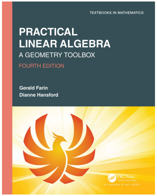
Practical Linear Algebra: A Geometry Toolbox

Fourth Edition

Chapter 14: General Linear Spaces

Gerald Farin & Dianne Hansford

A K Peters/CRC Press
www.farinhansford.com/books/pla



Outline

General Linear Spaces

All cubic polynomials over the interval [0,1] form a linear space
Some elements illustrated

Linear space = vector space
— Chapters 4 & 9: 2D & 3D

Here: higher dimensions
— Spaces can be abstract
— Powerful concept in dealing with real-life problems
- car crash simulations
- weather forecasts
- computer games

"General" refers to the dimension and abstraction

Basic Properties of Linear Spaces

$\mathcal{L}_n$: linear space of dimension n

Elements of $\mathcal{L}_n$ are vectors
— Denoted by boldface letters such as $\mathbf{u}$

Two operations defined on the elements of $\mathcal{L}_n$:
— Addition
— Multiplication by a scalar

Linearity property
Any *linear combination* of vectors results in a vector in the same space

$$\mathbf{w} = s\mathbf{u} + t\mathbf{v}$$

Both s and t may be zero $\Rightarrow$ every linear space has a zero vector in it

Basic Properties of Linear Spaces

Generalize linear spaces: include new kinds of vectors
— Objects in the linear space are not always in traditional vector format
— Key: the linearity property

Example: $\mathbb{R}^2$

$$\text{Elements of space:} \quad \mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$$

$$\Rightarrow \quad \mathbf{w} = 2\mathbf{u} + \mathbf{v} = \begin{bmatrix} 0 \\ 5 \end{bmatrix} \quad \text{is also in } \mathbb{R}^2$$

Example: Linear space $\mathcal{M}_{2\times 2}$ – the set of all 2×2 matrices
— Rules of matrix arithmetic guarantee the linearity property

Example: $\mathcal{V}_2$ – all vectors $\mathbf{w}$ in $\mathbb{R}^2$ that satisfy $w_2 \geq 0$
— $\mathbf{e}_1$ and $\mathbf{e}_2$ live in $\mathcal{V}_2$ — Is this a linear space?

Basic Properties of Linear Spaces

Objects in general linear spaces are not always in the traditional vector format $\Rightarrow$ favor *linear space* over *vector space*

To ensure that linearity property acts as we expect
a more detailed set of rules (axioms) might be helpful

1. If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathcal{L}$, then $\mathbf{u}+\mathbf{v}$ is in $\mathcal{L}$
2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
3. $\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{u}+\mathbf{v})+\mathbf{w}$
4. The zero vector $\mathbf{0}$ is in $\mathcal{L}$ such that $\mathbf{0}+\mathbf{u}=\mathbf{u}+\mathbf{0}=\mathbf{u}$
5. For each $\mathbf{u}$ in $\mathcal{L}$, there is a $-\mathbf{u}$, such that $\mathbf{u}+(-\mathbf{u})=(-\mathbf{u})+\mathbf{u}=\mathbf{0}$
6. If $\mathbf{u}$ is in $\mathcal{L}$, then $s\mathbf{u}$ is in $\mathcal{L}$
7. $s(\mathbf{u}+\mathbf{v})=s\mathbf{u}+s\mathbf{v}$
8. $(s+t)\mathbf{u}=s\mathbf{u}+t\mathbf{v}$
9. $\mathrm{s}(\mathrm{t}\mathbf{u})=(\mathrm{st})\mathbf{u}$
10. $1\mathbf{u}=\mathbf{u}$

Axioms satisfied $\Rightarrow$ tools of linear algebra available

In $\mathcal{L}_n$ define a set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ where $1 \leq r \leq n$

Vectors are linearly independent means

$$\mathbf{v}_1 = s_2\mathbf{v}_2 + s_3\mathbf{v}_3 + \ldots + s_r\mathbf{v}_r$$

Will *not* have a solution set $s_2, \ldots, s_r$

$\Rightarrow$ Zero vector can only be expressed in a trivial manner:

$$\text{If} \quad \mathbf{0} = s_1\mathbf{v}_1 + \ldots + s_r\mathbf{v}_r \quad \text{then } s_1 = \ldots = s_r = 0$$

If the zero vector *can* be expressed as a nontrivial combination of r vectors then these vectors are linearly dependent

Basic Properties of Linear Spaces

Subspace of $\mathcal{L}_n$ of dimension r:

Formed from all *linear combinations* of linearly independent $\mathbf{v}_1, \ldots, \mathbf{v}_r$

$\Rightarrow$ Subspace is spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_r$

