Linear Operators and Inverse Problems

Linear Operators

Definition: Take two linear spaces X, Y defined on the same scalars. A mapping $L : X \to Y$ is linear if

$$L(\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2) = \alpha_1 \; L(\mathbf{x}_1) + \alpha_2 \; L(\mathbf{x}_2) \qquad \forall \alpha_1, \alpha_2, \mathbf{x}_1, \mathbf{x}_2$$

Examples:

- $L(x(t)) = x(0.1)$
- $L(x(t)) = \frac{d}{dx} x(t) - 2x(t)$
- $L(x(t)) = \int_a^b x(t)\, \phi(t) dt$
- $L(x(t)) = \int_{-\infty}^t x(\tau) h(t - \tau) d\tau$
- $L(x(t)) = \int_{-\infty}^{\infty} K(t, \tau) x(\tau) d\tau$
- If $\mathbf{x}$ is a n-dimensional vector, then any matrix $\mathbf{A}$ is a linear operator $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m$

Space of Operators

- Consider two linear operators $L_1 : X \to Y$ and $L_2 : X \to Y$.
- Clearly $\alpha_1 L_1 + \alpha_2 L_2$ is a linear operator as well.
- $\Rightarrow$ Operators also make a vector space too!!!
- This explains why matrices and functions make vector spaces.
- If operators are vectors, we can try to define their norm.
- In general, we can define anything that satisfies the norm properties. But there are a few that are more useful than others.

Subordinate Norms

- We can defined norms using the underlying norms in spaces X and Y.
- When operator norms are dependent on other norms, they are called *subordinate norms*.
- The most useful are the so-called p-norms, defined as follows:

$$||L|| \triangleq \max \frac{||L(\mathbf{x})||_p}{||\mathbf{x}||_p}$$

- The p-norms show the maximum degree of amplification.
- Note the equivalence to:

$$||L|| = \max_{||\mathbf{x}||_p \leq 1} ||L(\mathbf{x})||_p$$

Properties of Operator p-norms

- They show maximum amplification, therefore:

$$||L(\mathbf{x})||_p \leq ||L||_p \; ||\mathbf{x}||_p$$

- The norm of identity operator is one: $||\mathbf{I}||_p = 1$.

- We say an operator norm has the **sub-multiplicative property** if for any two operators in the space:

$$||L_1 \, L_2|| \leq ||L_1|| \; ||L_2||$$

- The p-norm has the sub-multiplicative property because:

$$||L_1(L_2(\mathbf{x}))||_p \leq ||L_1||_p \, ||L_2(\mathbf{x})||_p \leq ||L_1||_p \, ||L_2||_p \, ||\mathbf{x}||_p$$

Matrix Norms

- We first try the ℓ_∞ norm:

$$\begin{aligned}
||\mathbf{A}||_\infty &= \max_{||\mathbf{x}||_\infty = 1} ||\mathbf{A}\mathbf{x}||_\infty \\
&= \max_{\max |x_n| = 1} (\max_i |\sum_j a_{ij} x_j|) \\
&\leq \max_{\max |x_n| = 1} (\max_i \sum_j |a_{ij} x_j|) \\
&\leq \max_{\max |x_n| = 1} (\max_i \sum_j |a_{ij}|) \\
&= \max_i \sum_j |a_{ij}|
\end{aligned}$$

- This is the largest row sum of absolute values.

The ℓ_1 Matrix Norm

- Now let's try the ℓ_1 norm:

$$
\begin{aligned}
||\mathbf{A}||_1 &= \max_{||\mathbf{x}||_1=1} ||\mathbf{A}\mathbf{x}||_1 \\
&= \max_{\sum |x_n|=1} \sum_i |\sum_j a_{ij}x_j| \\
&\le \max_{\sum |x_n|=1} \sum_i \sum_j |a_{ij}x_j| \\
&\le \max_{\sum |x_n|=1} \sum_j |x_j| \sum_i |a_{ij}| \\
&= \max_j \sum_i |a_{ij}|
\end{aligned}
$$

- The norm is the highest column sum of absolute values.

