Study guide for Midterm 2

▼ CH3

▼ • How can you solve a least-squares problem using the SVD (U and V are orthogonal, and Σ is square nonnegative and diagonal)?

Get the pseudo inverse

$$X^* = argmin||Ax - b||_2^2 = argmin[(Ax-b)^T(Ax-b)]$$

$$A = U\Sigma V^T$$
$$X^* = V\Sigma^+ U^T b$$

where \Simga ^+ contains the reciprocal of all nonzeros in \Simga.

Therefore the minimizer satisfies

$$U\Sigma V^T X^* \approx b$$
$$\Sigma y^* \approx d$$
$$where y^* = V^T x^* and d = U^T b$$

▼ Vector projection

vector projection of **a** onto **b** becomes

$$= \frac{a * b}{||b||} * \frac{b}{||b||}$$

▼ • On what factors does the conditioning of a linear least-squares problem depend on?

- The angle between span A and B
- The condition number for A

Whereas the conditioning of a square linear system $\boldsymbol{Ax} = \boldsymbol{b}$ depends only on the matrix $\boldsymbol{A}$, the conditioning of a least squares problem $\boldsymbol{Ax} \cong \boldsymbol{b}$ depends on the right-hand-side vector $\boldsymbol{b}$ as well as the matrix $\boldsymbol{A}$, and thus $\text{cond}(\boldsymbol{A})$ alone does not suffice to characterize sensitivity. In particular, if $\boldsymbol{b}$ lies near $\text{span}(\boldsymbol{A})$, then a small perturbation in $\boldsymbol{b}$ changes $\boldsymbol{y} = \boldsymbol{Pb}$ relatively little. But if $\boldsymbol{b}$ is nearly orthogonal to $\text{span}(\boldsymbol{A})$, on the other hand, then $\boldsymbol{y} = \boldsymbol{Pb}$ itself will be relatively small, so that a small change in $\boldsymbol{b}$ can cause a relatively large change in $\boldsymbol{y}$, and hence in the least squares solution $\boldsymbol{x}$. Thus, for a given $\boldsymbol{A}$, we would expect a least squares problem with a $\boldsymbol{b}$ that yields a large residual (i.e., a poor fit to the data) to be more sensitive than one with a small residual (i.e., a good fit to the data). An appropriate measure of the closeness of $\boldsymbol{b}$ to $\text{span}(\boldsymbol{A})$ is the ratio

$$\frac{\|\boldsymbol{Ax}\|_2}{\|\boldsymbol{b}\|_2} = \frac{\|\boldsymbol{y}\|_2}{\|\boldsymbol{b}\|_2} = \cos(\theta),$$

$$\frac{||\delta x||_2}{||x||_2} \leq \kappa(A) * \frac{||b||_2}{||Ax||_2} \frac{||\delta b||_2}{||b||_2}$$

▼ • What is QR factorization, when does it exist and is it unique?

If A is full-rank there exists an **orthogonal matrix** Q and a **unique upper-triangular matrix** R with a positive diagonal such that A = QR

Every matrix can have QR

▼ • What is a projection matrix?

Apply multiple times is the same as once

▼ • What are the classical and modified Gram-Schmidt processes? What can you say about their stability?

As for classical Gram-Schmidt:

the orthogonality of the computed qk will be lost due to rounding error

Additional storage for A and Q1

▼ • What is a Householder reflector matrix, what properties does it have and how is it computed?

Householder reflector matrix is both orthogonal and symmetric

$$H = I - 2 * \frac{vv^T}{vTv}$$
$$Ha = a - 2v * \frac{v^T a}{vTv}$$

$$Q^T = H_n ... H_1$$

▼ • What is the Householder QR factorization algorithm, and what can you say about its stability?

Householder method is generally the most efficient and accurate of the orthogonalization methods. It requires about mn^2-**n^**3 **multiplications and a similar number of additions**

Householder method can be expected to break down (in the back-substitution phase) only if cond(A) 1=mach or worse.

▼ • How many Householder iterations are needed for a QR factorization and what is the total cost?

n-1

It requires about mn^2-**n^**3 **multiplications and a similar number of additions**

▼ • What is a Givens rotation matrix, what properties does it have and how is it computed?

