Classical Numerical Analysis, Chapter 05

Abner J. Salgado and Steven M. Wise

asalgad1@utk.edu swise1@.utk.edu
University of Tennessee

Chapter 05, Part 1 of 2
Linear Least Squares Problem

Data Fitting

Suppose that we are given a table of values

$$(x_k, y_k), \quad k = 1, \ldots, n,$$

that is obtained, say, by a series of measurements.

Find a simple function — a linear function, for example, $y = c_1 x + c_0$, where $c_0, c_1 \in \mathbb{C}$ — that *fits* the data in some exact or approximate sense. If we demand that it matches the data exactly, then

$$y_k = c_1 x_k + c_0, \quad k = 1, \ldots, n.$$

This can also be expressed in vector form as

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_0 \end{bmatrix}.$$

Data Fitting

This is equivalent to the system of linear equations of the form $\mathsf{A}\mathbf{c} = \mathbf{y}$ with

$$\mathsf{A} = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \in \mathbb{C}^{n\times 2}, \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_0 \end{bmatrix} \in \mathbb{C}^2, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{C}^n.$$

Usually n is much larger than 2, so that we end up with an *overdetermined* system of equations, i.e., there are more equations than unknowns. There is no solution in general.

On the other hand, if we only approximately enforce the matching conditions

$$y_k \approx c_1 x_k + c_0, \quad k = 1, \ldots, n,$$

then it is not clear how to proceed.

Linear Least Squares: Full Rank Setting

Generalized and Least Squares Solutions

Idea: There may be no solution to $A\mathbf{x} = \mathbf{f}$. So, let us find $\mathbf{x}$ so that $\mathbf{f} - A\mathbf{x}$ is as small as possible in some norm.

Definition

Suppose that $\|\cdot\| : \mathbb{C}^m \to \mathbb{R}$ is a norm. Given $A \in \mathbb{C}^{m\times n}$, $m \geq n$, and $\mathbf{f} \in \mathbb{C}^m$ we say that $\mathbf{x} \in \mathbb{C}^n$ is a **weak** or **generalized** solution of the system $A\mathbf{x} = \mathbf{f}$ iff

$$\mathbf{x} \in \operatorname*{argmin}_{\mathrm{w}\in\mathbb{C}^n} \|\mathbf{r}(\mathbf{w})\| = \operatorname*{argmin}_{\mathrm{w}\in\mathbb{C}^n} \|\mathbf{f} - A\mathbf{w}\|.$$

We say that $\mathbf{x} \in \mathbb{C}^n$ is a **least squares** solution of the system $A\mathbf{x} = \mathbf{f}$ iff

$$\mathbf{x} \in \operatorname*{argmin}_{\mathrm{w}\in\mathbb{C}^n} \|\mathbf{r}(\mathbf{w})\|^2_{\ell^2(\mathbb{C}^m)} = \operatorname*{argmin}_{\mathrm{w}\in\mathbb{C}^n} \|\mathbf{f} - A\mathbf{w}\|^2_{\ell^2(\mathbb{C}^m)}.$$

When these minima exist and are unique, we replace $\in$ with $=$.

Full-Rank Coefficient Matrix $A \in \mathbb{C}^{m \times n}$, $m \geq n$

Lemma

Let $A \in \mathbb{C}^{m \times n}$ with $m \geq n$, then $A^H A$ is Hermitian positive definite (HPD) iff A is full rank, i.e., $\operatorname{rank}(A) = n$.

Proof.

Clearly the matrix $A^H A$ is Hermitian. By construction this matrix is also nonnegative definite since, for any $\mathbf{x} \in \mathbb{C}^n$,

$$(A^H A\mathbf{x}, \mathbf{x})_2 = (A\mathbf{x}, A\mathbf{x})_2 = \|A\mathbf{x}\|_2 \geq 0.$$

($\Longrightarrow$): Suppose that $A^H A$ is HPD and, to reach a contradiction, that $\operatorname{rank}(A) < n$. This is equivalent to say that there is a nonzero $\mathbf{x} \in \mathbb{C}^n$ such that $A\mathbf{x} = \mathbf{0}$. But then we must have $A^H A\mathbf{x} = \mathbf{0}$ and $(A^H A\mathbf{x}, \mathbf{x})_2 = 0$, contradicting the assumption that $A^H A$ is positive definite.

Proof, Cont.

($\Longleftarrow$): Let us now assume that A is of full rank. To reach a contradiction, suppose that $\mathsf{A}^{\mathsf{H}}\mathsf{A}$ is not positive definite. There must be a nonzero $\mathbf{x} \in \mathbb{C}^n_\star$ for which

$$0 = (\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{x}, \mathbf{x})_2 = (\mathsf{A}\mathbf{x}, \mathsf{A}\mathbf{x})_2 = \|\mathsf{A}\mathbf{x}\|_2^2,$$

which implies that

$$\mathsf{A}\mathbf{x} = \mathbf{0} \implies \mathbf{x} \in \ker(\mathsf{A}).$$

This contradicts the assumption that A has full rank. $\square$

Theorem (least squares: real case)

Let $\mathsf{A} \in \mathbb{R}^{m\times n}$*, with* $m \geq n$ *and* $\mathrm{rank}(\mathsf{A}) = n$*, and let* $\mathbf{f} \in \mathbb{R}^m$*. The vector* $\mathbf{x} \in \mathbb{R}^n$ *is the unique least squares solution to* $\mathsf{A}\mathbf{x} = \mathbf{f}$*, i.e.,*

$$\mathbf{x} = \underset{\mathbf{w}\in\mathbb{R}^n}{\mathrm{argmin}} \, \|\mathbf{f} - \mathsf{A}\mathbf{w}\|_{\ell^2(\mathbb{R}^m)}^2 , \tag{1}$$

iff $\mathbf{x}$ *is the unique solution to the normal equation*

$$\mathsf{A}^{\intercal}\mathsf{A}\mathbf{x} = \mathsf{A}^{\intercal}\mathbf{f}. \tag{2}$$

Proof.

