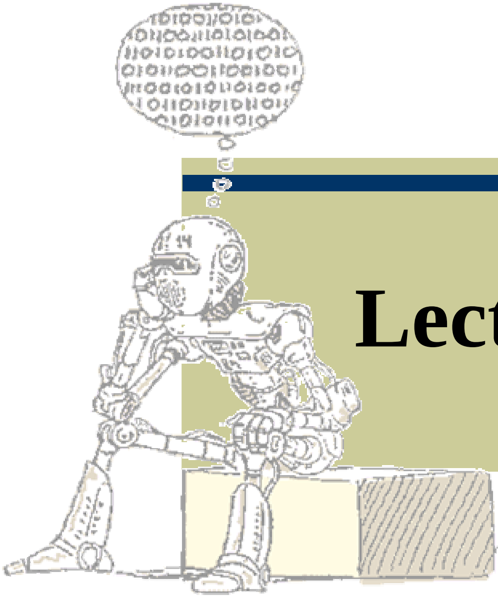


Lecture II: Linear Algebra Revisited

Overview

- Vector spaces, Hilbert & Banach Spaces, Metrics & Norms
- Matrices, Eigenvalues, Orthogonal Transformations, Singular Values
- Operators, Operator Norms, Function Spaces

Note: *We will need many of these concepts as basic tools to quantify and evaluate the performance of machine learning algorithms and also to come up with more efficient and effective solutions …*

Vectors

Usually denoted by lower case, bold letters, e.g. **x, y**

Operations :

- Multiplication by scalar ($\alpha\mathbf{x}$).
- Addition of vectors ($\mathbf{x}+\mathbf{y}$) – **x** and **y** have to be of same dimensions.
- Linear combination.
  - $\mathbf{u} = \alpha\mathbf{x}+\beta\mathbf{y}$ (**x** and **y** have to be of same dimensions).
- Angle between vectors. $\cos\theta = \dfrac{<\mathbf{v},\mathbf{w}>}{\|\mathbf{v}\|\,\|\mathbf{w}\|}$
- Linear independence.
  - When one vector cannot be written as a linear combination of other, then the vectors are said to be *linearly independent*.

Metric

Definition 1 (Metric/ *Distance*)

Denote by $\mathcal{X}$ *a space. Then* $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_0^+$ *is a metric on* $\mathcal{X}$ *if for all* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X}$

1. $d(\mathbf{x}, \mathbf{y}) = 0$ *is equivalent to* $\mathbf{x} = \mathbf{y}$

2. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$

3. $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$ *(Triangle Inequality)*

Example 1 (Trivial Metric)

For arbitrary $\mathcal{X}$ *define* $d(\mathbf{x}, \mathbf{y}) = 1$ *if* $\mathbf{x} \neq \mathbf{y}$ *and* $d(\mathbf{x}, \mathbf{y}) = 0$ *if* $\mathbf{x} = \mathbf{y}$.

Example 2 (Manhattan Distance)

For $\mathcal{X} = \mathbb{R}^n$ *define* $d(\mathbf{x}, \mathbf{y}) := \sum_{i=1}^{n} |x_i - y_i|$.

Vector Spaces

Definition 2 (Vector Spaces)

A space $\mathcal{X}$ *on which for all* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ *and for all* $\alpha \in \mathbb{R}$ *the following operations are defined:*

1. $\mathbf{x} + \mathbf{y} \in \mathcal{X}$ *(Addition)*
2. $\alpha \mathbf{x} \in \mathcal{X}$ *(Multiplication)*

Definition 3 (Cauchy Series)

Given a space $\mathcal{X}$, *a series* $\mathbf{x}_i \in \mathcal{X}$ *with* $i \in \mathbb{N}$ *is a Cauchy series if for any* ϵ *there exists an* n_0 *such that for all* $m, n \geq n_0$ *we have* $d(\mathbf{x}_m, \mathbf{x}_n) \leq \varepsilon$.

