We have seen several least squares problems thus far. For example, recall lectures 5 for fitting lines, lecture 6 for finding a vanishing point, lecture 7 for finding corners, and lecture 11 for image registration. We will see more in the upcoming lectures. For this reason it is good to have a more general understanding of least squares problems and how to solve them.

[We will not examine you on today's material. We present it mainly for completeness, and so that you can understand the fundamentals of the methods that will be being used in the next few lectures. As I (ML) mentioned in class, the next few lectures will be challenging because you need to understand both the problems being solved, as well as the mathematical techniques being used.]

Version 1: (linear regression)

Given an $m \times n$ matrix $\mathbf{A}$ with $m > n$ and a non-zero m-vector, $\mathbf{b} \neq \mathbf{0}$, minimize $\| \mathbf{Au} - \mathbf{b} \|_2$.

Note that the $\| \cdot \|_2$ is the L_2 norm. Minimizing the L_2 norm is equivalent to minimizing the sum squares of the elements of the vector $\mathbf{Au}$, i.e. the L_2 norm is just the square root of the sum of squares of the elements of $\mathbf{Au}$. Two examples were the vanishing point estimation problem of lecture 6, and the image registration problem of lecture 11.

Let's give a general solution to this problem. We want to solve for $\mathbf{u} \in \Re^n$. We expand:

$$\| \mathbf{Au} - \mathbf{b} \|^2 \ = \ (\mathbf{Au} - \mathbf{b})^T(\mathbf{Au} - \mathbf{b}) \ = \ \mathbf{u}^T\mathbf{A}^T\mathbf{Au} - 2\mathbf{b}^T\mathbf{Au} + \mathbf{b}^T\mathbf{b}$$

and then take partial derivatives with respect to the n $\mathbf{u}$ variables and set them to 0. If you write out the above matrix and vector products as summations of the various elements, you will get:

$$2\mathbf{A}^T\mathbf{Au} - 2\mathbf{A}^T\mathbf{b} = 0$$

or

$$\mathbf{A}^T\mathbf{Au} = \mathbf{A}^T\mathbf{b} \tag{1}$$

which are called the *normal equations*. Assume the columns of the $m \times n$ matrix $\mathbf{A}$ have full rank n, i.e. they are linearly independent. In that case, one can show $\mathbf{A}^T\mathbf{A}$ is invertible. Then, the solution is

$$\mathbf{u} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}.$$

What is the geometric interpretation of Eq. (1)? Since $\mathbf{b}$ is an m-dimensional vector and $\mathbf{A}$ is an $m \times n$ matrix, we can *uniquely* write $\mathbf{b}$ as a sum of a vector in the column space of $\mathbf{A}$ and a vector in the space orthogonal to the column space of $\mathbf{A}$. To minimize $\| \mathbf{Au} - \mathbf{b} \|$, by definition we find the $\mathbf{u}$ such that the distance from $\mathbf{Au}$ to $\mathbf{b}$ is as small as possible. This is done by choosing $\mathbf{u}$ such that $\mathbf{Au}$ is the component of $\mathbf{b}$ that lies in the column space of $\mathbf{A}$, that is, $\mathbf{Au}$ is the orthogonal projection of $\mathbf{b}$ to the column space of $\mathbf{A}$. Note that if $\mathbf{b}$ already belonged in the column space of $\mathbf{A}$ then $\| \mathbf{Au} - \mathbf{b} \|$ would be 0 and there would be an exact solution.

Pseudoinverse of A

The $n \times m$ matrix that maps $\mathbf{b}$ to $\mathbf{u}$ above is called the *pseudoinverse* of $\mathbf{A}$:

$$\mathbf{A}^+ \equiv (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T.$$

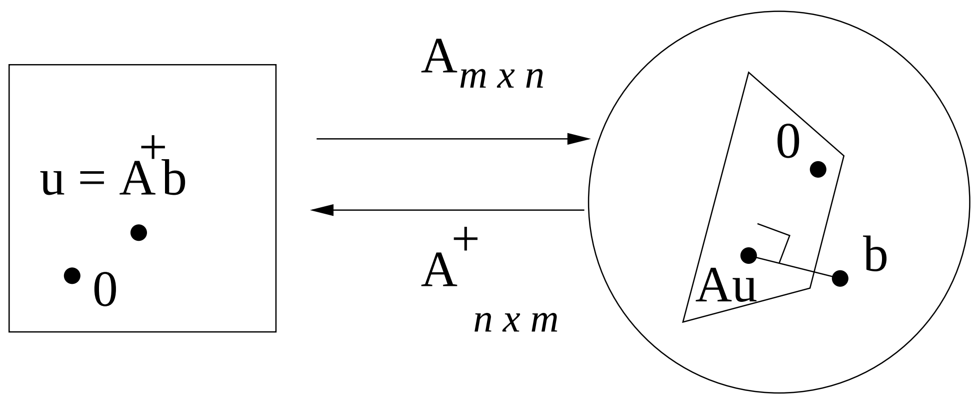
Recall we are assuming $m > n$, so that $\mathbf{A}$ maps from a lower dimensional space to a higher dimensional space. By inspection,

$$\mathbf{A}^+\mathbf{A} = \mathbf{I}.$$

Moreover,

$$\mathbf{A}\mathbf{A}^+ = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T.$$

The pseudoinverse maps in the reverse direction of $\mathbf{A}$, namely it maps $\mathbf{b}$ in an m-D space to some $\mathbf{u}$ in an n-D space. Rather than inverting the mapping $\mathbf{A}$, the pseudoinverse $\mathbf{A}^+$ only inverts the component of $\mathbf{b}$ that belongs to the column space of $\mathbf{A}$, and it nulls out any component of $\mathbf{b}$ that is orthogonal to the column space of $\mathbf{A}$. Specifically, $\mathbf{A}\mathbf{A}^+$ projects any vector $\mathbf{b} \in \Re^m$ onto the column space of $\mathbf{A}$, that is, removing from $\mathbf{b}$ the component that is orthogonal to the column space of $\mathbf{A}$.