The ℓ_2 Matrix Norm

- Finally, we look at the ℓ_2 norm

$$||\mathbf{A}||_2^2 = \max_{||\mathbf{x}||=1} ||\mathbf{A}\mathbf{x}||^2 = \max_{\mathbf{x}^H\mathbf{x}=1} \mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}$$

- This is constrained optimization. Write Lagrangian:

$$J = \mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x} - \lambda\mathbf{x}^H\mathbf{x}$$

$$\frac{d\,J}{d\mathbf{x}} = \mathbf{A}^H\mathbf{A}\mathbf{x} - \lambda\mathbf{x} = 0$$

$$\mathbf{A}^H\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

The maximum is via the largest eigenvalue of $\mathbf{A}^H\mathbf{A}$.

$$||\mathbf{A}|| = \sqrt{\lambda_1} = \sqrt{\rho(\mathbf{A}^H\mathbf{A})}$$

- This is known as the **spectral norm**

The Frobenius Norm

- **Definition:**

$$||\mathbf{A}||_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

- Not a p-norm
- It is a good measure of how close two matrices are together.
- Note: $||\mathbf{I}||_F = \sqrt{n}$.
- Equivalent definition:

$$||\mathbf{A}||_F = \sqrt{trace(\mathbf{A}^H \mathbf{A})}$$

- We can also write:

$$||\mathbf{A}||_F = \sqrt{vec(\mathbf{A})^H vec(\mathbf{A})}$$

Relationships between Matrix Norms

$$||\mathbf{A}||_2 \leq ||\mathbf{A}||_F \leq \sqrt{n}||\mathbf{A}||_2$$

$$\max |a_{ij}| \leq ||\mathbf{A}||_2 \leq \sqrt{mn} \max |a_{ij}|$$

$$\frac{1}{\sqrt{n}}||\mathbf{A}||_\infty \leq ||\mathbf{A}||_2 \leq \sqrt{m}||\mathbf{A}||_\infty$$

$$\frac{1}{\sqrt{m}}||\mathbf{A}||_1 \leq ||\mathbf{A}||_2 \leq \sqrt{n}||\mathbf{A}||_1$$

Bounded Operators

- **Definition:** An operator is *bounded* if its norm is finite, i.e., there exists some constant M such that for all $\mathbf{x}$

$$||L(\mathbf{x})|| \leq M\ ||\mathbf{x}||$$

- **Example:** The matrix spectral norm is bounded. For any matrix $\mathbf{A}$, we have $||\mathbf{A}||_2 \leq \sqrt{mn}\max|a_{ij}|$
- **Example:** Integral operators are bounded, for example,

$$L_\phi(x(t)) = \int_a^b x(t)\,\phi(t)\,dt \qquad \Rightarrow \qquad ||L||_2^2 \leq \int_a^b \phi^2(t)dt$$

- **Example:** The derivative operator on $C_1[-\pi,\pi]$ is unbounded. Take $L(x(t)) = dx/dt$, $x(t) = \sin\omega t$, we have $\frac{||L(x)||}{||x||} = \omega$.
We can choose $x(t)$ to make the ratio arbitrarily large.

Continuous Operators

- **Definition:** $L : X \to Y$ is continuous if for any $\epsilon > 0$ there exists a $\delta > 0$ such that for any $\mathbf{x}_0 \in X$

$$||\mathbf{x} - \mathbf{x}_0|| < \delta \implies ||L(\mathbf{x}) - L(\mathbf{x}_0)|| < \epsilon$$

- **Theorem:** A linear operator $L : X \to Y$ is is continuous if and only if it is bounded.

Proof: Assume L is bounded. For any vector $\mathbf{x}$ and increment $\mathbf{h}$ we have $||L(\mathbf{x}+\mathbf{h}) - L(\mathbf{h})|| = ||L(\mathbf{h})|| \leq M||\mathbf{h}||$. So for any ϵ there is an $\delta = \epsilon/M$ etc. etc.
Now assume the operator is continuous (around zero). $\exists\delta$ so that $||\mathbf{h}|| < \delta \Rightarrow ||L(\mathbf{h}) - L(0)|| < 1$. Then:

$$||L(\mathbf{x})|| = \left|\left|\frac{||\mathbf{x}||}{\delta}L(\delta\frac{\mathbf{x}}{||\mathbf{x}||})\right|\right| = \frac{||\mathbf{x}||}{\delta}\big|\big|L(\delta\frac{\mathbf{x}}{||\mathbf{x}||})\big|\big| \leq \frac{||\mathbf{x}||}{\delta}\cdot 1 = \frac{||\mathbf{x}||}{\delta}$$

Remarks

- **Theorem:** A linear operator $L : X \to Y$ on normed spaces. If X is finite dimensional, then L is continuous (and hence bounded). No further condition on Y is necessary.
- Not every linear operator is continuous (or bounded). Example: differentiation.
- Sometimes it is hard to prove continuity, but easy to show boundedness, or vice versa. Example: derivative operator.
- Topological definition of continuity on normed spaces: operators that map every bounded set to a bounded set. This leads to interesting formulations that are outside our scope.

