Test #5 Study Guide

November 29, 2015

Format

• Date, time, and place: Tuesday, December 8, 12:55-1:45 PM in LCB 225.
• Review session: Monday, December 7, 4:45-6:15 PM in JWB 333.
• You can always make an appointment with me by email.
• Sections covered: 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 7.1, 7.4.
• Study Quizzes 8 and 9, the recommended problems, and especially the concepts I’ve emphasized in lecture.
• One page of inventions and four pages of multiple-part questions. 10 points per page; 50 points total. Worth 10% of the final grade.

Key Concepts and Skills

Dot Product and Orthogonality

• Know the definition of dot product of two vectors in $\mathbb{R}^n$ and that $\vec{v} \cdot \vec{w} = \vec{v}^T \vec{w}$.
• Know the definition of length $\|\vec{v}\|$ of a vector $\vec{v}$ (distance of the point $\vec{v}$ from the origin), the definition of unit vector, and the formula $\vec{v} \cdot \vec{w} = \|\vec{v}\| \|\vec{w}\| \cos \theta$, where $\theta$ is the angle formed by the three points $\vec{v}$, $\vec{0}$, and $\vec{w}$.
• Be able to invent vectors with properties specified using dot products.
• Understand the definition of orthogonal vectors and the orthogonal complement $W^\perp$ of a subspace $W$ of $\mathbb{R}^n$.
• Be able to visualize orthogonal complements. For instance, the orthogonal complement of a line through the origin in $\mathbb{R}^2$ is the perpendicular line through the origin. The orthogonal complement of a line through the origin in $\mathbb{R}^3$ is the plane through the origin whose normal vector lies on that line. The orthogonal complement of a plane through the origin in $\mathbb{R}^3$ is the span of its normal vector.
• Understand why Row A and Nul A are orthogonal complements and why Col A and Nul A^T are orthogonal complements. Understand how these four subspaces fit into the picture of multiplication by A defining a linear transformation \( \mathbb{R}^n \xrightarrow{T} \mathbb{R}^m \). Be able to compute these four subspaces and sketch them.

• Be able to compute the \( \mathcal{B} \)-coordinates of a vector in \( \mathbb{R}^n \) when \( \mathcal{B} = \{ \vec{u}_1, \ldots, \vec{u}_n \} \) is an orthogonal basis.

• If \( U \) has orthonormal columns, then multiplication by \( U \) preserves dot products and therefore preserves lengths of vectors and angles between vectors. If \( U \) has orthonormal columns and is a square matrix, then \( U \) is called an orthogonal matrix and has the property \( U^T = U^{-1} \).

• Understand what it means to project a vector \( \vec{y} \) onto a subspace \( W \). Projection produces a decomposition \( \vec{y} = \hat{y} + \vec{z} \), where \( \hat{y} = \text{proj}_W \vec{y} \) is in \( W \) (more precisely, \( \hat{y} \) is the closest vector in \( W \) to \( \vec{y} \)) and \( \vec{z} \) is orthogonal to \( W \) (the length of \( \vec{z} \) is the distance of \( \vec{y} \) from \( W \)).

• Be able to compute projections when you have an orthogonal basis \( \{ \vec{u}_1, \ldots, \vec{u}_p \} \) of \( W \). Know where the formula for orthogonal projection comes from: think of writing \( \vec{y} = (c_1 \vec{u}_1 + \cdots + c_p \vec{u}_p) + \vec{z} \) and taking dot products of both sides by \( \vec{u}_i \) to get \( c_i = (\vec{y} \cdot \vec{u}_i)/(\vec{u}_i \cdot \vec{u}_i) \).

• Be able to use the Gram-Schmidt algorithm to orthogonalize a basis \( \{ \vec{x}_1, \ldots, \vec{x}_n \} \) for \( \mathbb{R}^n \). The algorithm amounts to modifying each \( \vec{x}_i \) by subtracting off the projection of \( \vec{x}_i \) onto the subspace spanned by \( \vec{x}_1, \ldots, \vec{x}_{i-1} \).

