E3101: A study guide and review, Short Form Version 1.3

Marc Spiegelman

December 8, 2007

1 The Short-form

Okay, here is the entire study guide in compact form

Solving $A\mathbf{x} = \mathbf{b}$ for $A \in \mathbf{R}^{n\times n}$

- Equation: $A\mathbf{x} = \mathbf{b}$
- Algorithm: Gaussian Elimination $[A\ \ \mathbf{b}] \rightarrow [U\ \ \mathbf{c}]$ and backsubstitute
- Factorization: $PA = LU$ (or $PA = LDU$ or $PA = LDL^T$ if $A^T = A$. Also $A = CC^T$ (Cholesky factorization for SPD matrices).

The Matrix Inverse A^{-1}

- Definition: $AA^{-1} = A^{-1}A = I$, solves $\mathbf{x} = A^{-1}\mathbf{b}$.
- Existence: A^{-1} exists iff Gaussian Elimination produces n pivots (i.e. n linearly independent columns).
- Uniqueness: if A is invertible, A^{-1} is unique and $\mathbf{x} = A^{-1}\mathbf{b}$ is unique
- Algorithm: Gauss-Jordan Elimination $[A\ \ I] \rightarrow [U\ \ E_d] \rightarrow [D\ \ E_uE_d] \rightarrow [I\ \ D^{-1}E_uE_d] = [I\ \ A^{-1}]$

Product Rules

- For Matrices
  - General AB: inner products must agree. In general $AB \neq BA$.
  - Inverse: $(AB)^{-1} = B^{-1}A^{-1}$ (if A,B both square and invertible)
  - Transpose: $(AB)^T = B^TA^T$ (all A and B such that AB exists.
  - Determinant: $|AB| = |A||B|$

- For vectors
  - $\mathbf{x}^T\mathbf{y}$: inner product, dot product, maps $\mathbf{R}^n \rightarrow \mathbf{R}$
  - $\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x}$
  - $\mathbf{x}^T\mathbf{x} = ||\mathbf{x}||^2$, where $||\mathbf{x}||$ is length of $\mathbf{x}$
  - $\mathbf{x}\mathbf{y}^T$: outer product. Is a rank 1 matrix. if $\mathbf{x} \in \mathbf{R}^m$ and $\mathbf{y} \in \mathbf{R}^n$ then $\mathbf{x}\mathbf{y}^T \in \mathbf{R}^{m\times n}$
  - $\mathbf{x}\mathbf{y}^T \neq \mathbf{y}\mathbf{x}^T$

General Solutions of $A\mathbf{x} = \mathbf{b}$ for $A \in \mathbf{R}^{m\times n}$

- Algorithm: Gauss-Jordan Elimination $[A\ \ \mathbf{b}] \rightarrow [R\ \ \mathbf{d}]$ where R is *reduced row echelon form*
- Identify rank of A (number of pivot columns) and label pivot and free columns.
- Check *existence* of solution ($\mathbf{d} \in C(R)$ implies $\mathbf{b} \in C(A)$).
- Solve $R\mathbf{x}_p = \mathbf{d}$ for the particular solution $\mathbf{x}_p$ (i.e. combination of pivot columns and **no** free columns that add to $\mathbf{d}$).
- Find special solutions as basis of $N(R) = N(A)$.
- General solution is $\mathbf{x} = \mathbf{x}_p + N\mathbf{c}$ if $\mathbf{b} \in C(A)$
- Note: $\mathbf{x}_p$ not usually entirely in $C(A^T)$ (i.e. $\mathbf{x}_p \neq x^+$). $\mathbf{x}^+ = A^+\mathbf{b} = A^+A\mathbf{x}_p$

The four fundamental subspaces of $A \in \mathbf{R}^{m\times n}$

- Definition of Basis: a minimum set of linearly independent vectors that span a vector space or subspace.
- Definition of Dimension: the number of basis vectors for any subspace.
- The four subspaces of a matrix A which is $m \times n$ with rank r

