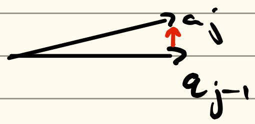
Last time

QR decomposition via Gram-Schmidt

$\Rightarrow$ works well if columns of A are "nearly" orthogonal

$\Rightarrow$ otherwise after orthogonalization

$\|q_j\| \approx 0$

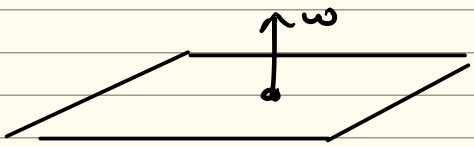


Householder reflections

Orthogonalization by reflection through hyperplanes

__Recall__: hyperplane is defined by an orthogonal vector $w \in \mathbb{R}^n$

all point $\{x \in \mathbb{R}^n : w^T x = 0\}$



Reflection through the hyperplane given by

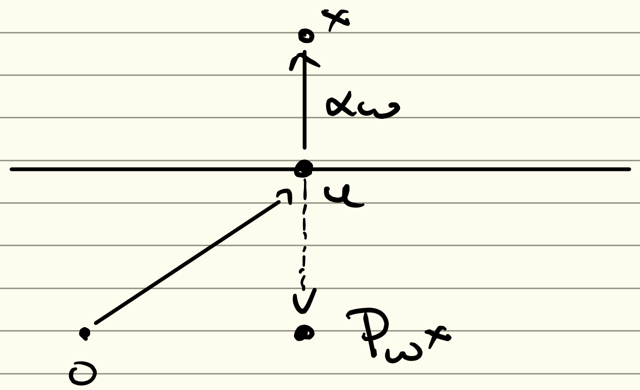
$$\boxed{P_w = I_n - \frac{2}{\|w\|_2^2} w w^T}$$

<u>Claim</u> : P_w is orthonormal $P_w^T = P_w$ ①
$P_w^2 = I$

For $x = \alpha w + u$ where $u \perp w$ ②
$\Rightarrow P_w x = -\alpha w + u$

① can be checked directly

②



We do not compute P_w explicitly as

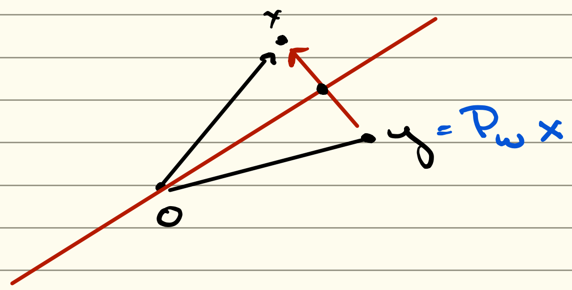
$$P_w x = x - \frac{2}{\|w\|_2^2} (w^T x) w$$

Cost : $4m + O(1)$ $\|w\|_2^2$ is only computed once

<u>Claim</u> : $x, y \in \mathbb{R}^m$ w/ $\|x\|_2 = \|y\|_2$ then for $w = x - y$

$P_w x = y$

<u>Observe</u> : $w = x - y \iff P_w x = y$



(x & y lie on the same sphere)

Hence we can map $x \mapsto \|x\|_2 e_1$ & $-\|x\|_2 e_1$
w/ Householder reflections

Lemma: $\|w_{\mp}\|_2 = 2\|x\|_2^2(\|x\|_2 \mp x_1)$

(choose so $\|x\|_2 \mp x_1$ is larger so

$-x_1$ if $x_1 < 0$; x_1 if $x_1 > 0$

Calculating QR via Householder

Change A into R w/ left $(m-1)$ Householder reflections

① $H_1 A \Rightarrow$ first column to $\alpha_1 e_1$

② 2^{nd} column of $H_1 A$ to $\alpha_2 e_2$

$\tilde{H}_2 = 1 \oplus H_2$

③ 3^{rd} column of $\tilde{H}_2 H_1 A$ to $\alpha_3 e_3$

$\hat{H}_3 = I_2 \oplus H_3$

$\vdots$

Cost $\quad R \approx 2nm^2 - \frac{2}{3}n^3$

Q: additional $4mn^2 - 2nm^2$

$$A = \begin{pmatrix} x & \cdots & x \\ \vdots & & \vdots \\ x & \cdots & x \end{pmatrix} \to H_1 \to$$

