MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering and Computer Science

6.241: Dynamic Systems—Fall 2003

RECITATION 2

Least Squares Solutions of Linear Equations

---

1 Outline

In this second recitation, we will address the following topics:

- Review of the projection theorem.
- Derivation of the projection operator on the range of a matrix, using completion of squares.
- Solutions of under and over constrained linear equations.

Notice that this recitation note is an alternative discussion of the least-squares problem which is more difficult than the one you can find in the lecture notes. We suggest that you go through it as an exercise that will give you a different view and more insight on the topic.

2 Projection Theorem

Consider the usual Euclidean space $\mathbb{R}^2$, an arbitrary point $\vec{y} \in \mathbb{R}^2$ and a line $\mathbb{S}$ that goes through the origin. We know that the point $\vec{y}^S \in \mathbb{S}$ that is closest to $\vec{y}$ satisfies $\vec{y} - \vec{y}^S \in \mathbb{S}^\perp$. This geometrically intuitive property can be generalized to other inner-product spaces. Besides the explanatory aid that we can get from it, the *orthogonality principle* is very useful on the solution of general minimum distance or *least squares* problems.

The generalization of the minimum distance problem is achieved through the following the projection theorem. It comprises two parts: the **existence and uniqueness** and the **the orthogonality principle**. In the lecture notes, the orthogonality principle is proven and subsequently used to derive explicit solutions to the least squares problem. By deriving those explicit solutions, the first part of the projection theorem (existence and uniqueness) is also solved.

Theorem 2.1 (Part 1- Existence and Uniqueness) *Let $\mathbb{V}$ be a Hilbert (complete [1] inner-product) space and $\mathbb{S} \subset \mathbb{V}$ be a finite dimensional subspace of $\mathbb{V}$. Then, for any given $\vec{y} \in \mathbb{V}$ the following problem has a unique solution:*

$$\vec{y}^S = arg \min_{\vec{s} \in \mathbb{S}} \|\vec{y} - \vec{s}\| \tag{1}$$

Theorem 2.2 (Part 2- Orthogonality principle) *Let $\mathbb{V}$ be a inner-product space and $\mathbb{S} \subset \mathbb{V}$ be a subspace of $\mathbb{V}$. Then for any given $\vec{y} \in \mathbb{V}$ the following two conditions are equivalent:*

---

[1] Complete means that we can use the norm of the space to define converging sequences and the notion of limit. You do not have to worry about this technicality because in this course we will only deal with complete spaces. So, you can just think of a Hilbert Space as an Inner-Product space.

- $\vec{y}^S = arg\min_{\vec{s}\in\mathbb{S}} \|\vec{y}-\vec{s}\|$
- $\vec{y} - \vec{y}^S \in \mathbb{S}^\perp$

In this recitation, we will use an alternative technique to solve the minimum distance problem for the case when $\mathbb{V} = \mathbb{R}^n$ is an inner-product space with inner product given by:

$$< \vec{v}_1, \vec{v}_2 >= \vec{v}_1' Q \vec{v}_2 \tag{2}$$

where Q is a symmetric and positive definite matrix. If Q is the identity matrix, then the inner product is the usual one for Euclidean spaces.

Definition 2.1 (Projection operator) *Given a Hilbert Space $\mathbb{V}$ and a finite dimensional subspace $\mathbb{S}$, define the projection operator $\mathcal{P}_\mathbb{S} : \mathbb{V} \Longrightarrow \mathbb{S}$ as:*

$$\mathcal{P}_\mathbb{S}(\vec{y}) = arg\min_{\vec{s}\in\mathbb{S}} \|\vec{y}-\vec{s}\| \tag{3}$$

The following are consequences of the projection theorem (You should try to prove them on your own):

- **Very important** Every vector $\vec{y} \in \mathbb{V}$ can be uniquely expressed in the form:

  $$\vec{y} = \vec{y}^S + \vec{y}^\perp \tag{4}$$

  where $\vec{y}^S = \mathcal{P}_\mathbb{S}(\vec{y}) \in \mathbb{S}$ and $\vec{y}^\perp \in \mathbb{S}^\perp$. Can you picture this in $\mathbb{R}^2$?

- We can use the projection theorem to prove that $\mathcal{P}_\mathbb{S}$ is linear! This is very important because we know that **if $\mathbb{V}$ is finite dimensional then $\mathcal{P}_\mathbb{S}$ has a matrix representation**.