Householder transformations introduce many zeros in a column at once. Although
generally good for efficiency, this approach can be too heavy-handed when greater
selectivity is needed in introducing zeros.

$$G = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}$$

$$c^2 + s^2 = 1$$

orthogonal

given introduce zero one at a time

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

$$c = \frac{a_1}{\sqrt{a_1^2 + a_2^2}}, s = \frac{a_2}{\sqrt{a_1^2 + a_2^2}}$$

▼ • How can Givens rotations be used to factorize a sparse matrix?

▼ • How does one solve linear least squares problems using a QR factorization?

$$\begin{bmatrix} R \\ O \end{bmatrix} x = Q^T b = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

▼ • What is Cholesky QR and how is it used for solving a linear least squares problem?

▼ • How can you solve a rank-deficient linear least squares problem?

QR with column pivoting or SVD(cost is the most expansive)

▼ Question in Quiz

▼ Quiz 7

Linear Algebra: Orthogonal Projectors

1 point

Consider the orthogonal projector matrix $\mathbf{P} \equiv \mathbf{q}\mathbf{q}^T$, where $\mathbf{q} \in \mathbb{R}^n$ is an arbitary real vector with unit length in the 2-norm (e.g., $\|\mathbf{q}\|_2 = 1$).

Mark the statements which are **guaranteed to be true** for an arbitrary vector $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$.

Select all that apply:

- [x] $\mathbf{I} - \mathbf{P}$ is a singular matrix
- [x] $(\mathbf{I} - \mathbf{P})^2 \equiv (\mathbf{I} - \mathbf{P})$
- [x] $\mathbf{Px} \in \text{nullspace}(\mathbf{I} - \mathbf{P})$
- [ ] $\mathbf{I} - \mathbf{P}$ has $n - 1$ non-zero eigenvalues
- [ ] $\text{colspace}(\mathbf{I} - \mathbf{P}) \cap \text{nullspace}(\mathbf{P}) = \{0\}$

Your answer is mostly correct. (80.0 %)

The correct answer is:

- $\mathbf{I} - \mathbf{P}$ is a singular matrix
- $\mathbf{Px} \in \text{nullspace}(\mathbf{I} - \mathbf{P})$
- $(\mathbf{I} - \mathbf{P})^2 \equiv (\mathbf{I} - \mathbf{P})$
- $\mathbf{I} - \mathbf{P}$ has $n - 1$ non-zero eigenvalues

The second and the third fully understand but not first and the fourth

▼ Quiz 8

Reflection Matrices in Householder Algorithm

1 point

Householder QR factorization algorithm uses elementary reflection matrices $\mathbf{H} \in \mathbb{R}^{n \times n}$ of the form

$$\mathbf{H} = \mathbf{I} - 2\frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|_2^2}$$

where $\mathbf{v} \in \mathbb{R}^n$ is some vector. Which of the following is **necessarily true** for $\mathbf{H}$?

Select all that apply:

- [x] If the angle between $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{v}$ as θ, or $\angle(\mathbf{x}, \mathbf{v}) = \theta$, then $\angle(\mathbf{Hx}, \mathbf{v}) = -\theta$.
- [x] $\|\mathbf{Hx}\|_2 = \|\mathbf{x}\|_2$
- [x] Computing $\mathbf{Hx}$ requires $O(n)$ operations.
- [ ] Computing $\mathbf{Hx}$ requires $O(n^2)$ operations.

▼ Cost analysis for various methods

- Normal equations (Reducing the problem into nxn when m>>n) mn^2 multiplication and a similar number of addition, cholesky requires n^3/6 multiplications
- The Householder method is the most efficient and accurate for the orthogonalization of dense linear least-squared problems (O(mn^2 - n^3)

m >> n, the Householder method requires about twice as much work as the normal equations method

- Cost of QR Factorization with Gram Schmidt is O(mn^2)
-

▼ Extra attention

$$\begin{aligned}
\frac{\|\Delta \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} &\leq \|\boldsymbol{A}^+\|_2 \frac{\|\Delta \boldsymbol{b}\|_2}{\|\boldsymbol{x}\|_2} \\
&= \operatorname{cond}(\boldsymbol{A}) \frac{\|\boldsymbol{b}\|_2}{\|\boldsymbol{A}\|_2 \cdot \|\boldsymbol{x}\|_2} \frac{\|\Delta \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2} \\
&\leq \operatorname{cond}(\boldsymbol{A}) \frac{\|\boldsymbol{b}\|_2}{\|\boldsymbol{A}\boldsymbol{x}\|_2} \frac{\|\Delta \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2} \\
&= \operatorname{cond}(\boldsymbol{A}) \frac{1}{\cos(\theta)} \frac{\|\Delta \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2}.
\end{aligned}$$