If this subspace equals whole space $\mathcal{L}_n$ then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ a basis for $\mathcal{L}_n$

If $\mathcal{L}_n$ is a linear space of dimension n

then any $n + 1$ vectors in it are linearly dependent

Next: two examples to practice terminology

Example: $\mathbb{R}^3$ and basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$

$$\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 7 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{is also in } \mathbb{R}^3$$

The four vectors $\mathbf{v}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are linearly dependent

Any *one* of four vectors forms a *one-dimensional subspace* of $\mathbb{R}^3$

Any *two* vectors here form a *two-dimensional subspace* of $\mathbb{R}^3$

Basic Properties of Linear Spaces

Example: $\mathbb{R}^4$

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 5 \\ 0 \\ -3 \\ 1 \end{bmatrix} \qquad \mathbf{v}_3 = \begin{bmatrix} 3 \\ 0 \\ -3 \\ 0 \end{bmatrix}$$

These vectors are linearly dependent since

$$\mathbf{v}_2 = \mathbf{v}_1 + 2\mathbf{v}_3 \quad \text{or} \quad \mathbf{0} = \mathbf{v}_1 - \mathbf{v}_2 + 2\mathbf{v}_3$$

Set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ contains only two linearly independent vectors

$\Rightarrow$ Any two of them spans a subspace of $\mathbb{R}^4$ of dimension two

Example: $\mathbb{R}^3$

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ -3 \end{bmatrix}$$

These four vectors are linearly dependent since

$$\mathbf{v}_3 = -\mathbf{v}_1 + 2\mathbf{v}_2 + \mathbf{v}_4$$

Any set of three of these vectors is a basis for $\mathbb{R}^3$

$A : \mathcal{L}_n \rightarrow \mathcal{L}_m$ means that linear map A that transforms $\mathcal{L}_n$ to $\mathcal{L}_m$

Linear map represented as an $m \times n$ matrix A

Preimage $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ in $\mathcal{L}_n$ mapped to image $A\mathbf{v}_1, A\mathbf{v}_2, A\mathbf{v}_3$ in $\mathcal{L}_m$

A preserves linear relationships means that

$$\mathbf{v}_1 = \alpha \mathbf{v}_2 + \beta \mathbf{v}_3 \quad \Rightarrow \quad A\mathbf{v}_1 = \alpha A\mathbf{v}_2 + \beta A\mathbf{v}_3$$

(Maps without this property are called nonlinear maps)

Suppose $A : [\mathbf{e}_1, \ldots, \mathbf{e}_n]$-system $\quad \rightarrow \quad [\mathbf{a}_1, \ldots, \mathbf{a}_n]$-system then

$$\mathbf{v}' = v_1 \mathbf{a}_1 + v_2 \mathbf{a}_2 + \ldots v_n \mathbf{a}_n \quad \text{is in the column space of } A$$

Example: $A : \mathbb{R}^2 \rightarrow \mathbb{R}^3$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 2 \end{bmatrix}$$

Given vectors in R^2 $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

mapped to vectors in $\mathbb{R}^3$ $\hat{\mathbf{v}}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ $\hat{\mathbf{v}}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$ $\hat{\mathbf{v}}_3 = \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix}$

$\mathbf{v}_i$ are *linearly dependent* since $\mathbf{v}_3 = 2\mathbf{v}_1 + \mathbf{v}_2$

Linear maps preserve linear relationships $\Rightarrow$ $\mathbf{v}'_3 = 2\mathbf{v}'_1 + \mathbf{v}'_2$

Matrix rank

- $m \times n$ matrix can be at most of rank $k = \min\{m, n\}$
- Rank equals number of linearly independent column vectors
- If $\text{rank}(A) = \min\{m, n\} \Rightarrow$ full rank
- If $\text{rank}(A) < \min\{m, n\} \Rightarrow$ rank deficient

- Linear map can never *increase* dimension

— Images of n basis vectors will span a subspace of dimension at most n

— See the last Example

- How to identify rank?