($\Longrightarrow$): Suppose that $\mathbf{x} \in \mathbb{R}^n$ solves the least squares problem and set $\mathbf{r} = \mathbf{f} - \mathsf{A}\mathbf{x}$. Now fix $\mathbf{y} \in \mathbb{R}^n$, and define

$$g(s) = \Phi(\mathbf{x} + s\mathbf{y}),$$

for all $s \in \mathbb{R}$, where

$$\Phi(\mathbf{z}) = \|\mathbf{r}(\mathbf{z})\|_{\ell^2(\mathbb{C}^m)}^2, \quad \mathbf{r}(\mathbf{z}) = \mathbf{f} - \mathsf{A}\mathbf{z}, \quad \forall\, \mathbf{z} \in \mathbb{C}^n.$$

Proof, Cont.

Then, for any $s \in \mathbb{R}$,

$$g(0) = \Phi(\mathbf{x}) \leq \Phi(\mathbf{x} + s\mathbf{y}) = g(s),$$

since $\mathbf{x}$ is a least squares solution. Calculating, we find

$$\begin{aligned} g(s) &= (\mathbf{r} - s\mathsf{A}\mathbf{y})^\intercal (\mathbf{r} - s\mathsf{A}\mathbf{y}) \\ &= \mathbf{r}^\intercal\mathbf{r} - s\mathbf{y}^\intercal\mathsf{A}^\intercal\mathbf{r} - s\mathbf{r}^\intercal\mathsf{A}\mathbf{y} + s^2\mathbf{y}^\intercal\mathsf{A}^\intercal\mathsf{A}\mathbf{y} \\ &= g(0) - 2s\,\mathbf{r}^\intercal\mathsf{A}\mathbf{y} + s^2\mathbf{y}^\intercal\mathsf{A}^\intercal\mathsf{A}\mathbf{y}. \end{aligned}$$

Since $\mathsf{A}^\intercal\mathsf{A}$ is symmetric positive definite, g is a positive quadratic function of one variable, with a global minimum at $s = 0$. Hence,

$$0 = \left.\frac{dg}{ds}\right|_{s=0} = -2\mathbf{r}^\intercal\mathsf{A}\mathbf{y},$$

for arbitrary $\mathbf{y} \in \mathbb{R}^n$. From this condition we conclude that $\mathbf{r}^\intercal\mathsf{A} = \mathbf{0}^\intercal$. This is equivalent to (2).

Proof, Cont.

($\Longleftarrow$): Now, suppose $\mathbf{x} \in \mathbb{R}^n$ solves the normal equation (2), and set $\mathbf{r} = \mathbf{f} - \mathsf{A}\mathbf{x}$. This implies that $\mathbf{r}^\intercal \mathsf{A}\mathbf{y} = 0$, for all $\mathbf{y} \in \mathbb{R}^n$, or, equivalently, $\mathbf{r} \in \mathrm{im}(\mathsf{A})^\perp$. Then

$$\begin{aligned}
\Phi(\mathbf{x}+\mathbf{y}) &= (\mathbf{r} - \mathsf{A}\mathbf{y})^\intercal (\mathbf{r} - \mathsf{A}\mathbf{y}) \\
&= \Phi(\mathbf{x}) - \mathbf{r}^\intercal \mathsf{A}\mathbf{y} - \mathbf{y}\mathsf{A}^\intercal \mathbf{r} + \mathbf{y}^\intercal \mathsf{A}^\intercal \mathsf{A}\mathbf{y} \\
&= \Phi(\mathbf{x}) + \mathbf{y}^\intercal \mathsf{A}^\intercal \mathsf{A}\mathbf{y} \\
&\geq \Phi(\mathbf{x}),
\end{aligned}$$

since $\mathbf{y}^\intercal \mathsf{A}^\intercal \mathsf{A}\mathbf{y} \geq 0$ for any $\mathbf{y}$. More importantly, since $\mathsf{A}^\intercal \mathsf{A}$ is symmetric positive definite (SPD),

$$\Phi(\mathbf{x}+\mathbf{y}) > \Phi(\mathbf{x}), \quad \forall\, \mathbf{y} \in \mathbb{R}^n_\star.$$

Thus $\mathbf{x} \in \mathbb{R}^n$ is a least squares solution. $\square$

Theorem (least squares: complex case)

Let $A \in \mathbb{C}^{m\times n}$*, with* $m \geq n$ *and* $\text{rank}(A) = n$*, and let* $\mathbf{f} \in \mathbb{C}^m$*. The vector* $\mathbf{x} \in \mathbb{C}^n$ *is the unique least squares solution to* $A\mathbf{x} = \mathbf{f}$*, i.e.,*

$$\mathbf{x} = \underset{\mathrm{w}\in\mathbb{C}^n}{\text{argmin}} \, \|\mathbf{f} - A\mathbf{w}\|^2_{\ell^2(\mathbb{C}^m)} \,, \tag{3}$$

iff $\mathbf{x}$ *is the unique solution to the normal equation*

$$A^{\mathrm{H}}A\mathbf{x} = A^{\mathrm{H}}\mathbf{f}. \tag{4}$$

Proof.

($\Longrightarrow$): Suppose that $\mathbf{x} \in \mathbb{C}^n$ solves the least squares problem and set $\mathbf{r} = \mathbf{f} - A\mathbf{x}$. Now fix $\mathbf{y} \in \mathbb{C}^n$, and define

$$g(s,t) = \Phi(\mathbf{x} + z\mathbf{y}) = \|\mathbf{r} - zA\mathbf{y}\|_2^2 \,,$$

for all $z = s + \mathfrak{i}t \in \mathbb{C}$, with $s,\ t \in \mathbb{R}$, $\mathfrak{i} = \sqrt{-1}$. Then, for any $s,\ t \in \mathbb{R}$,

$$g(0,0) = \Phi(\mathbf{x}) \leq \Phi(\mathbf{x} + z\mathbf{y}) = g(s,t),$$

since $\mathbf{x}$ is a least squares solution.

Proof, Cont.