Linear Functionals

- Linear operator whose value is a scalar, i.e., $f : X \to \mathbb{C}$.
- An important subclass of linear operators.
- **Example 1:** On $\mathbb{R}^n$, consider a fixed vector $\mathbf{c} = [c_1, \ldots, c_n]$. Then the following is a linear functional:

$$f(\mathbf{x}) = \mathbf{c}^t \mathbf{x}$$

- **Example 2:** On the space of functions over $[a, b]$, consider a fixed function $\phi(t)$. Then the following is a linear functional:

$$f(x(t)) = \int_a^b x(t)\ \phi(t) dt$$

Riesz Representation Theorem

- **Generalization:** In any Hilbert space, consider a fixed vector ϕ. Then the following is a linear functional:

$$f_\phi(\mathbf{x}) = < \mathbf{x}, \phi >$$

- **Theorem:** If X is a Hilbert space, then any continuous linear functional can be written as

$$f(\mathbf{x}) = < \mathbf{x}, \phi >$$

for some fixed $\phi \in X$.

- We have seen earlier that a functional is continuous iff it is bounded.

The Neumann Expansion

- Recall that if a real number $|x| < 1$, then,

$$1 + x + x^2 + \cdots = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} = (1-x)^{-1}$$

- **Theorem:** If $||\cdot||$ is a submultiplicative norm and $L : X \to X$ with $||L|| < 1$, then $(I - L)^{-1}$ exists and

$$(I-L)^{-1} = \sum_{k=0}^{\infty} L^k$$

Proof: If $||L|| < 1$ then $I - L$ is non-singular (why?)
Note that by submultiplicativity of the norm, $\lim_{k\to\infty} L^k = 0$.

$$(I-L)(I + L + L^2 + \cdots + L^{k-1}) = 1 - L^k$$

Then by $k \to \infty$ we get $(I-L)\sum_{k=0}^{\infty} L^k = I$

Adjoint Operators

- **Definition:** Consider linear operator $L : X \rightarrow Y$ on Hilbert spaces X, Y. The adjoint L^* is the operator $L^* : Y \rightarrow X$ such that for all $\mathbf{x} \in X, \mathbf{y} \in Y$,

$$< L\mathbf{x}, \mathbf{y} > = < \mathbf{x}, L^*\mathbf{y} >$$

- An operator is **self-adjoint** if $L^* = L$.
- **Example:** For a complex matrix $\mathbf{A}$, the adjoint is (show!)

$$\mathbf{A}^* = \mathbf{A}^H$$

- **Example:** Find the adjoint for the following linear operator on L_2, with the usual inner product.

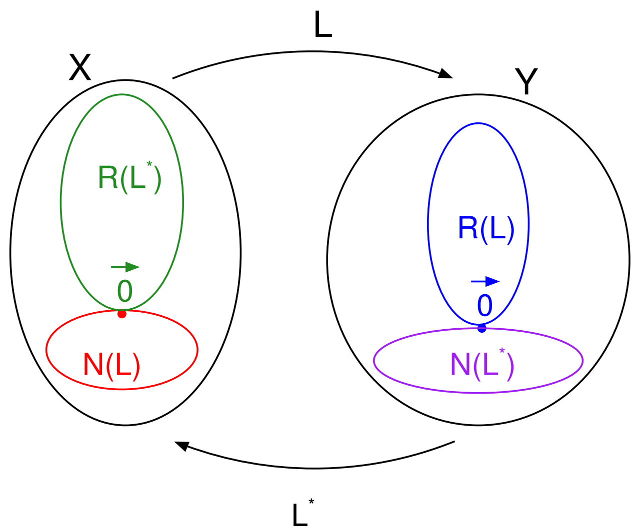
$$L(x(t)) = \int_a^b x(t)\phi(t)dt$$

Properties of Adjoints

- $L^{**} = L$
- $(L_1 + L_2)^* = L_1^* + L_2^*$
- $(\alpha L)^* = \overline{\alpha} L^*$
- $(L_1 L_2)^* = L_2^* L_1^*$
- If L has an inverse, then $(L^*)^{-1} = (L^{-1})^*$
- **Question:** Write the matrix adjoint with a weighted inner product?
- **Question:** Show that with the adjoint notation, the weighted least square solution is:

