Math 410 (Prof. Bayly) MINIMUM-LENGTH SOLUTIONS Summer 2005

The fundamental theorem of finite-dimensional linear algebra is that a system of simultaneous linear equations has either no solutions, one unique solution, or an infinite number of solutions. When an infinite number of solutions exist, an application-minded person would ask "Which one is the best?" It is often practical to ask, if there are several ways to solve a problem, which one does so in the most efficient or optimal way?

To quantify the concept of quality, we often define a "utility function" for which bigger values are better, or a "cost function" for which smaller values are more desirable. Finding the best is then to determine the solution that either maximizes the utility or minimizes the cost. In a real-world situation these functions depend on the situation, and the problem of finding the optimum can be quite complicated. There is a lot of current research activity on efficient techniques for finding optimum solutions. (Visit some of the Systems and Industrial Engineering professors!)

In Mathematics we're allowed to consider simple questions first, and in the context of linear algebra it's natural to define the "cost" of a vector as simply its length. It turns out that the problem of finding the minimum-length solution to a system that has infinitely-many solutions is solvable but not trivial. The answer ends up not only being a lot more useful than you might initially think for such a simplistic cost function, but we also learn some interesting mathematics. We'll see minimum-length solutions reappear in a different context in our later discussions of the Singular Value Decomposition and its applications.

Example 1. The most direct way to find the minimum-length solution of a system of equations is to find the general solution first, and then seek the one with minimum length. For example, consider $A\vec{x} = \vec{b}$, with

$$A = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 1 & -1 \end{pmatrix} \quad , \quad \vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad , \quad \vec{b} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}. \tag{1a}$$

We quickly find the solution family

$$\vec{x} = \begin{pmatrix} 2+z \\ -1-z \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \vec{x}_p + z\vec{n}, \tag{1b}$$

where $\vec{x}_p$ denotes the particular solution and $\vec{n}$ the null vector.

The length of the solution vector is a function of the free variable:

$$L(z) = \left[(2+z)^2 + (-1-z)^2 + z^2\right]^{1/2} = \left[3z^2 + 6z + 5\right]^{1/2}. \tag{2a}$$

It's a simple Calculus exercise to minimize $L(z)$ by finding the derivative

$$L'(z) = \frac{1}{2}\left[3z^2 + 6z + 5\right]^{-1/2}(6z+6), \tag{2b}$$

setting it to zero and solving for $z = -1$. Therefore the minimum-length solution is

$$\vec{x}_{ml} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}. \tag{3}$$

One thing to notice is that when we solved for the value of z at which $L' = 0$, all we did was set $6z + 6 = 0$, which is the derivative of the length squared $L^2(z) = 3z^2 + 6z + 5$. In retrospect it's clear that minimizing L^2 is equivalent to minimizing L, and we don't have to bother with the square root. It's also worth noting that L^2 is a simple quadratic, and the derivative of L^2 is a linear function of z, so it's very easy to find the minimum.

Example 2. Let's apply these observations to $A\vec{x} = \vec{b}$, where $\vec{x} = (\,x \quad y \quad z\,)^T$ again, $A = (\,1 \quad 1 \quad -2\,)$, and $\vec{b} = (\,3\,)$. Now the solution family has two free variables and corresponding null vectors:

$$\vec{x} = \begin{pmatrix} 3 - y + 2z \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} + y\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + z\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = \vec{x}_p + y\vec{n}_1 + z\vec{n}_2. \tag{4}$$

The length squared function is

$$L^2(y,z) = (3 - y + 2z)^2 + y^2 + z^2 = 2y^2 - 4yz + 5z^2 - 6y + 12z + 9. \tag{5a}$$

To find the minimum we calculate the *partial* derivatives of L^2 with respect to both y, z, and set both equal to zero. Explicitly

$$\frac{\partial L^2}{\partial y} = 4y - 4z - 6 \quad , \quad \frac{\partial L^2}{\partial z} = -4y + 10z + 12. \tag{5b}$$

Setting these to zero and solving yields $y = 1/2$, $z = -1$. Therefore the minimum-length solution is

$$\vec{x}_{ml} = \begin{pmatrix} 1/2 \\ 1/2 \\ -1 \end{pmatrix}. \tag{6}$$

Note again that L^2 is again a simple quadratic (albeit in two variables), its partial derivatives are linear functions, and finding the minimizing values of the free variables is just solving a pair of simultaneous linear equations.

It turns out that the pattern we have seen in these two examples always appears when we look for the minimum-length solution, whatever the number of free variables. Furthermore there is a direct way to find a system of linear equations whose solution is the desired minimum-length solution, without having to do any calculus.