• Understand why \( A\vec{x} = \vec{b} \) has a solution exactly when \( \vec{b} \) is in Col A. When \( A\vec{x} = \vec{b} \) has no solution, know how to compute \( \vec{x} \) so that \( A\vec{x} \) is as close to \( \vec{b} \) as possible. This amounts to solving \( A\vec{x} = \text{proj}_{\text{Col} A} \vec{b} \), which is equivalent to \( A^T A\vec{x} = A^T \vec{b} \).

• Know how to compute a line of best fit through some data points by solving for \( \vec{\beta} \) in the formula \( X^T X \vec{\beta} = X^T \vec{y} \) (in the book’s notation). The matrix \( X \) and the vector \( \vec{y} \) are constructed from the data points and the line is \( y = \beta_0 + \beta_1 x \).

**Diagonalization of Symmetric Matrices**

• Understand what it means for \( A \) to be symmetric.

• Know the key features of symmetric matrices that make them orthogonally diagonalizable: the eigenvalues are all real; the eigenspaces have full dimension; the eigenspaces are orthogonal.

• Be able to orthogonally diagonalize symmetric matrices. This will yield the nice formula \( A = PDP^T \), where we are using \( P^T = P^{-1} \) since we can choose \( P = [\vec{u}_1 \cdots \vec{u}_n] \) to be an orthogonal matrix. (Remember that the inverse of \( P \) is usually annoying to compute. No so here!)
• Understand why $P$ being orthogonal allows us to write the diagonalization as $A = \lambda_1 \tilde{u}_1 \tilde{v}_1^T + \cdots + \lambda_r \tilde{u}_r \tilde{v}_r^T$, where $r$ is the rank of $A$ and $\lambda_1, \ldots, \lambda_r$ are the nonzero eigenvalues. This expresses $A$ as the sum of $r$ rank 1 matrices.

Singular Value Decomposition (SVD)

• Be able to compute the SVD $A = U \Sigma V^T$ for any $m \times n$ matrix $A$ ($A$ need not be square!) using the following steps:

  (1) Compute $A^T A$, which is $n \times n$ and symmetric.

  (2) Orthogonally diagonalize $A^T A$ by computing the eigenvalues of $A^T A$ in descending order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ with corresponding orthonormal eigenvector basis $\{\tilde{v}_1, \ldots, \tilde{v}_n\}$ of $\mathbb{R}^n$. Let $V = [\tilde{v}_1 \cdots \tilde{v}_n]$ and let $\Sigma$ be the $m \times n$ matrix with the nonzero singular values $\sigma_1 = \sqrt{\lambda_1} \geq \cdots \geq \sigma_r = \sqrt{\lambda_r}$ on its “diagonal”.

  (3) Normalize each $A \tilde{u}_1, \ldots, A \tilde{u}_r$ to get $\tilde{u}_1, \ldots, \tilde{u}_r$ and extend to an orthonormal basis $\{\tilde{u}_1, \ldots, \tilde{u}_m\}$ of $\mathbb{R}^m$. Let $U = [\tilde{u}_1 \cdots \tilde{u}_m]$.

• Understand the geometric interpretation of the SVD: the linear transformation $\mathbb{R}^n \xrightarrow{A} \mathbb{R}^m$ is, up to orthogonal changes of coordinates, just multiplication by a diagonal non-square matrix $\Sigma$ of singular values.

• Understand the interpretation of the SVD as $A = \sigma_1 \tilde{u}_1 \tilde{v}_1^T + \cdots + \sigma_r \tilde{u}_r \tilde{v}_r^T$, where $r$ is the rank of $A$. The partial sum $A_k = \sigma_1 \tilde{u}_1 \tilde{v}_1^T + \cdots + \sigma_k \tilde{u}_k \tilde{v}_k^T$ is the best rank $k$ approximation of $A$.

• If $A$ is a matrix of data, be able to think of the $\tilde{v}_i^T$ as the row trends in $A$ sorted by their importance $\sigma_i$. Think of $\sigma_i \tilde{u}_i$ as the weights of those row trends in each of the rows of the matrix.

• Know that if the singular values decrease rapidly, then the data can be explained well by very few row trends, hence a low-rank approximation of $A$ will be very close to $A$. 

3