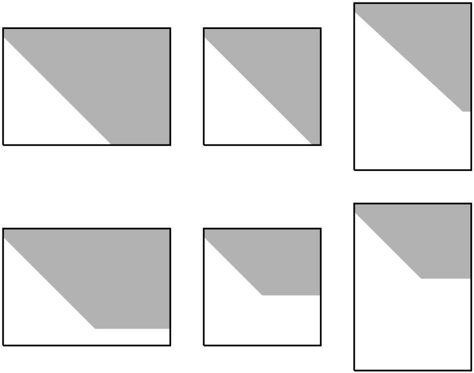
— Forward elimination to upper triangular form

— k nonzero rows $\Rightarrow$ rank is k

Rank scenarios for an $m \times n$ matrix in upper triangular form

$$m < n \qquad m = n \qquad m > n$$

Top row: full rank matrices



Bottom row: rank deficient matrices

Example: $\begin{bmatrix} 1 & 3 & 4 \\ 0 & 1 & 2 \\ 1 & 2 & 2 \\ -1 & 1 & 1 \end{bmatrix}$ Forward elimination $\Rightarrow$ $\begin{bmatrix} 1 & 3 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix}$

Rank 3 — full rank since $min\{4,3\} = 3$

Example: $\begin{bmatrix} 1 & 3 & 4 \\ 0 & 1 & 2 \\ 1 & 2 & 2 \\ 0 & 1 & 2 \end{bmatrix}$ Forward elimination $\Rightarrow$ $\begin{bmatrix} 1 & 3 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Rank 2 — rank deficient since $min\{4,3\} = 3 > 2$

Linear Maps

Review of features of linear maps from earlier chapters

$n \times n$ matrix A that is rank n is *invertible* $\Rightarrow$ *inverse matrix* A^{-1} exists

If A is invertible then it does not reduce dimension

$\Rightarrow$ *determinant* is non-zero

— Measures volume of nD parallelepiped defined by columns vectors

— Computed by

- transforming matrix to upper triangular
- determinant is the product of the diagonal elements
- if pivoting required: careful of sign

Inner Products

Inner product: a map from $\mathcal{L}_n$ to the reals $\mathbb{R}$ — denoted as $\langle \mathbf{v}, \mathbf{w} \rangle$

Properties:

Symmetry: $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$

Homogeneity: $\langle \alpha\mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{w}, \mathbf{v} \rangle$

Additivity: $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\mathbf{v}$ $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$

Positivity: $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$

Homogeneity and additivity properties combined:

$$\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{w} \rangle + \beta \langle \mathbf{v}, \mathbf{w} \rangle$$

Example: the *dot product* $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n$

Inner product space: a linear space with an inner product

Example: Define a "test" inner product in $\mathbb{R}^2$

$$\langle \mathbf{v}, \mathbf{w} \rangle = 4v_1 w_1 + 2v_2 w_2$$

Compare it to the dot product:

$$\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 4(1)(0) + 2(0)(1) = 0 \qquad \mathbf{e}_1 \cdot \mathbf{e}_2 = 0$$

Let $\mathbf{r} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ (unit vector)

$$\langle \mathbf{e}_1, \mathbf{r} \rangle = 4(1)(\frac{1}{\sqrt{2}}) + 2(0)(\frac{1}{\sqrt{2}}) = \frac{4}{\sqrt{2}} \qquad \mathbf{e}_1 \cdot \mathbf{r} = \frac{1}{\sqrt{2}}$$

Inner Products

Does the test inner product satisfy the necessary properties?

$$\text{Symmetry: } \langle \mathbf{v}, \mathbf{w} \rangle = 4v_1w_1 + 2v_2w_2 = 4w_1v_1 + 2w_2v_2 = \langle \mathbf{w}, \mathbf{v} \rangle$$

$$\text{Homogeneity: } \langle \alpha\mathbf{v}, \mathbf{w} \rangle = 4(\alpha v_1)w_1 + 2(\alpha v_2)w_2 = \alpha(4v_1w_1 + 2v_2w_2) = \alpha\langle \mathbf{v}, \mathbf{w} \rangle$$

$$\begin{aligned} \text{Additivity: } \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= 4(u_1 + v_1)w_1 + 2(u_2 + v_2)w_2 \\ &= (4u_1w_1 + 2u_2w_2) + (4v_1w_1 + 2v_2w_2) \\ &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

$$\text{Positivity: } \quad \langle \mathbf{v}, \mathbf{v} \rangle = 4v_1^2 + 2v_2^2 \geq 0 \text{ and } \langle \mathbf{v}, \mathbf{v} \rangle = 0 \text{ iff } \mathbf{v} = \mathbf{0}$$

Usefulness of this inner product? But it does satisfy the properties!

