorization: applying G-S to $A_{m\times n}$, we obtain $A = BT$ where $B_{m\times n}$ has orthonormal columns and T upper triangular with positive diagonal.

Modified G-S: auto-normalization:

$$u_k := x_k - \sum_{i<k} \frac{\langle x_k, u_i\rangle u_i}{\langle u_i, u_i\rangle}.$$

Least-Squares Problem

Idea: $Ax = b$ may or may not have a solution. If not, try to mnimize $\|b - Ax\|_2$.

Least-Squares: if $A^*(Ax - b) = 0$ then x solves the least-squares problem:

$$\|b - Ay\|_2^2 = \|b - Ax + A(x - y)\|_2^2 = \|b - Ax\|_2^2 + \|A(x - y)\|_2^2.$$

By assumption $b - Ax$ is orthogonal to the column space of A, in which there's $A(x - y)$.

Corollary: if $A = BT$ (B orthogonal, T triangular), then the least-squares solution is

$$Tx = (B^*B)^{-1}B^*b.$$

5.4 SVD & Pseudoinverses

SVD: any $A_{m\times n}$ can be factorized into

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V_{n\times n}^T$$

where U, V^T are unitary and D diagonal.

Pseudoinverse: if $A = U\Sigma V^T$ then the pseudoinverse A^+ is $A^+ = V\Sigma^+ U^T$ where Σ^+ takes reciprocal of nonzero diagonal entries.

Penrose: for $A_{m\times n}$, there exists at most one X satisfying (1) $AXA = A$, (2) $XAX = X$, (3) $(AX)^* = AX$, and (4) $(XA)^* = XA$ at the same time.

Corollary: A^+ is a (therefore *the*) matrix satisfying the Penrose properties.

More corollaries: let $A = U\Sigma V^T$.

(1) If Σ has r nonzero entries then $\text{rank}(A) = r$.

(2) $\{v_1, ..., v_r\}$ is an orthonormal basis for $\text{range}(A)$.

(3) $\{u_{r+1}, ..., u_n\}$ is an orthonormal basis for $\text{null}(A)$.

(4) $\|A\|_2 = \max|\sigma_i|$.

Minimal Solutions

Idea: further generalization of least-squares solution. Let $A_{m\times n}x = b$ be given. This system is consistent if there's a solution. Now we define the minimal solution:

(1) If the system is consistent and has a unique solution then it is the minimal solution.

(2) If consistent + a set of solutions, then the minimal solution is the one with the least Euclidean norm.

(3) If inconsistent + unique least-squared solution then it is the minimal solution.

(4) If inconsistent + a set of least-squared solution, then take the one with least Euclidean norm.

Corollary. $\rho := \inf\{\|Ax - b\|_2 : x \in \mathbb{C}^n\}$ can be obtained. Furthermore, among all that obtain this infimum, there exists an x that minimizes $\|x\|_2$ (i.e., minimal solutions always exist).

Minimal solution: the minimal solution of $Ax = b$ is given by $x = A^+b$.

6.1 Poly. Interpolation

Interpolation: given $(x_0, y_0), ..., (x_n, y_n)$ a set of $n+1$ nodes, construct a degree $\leqslant n$ polynomial p with $p(x_i) = y_i$.

Newton Form

Interpolation (Newton): there exists a unique polynomial of degree $\leqslant n$ that interpolates $(x_0, y_0), ..., (x_n, y_n)$ where the x_i's are distinct.

Proof sketch: induction. $p_0(x_0) = y_0$ *and*

$$p_k(x) := p_{k-1}(x) + c\prod_{i=0}^{k-1}(x - x_i)$$

where c is given by (to satisfy $p_k(x_k) = y_k$)

$$c := \frac{y_k - p_{k-1}(x_k)}{(x_k - x_0)...(x_k - x_{k-1})}.$$

Lagrange Form

Idea: (easy for us but hard for computers)

$$p_n(x) := y_0\ell_0(x) + ... + y_n\ell_n(x) = \sum_{i=0}^{n} y_i\ell_i(x).$$

Here $\ell_i(x_j) = \delta_{i,j}$. Because of this, ℓ can be characterized by

$$\ell_i(x) = c\prod_{j\neq i}(x - x_j)$$

and the condition $\ell_i(x_i) = 1$ demands

$$c = \prod_{j\neq i}(x_i - x_j)^{-1} \implies \ell_i(x) = \prod_{j\neq i}\frac{x - x_j}{x_i - x_j}.$$

The ℓ's are called the cardinal functions.