Name	Symbol	Dimension	Basis
Row Space	$C(A^T) \subset \mathbf{R}^n$	r	linearly independent rows of R
Null Space	$N(A) \subset \mathbf{R}^n$	$n - r$	special solutions of $A\mathbf{x} = \mathbf{0}$
Column Space	$C(A) \subset \mathbf{R}^m$	r	linearly independent columns of A
Left Null Space	$N(A^T) \subset \mathbf{R}^m$	$m - r$	special solutions of $A^T\mathbf{x} = \mathbf{0}$

- Orthogonality of the four subspaces:
  $C(A^T) \perp N(A)$ in $\mathbf{R}^n$ and $C(A) \perp N(A^T)$ in $\mathbf{R}^m$.

Projections, Q matrices, Least Squares

- Fundamental Equation: $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.
- Principal Algorithm: Gram-Schmidt Orthoganalization $A \rightarrow Q$

- Factorization: $A = QR$

Projections

- Project $\mathbf{b}$ onto the column space of A:
  1. solve $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ for the *least-squares solution* $\hat{\mathbf{x}}$. (symbolically $\hat{\mathbf{x}} = (A^T A)^{-1}A^T\mathbf{b}$ if A is full column rank... but you should use Gaussian Elimination to solve for $\hat{\mathbf{x}}$).
  2. The projection is $\mathbf{p} = A\hat{\mathbf{x}}$.
  3. The projection Matrix is $P = A(A^T A)^{-1}A^T$ s.t. $p = P\mathbf{b}$
- Projection onto a line is just the special case that $A = \mathbf{a}$
- Projection matrices are usually singular (unless A is square invertible where $P = I$.

Least-squares solutions and curve fitting

- Find the best solution (shortest error) to $\mathbf{Ax} = \mathbf{b}$ when A is full column rank and $\mathbf{b} \notin C(A)$. Just solve $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ for $\hat{\mathbf{x}}$ as above.
- The error $\mathbf{e} = A\hat{\mathbf{x}} - \mathbf{b}$ will be minimum such that $\mathbf{e} \in N(A^T)$ (i.e. $A^T\mathbf{e} = \mathbf{0}$).
- General function fitting through points:
  1. fit straight lines $y(x) = c_0 + c_1 x$ through n points
  2. fit general polynomials $y(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_p x^p$ through $n \geq p$ points.
  3. fit general linear combination of functions $y(x) = \sum_{i=0}^{n} c_i f_i(x)$
  4. In general, the problem set up is $A(\mathbf{x})\mathbf{c} = \mathbf{y}$ where $A(\mathbf{x})$ is a generalized vandermonde matrix with columns $f_i(\mathbf{x})$, $\mathbf{c}$ is the vector of unknown coefficients and $\mathbf{y}$ is a vector of data values.

Orthonormal Bases, Q matrices and Gram-Schmidt

- Q matrices (general $m \times n$) have n orthonormal columns such that $Q^T Q = I$.
- Gram-Schmidt takes $A \rightarrow Q$ by
  $$\mathbf{q}_1 = \frac{\mathbf{a}_1}{||\mathbf{a}_1||}$$
  $$\mathbf{b}_2 = \mathbf{a}_2 - \mathbf{q}_1(\mathbf{q}_1^T\mathbf{a}_2) \qquad \mathbf{q}_2 = \frac{\mathbf{b}_2}{||\mathbf{b}_2||}$$
  $$\mathbf{b}_3 = \mathbf{a}_3 - \mathbf{q}_1(\mathbf{q}_1^T\mathbf{a}_3) - \mathbf{q}_2(\mathbf{q}_2^T\mathbf{a}_3) \qquad \mathbf{q}_3 = \frac{\mathbf{b}_3}{||\mathbf{b}_3||}$$
  such that $C(Q) = C(A)$.
- $R = Q^T A$ is upper triangular.
- Least squares becomes $R\hat{\mathbf{x}} = Q^T\mathbf{b}$, $\mathbf{p} = QQ^T\mathbf{b}$, $P = QQ^T$.