Definition 4 (Completeness)

A space $\mathcal{X}$ *is complete if the limits of every Cauchy series are elements of* $\mathcal{X}$.
We call $\bar{\mathcal{X}}$ *the completion of* $\mathcal{X}$, *i.e. the union of* $\mathcal{X}$ *and all its limits of Cauchy series.*

Examples of Vector Spaces

Rational Numbers, **Real Numbers**, **Polynomials** are all Vector Spaces

Series

series (a_i) of numbers with $a_i \in \mathbb{R}$ and $i \in \mathbb{N}$ are clearly vector spaces.

Fourier Expansions

expansions via the discrete Fourier transform form a vector space where

$$f(x) = \sum_{j=1}^{n} s_j \sin(jx) + c_j \cos(jx)$$

Functions

many classes of functions, e.g., $f : [0, 1] \to \mathbb{R}$.

Counterexamples

- $f : [0, 1] \to [0, 1]$ does not form a vector space!
- $\mathbb{Z}$ is not a vector space, unless we only allow multiplications by integers.
- The alphabet $\{a, \ldots, z\}$ is not a vectorspace (still it can be an interesting mathematical object, e.g. when determining similarity of documents).

Norm & Banach Spaces

Definition 5 (Norm / *Length*)

Given a vector space $\mathcal{X}$, *a mapping* $\|\cdot\| : \mathcal{X} \to \mathbb{R}_0^+$ *is called a norm if for all* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ *and all* $\alpha \in \mathbb{R}$ *it satisfies*

1. $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = 0$
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ *(scaling)*
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ *(triangle inequality)*

A mapping $\|\cdot\|$ not satisfying (1) is called **pseudo norm**.
Note that a norm also introduces a **metric** via $d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|$.

Definition 6 (Banach Space)

A complete vector space $\mathcal{X}$ *together with a norm* $\|\cdot\|$.

Examples of Banach Spaces

ℓ_p Spaces

These are subspaces of $\mathbb{R}^{\mathbb{N}}$ with $\|\mathbf{x}\| := \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{\frac{1}{p}}$.

Not for all series x_i the sum converges, e.g., $x_i = \frac{1}{i}$ is in ℓ_2 but not in ℓ_1.

Function Spaces $L_p(\mathfrak{X})$

We replace sums by integrals over $\mathfrak{X}$ and obtain $\|f\| := \left(\int_{\mathfrak{X}} |f(x)|^p dx\right)^{\frac{1}{p}}$. Again, not for all f this integral is defined, i.e. they are not elements of the corresponding $L_p(\mathfrak{X})$.

Dot Products & Hilbert Spaces

Definition 7 (Dot Product/ *Inner Product*)

Given a vector space $\mathcal{X}$, *a mapping* $\langle \cdot, \cdot \rangle$ *with* $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *which for all* $\alpha \in \mathbb{R}$ *and* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X}$ *satisfies*

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ *(symmetry)*
2. $\langle \mathbf{x}, \alpha\mathbf{y} \rangle = \alpha\langle \mathbf{x}, \mathbf{y} \rangle$ *(linearity)*
3. $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$ *(additivity)*

Example :

$$\mathbf{v} = \begin{bmatrix} \mathbf{3} \\ \mathbf{4} \end{bmatrix}; \quad \|\mathbf{v}\| = \langle \mathbf{v}, \mathbf{v} \rangle^{\frac{1}{2}} = \sqrt{3^2 + 4^2} = 5$$

Definition 8 (Hilbert Space)

A complete vector space $\mathcal{X}$, *endowed with a dot product* $\langle \cdot, \cdot \rangle$.

The dot product automatically generates a norm (and a metric) by $\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Thus Hilbert spaces are special case of a Banach space.

Examples of Hilbert Spaces

Euclidean Spaces Use standard dot product for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ given by $\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^{m} x_i y_i$

Function Spaces ($L_2(X)$) Functions on X with $f : X \to \mathbb{C}$ for all $f \in \mathfrak{F}$. Here we can define the dot product for $f, g \in \mathfrak{F}$ by $\langle f, g \rangle := \int_X \overline{f(x)} g(x) dx$ Note that we take the complex conjugate of f. Also note that all we did was to replace the sum by an integral.

