**Review guide for exam 4**

This is a list of topic headings and main ideas that I think are important, to give you a rough study guide. Your main tool should be the text and homework exercises.

§5.1.Orthogonality. You should know about:

Dot product of vectors, length of a vector $||v|| = \sqrt{v \cdot v}$, unit vectors, orthogonal vectors, and orthonormal sets of vectors.
The orthogonal complement $V^{\perp}$ of a subspace $V$ of $\mathbb{R}^n$.

You should know that:
An orthonormal set of vectors in $\mathbb{R}^n$ are linearly independent.
If $V$ is a subspace of $\mathbb{R}^n$, then $V^{\perp}$ is also a subspace of $\mathbb{R}^n$.

You should know:
The formula for the orthogonal projection onto a subspace $V$ of $\mathbb{R}^n$, $proj_V$, if you are given an orthnormal basis $w_1, \ldots, w_r$ of $V$: $proj_V(x) = (x \cdot w_1)w_1 + \ldots + (x \cdot w_r)w_r$

You should know:
$||proj_V(x)|| \leq ||x||$, $||a+b||^2 = ||a||^2 + ||b||^2$ if $a \cdot b = 0$, the Cauchy=Schwartz inequality

$$|x \cdot y| \leq ||x|| ||y||$$

and that the angle $\theta$ between two vectors $x, y$ is determined by

$$\cos\theta = \frac{x \cdot y}{||x|| ||y||}.$$

§5.2 Gram-Schmidt process and the $QR$ factorization. You should know how to apply the Gram-Schmidt process to transform a linearly independent set of vectors $v_1, \ldots, v_r$ in $\mathbb{R}^n$ into an orthonormal set of vectors $w_1, \ldots, w_r$ such that

$$span(v_1, \ldots, v_i) = span(w_1, \ldots, w_i)$$

for all $i = 1, \ldots, r$. This is how you do it:

1. $w_1 = \frac{1}{||v_1||} \cdot w_1$

2.A. $v_2' = v_2 - proj_{w_1}(v_1) = v_2 - (w_1 \cdot v_1)w_1$
2.B. $w_2 = \frac{1}{||v_2'||} \cdot v_2'$

Repeat: Once you have constructed $w_1, \ldots, w_{m-1}$ this way (for $m - 1 < r$), make $w_m$ in two steps:
m.A. $v_m' = v_m - proj_{w_1, \ldots, w_{m-1}}(v_m) = v_m - [(w_1 \cdot v_m)w_1 + \ldots + (w_{m-1} \cdot v_m)w_{m-1}]$.
m.B. $w_m = \frac{1}{||v_m'||} \cdot v_m'$

The $QR$ factorization: The result of the Gram-Schmidt process can be written as a matrix equation. Start with a $n \times m$ matrix $M$ with linearly independent columns (these are the vectors $v_1, \ldots, v_m$). Let $Q$ be the $n \times m$ matrix with columns the orthonormal set of vectors $w_1, \ldots, w_m$ you get by applying the Gram-Schmidt process to the columns of $M$. Then there is an $m \times m$ matrix $R$ satisfying

a. $R$ is upper triangular: the $ij$ entry of $R$ is 0 if $i > j$
b. The diagonal entries of $R$ are all $> 0$.
c. $M = QR$.

To get $R$, you just keep track of your expressions for $v'_m$ and $w_m$ in the two-step Gram-Schmidt process, and use these to write

$$v_m = r_{1m}w_1 + r_{2m}w_2 + \ldots + r_{mm}w_m$$

Then $R$ is the matrix with $ij$ entry $r_{ij}$, in other words, the $m$th column of $R$ is $\begin{bmatrix} r_{1m} \\ r_{2m} \\ \vdots \\ r_{mm} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. If you want a "formula" for the $r_{ij}$, it's:

$$r_{ij} = \begin{cases} (w_i \cdot v_j) & \text{for } 1 \leq i < j \\ ||v'_j|| & \text{for } i = j \\ 0 & \text{for } i > j \end{cases}$$

Since every subspace $V$ or $\mathbb{R}^n$ has a basis, the Gram-Schmidt process gives you an orthonormal basis for $V$, so: Every subspace $V$ of $\mathbb{R}^n$ has an orthonormal basis.

§5.3. Orthogonal transformations and matrics. You should know

The definition of an orthogonal transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, and an orthogonal $n \times n$ matrix $A$: $||T(x)|| = ||x||$ or $||Ax|| = ||x||$ for all $x$ in $\mathbb{R}^n$.

If $T$ is orthogonal, then $T$ preserves dot products and angles between vectors: $T(x) \cdot T(y) = x \cdot y$.
An $n \times n$ matrix $A$ is an orthogonal matrix exactly when the columns of $A$ form an orthonormal basis of $\mathbb{R}^n$.

If $A$ is orthogonal, then $A$ is invertible and $A^{-1} = A^T$.