Properties of Projection matrices

- All projection matrices can be written as $P = QQ^T$ where $Q \in \mathbf{R}^{m\times n}$ contains an orthonormal basis for an n dimensional subspace of $\mathbf{R}^m$ S. $\mathbf{p} = P\mathbf{b}$ is the **orthogonal projection** of any vector $\mathbf{b}$ onto that subspace. $P^{\perp} = (I - QQ^T)$ is the projection matrix onto the orthogonal complement $S^{\perp}$.
- $P = QQ^T$ implies that
  - All projection matrices are square, symmetric and at least positive semi-definite.
  - $P^2 = P^3 = P^q = P$
  - Most projection matrices are singular (if $n < m$)
  - If $n = m$ (Q is square), $P = I$.
  - therefore the determinant $|P| = 0$ or 1.
  - Columns of Q are eigenvectors of P with eigenvalue $\lambda = 1$
  - vectors in $N(Q^T)$ are eigenvectors with eigenvalue $\lambda = 0$
  - The singular values of P are all 1.
  - The (economy sized) SVD of P is QI_nQ^T where I_n is the $n \times n$ identity matrix.

The Determinant (square matrices)

- $|A| = 0$ if A is singular, $|I| = 1$, $|P| = \pm 1$ for permutation matrices (swapping a row changes the sign of $|A|$). $|Q| = \pm 1$.
- $|tA| = t^n|A|$ for t a scalar (the determinant is linear by *rows*
- $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$
- $|A| = \pm|U|$ (determinant is the sum of the pivots)
- $|AB| = |A||B|$, $|A^{-1}| = 1/|A|$, $|A^T| = |A|$
- Find $|A|$ by cofactor expansion (and watch for signs)
- Find $|A - \lambda I|$ for eigenvalue problems.

Eigen problems (square matrices again)

- Equation: $A\mathbf{x} = \lambda\mathbf{x}$
- Algorithm:
  1. Find n Eigenvalues, λ as roots of n-th order polynomial $|A - \lambda I|$
  2. Find n Eigenvectors as the Null Space of $(A - \lambda_i I)$
- Tests:
  - $\sum_{i=1}^n \lambda_i = \mathrm{Tr}(A)$
  - $\Pi_{i=1}^n \lambda_i = |A|$

- Special cases:
  - 2×2 matrices: $\lambda = \frac{Tr \pm \sqrt{\mathrm{Tr}^2 - 4|A|}}{2}$
  - Singular matrices: any vector in $N(A)$ is a eigenvector with eigenvalue zero. Always $n - r$ zero eigenvalues.
  - Triangular matrices (including diagonal), eigenvalues are diagonal terms.
  - square Rank 1 matrix $A = \mathbf{u}\mathbf{v}^T$, $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$
    - $\mathbf{u}$ is an eigenvector with eigenvalue $\mathbf{u}^T\mathbf{v}$
    - any vector in $N(\mathbf{v}^T)$ is an eigenvector with eigenvalue 0
  - Projection matrices $P = QQ^T$ (see section above)
  - Markov Matrices (all columns sum to 1). The largest eigenvalue is 1 and all other eigenvalues are < 1. Example, the Google matrix.
- Factorizations:
  1. General: $AS = S\Lambda$
  2. Diagonalizable $A = S\Lambda S^{-1}$, $\Lambda = S^{-1}AS$.
  3. non-Diagonalizable $A = MJM^{-1}$ (not covered in class)
  4. A is symmetric ($A^T = A$): $A = Q\Lambda Q^T$.
- Diagonalization of matrices: A can be diagonalized if
  - all eigenvalues are distinct (no repeated roots)
  - Eigenvalues are repeated but Eigenvectors are Linearly independent
  - A is symmetric
- Symmetric Matrices
  - Have all real λ's
  - Have orthogonal eigenvectors (which can be chosen orthonormal)
  - Can always be diagonalized.
  - Positive Definite Symmetric matrices all have eigenvalues > 0
  - Positive Semi-Definite Symmetric matrices have $\lambda_i \geq 0$.
  - A^TA and AA^T are both symmetric and at least positive semi-definite.
  - Tests for Symmetric Positive Definite matrices
    - All pivots are positive (sign of pivots = sign of eigenvalues for $A^T = A$)
    - quadratic form $f(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$
- Applications
  1. Matrix Powers: if $A = S\Lambda S^{-1}$ then $A^n = S\Lambda^n S^{-1}$
  2. Iterative maps: $\mathbf{u}_{n+1} = A\mathbf{u}_n$ implies $\mathbf{u}_n = A^n\mathbf{u}_0$ or $\mathbf{u}_n = S\Lambda^n S^{-1}\mathbf{u}_0$ or $\mathbf{u}_n = c_1\lambda_1^n\mathbf{x}_1 + c_2\lambda_2^n\mathbf{x}_2 + \ldots + c_p\lambda_1^n\mathbf{x}_p$