Inner Products

Length

2-norm or Euclidean norm: $\|\mathbf{v}\|_2 = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$

(Subscript typically omitted for this "usual" norm)

Distance between two vectors

$$\mathrm{dist}(\mathbf{u}, \mathbf{v}) = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle} = \|\mathbf{u} - \mathbf{v}\|$$

Example: the dot product in $\mathbb{R}^n$

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

$$\mathrm{dist}(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \ldots + (u_n - v_n)^2}$$

Inner Products

Norm and distance for two inner products

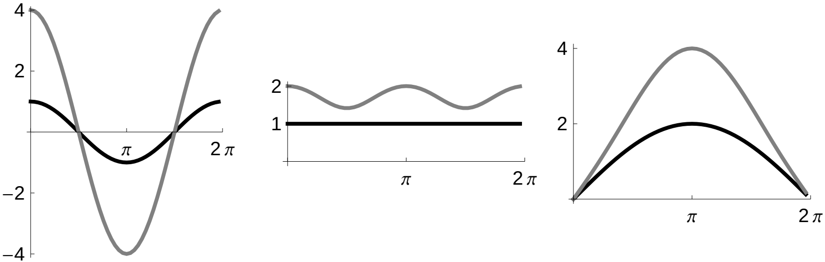
Test inner product	**Dot product**
$\langle \mathbf{v}, \mathbf{w} \rangle = 4v_1w_1 + 2v_2w_2$	$\langle \mathbf{v}, \mathbf{w} \rangle = v_1w_1 + v_2w_2$
$\\|\mathbf{e}_1\\| = \sqrt{\langle \mathbf{e}_1, \mathbf{e}_1 \rangle} = 4(1)^2 + 2(0)^2 = 4$	$\\|\mathbf{e}_1\\| = 1$
$\mathrm{dist}(\mathbf{e}_1, \mathbf{e}_2) = \sqrt{4(1-0)^2 + 2(0-1)^2} = \sqrt{6}$	$\mathrm{dist}(\mathbf{e}_1, \mathbf{e}_2) = \sqrt{2}$

Inner Products

Compare dot product and the test inner product $\langle \mathbf{v}, \mathbf{w} \rangle = 4v_1w_1 + 2v_2w_2$

Set of vectors: unit vectors $\mathbf{r}$ rotated through $[0, 2\pi]$

Black curve: dot product   Gray curve: test inner product



Left: inner product $\mathbf{e}_1 \cdot \mathbf{r}$ and $\langle \mathbf{e}_1, \mathbf{r} \rangle$

Middle: length $\sqrt{\mathbf{r} \cdot \mathbf{r}}$ and $\sqrt{\langle \mathbf{r}, \mathbf{r} \rangle}$

Right: distance $\sqrt{(\mathbf{e}_1 - \mathbf{r}) \cdot (\mathbf{e}_1 - \mathbf{r})}$ and $\sqrt{\langle (\mathbf{e}_1 - \mathbf{r}), (\mathbf{e}_1 - \mathbf{r}) \rangle}$

Inner Products

Orthogonality: $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for $\mathbf{v}, \mathbf{w}$ in $\mathcal{L}_n$

Orthogonal basis: $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis for $\mathcal{L}_n$
and all $\mathbf{v}_i$ are mutually orthogonal: $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for $i \neq j$

Mutually orthogonal and unit length: $\|\mathbf{v}_i\| = 1$

$\Rightarrow$ form an orthonormal basis

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Next section: the *Gram-Schmidt method*:
— Tool to transform a basis of a linear space into an orthonormal basis

Cauchy-Schwartz inequality — in the context of inner product spaces

$$\langle \mathbf{v}, \mathbf{w} \rangle^2 \leq \langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle$$

Equality holds if and only if **v** and **w** linearly dependent

Restate the Cauchy-Schwartz inequality

$$\langle \mathbf{v}, \mathbf{w} \rangle^2 \leq \|\mathbf{v}\|^2 \|\mathbf{w}\|^2$$

$$\left(\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \|\mathbf{w}\|} \right)^2 \leq 1$$

$$-1 \leq \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \|\mathbf{w}\|} \leq 1$$

Angle θ between **v** and **w**

$$\cos \theta = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \|\mathbf{w}\|}$$

Inner Products

Inner product properties suggest

$$\|\mathbf{v}\| \geq 0$$
$$\|\mathbf{v}\| = 0 \text{ if and only if } \mathbf{v} = 0$$
$$\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|$$

A fourth property is the triangle inequality:

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$$

(derived from the Cauchy-Schwartz inequality in Chapter 2)

Inner Products

General definition of a projection

Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ span a subspace $\mathcal{L}_k$ of $\mathcal{L}$

If $\mathbf{v}$ is a vector *not* in $\mathcal{L}_k$ then

$$P\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1 + \ldots + \frac{\langle \mathbf{v}, \mathbf{u}_k \rangle}{\langle \mathbf{u}_k, \mathbf{u}_k \rangle}\mathbf{u}_k$$

is $\mathbf{v}$'s orthogonal projection into $\mathcal{L}_k$

Gram-Schmidt Orthonormalization

Every inner product space has an orthonormal basis

Given: orthonormal vectors $\mathbf{b}_1, \ldots, \mathbf{b}_r$ that form basis of subspace $\mathcal{S}_r$ of $\mathcal{L}_n$ where $n > r$