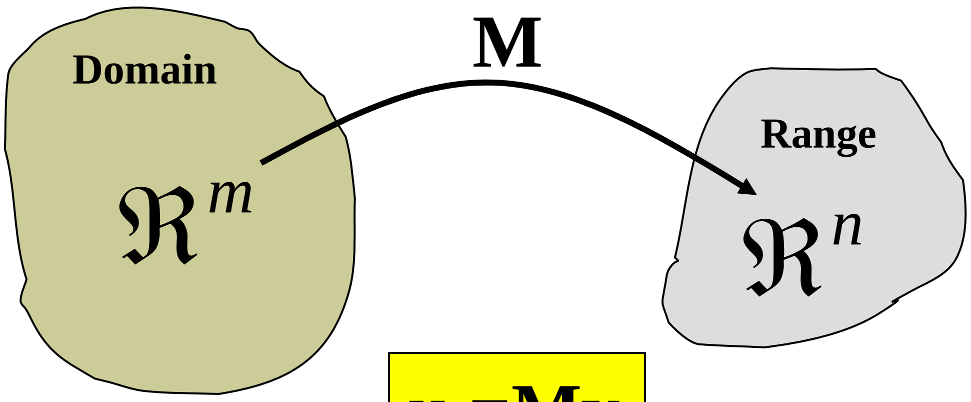
ℓ_2 (Infinite) series of real numbers, $\ell_2 \subset \mathbb{R}^{\mathbb{N}}$. We define a dot product for $\mathbf{x}, \mathbf{y} \in \ell_2$ by $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{\infty} x_i y_i$

Polarization Inequality We can recover the dot product from the norm $\|\mathbf{x}\|$ by computing $\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 = 2\langle \mathbf{x}, \mathbf{y} \rangle$.

Matrices

In the following we assume that a matrix $M \in \mathbb{R}^{m \times n}$ corresponds to a linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ and is given by its entries $M_{ij} \in \mathbb{R}$.

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}$$



$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

$$\boxed{\mathbf{u} = \mathbf{M}\mathbf{v}}$$

Review:

- Addition of Matrices
- Multiplication of matrices by scalars, vectors and matrices.
- Domain and Range of a Matrix

Special matrices

Square and Diagonal Matrix

A **square** matrix has equal number of rows and columns. A **diagonal** matrix has all off-diagonal elements zero.

Symmetric Matrix

A symmetric matrix $M \in \mathbb{R}^{m \times m}$ satisfies $M_{ij} = M_{ji}$.

Anti-symmetric Matrix

An antisymmetric matrix $M \in \mathbb{R}^{m \times m}$ satisfies $M_{ij} = -M_{ji}$.

Orthogonal Matrix

A matrix $M \in \mathbb{R}^{m \times m}$ with $M^\top M = \mathbf{1}$ is called an orthogonal matrix (if $M \in \mathbb{C}^{m \times m}$ it is called unitary). This means $M^\top = M^{-1}$. (Often denoted as **O**(m))

Matrix Concepts

Rank of a Matrix

Denote by I the image of $\mathbb{R}^m$ under $M \in \mathbb{R}^{m \times n}$. Since M is a linear map, we can find a I as a linear combination of vectors. $\text{rank}(M)$ is the smallest number of such vectors that span I.

Range, Domain and Null Space

Range of **M**, denoted by $\mathcal{R}(\mathbf{M})$, is the space of all vectors that can be obtained by the operation of **M** on the vectors in the domain of **M** denoted by $\mathcal{D}(\mathbf{M})$.

$$\boxed{v \in R(\mathbf{M}) \; iff \; \exists u \in D(\mathbf{M}) \; s.t. \; u\mathbf{M} = v}$$

Null Space of **M**, denoted by $\mathcal{N}(\mathbf{M})$, is a subspace of all the vectors in the domain of **M** $\mathcal{D}(\mathbf{M})$ that map to the zero (null) vector in $\mathcal{R}(\mathbf{M})$ when operated upon by the matrix **M**.

$$\boxed{v \in N(\mathbf{M}) \; iff \; v\mathbf{M} = \mathbf{0}}$$

Matrix Invariants

Trace

$\mathrm{tr}M := \sum_{i=1}^{m} M_{ii}$ One can show $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ and thus for symmetric matrices

$$\mathrm{tr}M = \mathrm{tr}(O^{\top}\Lambda O) = \mathrm{tr}(\Lambda O O^{\top}) = \mathrm{tr}\Lambda = \sum_{i=1}^{m} \lambda_i$$