General theory. Let's consider the case of one free variable. The general solution is $\vec{x} = \vec{x}_p + z\vec{n}$ where $\vec{x}_p$ and $\vec{n}$ depend on the matrix A and right-hand side vector $\vec{b}$. The length squared function can be written

$$L^2(z) = \vec{x}^T\vec{x} = (\vec{x}_p + z\vec{n})^T(\vec{x}_p + z\vec{n}) = z^2\vec{n}^T\vec{n} + 2z\vec{n}^T\vec{x}_p + \vec{x}_p^T\vec{x}_p. \qquad (7a)$$

Its derivative is

$$\frac{dL^2}{dz^2} = 2\vec{n}^T\vec{n} + 2\vec{n}^T\vec{x}_p, \qquad (7b)$$

which is zero when $z = -\vec{n}^T\vec{x}_p/\vec{n}^T\vec{n}$. Therefore the minimum length solution is

$$\vec{x}_{ml} = \vec{x}_p - \frac{\vec{n}^T\vec{x}_p}{\vec{n}^T\vec{n}}\vec{n}. \qquad (8)$$

There is nothing new (yet) about the result in (8), except you may recognize it as the formula for the projection of $\vec{x}_p$ orthogonal to $\vec{n}$. If you have not learned about orthogonal projections, you will very soon - if you're psychic you can precognize* it from the future instead of recognizing it from the past. If neither, then just observe that

$$\vec{n}^T\vec{x}_{ml} = \vec{n}^T\left(\vec{x}_p - \frac{\vec{n}^T\vec{x}_p}{\vec{n}^T\vec{n}}\vec{n}\right) = \vec{n}^T\vec{x}_p - \frac{\vec{n}^T\vec{x}_p}{\vec{n}^T\vec{n}}\vec{n}^T\vec{n} = 0. \qquad (9)$$

The important conclusion is that *the minimum length solution $\vec{x}_{ml}$ is orthogonal to the null vector $\vec{n}$.*

A similar calculation involving partial derivatives shows that whatever the number of free variables, THE MINIMUM LENGTH SOLUTION IS ALWAYS ORTHOGONAL TO EVERY NULL VECTOR OF THE MATRIX! (You should go back right now and verify this assertion for the examples we did above.) This is the crucial observation that knocks down the calculus problem of minimizing a quadratic function to a problem of simultaneous linear equations.

There's still a bit of thinking that has to be gone through before we're home, which is connected to the problem of what it means for a vector to be orthogonal to the null vectors of a matrix. If you think about where the null vectors of a matrix A come from, you recall that each null vector $\vec{n}$ has the property that $A\vec{n} = \vec{0}$. And the operation of

* Recognize is to recognition as precognize is to precognition.

matrix-vector multiplication simply consists of dot-multiplying the row vectors of A with the column vector $\vec{n}$. So each row vector of A dotted with the null vector $\vec{n}$ gives zero, i.e. the null vector is orthogonal to each row vector of A. Or, since we often prefer to think of column vectors, each null vector of A is orthogonal to each column vector of A^T.

The conclusion of the previous paragraph can be restated as saying that each column vector of A^T is orthogonal to each null vector of A. It turns out, via a somewhat nontrivial proof†, that *any* vector that is orthogonal to each null vector of A *must* be a linear combination of the column vectors of A^T.

The implication is that $\vec{x}_{ml}$ must be a linear combination of the column vectors of A^T. It's clearly the case in our second example, where $\vec{x}_{ml} = \frac{1}{2}A^T$, and you can check that in the first example $\vec{x}_{ml}$ is $-1/5$ times the first column of A^T plus $3/5$ times the second. We can use this fact to get a practical and efficient way to find minimum length solutions in general.

Practical method. An algebraic way to express the fact that the minimum length solution is a linear combination of the column vectors of A^T is $\vec{x}_{ml} = A^T\vec{u}$, where the components of $\vec{u}$ are the coefficients in the desired linear combination. Since the original problem was $A\vec{x} = \vec{b}$, with $\vec{x}_{ml}$ being simply the shortest possible solution, the coefficient vector $\vec{u}$ must satisfy $A(A^T\vec{u}) = \vec{b}$. Or, using the associative property of matrix multiplication,

$$(AA^T)\vec{u} = \vec{b}. \tag{10}$$

Therefore, if (10) can be solved for $\vec{u}$, all we need to do is multiply $\vec{u}$ by A^T to obtain the minimum length solution $\vec{x}_{ml}$. No calculus, no free variables or null vectors! We still have to solve a linear system, and do a couple of extra multiplications, but conceptually at least we have a much simpler procedure.