2.1 Derivation of the projection operator for $\mathbb{V} = \mathbb{R}^n$ (Matrix representation)

We will use a technique called completion of squares to compute the matrix representation of $\mathcal{P}_\mathbb{S}$.

Let us start by considering the following:

- If $\mathbb{S}$ is a subspace of $\mathbb{V} = \mathbb{R}^n$, we can form a basis $\mathbb{B} = \{\vec{b}_1, \ldots, \vec{b}_m\}$, where $m \leq n$ is the dimension of $\mathbb{S}$.

- We can also construct the following matrix $B \in \mathbb{R}^{n\times m}$:

  $$B = \left[\vec{b}_1 \ldots \vec{b}_m\right] \tag{5}$$

  where we assume that $\vec{b}_1 \ldots \vec{b}_m$ are the elements of the base $\mathbb{B}$ represented as column vectors.

- It is also clear that $\mathbb{S} = span(\mathbb{B}) = \mathcal{R}(B)$.

- Consequently you should convince yourself that we can restate the minimum distance problem as:

$$arg\min_{\vec{s}\in\mathbb{S}} \|\vec{y}-\vec{s}\| = B\left(arg\min_{\vec{\alpha}\in\mathbb{R}^m} \|\vec{y}-B\vec{\alpha}\|\right) \tag{6}$$

In order to solve the minimum distance problem above, we write:

$$arg \min_{\vec{\alpha}\in\mathbb{R}^m} \|\vec{y}-B\vec{\alpha}\| = arg \min_{\vec{\alpha}\in\mathbb{R}^m} \|\vec{y}-B\vec{\alpha}\|^2 = arg \min_{\vec{\alpha}\in\mathbb{R}^m} <\vec{y}-B\vec{\alpha}, \vec{y}-B\vec{\alpha}> \tag{7}$$

but, we also know that:

$$arg \min_{\vec{\alpha}\in\mathbb{R}^m} <\vec{y}-B\vec{\alpha}, \vec{y}-B\vec{\alpha}> = arg \min_{\vec{\alpha}\in\mathbb{R}^m} (\vec{y}-B\vec{\alpha})'Q(\vec{y}-B\vec{\alpha}) \tag{8}$$

This way we have to solve:

$$\vec{\alpha}^* = arg \min_{\vec{\alpha}\in\mathbb{R}^m} (\vec{y}-B\vec{\alpha})'Q(\vec{y}-B\vec{\alpha}) \tag{9}$$

The following is a useful formula known as completion of squares:

$$\begin{aligned}(\vec{y}-B\vec{\alpha})'Q(\vec{y}-B\vec{\alpha}) &= \\ &\vec{y}'Q\vec{y} - \vec{y}'QB(B'QB)^{-1}QB'\vec{y} + \left(\vec{\alpha}-(B'QB)^{-1}B'Q\vec{y}\right)'(B'QB)\left(\vec{\alpha}-(B'QB)^{-1}B'Q\vec{y}\right)\end{aligned} \tag{10}$$

Now notice that in the equality above, the first two terms on the right hand side are fixed. The last term is the only one that depends on the optimization variable $\vec{\alpha}$. Since $(B'QB)$ is also positive definite, we conclude that the minimum is achieved by choosing $\vec{\alpha}$ such that the last term is zero:

$$\vec{\alpha}^* = (B'QB)^{-1}B'Q\vec{y} \tag{11}$$

That gives the final expression for the projection operator $\mathcal{P}_{\mathbb{S}}$, with $\mathbb{S} = range(B)$:

$$\mathcal{P}_{\mathbb{S}}(\vec{y}) = B\vec{\alpha}^* = B(B'QB)^{-1}B'Q\vec{y} \tag{12}$$

As we expected the projection operator has a matrix representation given by:

$$\mathcal{P}_{\mathbb{S}}(\vec{y}) = P_{\mathbb{S}}\vec{y} \tag{13}$$

where $P_{\mathbb{S}} \in \mathbb{R}^{n\times n}$ is the following matrix:

$$P_{\mathbb{S}} = B(B'QB)^{-1}B'Q \tag{14}$$

2.2 (Just a Side) Deriving the projection operator by means of orthogonal projection

We can show that:

$$\vec{y}^S = \sum_{i=1}^{m} <\vec{y}, \vec{z}_i> \frac{\vec{z}_i}{\|\vec{z}_i\|} \tag{15}$$

provided that $\{\vec{z}_1, \ldots, \vec{z}_m\} = \mathbb{Z}^S$ is an orthogonal basis for $\mathbb{S}$. Does this remind you the formula of the Fourier transform?