Properties of Trace : $tr(a\mathbf{A}) = a\,tr(\mathbf{A})$

$$tr(\mathbf{A}+\mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$$

$$tr(\mathbf{AB}) = tr(\mathbf{BA})$$

Determinant

Determinant can be written as the product of the eigenvalues : $\det M = \prod_{i=1}^{m} \lambda_i$

Note: ***Trace and Determinant are invariant under orthogonal transformation***

Matrix Norms

Operator Norm

The norm of a linear operator A between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$ is defined as

$$\|A\| := \max_{\mathbf{x}\in\mathcal{X}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$$

This clearly satisfies all conditions of a norm:

- $\|\alpha A\| = \max_{\mathbf{x}\in\mathcal{X}} \frac{\|\alpha A\mathbf{x}\|}{\|\mathbf{x}\|} = |\alpha|\|A\|$.
- $\|A+B\| = \max_{\mathbf{x}\in\mathcal{X}} \frac{\|(A+B)\mathbf{x}\|}{\|\mathbf{x}\|} \leq \max_{\mathbf{x}\in\mathcal{X}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} + \max_{\mathbf{x}\in\mathcal{X}} \frac{\|B\mathbf{x}\|}{\|\mathbf{x}\|} = \|A\| + \|B\|$
- $\|A\| = 0$ implies $\max_{\mathbf{x}\in\mathcal{X}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = 0$ and thus $A\mathbf{x} = 0$ for all $\mathbf{x}$. This means that $A = 0$.

Frobeius Norm

For a matrix $M \in \mathbb{R}^{m\times n}$ we can define a norm in analogy to the vector norm by

$$\|M\|^2_{\mathrm{Frob}} = \sum_{i=1}^{m}\sum_{j=1}^{n} M_{ij}^2$$

Eigensystems

Definition 9 (Eigenvalues/ Eigenvectors)

Denote by $M \in \mathbb{R}^{m \times m}$ *matrix, then an eigenvalue* $\lambda \in \mathbb{R}$ *and eigenvector* $\mathbf{x} \in \mathbb{R}^m$ *satisfy*

$$M\mathbf{x} = \lambda\mathbf{x}$$

IMP: ***Defined only for square Matrices***

- **Intuitive Explanation.**
  - A square matrix **M** is a mapping from ***n***-dim to ***n***-dim space.
  - Most vectors change *both* direction and length when undergoing this mapping transformation.
  - Those vectors which ***only change length*** (i.e., multiplying them by a matrix is similar to multiplying by a scalar) are called eigenvectors and the eigenvalue indicates how much they are shortened or lengthened.

Eigensystems II

Eigenvectors/Eigenvalues of Symmetric Matrices

- All *eigenvalues* of symmetric matrices are **real**
- Symmetric matrices are fully **diagonalizable**, i.e. we can find *m* eigenvectors
- All *eigenvectors* of symmetric matrices M with different eigenvalues are **mutually orthogonal** (Prove !!)

Decomposition of Symmetric Matrices

We can decompose symmetric $M \in \mathbb{R}^{m \times m}$ into $O^\top \Lambda O$ where $O \in SO(n)$ and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$.

Example

$M = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ has eigenvalues $(-1, 3)$ and eigenvectors $v_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

Positive Matrices

Definition 10 (Positive Definite Matrices)

A matrix $M \in \mathbb{R}^{m \times m}$ for which for all $\mathbf{x} \in \mathbb{R}^m$ we have

$$\mathbf{x}^\top M \mathbf{x} \geq 0 \text{ if } \mathbf{x} \neq 0$$

This matrix has only positive eigenvalues since for all eigenvectors $\mathbf{x}$ we have $\mathbf{x}^\top M \mathbf{x} = \lambda \mathbf{x}^\top \mathbf{x} = \lambda \|\mathbf{x}\|^2 > 0$ and thus $\lambda > 0$.