$$A_1 = \begin{pmatrix} x & x & \cdots & x \\ 0 & \vdots & & \vdots \\ \vdots & & & \\ 0 & x & \cdots & x \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & H_2 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} x & x & x & \cdots & x \\ 0 & x & \vdots & & \vdots \\ 0 & 0 & & & \\ \vdots & \vdots & & & \\ 0 & 0 & x & \cdots & x \end{pmatrix} \to \begin{pmatrix} I_2 & 0 \\ 0 & H_3 \end{pmatrix} \to A_3 = \begin{pmatrix} x & x & x & x & \cdots & x \\ 0 & x & x & \vdots & & \vdots \\ \vdots & 0 & x & \vdots & & \vdots \\ \vdots & & 0 & & & \\ 0 & \cdots & 0 & x & \cdots & x \end{pmatrix}$$

$$\Downarrow$$

$$\begin{pmatrix} I_3 & 0 \\ 0 & H_4 \end{pmatrix}$$

Finding Eigenvalues

$$\boxed{Ax = \lambda x}$$

Recall

For a square matrix $A \in \mathbb{R}^{n \times n}$, λ is an eigenvalue (w/ $\lambda \in \mathbb{C}$) if there exists a vector $v \in \mathbb{R}^n$ st.

$$Av = \lambda v$$

A & B are said to be similar if there exists an invertible matrix $S \in \mathbb{R}^{n \times n}$ s.t.

$$\underline{B = SAS^{-1}}$$

Claim: the eigenvalues of A are the zeros of the characteristic polynomial

$$p_A(x) = \det(A - \lambda I)$$

Pf: $Av = \lambda v \Leftrightarrow v \in \ker(A - \lambda I)$

hence $0 \in \det(A - \lambda I)$

If $Av = \lambda v \Rightarrow (SAS^{-1})(Sv) = SAv = S\lambda v$

$$= \lambda Sv$$

$$\therefore B(Sv) = \lambda Sv$$

Distinguish between geometric $g_A(\lambda)$ & algebraic $a_A(\lambda)$

multiplicities of λ

geometric $\Rightarrow \ker(A-\lambda I)$

algebraic $\Rightarrow$ degree of λ as a zero of $p_A(\lambda)$

$$\boxed{1 = g_A(\lambda) \le a_A(\lambda)}$$

Similarity of matrices

Example

$A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ has an eigenvalue of 2 where

$$g_A(2) = 1$$

$$a_A(2) = 2$$

Lemma: If A & B are similar they have the same eigenvalues & Sv is an eigenvector of B

Pf:

$$\det(A - \lambda I) = \det(S^{-1}S(A-\lambda I))$$

$$= \det(S(A-\lambda I)S^{-1}) \quad \text{since } \det(XY) = \det(YX)$$

$$= \det(S^{-1}AS - \lambda I)$$

$$= \det(B - \lambda I)$$

Lemma if $A \in \mathbb{R}^{n \times n}$ has unique eigenvalues $\lambda_1 \dots \lambda_n$ then the eigenvectors span $\mathbb{R}^n$

Diagonalization

If we express x in terms of eigenvectors v_i

$$x = \sum_{i=1}^{n} \beta_i v_i$$

then

$$Ax = A\left(\sum_i \beta_i v_i\right) = \sum_i \beta_i A v_i = \sum_i \beta_i \lambda_i v_i$$