3. Dynamical Systems: $\frac{d\mathbf{u}}{dt} = A\mathbf{u}$. General solution $\mathbf{u}(t) = Se^{\Lambda t}S^{-1}\mathbf{u}_0$ or $\mathbf{u}(t) = c_1e^{\lambda_1 t}\mathbf{x}_1 + c_2e^{\lambda_2 t}\mathbf{x}_3 c + \ldots + c_n e^{\lambda_n t}\mathbf{x}_n$
4. Matrix exponential e^A. If A is diagonalizable $e^A = Se^{\Lambda}S^{-1}$. In general

$$e^A = I + A + A^2/2 + \ldots A^n/n! + \ldots$$

The Singular value decomposition: SVD $A = U\Sigma V^T$

- All matrices ($A \in \mathbf{R}^{m\times n}$) can be written as $U\Sigma V^T$ (or columwise as $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ where
  1. $U \in \mathbf{R}^{m\times m}$, $\Sigma \in \mathbf{R}^{m\times n}$ and $V \in \mathbf{R}^{n\times n}$
  2. $U^TU = I$, $V^TV = I$
  3. Σ is a "diagonal" matrix with $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r > 0$ where r is the rank (which is $\leq \min(m,n)$)
  4. The columns of V are eigenvectors of A^TA, the columns of U are eigenvectors of AA^T and the "singular values" $\sigma_i = \sqrt{\lambda_i}$ where λ_i are the eigenvalues of A^TA (or AA^T, they're the same) sorted from largest to smallest.
- A recipe for finding the SVD
  1. Find eigenvalues and eigenvectors of A^TA.
  2. Sort the eigenvalues largest to smallest and set $\sigma_i = \sqrt{\lambda_i}$
  3. Make the corresponding eigenvectors orthonormal.
  4. For all $\sigma_i > 0$ calculate $\mathbf{u}_i = A\mathbf{v}_i/\sigma_i$
  5. For all $\sigma_i = 0$ the $\mathbf{u}_i$ are found from an *orthonormal* basis for the left Null-space $N(A^T)$.
- Properties of the SVD
  - The columns of U and V contain orthonormal bases for the 4 fundamental subspaces.
    1. The first r columns of U are a basis for the column space $C(A)$
    2. The last $m - r$ columns of U are a basis for the left-Null space $N(A^T)$
    3. The first r columns of V are a basis for the row space $C(A^T)$
    4. The last $n - r$ columns of V are a basis for the Null Space $N(A)$.
  - if A is invertible, $A^{-1} = V\Sigma^{-1}U^T$.
  - if A has a null-space the Pseudo-inverse $A^+ = V\Sigma^+U^T$ where $\Sigma^+ = \Sigma^{-1}$ for all $\sigma_i > 0$ and equals 0 where $\sigma_i = 0$.
  - $\mathbf{x}^+ = A^+\mathbf{b}$ is the *shortest* least squares solution (i.e is entirely in the row space of A).
  - $P = AA^+ = QQ^T$ is the projection matrix onto $C(A)$ (i.e. where Q contains the first r columns of U.

  - $P = A^+A$ is the projection matrix onto $C(A^T)$
- Applications of the SVD
  - Least-squares by pseudo-inverse $\mathbf{x}^+ = A^+\mathbf{b}$
  - Total-Least squares (best fit line/plane with orthogonal errors)
  - EOF/Principal component analysis.