Version 2: (total least squares)

We saw a second version of least squares problem. An example was fitting a line to a set of points in the plane, where we were minimizing the perpendicular distance to the line. Here is the general formulation of this problem. We will see several other problems of this form in upcoming lectures.

Given an $m \times n$ matrix $\mathbf{A}$, where $m > n$, find a unit length vector $\mathbf{u}$ that minimizes $\| \mathbf{Au} \|$.

Note that if we don't restrict the minimization to be for $\mathbf{u}$ of unit length, then the minimum is achieved when $\mathbf{u} = \mathbf{0}$ which would be uninteresting.

We are trying to minimize $\mathbf{u}^T\mathbf{A}^T\mathbf{A}\mathbf{u}$. By inspection, if $\mathbf{u}$ were one of the eigenvectors of $\mathbf{A}^T\mathbf{A}$, i.e. $\mathbf{A}^T\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$, then we would choose the $\mathbf{u}$ with the smallest eigenvalue λ, that is,

$$\mathbf{u}^T\mathbf{A}^T\mathbf{A}\mathbf{u} \;=\; \lambda\mathbf{u}^T\mathbf{u} \;=\; \lambda.$$

What about for more general $\mathbf{u}$? For this, we need to understand a bit more about our matrix $\mathbf{A}$ and the matrix $\mathbf{A}\mathbf{A}^T$.

The $n \times n$ matrix $\mathbf{A}^T\mathbf{A}$ is symmetric and positive semi-definite.[1] One can show using basic linear algebra $\mathbf{A}^T\mathbf{A}$ has an orthonormal and complete set of eigenvectors, which we will denote by $\mathbf{v}_i$. Therefore, any vector $\mathbf{u}$ can be written as a sum of these eigenvectors $\mathbf{u} = \sum_i a_i \mathbf{v}_i$. It is then straightforward to show that

$$\mathbf{u}^T\mathbf{A}^T\mathbf{A}\mathbf{u} = \sum_i a_i \mathbf{v_i}^T \mathbf{A}^T\mathbf{A}\mathbf{v_i} = \sum_i a_i \lambda_\mathbf{i}$$

and this quantity is minimized when λ_k is the smallest eigenvalue and $a_k = 1$ and $a_i = 0$ for $i \neq k$. That is, the unit length vector $\mathbf{u}$ that minimizes $\mathbf{u}^T\mathbf{A}^T\mathbf{A}\mathbf{u}$ is the eigenvector of $\mathbf{A}^T\mathbf{A}$ with smallest eigenvalue.

SVD (Singular Value Decomposition)

One can solve such a minimization problem by directly finding the eigenvectors and eigenvalues of $\mathbf{A}^T\mathbf{A}$. Alternatively, one can find them as follows.

Let $\mathbf{V}$ be an $n \times n$ matrix whose columns are the orthonormal eigenvectors $\mathbf{v}_i$ of $\mathbf{A}^T\mathbf{A}$. Since the eigenvalues λ_i are non-negative, we can write them as σ_i^2, that is, $\sigma_i = \sqrt{\lambda_i}$. We can define the $n \times n$ diagonal matrix $\mathbf{\Sigma}$ such that $\Sigma_{ii} = \sigma_i$ on the diagonal. The elements σ_i are called the *singular values* of $\mathbf{A}$. Note that $\mathbf{\Sigma}^2$ is an $n \times n$ diagonal matrix, and

$$\mathbf{A}^T\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Sigma}^2.$$

Next, define a matrix $m \times n$ matrix $\mathbf{U}$ such that:

$$\mathbf{U} = \mathbf{A}\mathbf{V}\mathbf{\Sigma}^{-1}. \tag{2}$$

where $\mathbf{\Sigma}^{-1}$ is the inverse of $\mathbf{\Sigma}$, that is, the $n \times n$ diagonal matrix with elements $\frac{1}{\sigma_i}$. Then

$$\mathbf{U}^T\mathbf{U} = \mathbf{\Sigma}^{-1}\mathbf{V}^\mathbf{T}\mathbf{A}^\mathbf{T}\mathbf{A}\mathbf{V}\mathbf{\Sigma}^{-1} = \mathbf{\Sigma}^{-1}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^2\mathbf{\Sigma}^{-1} = \mathbf{I}.$$

So, we see that $\mathbf{U}$ is an orthonormal matrix, i.e. the columns of $\mathbf{U}$ are orthogonal and of unit length. One usually writes (2) as follows, by using the fact that $\mathbf{U}^{-1} = \mathbf{U}^T$ and $\mathbf{V}^{-1} = \mathbf{V}^T$:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathbf{T}$$

which is called the *singular value decomposition* of $\mathbf{A}$. So if you can compute the SVD of $\mathbf{A}$, that is, if you have the matrices $\mathbf{U}$ and $\mathbf{V}$ and $\mathbf{\Sigma}$ then you can choose the eigenvector (column of $\mathbf{V}$) with smallest eigenvalue (corresponding smallest singular value).

Note that Matlab has a function `svd` which computes the singular value decomposition. One can use `[U, S, V] = svd(A)`.

[1] It is obvious that $\mathbf{A}^T\mathbf{A}$ is symmetric, and it is easy to see that the eigenvalues of $\mathbf{A}^T\mathbf{A}$ are non-negative since $\mathbf{u}^T\mathbf{A}^T\mathbf{A}\mathbf{u} \geq 0$ for any real $\mathbf{u}$.