So A behaves like a diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

; $A = SDS^{-1}$ where $S = [v_1, \dots, v_n] \in \mathbb{R}^{n \times n}$

Jordan & Schur Form

If we do not have n unique eigenvalues we can write its Jordan form $J(A)$

ie. a matrix can be written as a direct sum of Jordan blocks

$$J(\lambda) = \begin{pmatrix} \lambda & 1 & & & 0 \\ & \lambda & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & \lambda \end{pmatrix}$$

Jordan form is not numerically stable (instead we use Schur form)

Claim: $A \in \mathbb{R}^{n\times n}$ then there exists Schur form

1. Unitary matrix $Q \in \mathbb{C}^{n\times n}$ $(QQ^* = I_n)$ so that

   Q^*AQ is upper triangular

2. Orthogonal matrix $U \in \mathbb{R}^{n\times n}$ $(UU^T = I_n)$

   so that U^TAU is a block upper diagonal matrix (where blocks are 2x2 or 1x1)

Pf: See extra handout

Finding eigenvalues

- power method: largest ei euvalue (in absolute value)
- inverse iteration: eigenvector for eigenvalue closest to $\sigma \in \mathbb{R}$

- orthogonal iteration: find whole subspace for top p-eigenvalue

- QR iteration: variant of orthogonal iteration

- orthogonal iteration: find whole subspace for eigenvalue λ

Finding Dominant Eigenvalues

Def: λ_1 is dominant if

$$|\lambda_1| > |\lambda_i| \quad \text{for } i = 2 \dots n$$

Often we only care about the dominant eigenvalue; e.g. PageRank
MCMC

Power Method

$i = 0 \quad \hat{\lambda}_0 = x_0^T A x_0$

while $\|A x_i - \hat{\lambda}_i x_i\| > tol$:

$$y_{i+1} = A x_i$$

$$x_{i+1} = \frac{y_{i+1}}{\|y_{i+1}\|_2}$$

$$\hat{\lambda}_{i+1} = x_{i+1}^T A x_{i+1}$$

$i+1$

Claim: The power method converges to λ_1 & the speed of convergence depends on

$$\frac{|\lambda_2|}{|\lambda_1|}$$

Pf: Write x_0 in terms of v_i

$$A^k x_0 = A^k\left(\sum_i \beta_i v_i\right) = A^{k-1}\left(\sum_i \beta_i \lambda_i v_i\right)$$

$$= A^{k-2}\left(\sum_i \beta_i \lambda_i^2 v_i\right) =$$

$$\vdots$$

$$= \sum_i \beta_i \lambda_i^k v_i = \lambda_1^k\left(\beta_1 v_1 + \sum_{i=2}^n \left(\frac{\lambda_i}{\lambda_1}\right)^k \beta_i v_i\right)$$

$$\frac{A x_{k-1}}{\|A x_{k-1}\|} = \frac{A^k x_0}{\|A^k x_0\|_2} = \frac{\beta_1 v_1 + \sum_{i=2}^n \left(\frac{\lambda_i}{\lambda_1}\right)^k \beta_i v_i}{\left\|\beta_1 v_1 + \sum_{i=2}^n \left(\frac{\lambda_i}{\lambda_1}\right)^k \beta_i v_i\right\|_2}$$

Since $|\lambda_i| < |\lambda_1|$ for $i > 1$ $\left(\frac{\lambda_i}{\lambda_1}\right)^k \to 0$

PageRank

Google's initial search algorithm

Before: Primarily based on keywords

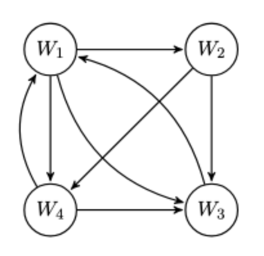
Idea: Represent the internet as a directed graph
Each webpage $\Rightarrow$ vertex
if page i links to page j $\Rightarrow$ make an edge $i \to j$

Total # of links out of page $i = T_i$

Adjacency matrix H:

$$H_{ij} = \begin{cases} \frac{1}{T_i} & \text{if there is a link } i \to j \\ 0 & \text{else} \end{cases}$$

