These are brief notes for the lecture on Monday November 30, 2009: they are not complete, but they are a guide to what I want to say on those days. They are guaranteed to be incorrect.

6.4. Gram-Schmidt Orthonormalization

The Gram-Schmidt algorithm constructs an orthogonal or orthonormal basis for any subspace $W \neq \{\underline{0}\}$ of $\mathbb{R}^n$. We start with a couple of examples before developing the algorithm:

Example: Let $W = \mathrm{Span}(\underline{x}_1, \underline{x}_2)$ where $\underline{x}_1 = \begin{pmatrix} 3 \\ 6 \\ 0 \end{pmatrix}$ and $\underline{x}_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$. Find an orthogonal basis for W.

Now find an orthonormal basis for W:

Example: Let $W = \text{Span}(\underline{x}_1, \underline{x}_2, \underline{x}_3)$ where $\underline{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, $\underline{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ and $\underline{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$. Clearly the vectors are linearly independent, so W is a three-dimensional subspace of $\mathbb{R}^4$. Find an orthogonal basis for W.

Now find an orthonormal basis for W:

Let's now analyze the theory of this. Suppose that we have a subspace $W < \mathbb{R}^n$, and that we have an orthogonal basis $\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k\}$ for W. Now suppose that $\underline{v}$ is a vector in $\mathbb{R}^n$ but $\underline{v} \notin W$.

Since $\underline{v} \notin W$, $\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k\}, \underline{v}$ spans a strictly bigger space than W. If we now compute the projections onto each of the vectors $\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k\}$, and subtract from $\underline{v}$, then we obtain a vector

$$\underline{u}_{k+1} = \underline{v} - c_1\underline{u}_1 - c_2\underline{u}_2 - \cdots - c_k\underline{u}_k$$

which is orthogonal to each of the vectors in $\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k\}$. It is clear that this spans $\text{Span}(\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k, \underline{v}\})$.

Now, suppose that we have a basis $\{\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n\}$ for $\mathbb{R}^n$, which is not necessarily orthogonal. We can convert this into an orthogonal basis for $\mathbb{R}^n$ as follows.

Let $W_k = \text{Span}\{\underline{v}_1, \underline{u}_2, \ldots, \underline{v}_k\})$ be the span of the first k vectors in the basis. We will assume that we've already constructed an orthogonal basis $\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k\}$ for W_k: if we now proceed as above, with $\underline{v} = \underline{v}_{k+1}$, then $\underline{u}_{k+1}$ is orthogonal to $\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k\}$, and $\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k, \underline{k+1}\}$ is a basis for W_{k+1}.

THEOREM 11. *Suppose $W \leq \mathbb{R}^n$ has a basis $\{\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_p\}$. Then the procedure above produces an orthogonal basis for W.*

If we wish to find an orthonormal basis, then we can either proceed as above, and normalize all the vectors at the end, or we can normalize along the way. Working by hand, it is usually better to normalize at the end (it saves carrying along ugly square roots). Working numerically on a computer, it is better to normalize along the way.

THEOREM 12 (QR Factorization). *Suppose that A is a $m \times n$ matrix with linearly independent columns. Then there exist matrices Q and R so that*

(1) *$A = QR$.*

(2) *Q is $m \times n$ and the columns of Q form an orthonormal basis for Col(A), the column space of A.*

(3) *R is an upper triangular, square matrix with positive entries on the diagonal.*

Proof sketch: The columns of A are linearly independent, so they form a basis for Col(A). Convert them to an orthonormal basis via the Gram Schmidt algorithm. Check that $A = QR$ with Q and R as claimed follows from the way we construct the orthonormal basis.

Note: this shows the existence of such a factorization. In fact, there are others, often numerically better for computational purposes.

6.5. Least Squares Problems

Problems often arise in practice in which, because of an approximation somewhere, we want to solve an inconsistent system $A\underline{x} = \underline{b}$. Since there is no solution, this might seem like a hopeless task. However, we can take a slightly different perspective: a solution to

$$A\underline{x} = \underline{b}$$

is also a vector which minimizes

$$\|\underline{b} - A\underline{x}\|$$

(indeed, since this is a length, it is non-negative, so clearly if it is zero, it is as small as it can be!)

Now, if there is no solution to $A\underline{x} = \underline{b}$ then the minimum value will not be zero: but we can still look for a minimizing vector $\hat{\underline{x}}$: that is, a vector $\hat{\underline{x}} \in \mathbb{R}^n$ so that

$$\|\underline{b} - A\hat{\underline{x}}\| \leq \|\underline{b} - A\underline{x}\|$$

for all $\underline{x} \in \mathbb{R}^n$. Such a vector is called a *least squares solution* of $A\underline{x} = \underline{b}$.

Now, since $A\underline{x}$ is a linear combination of the columns of A, we are looking for a vector $\hat{\underline{b}} \in \mathrm{Col}(A)$ which is closest to $\underline{b}$. Clearly, the vector which does this is the projection $\hat{\underline{b}}$ of $\underline{b}$ onto $\mathrm{Col}(A)$.

Note that if $\hat{\underline{b}} = A\hat{\underline{x}}$ is this projection, then $\underline{b} - \hat{\underline{b}}$ is orthogonal to each column of A.

This implies that $A^T(\underline{b} - \hat{\underline{b}}) = \underline{0}$ (since the rows of A^T are precisely the columns of A). Hence

$$A^T(\underline{b} - A\hat{\underline{x}}) = \underline{0}$$

so that

$$A^T\underline{b} - A^T A\hat{\underline{x}} = \underline{0}$$

or

$$A^T\underline{b} = A^T A\hat{\underline{x}}.$$

This shows the following:

Theorem 13. *The set of least squares solutions to $A\underline{x} = \underline{b}$ coincides with the (non-empty) set of solutions to $A^T A\underline{x} = A^T\underline{b}$.*

The fact that there exist such solutions to the latter equation follows from the fact that there are least-squares solutions to the original equation.

Theorem 14. *The matrix $A^T A$ is invertible if and only if the columns of A are linearly independent. In this case, the equation $A\underline{x} = \underline{b}$ has a unique least squares solution, and it is given by*

$$\hat{\underline{x}} = (A^T A)^{-1} A^T \underline{b}$$

Theorem 15. *Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR-factorization of A. Then for each $\underline{b} \in \mathbb{R}^m$, the equation $A\underline{x} = \underline{b}$ has a unique least-squares solution, given by*

$$\hat{\underline{x}} = R^{-1} Q^T \underline{b}.$$