Also, note that every inner-product space with a basis $\{\vec{b}_1, \ldots, \vec{b}_m\} = \mathbb{B}^S$ has also an orthogonal basis $\{\vec{z}_1, \ldots, \vec{z}_m\} = \mathbb{Z}^S$. Such basis can be constructed using the following algorithm, called Gram-Schmidt: (check it yourself)

$$\vec{z}_i = \begin{cases}\vec{b}_1 & \text{if i=1} \\ \vec{b}_i - \sum_{j=1}^{i-1} <\vec{b}_i, \vec{z}_j> \vec{z}_j & \text{Otherwise}\end{cases} \tag{16}$$

3 Using the projection operator to solve under and over constrained linear equations

We will use the projection operator $\mathcal{P}_{\mathbb{S}}$ and its matrix representation $P_{\mathbb{S}}$ to derive the solutions of the under and over-constrained linear equations.

We start by recalling that:

- $\mathbb{V} = \mathbb{R}^n$ is equipped with the following inner product:

$$< \vec{v}_1, \vec{v}_2 >= \vec{v}_1' Q \vec{v}_2 \tag{17}$$

- $\mathbb{S} \subset \mathbb{R}^n$ is a subspace of dimension $m \leq n$
- We can select a matrix $B \in \mathbb{R}^{n \times m}$ such that $\mathbb{S} = range(B)$
- $\vec{y}^S = arg\min_{\vec{s} \in \mathbb{S}} \|\vec{y} - \vec{s}\| \Longleftrightarrow \vec{y} - \vec{y}^S \in \mathbb{S}^{\perp}$
- $\vec{y}^S = P_{\mathbb{S}} \vec{y}$, where $P_{\mathbb{S}} = B(B'QB)^{-1}B'Q$

3.1 Over-constrained Linear Equations

The over-constrained problem appears when we have a matrix $A \in \mathbb{R}^{n \times m}$, but $rank(A) = m < n$. In that situation, the linear equation $\vec{y} = A\vec{x}$ may not have a solution. To see that take $\vec{y} \notin \mathcal{R}(A)$. For a given $\vec{y} \in \mathbb{R}^n$, the best we can do is to find $\vec{x}$ that leads to the minimum $\|\vec{y} - A\vec{x}\|$.

Notice that we already solved this problem when computing the projection operator and the optimal solution is:

$$\vec{x}^* = arg\min_{\vec{x} \in \mathbb{R}^m} \|\vec{y} - A\vec{x}\| = (A'QA)^{-1}A'Q\vec{y} \tag{18}$$

You can also show that the same solution would be obtained by requiring that $\vec{y} - A\vec{x}^* \in \mathcal{R}(A)^{\perp}$. That leads to the equation:

$$P_{\mathcal{R}(A)}(\vec{y} - A\vec{x}^*) = \vec{0} \tag{19}$$

Indeed, by substituting $P_{\mathcal{R}(A)} = A(A'QA)^{-1}A'Q$, we get:

$$A((A'QA)^{-1}A'Q\vec{y} - \vec{x}^*) = \vec{0} \tag{20}$$

Given that $rank(A) = m$, we conclude that $\mathcal{N}(A) = \{\vec{0}\}$ and the only solution to the equation above is:

$$\vec{x}^* = (A'QA)^{-1}A'Q\vec{y} \tag{21}$$

3.2 Under-constrained linear equations

Consider the following linear equation:

$$\vec{y} = A\vec{x} \tag{22}$$

where $A \in \mathbb{R}^{n \times m}$, with $m > n$ and $rank(A) = n$.