Example



$$H = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}$$

The weight of each $x_k = \sum_{i \in I_k} \frac{x_i}{T_i}$ ①

(weight coming into v_k)

We can rewrite ① as $x = Hx$

i.e. find the eigenvector of H corresponding to $\lambda(H) = 1$

Q. Does it always exist?

A. Not necessarily (but it does with a small change)

Definitions

$A \in \mathbb{R}^{n \times n}$

① is positive if $a_{ij} \geq 0 \ \forall\, i, j$

② reducable if $\exists$ a permutation of row/columns into upper triangular form

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

③ Row stochastic $\sum_j a_{ij} = 1 \quad \forall i$

④ Column stochastic $\sum_i a_{ij} = 1 \quad \forall j$

Doubly stochastic = column + row

Perron-Frobenius Theorem

If A is an irreducable, positive, row or column stochastic matrix, then the largest eigenvalue is 1 (& its eigenvector is non-negative)

⇒ used for Markov chains (stationary dist.)

For PageRank - graph not connected, may have dead ends, not stochastic

Fix H: $\hat{H} = \alpha H + (1-\alpha)\frac{1}{N} S$

for $\alpha \in (0,1)$, N is the # of webpages & S is a matrix of all 1's (fully connected w/ some small probability)

$$\boxed{\text{We can solve } \hat{H}x = x}$$

- Speed of convergence depends on α
  - α - small ⇒ graph structure important
  - α - large ⇒ fast convergence

Book

Google's PageRank & Beyond: The Science of Search Engine Ranking

Wikifier

Inverse Iteration

If know the approximate value of an eigenvalue σ we can make it the dominant eigenvalue by subtracting σI & inverting

$$\boxed{B := (A - \sigma I)^{-1}}$$

If λ_i are eigenvalues of A then

$\dfrac{1}{\lambda_i - \sigma_i}$ are eigenvalues of B

* eigenvectors are the same

Alg

$i = 0, \ \tilde{\lambda}_0 = x_0^T A x_0$

while $\| A x_i - \tilde{\lambda}_i x_i \| > tol$

Solve $(A - \sigma I) y_{i+1} = x_i$

$x_{i+1} = \dfrac{y_{i+1}}{\|y_{i+1}\|}$

$\tilde{\lambda}_{i+1} = x_{i+1}^T A x_{i+1}$

$i+1$

Orthogonal Iteration

Search for p-dim subspace at the same time

When might this be useful? e.g. low rank approximation

Input: A, Z_0 - orthogonal $n \times p$ matrix

Output: $\lambda_1 \dots \lambda_p$

$i = 0$ $\quad A_0 = Z_0^T A Z_0$ $\quad L_0$ - lower triang. part of A_0

while $\dfrac{\|L_i\|_F}{\|A_i\|_F} > tol$

$$Y_{i+1} = A Z_i$$

$$Q_{i+1} R_{i+1} = Y_{i+1}$$

$$Z_{i+1} = Q_{i+1}$$

$$A_{i+1} = Z_{i+1}^T A Z_{i+1}$$; L_{i+1} : lower triang. part of A_{i+1}

diag of A_i are $\lambda_1 \dots \lambda_p$

Claim: Let $|\lambda_1| > \dots > |\lambda_n|$

$\Rightarrow \operatorname{Span}(Z_i) = \operatorname{Span}(v_1 \dots v_p)$

Pf idea: Since absolute values are distinct each direction is well-defined.

Claim : 1) if $A_0 = I_0$, then $Z_i^T A Z_i$ converges to the Schur form of A with λ_i in descending order of absolute value on the diagonal

2) Speed on convergence of the j-th entry depends only on

$$\min\left(\frac{|\lambda_j|}{|\lambda_{j-1}|}, \frac{|\lambda_{j+1}|}{|\lambda_j|}\right)$$