Induced Norm and Metrics

Every positive definite matrix induces a norm via

$$\|\mathbf{x}\|_M^2 := \mathbf{x}^\top M \mathbf{x}$$

Mahalanobis distance

- Linearity is obvious, so is uniqueness
- The triangle inequality can be seen by writing

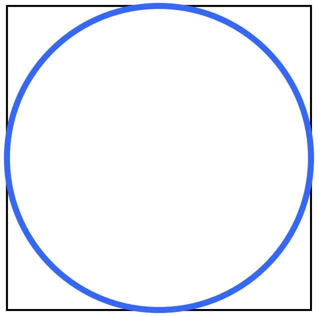
$$\|\mathbf{x} + \mathbf{x}'\|_M^2 = (\mathbf{x} + \mathbf{x}')^\top M^{\frac{1}{2}} M^{\frac{1}{2}} (\mathbf{x} + \mathbf{x}') = \|M^{\frac{1}{2}}(\mathbf{x} + \mathbf{x}')\|^2$$

and using the triangle inequality for $M^{\frac{1}{2}}\mathbf{x}$ and $M^{\frac{1}{2}}\mathbf{x}'$.

Mahalanobis Distance

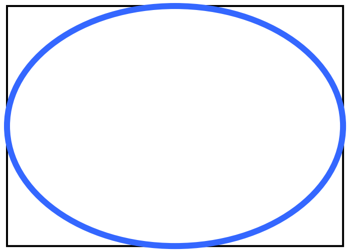
$$d(\mathbf{x},\mathbf{y})_{\mathbf{M}} = \langle \mathbf{x},\mathbf{y} \rangle_{\mathbf{M}} = \mathbf{x}^T \mathbf{M} \mathbf{y}$$

$$\mathbf{M} = \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

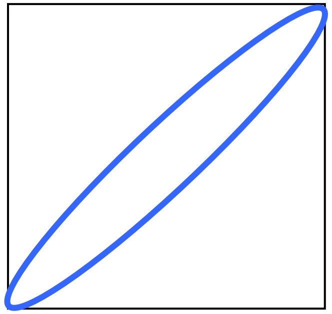


= Euclidean Distance

$$\mathbf{M} \neq \mathbf{I} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mathbf{M} = \begin{bmatrix} 2 & 1.4 \\ 1.4 & 1 \end{bmatrix}$$



Singular Value Decomposition

Note: Eigenvalue/Eigenvector decompositions are valid only for *square matrices*

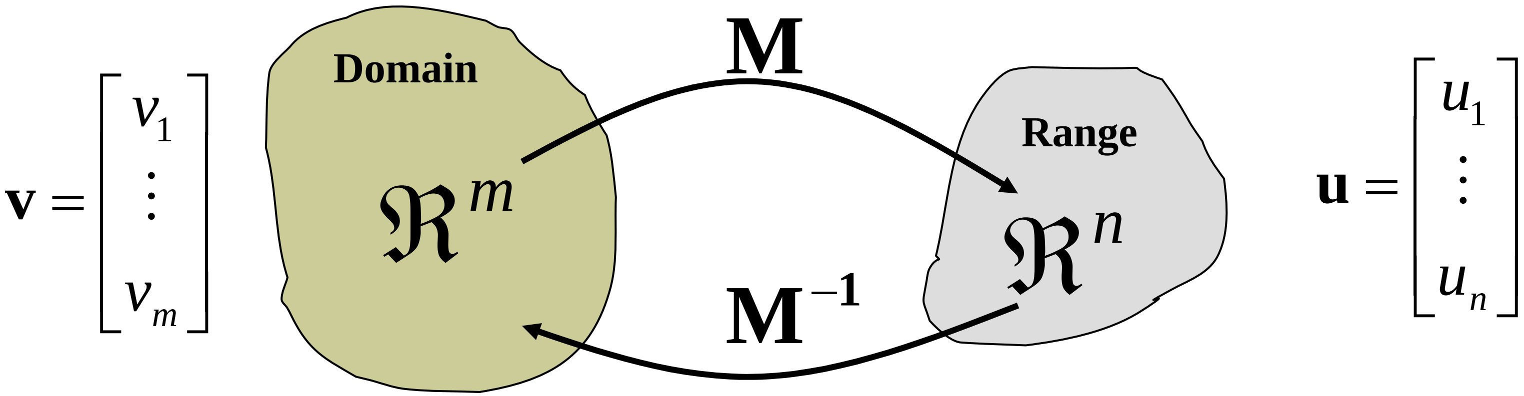
Do we have some decomposition for rectangular matrices ??