$$
\begin{aligned}
\frac{\|\Delta \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \quad & \lessapprox \quad \|(\boldsymbol{A}^T\boldsymbol{A})^{-1}\|_2 \cdot \|\boldsymbol{E}\|_2 \frac{\|\boldsymbol{r}\|_2}{\|\boldsymbol{x}\|_2} + \|\boldsymbol{A}^+\|_2 \cdot \|\boldsymbol{E}\|_2 \\
& = \quad [\operatorname{cond}(\boldsymbol{A})]^2 \frac{\|\boldsymbol{E}\|_2}{\|\boldsymbol{A}\|_2} \frac{\|\boldsymbol{r}\|_2}{\|\boldsymbol{A}\|_2 \cdot \|\boldsymbol{x}\|_2} + \operatorname{cond}(\boldsymbol{A}) \frac{\|\boldsymbol{E}\|_2}{\|\boldsymbol{A}\|_2} \\
& \leq \quad \left([\operatorname{cond}(\boldsymbol{A})]^2 \frac{\|\boldsymbol{r}\|_2}{\|\boldsymbol{A}\boldsymbol{x}\|_2} + \operatorname{cond}(\boldsymbol{A})\right) \frac{\|\boldsymbol{E}\|_2}{\|\boldsymbol{A}\|_2} \\
& = \quad \left([\operatorname{cond}(\boldsymbol{A})]^2 \tan(\theta) + \operatorname{cond}(\boldsymbol{A})\right) \frac{\|\boldsymbol{E}\|_2}{\|\boldsymbol{A}\|_2}.
\end{aligned}
$$

QR Factorization

1 point

Given QR factorization of

$$\mathbf{A} = \mathbf{Q}\begin{bmatrix}\mathbf{R} \\ \mathbf{0}\end{bmatrix}$$

where $\mathbf{A}$ is an $m \times n$ matrix with $m > n$ and $\mathbf{A}$ has full column rank.

Which of the following is true?

Select all that apply:

- [x] If we partition $\mathbf{Q} = [\mathbf{Q}_1\ \mathbf{Q}_2]$, where $\mathbf{Q}_1$ contains the first n columns and $\mathbf{Q}_2$ contains the remaining $m-n$ columns of $\mathbf{Q}$, then we know that columns of $\mathbf{Q}_1$ form an orthonormal basis for $\operatorname{span}(\mathbf{A})$.
- [x] To minimize the residual norm $\|\mathbf{r}\|_2^2$, we have to solve a triangular linear system $\mathbf{R}\mathbf{x} = \mathbf{c}_1$, where $\mathbf{Q}^T\mathbf{b} = \begin{bmatrix}\mathbf{c}_1 \\ \mathbf{c}_2\end{bmatrix}$.
- [x] A matrix vector product with $\mathbf{Q}$ always preserves the Euclidean norm (2-norm) of the vector.
- [x] If we partition $\mathbf{Q} = [\mathbf{Q}_1\ \mathbf{Q}_2]$, where $\mathbf{Q}_1$ contains the first n columns and $\mathbf{Q}_2$ contains the remaining $m-n$ columns of $\mathbf{Q}$, then we know that columns of $\mathbf{Q}_2$ form an orthonormal basis for $\operatorname{span}(\mathbf{A})^\perp$ (the complement of the column space of $\mathbf{A}$).

▼ CH4

▼ • What is an eigenvector? an eigenvalue of a matrix? (i.e. know the definition)

characteristic of the matrix

eigenvector to eigenvalues Raley quo

▼ • What is a similarity transformation?

▼ • What is the relationship between the SVD and the eigenvalue decomposition? (Really important)

Same for SPD (square root of \lambda(A^TA) is

▼ • When are eigenvectors linearly independent?

independent → full rank matrix

The matrix is nondefective

▼ • What are the Jordan and Schur forms?

not all square matrix can be diagonalized. (algebratically same but geometrically defective)

all can be transformed into Jordan form (have

Real schur forms → block diagonal complex conjugate pair

Why Schur forms? Any square matrix coverted into Schur form

▼ • What is a normal matrix? a defective matrix? a diagonalizable matrix?

A^HA = AA^H normal matrix

linear dependent eigenvector → defective

diagonalizable → symmetric SPD?

▼ • What is an eigenvalue multiplicity? what is a complex eigenvalue pair?

multiplicity refers to the

+_ i

▼ • What is power iteration? What can be obtained using power iteration?