Now, we partition A in the usual column form:

$$A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_m \end{bmatrix} \tag{23}$$

Since $\vec{a}_i \in \mathbb{R}^n$ and $m > n$, we conclude that not all the m columns can be linearly independent. Linear dependence means that we can form a linear combination of the columns of A that gives the zero vector, or equivalently we can state:

$$dim(\mathcal{N}(A)) \geq 1 \tag{24}$$

This means that if $\vec{x}_1$ is a solution ($\vec{y} = A\vec{x}_1$), then we can choose a non-zero vector $\vec{v}_0 \in \mathcal{N}(A)$ and get another solution $\vec{x}_2 = \vec{x}_1 + \vec{v}_0$. To check that, note that $A\vec{x}_2 = A\vec{x}_1 + A\vec{v}_0 = A\vec{x}_1 + \vec{0} = \vec{y}$. That means that the undetermined linear equation has a **non-unique solution**. In order to select the solution with the smallest norm, we solve the following problem:

$$\vec{x}^* = arg \min_{\vec{x}:A\vec{x}=\vec{y}} \|\vec{x}\| \tag{25}$$

For reasons that will be clear latter, we rewrite the linear equation as:

$$\vec{y} = D'Q\vec{x} \text{ with } D = Q^{-1}A' \tag{26}$$

Still, since Q is positive definite, it is invertible and $rank(D) = rank(A)$. Also $D \in \mathbb{R}^{n\times m}$ and $dim(\mathcal{N}(D)) = dim(\mathcal{N}(A)) \geq 1$, meaning that we still have non-unique solutions.

From the definition of the inner-product, we know that we can equivalently express $\vec{y} = A\vec{x} = D'Q\vec{x}$ as:

$$\vec{y} = \begin{bmatrix} < \vec{d}_1, \vec{x} > \\ \vdots \\ < \vec{d}_n, \vec{x} > \end{bmatrix} \tag{27}$$

where $\vec{d}_i \in \mathbb{R}^m$ **are the columns of** D. This can be used to show that $\mathcal{N}(A)$ comprises the vectors which are orthogonal to the columns of D, according to the inner product $< \vec{v}_1, \vec{v}_2 >= \vec{v}_1'Q\vec{v}_2$. Since the set of orthogonal vectors is a subspace, we can state:

$$\mathcal{N}(A) = \mathcal{R}(D)^{\perp} \tag{28}$$

Now, we can use the projection theorem to show that any solution $\vec{x}$ can be written as:

$$\vec{x} = \vec{x}^* + \vec{x}^{\perp} \tag{29}$$

where $\vec{x}^* \in \mathcal{R}(D)$ and $\vec{x}^{\perp} \in \mathcal{N}(A) = \mathcal{R}(D)^{\perp}$. Notice that, since $\vec{x}^*$ is not in $\mathcal{N}(A)$ we cannot change it and still satisfy $\vec{y} = A\vec{x} = A\vec{x}^*$. Our only degree of freedom is on the choice of $\vec{x}^{\perp}$. But notice that $< \vec{x}^*, \vec{x}^{\perp} >= 0$, implying:

$$\|\vec{x}\|^2 = \|\vec{x}^*\|^2 + \|\vec{x}^{\perp}\|^2 \tag{30}$$

So, our choice is obvious!. If we want to minimize the norm and $\vec{x}^*$ is fixed, the optimal solution is $\vec{x}^{\perp} = \vec{0}$. That can also be stated as:

$$\vec{x}^* = arg \min_{\vec{x}:A\vec{x}=\vec{y}} \|\vec{x}\| \in \mathbb{R}(D) \tag{31}$$

Now, to compute $\vec{x}^*$ we use the fact that:

$$\vec{x}^* \in \mathcal{R}(D) \implies P_{\mathcal{R}(D)}\vec{x}^* = \vec{x}^* \tag{32}$$

By substitution on the formula, we get that the projection operator is $P_{\mathcal{R}(D)} = D(D'QD)^{-1}D'Q$. Using $D = Q^{-1}A'$, we get $P_{\mathcal{R}(D)} = Q^{-1}A'(AQ^{-1}A')^{-1}A$. Direct substitution on the equation above gives:

$$P_{\mathcal{R}(D)}\vec{x}^* = \vec{x}^* \implies Q^{-1}A'(AQ^{-1}A')^{-1}A\vec{x}^* = \vec{x}^* \tag{33}$$

Now, we just have to substitute $\vec{y} = A\vec{x}^*$ to get (finally):

$$\vec{x}^* = Q^{-1}A'(AQ^{-1}A')^{-1}\vec{y} \tag{34}$$