Singular value Decomposition (SVD)

Without loss of generality assume $m \geq n$ For $M \in \mathbb{R}^{m \times n}$ we may write M as $U \Lambda O$ where $U \in \mathbb{R}^{m \times n}$, $O \in \mathbb{R}^{n \times n}$, and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

Furthermore $O^\top O = O O^\top = U^\top U = \mathbf{1}$.

Matrix Inverse



$$\mathbf{M}^{-1}\mathbf{M} = \mathbf{I}; \quad \mathbf{M}\mathbf{M}^{-1} = \mathbf{I}$$

Note: A regular inverse exists only for square matrices with linearly independent column vectors

Interpretation: *We need a one-to-one mapping to uniquely go from one element of a space to another and back. Square matrices and linearly independent columns ensure this !!*

Pseudoinverse

For rectangular matrices and square matrices with linearly dependent columns, there exists the **pseudoinverse** or **generalized inverse** which performs the inverse mapping. In general, these inverses are not unique.

The Moore-Penrose Pseudoinverse ($\mathbf{M}^{+}$)

$$\mathbf{M}^{+} = (\mathbf{M}^{T}\mathbf{M})^{-1}\mathbf{M}^{T} \ \ or\ \mathbf{M}^{T}(\mathbf{M}\mathbf{M}^{T})^{-1}$$

The above generalized inverse is called the Moore-Penrose Psuedoinverse and is ***unique****. Among the multiple inverse solutions, it chooses the one with the* ***minimum norm****.*

Operators

Linear Operators

Generalization of matrix – a mapping from one Banach space to another. Norms and eigenvalues/eigenvectors are defined as for matrices; so are Range & Null Spaces.

Notation

A : F→G denotes a linear operator **A** mapping from space F to space G.

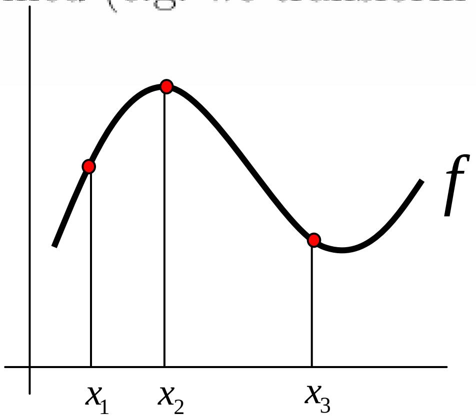
A Matrix-Operator Correspondence

Matrix Transpose $\mathbf{A}^T$	Adjoint Operator $\langle \mathbf{A}f, g\rangle = \langle f, \mathbf{A}^* g\rangle$ *for all* $f \in F, g \in G.$
Symmetric Matrix $\mathbf{A} = \mathbf{A}^T$	Self Adjoint Operator $\langle \mathbf{A}f, g\rangle = \langle \mathbf{A}^* f, g\rangle$ *for all* $f \in F, g \in G.$
Orthogonal Matrix $\mathbf{A}^{-1} = \mathbf{A}^T$	Isometry $\langle f, g\rangle = \langle \mathbf{A}f, \mathbf{A}g\rangle$ *for all* $f \in F, g \in G.$

Linear Operators - *Examples*

Input transformation

Consider class of functions $f:\mathbb{R}^n \to \mathbb{R}$. A linear operator on such functions could be $A: f(\cdot) \to f(a(\cdot))$, i.e. the argument of f is transformed (e.g. we transform the images before feeding them into a classifier).



Sampling

Sampling from a function to yield scalar outputs.
(We will see later why this is so !!!)

Fourier Transform

We map f into its Fourier transform. This leads to

$$f \to \tilde{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx$$

This map is an isometry, since

$$\|f\|^2 = \int_{\mathbb{R}} |f(x)|^2 dx = \int_{\mathbb{R}} |\tilde{f}(\omega)|^2 d\omega = \|\tilde{f}\|^2$$

Range and Null Space of Operators

Recollect: Definition of Range, Domain and Null space of a matrix