Calculating, we find

$$\begin{aligned} g(s,t) &= (\mathbf{r} - z\mathsf{A}\mathbf{y})^{\mathsf{H}}(\mathbf{r} - z\mathsf{A}\mathbf{y}) \\ &= \mathbf{r}^{\mathsf{H}}\mathbf{r} - \bar{z}\mathbf{y}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathbf{r} - z\mathbf{r}^{\mathsf{H}}\mathsf{A}\mathbf{y} + |z|^2\mathbf{y}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{y} \\ &= g(0,0) - 2\,\Re\left(z\mathbf{r}^{\mathsf{H}}\mathsf{A}\mathbf{y}\right) + \left(s^2 + t^2\right)\mathbf{y}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{y} \\ &= g(0,0) - 2s\,\Re\left(\mathbf{r}^{\mathsf{H}}\mathsf{A}\mathbf{y}\right) + 2t\,\Im\left(\mathbf{r}^{\mathsf{H}}\mathsf{A}\mathbf{y}\right) + \left(s^2 + t^2\right)\mathbf{y}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{y}\ . \end{aligned}$$

Since $\mathsf{A}^{\mathsf{H}}\mathsf{A}$ is Hermitian positive definite, g is a positive quadratic function of two variables. Hence,

$$0 = \left.\frac{\partial g}{\partial s}\right|_{s,t=0} = -2\,\Re\left(\mathbf{r}^{\mathsf{H}}\mathsf{A}\mathbf{y}\right), \quad 0 = \left.\frac{\partial g}{\partial t}\right|_{s,t=0} = 2\,\Im\left(\mathbf{r}^{\mathsf{H}}\mathsf{A}\mathbf{y}\right),$$

for arbitrary $\mathbf{y} \in \mathbb{C}^n$. From these two conditions we conclude that $\mathbf{r}^{\mathsf{H}}\mathsf{A} = \mathbf{0}^{\intercal}$. Hence $\mathbf{r} \in \operatorname{im}(\mathsf{A})^{\perp}$, as desired.

Proof, Cont.

($\Longleftarrow$): This step is more or less the same as in the real case. Suppose $\mathbf{x} \in \mathbb{C}^n$ satisfies $\mathbf{r} \in \operatorname{im}(\mathsf{A})^{\perp}$, and set $\mathbf{r} = \mathbf{f} - \mathsf{A}\mathbf{x}$. Then

$$\begin{aligned}
\Phi(\mathbf{x}+\mathbf{y}) &= (\mathbf{r} + \mathsf{A}\mathbf{y})^{\mathsf{H}} (\mathbf{r} + \mathsf{A}\mathbf{y}) \\
&= \Phi(\mathbf{x}) + \mathbf{r}^{\mathsf{H}}\mathsf{A}\mathbf{y} + \mathbf{y}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathbf{r} + \mathbf{y}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{y} \\
&= \Phi(\mathbf{x}) + \mathbf{y}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{y} \\
&\geq \Phi(\mathbf{x}),
\end{aligned}$$

since $\mathbf{y}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{y} \geq 0$ for any $\mathbf{y}$. And, since $\mathsf{A}^{\mathsf{H}}\mathsf{A}$ is HPD,

$$\Phi(\mathbf{x}+\mathbf{y}) > \Phi(\mathbf{x}), \quad \forall\, \mathbf{y} \in \mathbb{C}^n_{\star}.$$

Thus $\mathbf{x} \in \mathbb{C}^n$ is a least squares solution. $\square$

Least Squares in Other Inner Products

Instead of doing our least squares computations using the norm $\|\cdot\|_{\ell^2(\mathbb{C}^m)}$, we could, in fact, use any norm that on $\mathbb{C}^n$ that arises from an inner product. Suppose that $(\cdot\,,\,\cdot)_w : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ is an inner product and $\|\cdot\|_w$ is the norm induced by this inner product. We could pose the least squares problem with respect to this inner product.

As an example, suppose that $\mathsf{B} \in \mathbb{C}^{n\times n}$ is HPD, and consider the inner product $(\cdot\,,\,\cdot)_{\mathsf{B}}$, defined by

$$(\mathbf{x}, \mathbf{y})_{\mathsf{B}} = \mathbf{y}^{\mathsf{H}}\mathsf{B}\mathbf{x}, \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{C}^n.$$

Define

$$\Phi_{\mathsf{B}}(\mathbf{z}) = \|\mathbf{r}(\mathbf{z})\|_{\mathsf{B}}^2, \quad \mathbf{r}(\mathbf{z}) = \mathbf{f} - \mathsf{A}\mathbf{z}, \quad \forall\, \mathbf{z} \in \mathbb{C}^n. \tag{5}$$

What is the analogue of the normal equations for this case?

Projection Matrices

Projection Matrices

To study the least squares problem in the case that the coefficient matrix $\mathsf{A} \in \mathbb{C}^{m\times n}$, $m \geq 0$ is rank-deficient, that is, $\operatorname{rank}(\mathsf{A}) = r < n$, it helps to introduce projection matrices.

Definition

The square matrix $\mathsf{P} \in \mathbb{C}^{n\times n}$ is called a **projection matrix** iff it is idempotent, that is,

$$\mathsf{P}^2 = \mathsf{P}.$$

The Image of a Projection Matrix P

Proposition

Let $\mathsf{P} \in \mathbb{C}^{n\times n}$ *be a projection matrix and* $\mathbf{v} \in \mathrm{im}(\mathsf{P})$. *Then*

$$\mathsf{P}\mathbf{v} = \mathbf{v}.$$

Proof.

If $\mathbf{v} \in \mathrm{im}(\mathsf{P})$, then there is a vector $\mathbf{w} \in \mathbb{C}^n$ such that $\mathsf{P}\mathbf{w} = \mathbf{v}$. Then

$$\mathsf{P}\mathbf{v} = \mathsf{P}\left(\mathsf{P}\mathbf{w}\right) = \mathsf{P}^2\mathbf{w} = \mathsf{P}\mathbf{w} = \mathbf{v}. \qquad \square$$

Theorem (properties of a projection matrix)

Suppose that $\mathrm{P} \in \mathbb{C}^{n\times n}$ *is a projection matrix. Then* $\mathrm{I}_n - \mathrm{P}$ *is also a projection matrix and*

1. $\mathrm{im}(\mathrm{I}_n - \mathrm{P}) = \ker(\mathrm{P})$;
2. $\ker(\mathrm{I}_n - \mathrm{P}) = \mathrm{im}(\mathrm{P})$;
3. $\mathrm{im}(\mathrm{P}) \cap \ker(\mathrm{P}) = \{\mathbf{0}\}$;
4. $\mathrm{im}(\mathrm{I}_n - \mathrm{P}) \cap \ker(\mathrm{I}_n - \mathrm{P}) = \{\mathbf{0}\}$.