$$\hat{\mathbf{x}} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{b}$$

Range and Null Spaces of L and L^*



Four Fundamental Subspaces

Theorem: Let $L : X \to Y$ on Hilbert spaces, then:

$$\big[R(L)\big]^{\perp} = N(L^*) \qquad \big[R(L^*)\big]^{\perp} = N(L)$$

conversely:

$$\big[N(L)\big]^{\perp} = \overline{R(L^*)} \qquad \big[N(L^*)\big]^{\perp} = \overline{R(L)}$$

If the ranges are closed, we have:

$$\big[N(L)\big]^{\perp} = R(L^*) \qquad \big[N(L^*)\big]^{\perp} = R(L)$$

Four Fundamental Subspaces (2)

Proof: Let $\mathbf{n} \in N(L^*)$, then $\forall \mathbf{y} \in R(L)$,

$$< \mathbf{y}, \mathbf{n} >=< L(\mathbf{x}), \mathbf{n} >=< \mathbf{x}, L^*(\mathbf{n}) >=< \mathbf{x}, 0 >= 0$$

$\mathbf{n} \in R(L)^\perp$ so $N(L^*) \subset R(L)^\perp$.

Now let $\mathbf{y}_0 \in R(L)^\perp$, then

$$< L(\mathbf{x}), \mathbf{y}_0 >= 0 \qquad \Rightarrow \qquad < \mathbf{x}, L^*(\mathbf{y}_0) >= 0 \qquad \forall \mathbf{x} \in X$$

we must have $L^*(\mathbf{y}_0) = 0$ or $\mathbf{y}_0 \in N(L^*)$, i.e., $R(L) \subset N(L^*)$.
The two parts prove $R(L)^\perp = N(L^*)$.

The counterparts are proved by symmetry, and via taking orthogonal complements.

Solutions to Linear Equations

Existence and uniqueness of the solution of linear operator equation:

$$L(\mathbf{x}) = \mathbf{b}$$

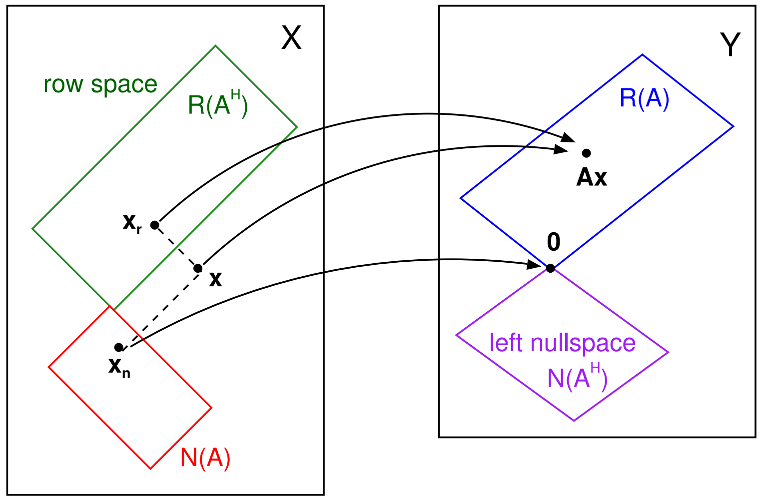
- **Fredholm Alternative Theorem:** The above equation has a solution iff $< \mathbf{b}, \mathbf{v} >= 0$ for all $\mathbf{v} \in N(L^*)$

  $$\mathbf{b} \in R(L) \ \Rightarrow \ \mathbf{b} \perp N(L^*)$$

  For matrices, $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution iff $\mathbf{b}^H\mathbf{v} = 0$ for all $\mathbf{v}$ s.t. $\mathbf{A}^H\mathbf{v} = \mathbf{0}$.

- The solution to the above equation (if it exists) is unique if $N(L)$ is trivial, i.e., $N(L) = \{\mathbf{0}\}$

Fundamental Subspaces for Matrices



Least Squares and the Adjoint

Wish to find $\hat{\mathbf{x}}$ to minimize $||L(\mathbf{x}) - \mathbf{b}||$.