Vandermonde Matrix

Idea: we want to find

$$p_n(x) := a_0 + a_1x + ...a_nx^n = \sum_{i=0}^{n} a_ix^i$$

that interpolates the data.

$$\begin{bmatrix} 1 & x_0 & \cdots & x_0^n \\ 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^n \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

The matrix on the left is a Vandermonde matrix. It is invertible if x_i's are distinct.

Warning. This matrix is often ill-conditioned if some $|x_i| > 1$. Thus, it is not ideal for computer computations.

Errors & Chebyshev

Errors. If $f \in C^{n+1}[a,b]$ and p of degree $\leqslant n$ interpolates f at $n+1$ points. Then for each $x \in [a,b]$ there exists $\xi_x \in [a,b]$ with

$$f(x) - p(x) = \frac{1}{(n+1)!}f^{(n+1)}(\xi_x)\prod_{i=0}^{n}(x - x_i).$$

Chebyshev polynomials: define iteratively

$$\begin{cases} T_0(z) = 1 \quad T_1(x) = x \\ T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \end{cases}$$

Properties:

(1) $T_n(x) = \cos(n\cos^{-1}(x))$ for $x \in [-1,1]$

(2) $|T_n(x) \leqslant 1|$ for $x \in [-1,1]$.

(3) $T_n(\cos(k\pi/n)) = \cos(k\pi) = (-1)^k$ for $0 \leqslant k \leqslant n$.

(4) T_n is a degree-n poly with leading term $2^{n-1}x^n$, so $2^{1-n}T_n$ is monic.

(5) If p is monic with $\deg(p) = n$ then $\|p\|_\infty = \sup_{-1\leqslant x\leqslant 1}|p(x)| \geqslant 2^{1-n} = \|2^{1-n}T_n\|_\infty$.

Corollary. From above, the last term in the "error" theorem attains minimum if the polynomial is $2^{-n}T_{n+1}$. If so, the nodes are

$$x_i = \cos\left(\frac{(2k+1)\pi}{2k+2}\right) \qquad 0 \leqslant k \leqslant n.$$

6.2 Divided Differences

Idea: design a specific algorithm to obtain the coefficients for the Newton form.

Process: for simplicity consider

$$p_2(x) = c_0q_0(x) + c_1q_1(x) + c_2q_2(x)$$

where $(x_0, y_0), (x_1, y_1)$, and (x_2, y_2) are given. By definition $q_0(x) = 1, q_1(x) = x - x_0$, and $q_2(x) = (x - x_0)(x - x_1)$.

(1) Solve $p_0(x) = c_0q_0(x)$. $q_0 = y_0 =: f[x_0]$.

(2) Solve $p_1(x) = y_0 + c_1(x - x_0)(x - x_1)$. This gives
$$c_1 = \frac{y_1 - y_0}{x_1 - x_0} =: f[x_0, x_1].$$

(3) Finally, solve for p_2 and get
$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}.$$

Divided differences: in Newton form, if $p_n(x) = \sum_{i=0}^{n} c_iq_i(x)$ then

$$c_i = f[x_0, ..., x_i]$$

and these are called the divided differences.

Corollary. Divided differences are symmetric. If $\{x_0, ..., x_n\} = \{z_0, ..., z_n\}$,

$$f[x_0, ..., x_n] = f[z_0, ..., z_n].$$

Recursive divided difference:

$$f[x_0, ..., x_n] = \frac{f[x_1, ..., x_n] - f[x_0, ..., x_{n-1}]}{x_n - x_0}.$$

This gives a systematic way to compute all the divided differences. Once we obtain $f[x_i]$ from step (1) above, we can repeatedly use this theorem to compute more.

Corollary. If p is a polynomial of degree $\leqslant n$ that interpolates f on $x_0, ..., x_n$, then for a different point t,

$$f(t) - p(t) = f[x_0, ..., x_n, t]\prod_{i=0}^{n}(t - x_i).$$