You should also know the basic properties of taking the transpose of a matrix:

$(A^T)_{ij} = A_{ji}$, $(AB)^T = B^T A^T$, and if $A$ is an $m \times n$ matrix, $x$ in $\mathbb{R}^n$ and $y$ in $\mathbb{R}^m$, then $(A^T y) \cdot x = y \cdot (Ax)$

If $w_1, \ldots, w_m$ are an orthonormal basis of a subspace $V$ of $\mathbb{R}^n$, and $A = \begin{bmatrix} w_1 & \ldots & w_m \end{bmatrix}$, then the matrix of the orthogonal projection onto $V$ is $AA^T$.

§5.4. Least squares and data fitting. You should know

If $A$ is an $m \times n$ matrix, then $im(A)$ and $ker(A^T)$ are both subspaces of $\mathbb{R}^m$. In fact

$$(im(A))^{\perp} = ker(A^T).$$

Similarly, the subspaces $(im(A^T))^{\perp},\ ker(A)$ of $\mathbb{R}^n$ are equal:

$$(im(A^T))^{\perp} = ker(A).$$

For a subspace $V$ of $\mathbb{R}^n$,
1. $\dim V + \dim V^{\perp} = n$
2. $(V^{\perp})^{\perp} = V$.
3. $V \cap V^{\perp} = \{0\}$

Applying this to $V = im(A)$ for some $m \times n$ matrix $A$ tells us:

$$rank(A) = rank(A^T).$$

*Least squares approximation.* This relies on an important fact: Let $V$ be a subspace of $\mathbb{R}^n$. Then for $x$ in $\mathbb{R}^n$ the vector $y = proj_V(x)$ is the vector in $V$ that is closest to $x$.

Suppose you are faced with an inconsistent system of linear equations: $Ax = b$. Well, this means there is no solution, but you could try to find an $x$ so that $Ax$ is as close as possible to $b$. Now $Ax$ is just some element in $im(A)$, and the closest you can get to $b$ and still be in $im(A)$ is $b^* := proj_{im(A)}(b)$. Since $b^*$ is in $im(A)$, the system $Ax = b^*$ is consistent. A solution $x^*$ to $Ax = b^*$ is called a *least squares solution* to the original (inconsistent) system $Ax = b$.

Since $(imA)^{\perp} = ker(A^T)$, and since $b - b^*$ is in $im(A)^{\perp}$, we have $A^T(b = b^*) = 0$, so $A^T b = A^T b^*$. Thus

$$Ax^* = b^* \implies A^T A x^* = A^T b^* \implies A^T A x^* = A^T b.$$

In other words, a least squares solution $x^*$ to $Ax = b$ is an *actual* solution to $A^T A x = A^T b$. In fact, the opposite also holds, so you can find a least squares solution by solving $A^T A x = A^T b$, without going to the trouble of finding $b^*$. The equation $A^T A x = A^T b$ is called the *normal equation* of $Ax = b$.

Example: Find the quadratic function $f(t) = a_0 + a_1 t + a_2 t^2$ whose graph is the best least squares fit to passing through the points $(0,0)$, $(1,1)$ $(2,3)$ and $(3,1)$. To solve: first write the equations you want to satisfy: $f(0) = 0$, $f(1) = 1$, $f(2) = 3$, $f(3) = 1$ as a matrix equation for the coefficients $a_0, a_1, a_2$. For this, note that you can write $f(t)$ as

$$f(t) = \begin{bmatrix} 1 & t & t^2 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

so our problem is to find the least squares solution of the matrix equation

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 1 \end{bmatrix}.$$

The normal equation is

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \\ 1 \end{bmatrix}.$$

or

$$\begin{bmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \\ 22 \end{bmatrix}.$$

Using Gauss-Jordan elimination, the solution is $a_0 = -0.25$, $a_1 = 2.75$, $a_2 = -0.75$, so the best least squares function is $f(t) = -0.25 + 2.75t - 0.75t^2$.

There is a refinement: if $\ker A = 0$, then $A^T A$ is automatically invertible, so there is a s solution to the normal equation for $Ax = b$:

$$x^* = (A^T A)^{-1} A^T \cdot b.$$

Finally, we started the whole discussion by requiring that $Ax^* = proj_{im(A)}(b)$. Thus $proj_{im(A)}(b) = A((A^T A)^{-1} A^T b$. If now $V$ is any subspace of $\mathbb{R}^n$ with basis $v_1, \ldots, v_m$, and if we take $A = [v_1, \ldots, v_m]$, then $V = im(A)$ and we have an explicit formula for the matrix of $proj_V$, without going to the trouble of finding an orthonormal basis of $V$:

$$\text{The matrix of } proj_V \text{ is } A(A^T A)^{-1} A^T.$$

Of course, it may be a lot of trouble to find the inverse of $A^T A$.