Theorem: The vector $\hat{\mathbf{x}}$ minimizes $||L(\mathbf{x}) - \mathbf{b}||$ iff

$$L^*(L(\hat{\mathbf{x}})) = L^*(\mathbf{b})$$

Proof: Take $L(\hat{\mathbf{x}}) = \hat{\mathbf{b}}$. Then by orthogonality:

$$\mathbf{b} - \hat{\mathbf{b}} \in R(L)^{\perp}$$

$$\mathbf{b} - \hat{\mathbf{b}} \in N(L^*)$$

$$L^*(\mathbf{b} - \hat{\mathbf{b}}) = \mathbf{0}$$

$$L^*(\hat{\mathbf{b}}) = L^*(\mathbf{b})$$

$$L^*(L(\hat{\mathbf{x}})) = L^*(\mathbf{b})$$

If the composite operator L^*L is invertible, then:

$$\hat{\mathbf{x}} = (L^*L)^{-1}L^*(\mathbf{b})$$

Recall that for matrix equations $\mathbf{Ax} = \mathbf{b}$, the LS solution is:

$$\hat{\mathbf{x}} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{b}$$

But for general LS problems we had to write the Grammian.

We now see that the simple formula can be extended to all operators using the notion of the adjoint.

NOTE: If (L^*L) is not invertible, this method does not work.

Exercise: Calculate L^* and $(L^*L)^{-1}$ for the polynomial interpolation problem.

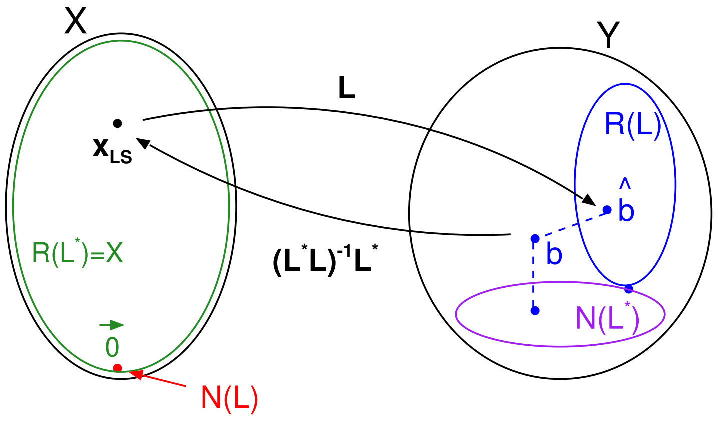
Least Squares and the Left Inverse

- If $N(L)$ is trivial, L is **one-to-one**
- Then, L^*L is invertible (why?)
- Then the least squares solution to $L(\mathbf{x}) = \mathbf{b}$ is unique.

$$\hat{\mathbf{x}} = (L^*L)^{-1}L^*\ \mathbf{b}$$

- **Note:** The operator above is the left inverse of L.
- The LS solution:
  - First removes any component in the null-space of L^*.
  - Then inverts the effect of the application of L^*L.

Least Squares and Fundamental Subspaces



Min-Norm and the Adjoint

$$\min ||\mathbf{x}||$$
$$\text{s.t.} \quad L(\mathbf{x}) = \mathbf{b}$$

Theorem: The min-norm solution is in $R(L^*)$, i.e.,

$$\mathbf{x}_{MN} = L^*(\mathbf{c})$$

for some $\mathbf{c}$ so that we have $L(\mathbf{x}) = L(L^*(\mathbf{c})) = \mathbf{b}$.

Proof: Simply decompose any feasible $\mathbf{x}$ into components in $R(L^*)$ and $N(L)$. Clearly $||\mathbf{x}||$ is minimized if the component in $N(L)$ is zero.