Proof.

We will prove the first property and leave the remaining ones as an exercise. Suppose that $\mathbf{x} \in \mathrm{im}(\mathrm{I}_n - \mathrm{P})$. Then there is a vector $\mathbf{y} \in \mathbb{C}^n$ such that $(\mathrm{I}_n - \mathrm{P})\mathbf{y} = \mathbf{x}$. So,

$$\mathrm{P}\mathbf{x} = \mathrm{P}(\mathrm{I}_n - \mathrm{P})\mathbf{y} = \mathrm{P}\mathbf{y} - \mathrm{P}^2\mathbf{y} = \mathbf{0}$$

Consequently $\mathrm{P}\mathbf{x} = \mathbf{0}$, and $\mathbf{x} \in \ker(\mathrm{P})$.

Proof, Cont.

Suppose now that $\mathbf{x} \in \ker(\mathsf{P})$. Then $\mathsf{P}\mathbf{x} = \mathbf{0}$. Therefore,

$$(\mathsf{I}_n - \mathsf{P})\mathbf{x} = \mathbf{x},$$

which implies that $\mathbf{x} \in \operatorname{im}(\mathsf{I}_n - \mathsf{P})$. $\square$

Sums of Subspaces

Definition

Let $S_1, S_2 \subseteq \mathbb{C}^n$ be subspaces. Recall, that we write $S_1, S_2 \leq \mathbb{C}^n$, for short. Then,

$$S_1 + S_2 = \{\mathbf{w} \in \mathbb{C}^n \mid \mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2, \ \exists\, \mathbf{v}_i \in S_i, \ i = 1, 2\} .$$

Proposition (property of the sum)

Let $S_1, S_2 \leq \mathbb{C}^n$. Then $S_1 + S_2 \leq \mathbb{C}^n$.

Proof.

Exercise. □

Complementary Subspaces

Definition

Suppose $S_1, S_2 \leq \mathbb{C}^n$. If $S_1 + S_2 = \mathbb{C}^n$ and $S_1 \cap S_2 = \{\mathbf{0}\}$, then we call S_1 and S_2 **complementary** subspaces, and we write $S_1 \oplus S_2 = \mathbb{C}^n$.

Theorem (decomposition)

Suppose that $\mathsf{P} \in \mathbb{C}^{n\times n}$ *is a projection matrix. Then* $\operatorname{im}(\mathsf{P}) \oplus \ker(\mathsf{P}) = \mathbb{C}^n$, *i.e.,* $\operatorname{im}(\mathsf{P})$ *and* $\ker(\mathsf{P})$ *are complementary.*

Proof.

This follows from a previous theorem and the trivial decomposition

$$\mathbf{v} = \mathsf{P}\mathbf{v} + \mathbf{v} - \mathsf{P}\mathbf{v} = \underbrace{\mathsf{P}\mathbf{v}}_{\in \operatorname{im}(\mathsf{P})} + \underbrace{(\mathsf{I} - \mathsf{P})\mathbf{v}}_{\in \ker(\mathsf{P})}. \qquad \square$$

Constructing a Projection

Theorem

Let $S_1, S_2 \leq \mathbb{C}^n$ be complementary subspaces. Then, there is a projection matrix $\mathsf{P} \in \mathbb{C}^{n \times n}$ *such that*

$$S_1 = \mathbf{im}(\mathsf{P}) \quad \textit{and} \quad S_2 = \mathbf{ker}(\mathsf{P}).$$

Proof.

Suppose that

$$B_i = \{\mathbf{w}_1^{(i)}, \ldots, \mathbf{w}_{k_i}^{(i)}\} \subset S_i,$$

is a basis for S_i, $i = 1, 2$. Then, it is left to the reader to prove that the set

$$B = B_1 \cup B_2$$

is a basis for $\mathbb{C}^n = S_1 \oplus S_2$.

Proof, Cont.

Now, define a mapping $\mathsf{P} : \mathbb{C}^n \to \mathbb{C}^n$ such that

$$\mathsf{P}(\mathbf{w}_j^{(1)}) = \mathbf{w}_j^{(1)}, \quad j = 1, \ldots, k_1,$$

and

$$\mathsf{P}(\mathbf{w}_j^{(2)}) = \mathbf{0}, \quad j = 1, \ldots, k_2.$$

We require P to be linear, so that

$$\begin{aligned}\mathsf{P}\left(\sum_{j=1}^{k_1} c_j^{(1)}\mathbf{w}_j^{(1)} + \sum_{j=1}^{k_2} c_j^{(2)}\mathbf{w}_j^{(2)}\right) &= \sum_{j=1}^{k_1} c_j^{(1)}\mathsf{P}(\mathbf{w}_j^{(1)}) + \sum_{j=1}^{k_2} c_j^{(2)}\mathsf{P}(\mathbf{w}_j^{(2)}) \\ &= \sum_{j=1}^{k_1} c_j^{(1)}\mathbf{w}_j^{(1)}.\end{aligned}$$

It is now straightforward to prove that $\mathsf{P}^2 = \mathsf{P}$ and $\mathbf{im}(\mathsf{P}) = S_1$ and $\mathbf{ker}(\mathsf{P}) = S_2$, as required. $\square$

Orthogonal Subspaces

Definition

Two subspaces $S_1, S_2 \leq \mathbb{C}^n$ are called **orthogonal** iff

$$(\mathbf{v}_1, \mathbf{v}_2)_{\ell^2(\mathbb{C}^n)} = (\mathbf{v}_1, \mathbf{v}_2)_2 = \mathbf{v}_2^{\mathsf{H}}\mathbf{v}_1 = 0,$$

for all $\mathbf{v}_1 \in S_1$ and $\mathbf{v}_2 \in S_2$.