Let's see how it works. In example 1, $AA^T = \begin{pmatrix} 14 & 3 \\ 3 & 6 \end{pmatrix}$, so the system (10) becomes

$$\begin{pmatrix} 14 & 3 \\ 3 & 6 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \qquad therefore \qquad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -1/5 \\ 3/5 \end{pmatrix}. \tag{11}$$

These are exactly the coefficients that we observed above give $\vec{x}_{ml}$ as a combination of the columns of A^T, or more concisely,

$$\vec{x}_{ml} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = A^T\vec{u} = \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} -1/5 \\ 3/5 \end{pmatrix}.$$

† There's magic going on here, folks, and one of the rewards of studying rigorous pure mathematics is the power you acquire to wield that magic.

Example 2 goes even faster. $AA^T = (6)$, so (10) is nothing but the scalar equation $6u = 3$, whose solution $u = 1/2$ is the desired multiple we identified above.

To summarize, a practical way to find the minimum length solution of $A\vec{x} = \vec{b}$ is:

(1) Calculate AA^T,

(2) Solve $AA^T\vec{u} = \vec{b}$ for $\vec{u}$,

(3) Calculate $\vec{x}_{ml} = A^T\vec{u}$.

NOTE: On homework and/or exams we may have you derive the general solution to $A\vec{x} = \vec{b}$, with free variables and null vectors, as the first part of a problem, and ask you to find the min length solution for the second part. Or we might actually give you the general solution as part of the problem. If so, i.e. if you already have the general solution, it might be less work to use the calculus approach, especially if there is only one free variable. We advise you to become expert with both techniques, so that you can choose the best method under any circumstances.

Remarks. You may wonder whether the new linear system $AA^T\vec{u} = \vec{b}$ has no solutions, one unique solution, or an infinite number of solutions. It turns out (if you do the rigorous theory) that if the original system $A\vec{x} = \vec{b}$ has at least one solution, then so does $AA^T\vec{u} = \vec{b}$. However it is quite possible for $AA^T\vec{u} = \vec{b}$ to have an infinite number of solutions for $\vec{u}$, for example if

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \end{pmatrix} \quad , \quad \vec{b} = \begin{pmatrix} 6 \\ 12 \end{pmatrix},$$

then

$$AA^T = \begin{pmatrix} 6 & 12 \\ 12 & 24 \end{pmatrix}, \quad therefore \quad \vec{u} = \begin{pmatrix} 1-2v \\ v \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + v\begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Does this mean that there are infinitely many minimum length solutions? That would seem to contradict our expectation that we are finding "the one" with smallest length. If we follow through to find

$$\vec{x}_{ml} = A^T\vec{u} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ -1 & -2 \end{pmatrix}\begin{pmatrix} 1-2v \\ v \end{pmatrix} = \begin{pmatrix} 1-2v+2v \\ 2-4v+4v \\ -1+2v-2v \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix},$$

however, we find that the free variable in $\vec{u}$ disappears. It turns out that our expectation is indeed correct; the minimum length solution is always unique even if $AA^T\vec{u} = \vec{b}$ has an infinite number of solutions.

It frequently happens that AA^T is invertible, in which case we can write $\vec{u} = (AA^T)^{-1}\vec{b}$ and therefore $\vec{x}_{ml} = A^T(AA^T)^{-1}\vec{b}$. The matrix $A^r = A^T(AA^T)^{-1}$ plays a similar role to the matrix A^{-1} when A itself is invertible. In fact A^r has the property that when it multiplies A on the right it yields $AA^r = I$, so we sometimes call it the *right semi-inverse* of A. However, A^rA is not necessarily equal to I, so a right semi-inverse is not necessarily a genuine inverse.

Preview of coming attraction. A matrix A may have the property that A^TA is invertible, even if neither A nor AA^T is invertible. Then the matrix $A^l = (A^TA)^{-1}A^T$ has the property that $A^lA = I$, and is called the *left semi-inverse* of A. The left semi-inverse arises when the system $A\vec{x} = \vec{b}$ has no solutions, and we seek the *least squares best approximation*, as described in chapter 4 of Olver and Shakiban. When the left semi-inverse exists it gives the least-squares best approximate solution as $\vec{x}ls = A^l\vec{b}$.

As mentioned, the left semi-inverse (if it exists) is not necessarily a genuine inverse. If a matrix does have both right and left semi-inverses, then it must be invertible, and the one-sided semi-inverses are both equal to the genuine inverse.

Distant future. When we study the Singular Value Decomposition (SVD, see chapter 8 in Olver and Shakiban) we will learn about a matrix A^+ called the *pseudoinverse* of A. The pseudoinverse always exists for any matrix A whatever (as long as at least one entry is nonzero), and coincides with whichever of the right and/or left and/or genuine inverses happen to exist. Given the system $A\vec{x} = \vec{b}$, we can therefore always find $\vec{x}lsml = A^+\vec{b}$, the minimum length least squares best approximate solution.

Not only does the SVD give us a completely guaranteed foolproof "solution", whatever the properties of the original problem, it has a lot of properties that are useful in many other situations. Watch this space!