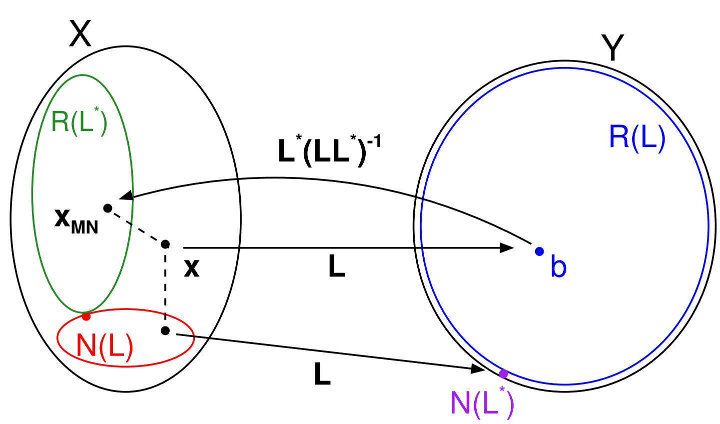
Min-Norm and the Right Inverse

- If $N(L^*)$ is trivial, then $R(L) = Y$.
- We say L is "**onto**"
- Then all $L(\mathbf{x}) = \mathbf{b}$ have feasible solutions
- In that case, LL^* is invertible (why?)
- In that case, the min-norm solution is written as:

$$\mathbf{x}_{MN} = L^*(LL^*)^{-1}\mathbf{b}$$

- **Note:** The operator above is the right inverse of L

MN and Fundamental Subspaces



More on Inverses

- We see that LS and MN problems are closely related to inverting operators.
- We have seen partial solutions when either $N(L)$ or $N(L^*)$ were trivial.
- Correspondingly, we have solved the LS and MN problems individually.
- Natural question: what happens when both $N(L)$ and $N(L^*)$ are non-trivial?
- Then, if we try to find LS solution to $\mathbf{Ax} = \mathbf{b}$, we might find there are many!!! The we want to find the MN-LS solution.
- In this case, neither LL^* nor L^*L are invertible.
- This will lead to a more general pseudo-inverse.

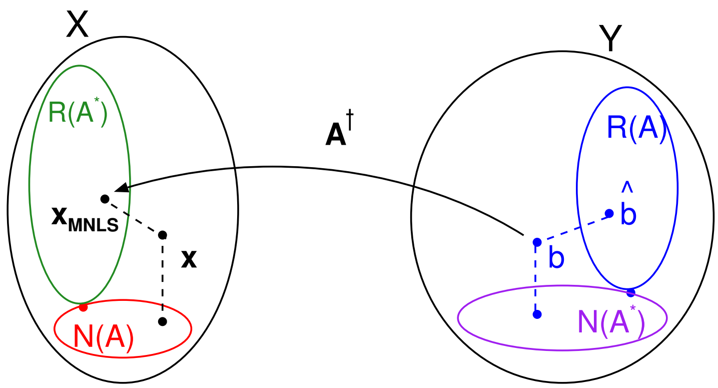
Minimum-Norm Least-Squares Solutions

- Let's concentrate on matrix equations $\mathbf{Ax} = \mathbf{b}$.
- Projecting $\mathbf{b}$ onto range of $\mathbf{A}$ will not result in a unique solution if $\mathbf{A}$ is not full-rank.
- Then we can look for a MNLS solution.
- **Corollary:** $\mathbf{x}_{\text{MNLS}}$ satisfies the normal equations.

  Proof: $\mathbf{x}_{\text{MNLS}}$ is also (one of) the least squares solutions, therefore must satisfy the normal equations. This means $\mathbf{A}\mathbf{x}_{\text{MNLS}}$ is the projection of $\mathbf{b}$ on $R(\mathbf{A})$.
- **Corollary:** $\mathbf{x}_{\text{MNLS}}$ is in the range of $\mathbf{A}^H$

  Proof: Any component in $N(\mathbf{A})$ would only add to the norm.

MNLS Solution and Fundamental Subspaces



The Moore-Penrose Pseudo-Inverse

We continue to restrict our attention to matrices.

Definition: A $n \times m$ matrix $\mathbf{A}^\dagger$ is the Moore-Penrose pseudo-inverse of $m \times n$ matrix $\mathbf{A}$ if:

- $\mathbf{A}\mathbf{A}^\dagger$ is the projection matrix for $R(\mathbf{A})$
- $\mathbf{A}^\dagger\mathbf{A}$ is the projection matrix for $R(\mathbf{A}^\dagger)$

Agenda:

- ◦ To show that $\mathbf{A}^\dagger$ exists and is unique.
- ◦ To show that it is the left or right inverse when $\mathbf{A}$ is full-rank
- ◦ To show that it provides the MNLS solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Properties of $\mathbf{A}^\dagger$

Theorem: $R(\mathbf{A}^\dagger) = R(\mathbf{A}^H)$.