Proposition (orthogonality)

If $S_1, S_2 \leq \mathbb{C}^n$ are orthogonal subspaces, then $S_1 \cap S_2 = \{\mathbf{0}\}$.

Proof.

To get a contradiction, suppose there is a non-zero vector, $\mathbf{v}$ in the intersection. Thus,

$$\|\mathbf{v}\|_2^2 = \mathbf{v}^{\mathsf{H}}\mathbf{v} > 0.$$

But, since $\mathbf{v} \in S_1$ and $\mathbf{v} \in S_2$, the orthogonality property demands that $\mathbf{v}^{\mathsf{H}}\mathbf{v} = 0$, a contradiction. □

Orthogonal Decomposition

Proposition

Suppose that $L \leq \mathbb{C}^n$ has dimension $1 \leq k < n$. Then $L^\perp$ is a complementary subspace of dimension $n - k$:

$$L \oplus L^\perp = \mathbb{C}^n.$$

Furthermore, the decomposition of any $\mathbf{w} \in \mathbb{C}^n$ into

$$\mathbf{w} = \mathbf{x} + \mathbf{y}, \quad \mathbf{x} \in L, \quad \mathbf{y} \in L^\perp,$$

is unique.

Proof.

L and $L^\perp$ are orthogonal subspaces, and therefore $L \cap L^\perp = \{\mathbf{0}\}$. Let $B = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ be an orthonormal basis for L. Using basis completion and the Gram-Schmidt process, we can find the orthonormal set $T = \{\mathbf{w}_{k+1}, \ldots, \mathbf{w}_n\}$, such that $B \cup T$ is an orthonormal basis for $\mathbb{C}^n$. We claim that T is an orthonormal basis for $L^\perp$.

Proof, Cont.

Now, let $\mathbf{w} \in \mathbb{C}^n$ be arbitrary. Define

$$\mathbf{x} = \sum_{i=1}^{k} \alpha_i \mathbf{w}_i \quad \text{and} \quad \mathbf{y} = \sum_{i=k+1}^{n} \alpha_i \mathbf{w}_i,$$

where

$$\alpha_i = \mathbf{w}_i^{\mathsf{H}} \mathbf{w} = (\mathbf{w}, \mathbf{w}_i)_{\ell^2(\mathbb{C}^n)}, \quad i = 1, \ldots, n.$$

It is not too hard to see that $\mathbf{w} = \mathbf{x} + \mathbf{y}$. To see that such decompositions are unique. Suppose that

$$\mathbf{w} = \mathbf{x}_1 + \mathbf{y}_1 = \mathbf{x}_2 + \mathbf{y}_2, \quad \mathbf{x}_i \in L, \quad \mathbf{y}_i \in L^{\perp}, \quad i = 1, 2.$$

Then

$$L \ni \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{y}_2 - \mathbf{y}_1 \in L^{\perp}.$$

This implies that, since $L \cap L^{\perp} = \{\mathbf{0}\}$,

$$\mathbf{0} = \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{y}_2 - \mathbf{y}_1 = \mathbf{0}.$$

□

Definition (orthogonal projection)

The matrix $\mathsf{P} \in \mathbb{C}^{n\times n}$ is called an **orthogonal projection** iff $\mathsf{P} = \mathsf{P}^2$ and $\mathbf{im}(\mathsf{P})$ and $\mathbf{ker}(\mathsf{P})$ are orthogonal subspaces.

Theorem (characterization of orthogonal projection)

Let $\mathsf{P} \in \mathbb{C}^{n\times n}$ *be a projection matrix. Then* P *is an orthogonal projection iff* $\mathsf{P} = \mathsf{P}^{\mathsf{H}}$*.*

Proof.

Suppose that $\mathsf{P}^{\mathsf{H}} = \mathsf{P}$. Let $\mathbf{v}_1 \in \mathbf{ker}(\mathsf{P})$ and $\mathbf{v}_2 \in \mathbf{im}(\mathsf{P})$ be arbitrary. Then

$$\mathsf{P}\mathbf{v}_2 = \mathbf{v}_2, \qquad (\mathsf{I}_n - \mathsf{P})\mathbf{v}_1 = \mathbf{v}_1,$$

and

$$\mathbf{v}_2^{\mathsf{H}}\mathbf{v}_1 = (\mathsf{P}\mathbf{v}_2)^{\mathsf{H}}(\mathsf{I}_n - \mathsf{P})\mathbf{v}_1 = \mathbf{v}_2^{\mathsf{H}}\mathsf{P}^{\mathsf{H}}(\mathsf{I}_n - \mathsf{P})\mathbf{v}_1 = \mathbf{v}_2^{\mathsf{H}}(\mathsf{P} - \mathsf{P}^2)\mathbf{v}_1 = 0.$$

Thus $\mathbf{im}(\mathsf{P})$ and $\mathbf{ker}(\mathsf{P})$ are orthogonal subspaces, which implies that P is an orthogonal projection.

Proof, Cont.

Suppose now that P is an orthogonal projection. Set

$$S_1 = \mathrm{im}(\mathsf{P}), \qquad S_2 = \ker(\mathsf{P}),$$

with

$$\dim(S_1) = k < n, \qquad \dim(S_2) = n - k.$$

We want to prove that $\mathsf{P}^{\mathsf{H}} = \mathsf{P}$, using the fact that S_1 and S_2 are orthogonal subspaces of $\mathbb{C}^n$. Let

$$B_1 = \{\mathbf{q}_1, \ldots, \mathbf{q}_k\} \subset S_1, \qquad B_2 = \left\{\mathbf{q}_{k+1}, \ldots, \mathbf{q}_n\right\} \subset S_2,$$

be orthonormal bases for the respective spaces. This is always possible thanks to basis completion and the Gram–Schmidt process. We leave it as an exercise for the reader to prove that $B = B_1 \cup B_2$ is an orthonormal basis for $\mathbb{C}^n$. (Use the fact that S_1 and S_2 are complementary orthogonal subspaces.)