- The situation for power iteration not work:
    1. T**he starting vector x_0 may have no component in the dominant eigenvector v_1. But will not be a problem when rounding error is introduced by finite arthritic**
    2. There maybe more than one eigenvalue having the same (maximum) modulus, in which case the iteration may converge into a vector that i**s linear combination of the corresponding eigenvectors.**
    3. For a real matrix and real starting vector, the iteration cannot converge to a complex vector

shifting will accelerate the covergence:

$$|\frac{\lambda_2 - \theta}{\lambda_1 - \theta}| < |\frac{\lambda_2}{\lambda_1}|$$

▼ • What is normalized power iteration? What problem does it address?

Normalize to avoid underflow and overflow.

▼ • Given an approximate eigenvector, how can you estimate eigenvalues?

Rayleigh quotient

▼ • What is the Rayleigh Quotient? Rayleigh Quotient iteration?

$$\lambda = \frac{x^T Ax}{xTx}$$

▼ • What is inverse iteration and inverse shift iteration? Which eigenvalue does it find?

the conditioning is the same as A

▼ • What is the conditioning of eigenvalues and eigenvectors in an eigenvalue problem?

u is an eigenvalue of the perturbed matrix A + E and X is a nonsigular matrix that X^{-1}AX = D = diag(\lambda_1, …, \lambda_n)

$$|\mu - \lambda_k| \leq cond_2(X)||E||_2$$

Also considering the perturbation in eigenvectors,

$$(A + E)(x + \Delta x) = (\lambda + \Delta\lambda)(x + \Delta x)$$

let x and y be the right and left eigenvectors

$$\Delta\lambda \approx \frac{y^H Ex}{y^H x}$$

Taking norms yields the bound:

$$|\Delta\lambda| \leq \frac{1}{cos(\theta)}||E||_2$$

where \theta is the angle between x and y

In particular, the eigenvalues of real symmetric and complex Hermitian matrices are always well-conditioned since the right and left eigenvectors

are the
same, so that cos(\theta) = 1.

▼ • What are the eigenvalues of an upper triangular matrix and how do you compute the eigenvectors?

The diagonal line is the eigenvalues

▼ • What is orthogonal iteration? QR iteration? how are they related?

Problems of simple simultaneous iteration

- columns need to be rescaled at each iteration
- the columns of X_k form an increasingly ill-conditioned basis for the subspace since orthogonality is lost.

orthogonal is ??

▼ • How can one reduce a matrix to Hessenberg form and why is it helpful?

Reduction to upper-Hessenberg costs O(n^3)

As for Hessenberg → cost per QR is O(n^2)

As for tridiagonal → cost per QR is O(n)

▼ • How can one incorporate shifting into QR iteration?

$$Q_i R_i = A_i - \sigma_i I$$
$$A_{i+1} = R_i Q_i + \sigma_i I$$

The aim is to accelerate the convergence rate. The smallest eigenvalue (bottom right element if A_i is triangular) and n-1 columns and so on.

▼ • What is deflation?

- Method 1

$$HAH^{-1} = \begin{bmatrix} \lambda_1 & b^T \\ 0 & B \end{bmatrix}$$

$$Hx_1 = \alpha e_1$$

- Method 2 A - x_1u_1^T has eigenvalues 0, \lambda_2, …, \lambda_n

Possible choices for u_1 includes:

- u_1 = \lambda_1 x_1 if A is symmetric
- u_1 = \lambda_1 y_1 where y_1 is the corresponding left eigenvector (i.e. A^Ty_1 = \lambda_1 y_1 normalized so that y_1^Tx_1 = 1
- u_1 = A^Te_k if x_1 is normalized so that ||x||_{\inf} = 1 and the kth component of x_1 is 1

▼ • What is a Krylov subspace? how is it related to a Companion matrix?

$$K_n = [A^0 x A x A^2 x \ldots A^{n-1} x]$$
$$K_n^{-1} A K_n$$
$$[e_1 e_2 \ldots . e_n K_n^{-1} x_n] := companionmatrix$$

▼ • What is the Arnoldi method? the Lanczos method?

Lanczos is optimized to tridiagonal matrix

▼ Cost analysis

The cost for Rayleigh Quotient Iteration is O(n^3) per step

The complexity for inverse iteration is O(n^2) per step

QR iteration for Hessenberg matrix is O(n^2)

QR iteration for Tridiagonal matrix is O(n)

▼