Proof: Since $\mathbf{A}^\dagger\mathbf{A}$ is a projection on $R(\mathbf{A}^\dagger)$ then it is symmetric and $R(\mathbf{A}^\dagger) = R(\mathbf{A}^\dagger\mathbf{A})$ (why?)

$$\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}^H(\mathbf{A}^\dagger)^H$$

So $R(\mathbf{A}^\dagger)$ must be a subset of $R(\mathbf{A}^H)$. On the other hand $\mathbf{A}\mathbf{A}^\dagger$ is a projection on $R(A)$ therefore $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$, which implies

$$\mathbf{A}^H = (\mathbf{A}^\dagger\mathbf{A})^H\mathbf{A}^H = \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^H$$

So $R(\mathbf{A}^H)$ must be a subset of $R(\mathbf{A}^\dagger\mathbf{A})$.

Together it follows that $R(\mathbf{A}^H) = R(\mathbf{A}^\dagger\mathbf{A}) = R(\mathbf{A}^\dagger)$

$\mathbf{A}^\dagger$ and the MNLS Solution

Theorem: The mapping from $\mathbf{b}$ to $\mathbf{x}_{\text{MNLS}}$ is a linear mapping and is given by the pseudo-inverse:

$$\mathbf{x}_{\text{MNLS}} = \mathbf{A}^\dagger\mathbf{b}$$

Proof: Observe that $R(\mathbf{A}^\dagger) = R(\mathbf{A}^H)$, therefore $\mathbf{A}^\dagger\mathbf{b} \in R(\mathbf{A}^H)$. Furthermore, $\mathbf{A}^\dagger\mathbf{b}$ satisfies the normal equations because:

$$\begin{aligned}\mathbf{A}^H\mathbf{A}(\mathbf{A}^\dagger\mathbf{b}) &= \mathbf{A}^H(\mathbf{A}\mathbf{A}^\dagger)\mathbf{b} = \mathbf{A}^H(\mathbf{A}\mathbf{A}^\dagger)^H\mathbf{b} \\ &= (\mathbf{A}\mathbf{A}^\dagger\mathbf{A})^H\mathbf{b} \\ &= \mathbf{A}^H\mathbf{b}\end{aligned}$$

Therefore $\mathbf{A}^\dagger\mathbf{b}$ is the minimum-norm least-squares solution.

Since the min-norm least-square solution exists for every equation, the above theorem establishes the existence of the pseudo-inverse.

Uniqueness of the Pseudo-Inverse

Moore's uniqueness Lemma: The pseudo-inverse, if it exists, is unique.

Proof: If there are two speudo-inverses, they are both equal to the MNLS solution, which is unique, therefore the pseudo-inverses must be identical.

Penrose Conditions

Theorem: $\mathbf{B}$ is the pseudo-inverse of $\mathbf{A}$ iff it satisfies the Penrose conditions:

$$(\mathbf{AB})^H = \mathbf{AB} \qquad \mathbf{ABA} = \mathbf{A}$$
$$(\mathbf{BA})^H = \mathbf{BA} \qquad \mathbf{BAB} = \mathbf{B}$$

Proof: Necessity follows from definition of pseudo-inverse.

For sufficiency: Condition 1 shows $\mathbf{AB}$ is symmetric and idempotent, i.e. projection with range $R(\mathbf{AB})$ which is a subset of $R(\mathbf{A})$. But $\mathbf{ABA} = \mathbf{A}$ shows $R(\mathbf{A}) \subset R(\mathbf{AB})$, so $R(\mathbf{A}) = R(\mathbf{AB})$, and $\mathbf{AB}$ is the projection on $R(\mathbf{A})$. Similarly, the other two conditions show $\mathbf{BA}$ is the projection on $R(\mathbf{B})$. This is the definition of pseudo-inverse.

Fundamental Subspaces and Inverses

- $R(\mathbf{A}) = R(\mathbf{A}\mathbf{A}^*)$
- $N(\mathbf{A}) = N(\mathbf{A}^*\mathbf{A})$
- **Corollary:** $\mathbf{A}^H\mathbf{A}\mathbf{B} = \mathbf{A}^H\mathbf{A}\mathbf{C}$ iff $\mathbf{A}\mathbf{B} = \mathbf{A}\mathbf{C}$
- If $\mathbf{A}$ has full column rank, then $\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$
- If $\mathbf{A}$ has full row rank, then $\mathbf{A}^\dagger = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1}$