Proof, Cont.

By our construction,

$$\mathsf{P}\mathbf{q}_j = \begin{cases} \mathbf{q}_j & j = 1, \dots, k, \\ \mathbf{0} & j = k+1, \dots, n. \end{cases}$$

Set

$$\mathsf{Q} = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \in \mathbb{C}^{n \times n}.$$

Then

$$\mathsf{P}\mathsf{Q} = \begin{bmatrix} | & & | & | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_k & \mathbf{0} & \cdots & \mathbf{0} \\ | & & | & | & & | \end{bmatrix} \in \mathbb{C}^{n \times n},$$

and

$$\mathsf{Q}^{\mathsf{H}}\mathsf{P}\mathsf{Q} = \begin{bmatrix} \mathsf{I}_k & \mathsf{O} \\ \mathsf{O} & \mathsf{O}_{n-k} \end{bmatrix} = \mathbf{\Sigma} \in \mathbb{C}^{n \times n}.$$

Proof, Cont.

Consequently,

$$P = Q\boldsymbol{\Sigma}Q^{H},$$

and $P^{H} = P$. $\square$

Theorem (special projectors)

Let $k \in \{1, \ldots, n\}$ and suppose that the collection of vectors $\{\mathbf{q}_1, \ldots, \mathbf{q}_k\} \subset \mathbb{C}^n$ is orthonormal. Define

$$\hat{Q} = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_k \\ | & & | \end{bmatrix} \in \mathbb{C}^{n\times k},$$

then the matrices

$$P = \hat{Q}\hat{Q}^{H}, \quad I_n - P = I_n - \hat{Q}\hat{Q}^{H}$$

are orthogonal projectors.

Rank–one Projections

Definition

Suppose that $\mathbf{q} \in \mathbb{C}^n$, $\|\mathbf{q}\|_2 = 1$. The matrix $\mathsf{P} = \mathbf{q}\mathbf{q}^{\mathsf{H}}$ is called a **rank–one orthogonal projection**. The complement, $\mathsf{I}_n - \mathbf{q}\mathbf{q}^{\mathsf{H}}$, is called a **rank-$(n-1)$ orthogonal projection**.

Theorem

The identity matrix $\mathsf{I}_n \in \mathbb{C}^{n\times n}$ is a sum of n rank-one orthogonal projection matrices.

Proof.

Write $\mathsf{I}_n = \mathsf{U}\mathsf{U}^{\mathsf{H}}$, where U is unitary. Then,

$$\mathsf{I}_n = \mathsf{U}\mathsf{I}_n\mathsf{U}^{\mathsf{H}} = \sum_{i=1}^{n} \mathsf{U}\mathbf{e}_i\mathbf{e}_i^{\intercal}\mathsf{U}^{\mathsf{H}} = \sum_{i=1}^{n} \mathbf{u}_i\mathbf{u}_i^{\mathsf{H}}.$$

□

The 2-Norm of a Projection Matrix

Theorem

Let $\mathsf{P} \in \mathbb{C}^{n\times n}$ *be a non–zero projection matrix. Then* $\|\mathsf{P}\|_2 \geq 1$*. Moreover,* $\|\mathsf{P}\|_2 = 1$ *iff* P *is an orthogonal projection, that is* $\mathsf{P}^{\mathsf{H}} = \mathsf{P}$*.*

Proof.

Suppose that P is a projection matrix, that is, $\mathsf{P}^2 = \mathsf{P}$. Then, using the sub–multiplicativity of the 2–norm

$$\|\mathsf{P}\|_2 = \left\|\mathsf{P}^2\right\|_2 \leq \|\mathsf{P}\|_2 \|\mathsf{P}\|_2 .$$

Since P is not the zero matrix, $\|\mathsf{P}\|_2 > 0$, and $\|\mathsf{P}\|_2 \geq 1$.

Now, for the second part, we have two directions to prove. Assume first that $\mathsf{P}^2 = \mathsf{P}$ and $\mathsf{P}^{\mathsf{H}} = \mathsf{P}$.

Proof, Cont.

In the proof of a previous theorem, we showed that

$$\mathsf{P} = \mathsf{Q}\boldsymbol{\Sigma}\mathsf{Q}^{\mathsf{H}},$$

where $\mathsf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_n] \in \mathbb{C}^{n\times n}$ is unitary and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathsf{I}_k & \mathsf{O} \\ \mathsf{O} & \mathsf{O}_{n-k} \end{bmatrix},$$

for some $1 \le k < n$. In other words, $\boldsymbol{\Sigma} = \mathbf{diag}[\sigma_1, \ldots, \sigma_k, 0, \ldots, 0]$, where $\sigma_i = 1$, $1 \le i \le k$. Recall that $\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$ is a basis for $\mathbf{im}(\mathsf{P})$ and $\{\mathbf{q}_{k+1}, \ldots, \mathbf{q}_n\}$ is a basis for $\mathbf{ker}(\mathsf{P})$. In any case, P is Hermitian positive semi–definite, $\mathsf{P} = \mathsf{Q}\boldsymbol{\Sigma}\mathsf{Q}^{\mathsf{H}}$ is a unitary diagonalization of P, and

$$\|\mathsf{P}\|_2 = \sigma_1 = \rho(\mathsf{P}) = 1.$$

Proof, Cont.

The other implication is proved by contrapositive. Let us assume that $\mathsf{P}^2 = \mathsf{P}$, but $\mathsf{P}^H \neq \mathsf{P}$. We want to show that this implies that

$$\|\mathsf{P}\|_2 > 1.$$

Since $\mathsf{P}^H \neq \mathsf{P}$, $\mathbf{ker}(\mathsf{P}) \cap \mathbf{im}(\mathsf{P}) = \{\mathbf{0}\}$, but

$$\mathbf{ker}(\mathsf{P}) \not\perp \mathbf{im}(\mathsf{P}).$$

So, there is some non–zero vector $\mathbf{v}_1 \in \mathbf{im}(\mathsf{P})$ and some non–zero vector $\mathbf{v}_2 \in \mathbf{ker}(\mathsf{P})$ such that

$$(\mathbf{v}_1, \mathbf{v}_2)_2 = \mathbf{v}_2^H \mathbf{v}_1 \neq 0.$$

Set

$$\mathbf{v} = \mathbf{v}_1 + \alpha \mathbf{v}_2, \quad \alpha \in \mathbb{C}.$$

Then $\mathsf{P}\mathbf{v} = \mathbf{v}_1$. Now, we want to choose $\alpha \in \mathbb{C}$ so that

$$\|\mathsf{P}\mathbf{v}\|_2 > \|\mathbf{v}\|_2 > 0.$$

Proof, Cont.

Indeed, if such an $\alpha \in \mathbb{C}$ exists, then

$$\|\mathsf{P}\|_2 = \sup_{x \in \mathbb{C}^n_\star} \frac{\|\mathsf{P}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \geq \frac{\|\mathsf{P}\mathbf{v}\|_2}{\|\mathbf{v}\|_2} > 1.$$

Since

$$\mathsf{P}\mathbf{v} = \mathsf{P}(\mathbf{v}_1 + \alpha \mathbf{v}_2) = \mathsf{P}\mathbf{v}_1 + \alpha \mathsf{P}\mathbf{v}_2 = \mathbf{v}_1,$$

it follows that

$$\|\mathbf{v}_1\|_2^2 = \|\mathsf{P}\mathbf{v}\|_2^2$$

and

$$\|\mathbf{v}\|_2^2 = \|\mathbf{v}_1 + \alpha \mathbf{v}_2\|_2^2 = \|\mathsf{P}\mathbf{v}\|_2^2 + 2\Re\left(\alpha \mathbf{v}_1^{\mathsf{H}} \mathbf{v}_2\right) + |\alpha|^2 \|\mathbf{v}_2\|_2^2 = \|\mathsf{P}\mathbf{v}\|_2^2 + T,$$

where

$$T = 2\Re\left(\alpha \mathbf{v}_1^{\mathsf{H}} \mathbf{v}_2\right) + |\alpha|^2 \|\mathbf{v}_2\|_2^2 .$$

Proof, Cont.

Therefore, it suffices to choose $\alpha \in \mathbb{C}$ so that $T < 0$ for, in that case,

$$\|\mathbf{v}\|_2^2 < \|P\mathbf{v}\|_2^2 .$$

The key is to choose $\alpha \in \mathbb{C}$ such that $\mathbf{v} \perp \mathbf{v}_2$, that is, $\mathbf{v}_2^H\mathbf{v} = 0$. This is equivalent to

$$\begin{aligned}
& \mathbf{v}_2^H(\mathbf{v}_1 + \alpha\mathbf{v}_2) = 0 \\
\iff \quad & \mathbf{v}_2^H\mathbf{v}_1 = -\alpha\mathbf{v}_2^H\mathbf{v}_2 \\
\iff \quad & \bar{\alpha}\mathbf{v}_2^H\mathbf{v}_1 = -|\alpha|^2\,\|\mathbf{v}_2\|_2^2 \in \mathbb{R} \\
\iff \quad & \bar{\alpha}\overline{\mathbf{v}_1^H\mathbf{v}_2} = -|\alpha|^2\,\|\mathbf{v}_2\|_2^2 \in \mathbb{R} \\
\iff \quad & \alpha\mathbf{v}_1^H\mathbf{v}_2 = -|\alpha|^2\,\|\mathbf{v}_2\|_2^2 \in \mathbb{R}.
\end{aligned}$$

In this case,

$$T = -|\alpha|^2\,\|\mathbf{v}_2\|_2^2 \in \mathbb{R}.$$

Thus, the result follows upon choosing $\alpha = -\frac{\mathrm{v}_2^H\mathrm{v}_1}{\|\mathrm{v}_2\|_2^2}$. $\square$

Rank–Deficient Case

Theorem (general least squares)

Suppose that $m \geq n$, *and the matrix* $\mathsf{A} \in \mathbb{C}^{m\times n}$ *is such that* $\mathsf{rank}(\mathsf{A}) \leq n$, *i.e.,* A *may be rank deficient. Let* $\mathbf{f} \in \mathbb{C}^m$ *be given. Then the normal equations,*

$$\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{x} = \mathsf{A}^{\mathsf{H}}\mathbf{f}, \tag{6}$$

always have at least one solution, and, for any two solutions, $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^n$ *— in the case that there are multiple solutions — we find that*

$$\mathbf{r}(\mathbf{x}_1) = \mathbf{r}(\mathbf{x}_2),$$

where, for $\mathbf{w} \in \mathbb{C}^n$, *we defined*

$$\mathbf{r}(\mathbf{w}) = \mathbf{f} - \mathsf{A}\mathbf{w}.$$

In other words, the residual is always unique.

Theorem (Cont.)

Furthermore, the following are equivalent

1. $\mathbf{x}_o \in \mathbb{C}^n$ *is a solution to*

$$\mathbf{x} \in \operatorname*{argmin}_{\mathbf{w} \in \mathbb{C}^n} \Phi(\mathbf{w}), \quad \Phi(\mathbf{w}) = \|\mathbf{r}(\mathbf{w})\|_2^2 . \tag{7}$$

2. $\mathbf{x}_o \in \mathbb{C}^n$ *is a solution to the normal equations* (6).
3. $\mathbf{x}_o \in \mathbb{C}^n$ *has the property that*

$$\mathbf{r}(\mathbf{x}_o) \perp \operatorname{im}(A).$$

Proof.

First, let us prove that the normal equations (6) have a solution. Set $L = \mathbf{im}(\mathsf{A})$. Then L and $L^\perp$ are complementary, orthogonal, subspaces of $\mathbb{C}^m$:

$$\mathbb{C}^m = L \oplus L^\perp.$$

Therefore the decomposition

$$\mathbf{f} = \mathbf{s} + \mathbf{r}, \quad \mathbf{s} \in L = \mathbf{im}(\mathsf{A}), \quad \mathbf{r} \in L^\perp$$

is unique. Since $\mathbf{s} \in \mathbf{im}(\mathsf{A})$, there is at least one vector $\mathbf{x}_o \in \mathbb{C}^n$ such that

$$\mathsf{A}\mathbf{x}_o = \mathbf{s}.$$

Since $\mathbf{r} \in L^\perp$, $\mathbf{r}^{\mathsf{H}}\mathsf{A}\mathbf{x} = 0$, for all $\mathbf{x} \in \mathbb{C}^n$. This implies that

$$\mathsf{A}^{\mathsf{H}}\mathbf{r} = \mathbf{0} \in \mathbb{C}^n.$$

Proof, Cont.

Recall that,

$$\mathbf{f} = \mathsf{A}\mathbf{x}_o + \mathbf{r},$$

which implies that

$$\mathbf{r} = \mathbf{f} - \mathsf{A}\mathbf{x}_o = \mathbf{r}(\mathbf{x}_o).$$

Hence

$$\mathsf{A}^{\mathsf{H}}\mathbf{r}(\mathbf{x}_o) = \mathbf{0} \in \mathbb{C}^n,$$

which is equivalent to the normal equations. Thus, the normal equations have at least one solution.

Suppose that $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^n$ are solutions to the normal equations (6). Recall that the decomposition

$$\mathbf{f} = \mathbf{s} + \mathbf{r}, \quad \mathbf{s} \in L = \mathrm{im}(\mathsf{A}), \quad \mathbf{r} \in L^{\perp}$$

is unique. But

$$\mathbf{f} = \mathsf{A}\mathbf{x}_1 + \mathbf{r}(\mathbf{x}_1) = \mathsf{A}\mathbf{x}_2 + \mathbf{r}(\mathbf{x}_2)$$

So $\mathbf{s} = \mathsf{A}\mathbf{x}_1 = \mathsf{A}\mathbf{x}_2$ and $\mathbf{r} = \mathbf{r}(\mathbf{x}_1) = \mathbf{r}(\mathbf{x}_2)$.

Proof, Cont.

$(2 \iff 3)$: This follows from the calculations we carried out above.

$(2 \implies 1)$: This argument is similar to previous ones. Suppose that $\mathbf{x}_o \in \mathbb{C}^n$ is a solution to the normal equation (6). Let $\mathbf{w} \in \mathbb{C}^n$ be arbitrary. Then

$$\begin{aligned}\Phi(\mathbf{x}_o + \mathbf{w}) &= (\mathbf{r}(\mathbf{x}_o) - \mathsf{A}\mathbf{w})^{\mathsf{H}} (\mathbf{r}(\mathbf{x}_o) - \mathsf{A}\mathbf{w}) \\ &= \Phi(\mathbf{x}_o) - (\mathbf{r}(\mathbf{x}_o))^{\mathsf{H}}\mathsf{A}\mathbf{w} - \mathbf{w}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathbf{r}(\mathbf{x}_o) + \mathbf{w}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{w} \\ &= \Phi(\mathbf{x}_o) + \mathbf{w}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{w} \\ &\geq \Phi(\mathbf{x}_o),\end{aligned}$$

since $\mathsf{A}^{\mathsf{H}}\mathsf{A}$ is Hermitian positive semi–definite. Note that we cannot claim that $\Phi(\mathbf{x}_o + \mathbf{w}) > \Phi(\mathbf{x}_o)$, for all $\mathbf{w} \in \mathbb{C}^n_\star$, since we do not know that $\mathsf{A}^{\mathsf{H}}\mathsf{A}$ is HPD. However, we can still assert that $\mathbf{x}_o$ is a minimizer, though it might not be unique.

Proof, Cont.

$(1 \implies 3)$: Suppose that $\mathbf{x}_o \in \mathbb{C}^n$ is a solution to (7). We want to show that $\mathbf{r}(\mathbf{x}_o) \perp \text{im}(\mathsf{A})$. To get a contradiction, suppose that $\mathbf{r}(\mathbf{x}_o) \not\perp \text{im}(\mathsf{A})$. If this is the case, there is some $\mathbf{q} \in \text{im}(\mathsf{A})$, $\mathbf{q} \neq \mathbf{0}$, such that $\mathbf{q}^{\mathsf{H}}\mathbf{r}(\mathbf{x}_o) \neq 0$. Since $\mathbf{q} \in \text{im}(\mathsf{A})$, there is a vector $\mathbf{w} \in \mathbb{C}^n$ such that $\mathsf{A}\mathbf{w} = \mathbf{q}$. Since $\mathbf{x}_o$ is a minimizer of Φ, for any $\alpha \in \mathbb{C}$

$$\begin{aligned}
\|\mathbf{r}(\mathbf{x}_o)\|_2^2 &= \Phi(\mathbf{x}_o) \\
&\leq \Phi(\mathbf{x}_o + \alpha\mathbf{w}) \\
&= \Phi(\mathbf{x}_o) - \alpha(\mathbf{r}(\mathbf{x}_o))^{\mathsf{H}}\mathsf{A}\mathbf{w} - \bar{\alpha}\mathbf{w}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathbf{r}(\mathbf{x}_o) + |\alpha|^2\mathbf{w}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{w} \\
&= \|\mathbf{r}(\mathbf{x}_o)\|_2^2 - 2\Re\left(\bar{\alpha}\mathbf{q}^{\mathsf{H}}\mathbf{r}(\mathbf{x}_o)\right) + |\alpha|^2\mathbf{q}^{\mathsf{H}}\mathbf{q}.
\end{aligned}$$

Thus, for all $\alpha \in \mathbb{C}$,

$$2\Re\left(\bar{\alpha}\mathbf{q}^{\mathsf{H}}\mathbf{r}(\mathbf{x}_o)\right) \leq |\alpha|^2\mathbf{q}^{\mathsf{H}}\mathbf{q}.$$

Now, set

$$\alpha = \frac{\mathbf{q}^{\mathsf{H}}\mathbf{r}(\mathbf{x}_o)}{\mathbf{q}^{\mathsf{H}}\mathbf{q}}$$

to get a contradiction. □