Find: $\mathbf{b}_{r+1}$ orthogonal to the given $\mathbf{b}_i$

Let $\mathbf{u}$ be an arbitrary vector in $\mathcal{L}_n$, but not in $\mathcal{S}_r$

$\mathbf{u}$'s *orthogonal projection* into $\mathcal{S}_r$:

$$\hat{\mathbf{u}} = \mathrm{proj}_{\mathcal{S}_r}\mathbf{u} = \langle \mathbf{u}, \mathbf{b}_1 \rangle \mathbf{b}_1 + \ldots + \langle \mathbf{u}, \mathbf{b}_r \rangle \mathbf{b}_r$$

Gram-Schmidt Orthonormalization

Check orthogonality: for example $\langle \mathbf{u} - \hat{\mathbf{u}}, \mathbf{b}_1 \rangle = 0$

$$\langle \mathbf{u} - \hat{\mathbf{u}}, \mathbf{b}_1 \rangle = \langle \mathbf{u}, \mathbf{b}_1 \rangle - \langle \mathbf{u}, \mathbf{b}_1 \rangle \langle \mathbf{b}_1, \mathbf{b}_1 \rangle - \ldots - \langle \mathbf{u}, \mathbf{b}_r \rangle \langle \mathbf{b}_1, \mathbf{b}_r \rangle$$

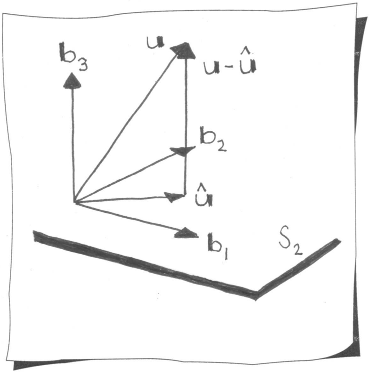
$$\Rightarrow \quad \mathbf{b}_{r+1} = \frac{\mathbf{u} - \mathrm{proj}_{\mathcal{S}_r} \mathbf{u}}{\| \cdot \|}$$

Repeat to form an orthonormal basis for all of $\mathcal{L}_n$

Key tools: projections and vector decomposition

Gram-Schmidt Orthonormalization

Build the orthonormal basis:



$\mathcal{S}_2$ is depicted as $\mathbb{R}^2$

Given: basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathcal{L}_n$

$$\mathbf{b}_1 = \frac{\mathbf{v}_1}{\|\cdot\|}$$

$$\mathbf{b}_2 = \frac{\mathbf{v}_2 - \mathrm{proj}_{\mathcal{S}_1}\mathbf{v}_2}{\|\cdot\|} = \frac{\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{b}_1\rangle \mathbf{b}_1}{\|\cdot\|}$$

$$\mathbf{b}_3 = \frac{\mathbf{v}_3 - \mathrm{proj}_{\mathcal{S}_2}\mathbf{v}_3}{\|\cdot\|}$$

$$= \frac{\mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{b}_1\rangle \mathbf{b}_1 - \langle \mathbf{v}_3, \mathbf{b}_2\rangle \mathbf{b}_2}{\|\cdot\|}$$

$$\vdots$$

Gram-Schmidt Orthonormalization

Example: $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \quad \mathbf{v}_4 = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}$

Form an orthonormal basis $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4$

$$\mathbf{b}_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \quad \mathbf{b}_2 = \begin{bmatrix}0\\1/\sqrt{3}\\1/\sqrt{3}\\1/\sqrt{3}\end{bmatrix} \quad \mathbf{b}_3 = \begin{bmatrix}0\\2/\sqrt{6}\\-1/\sqrt{6}\\-1/\sqrt{6}\end{bmatrix}$$

$$\mathbf{b}_4 = \frac{\mathbf{v}_4 - \langle \mathbf{v}_4, \mathbf{b}_1 \rangle \mathbf{b}_1 - \langle \mathbf{v}_4, \mathbf{b}_2 \rangle \mathbf{b}_2 - \langle \mathbf{v}_4, \mathbf{b}_3 \rangle \mathbf{b}_3}{\| \cdot \|} = \begin{bmatrix}0\\0\\1/\sqrt{2}\\-1/\sqrt{2}\end{bmatrix}$$

Check: $\begin{vmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4\end{vmatrix} = 1$

Gram-Schmidt Orthonormalization

Sometimes an *orthogonal set* of vectors is more desirable than an *orthonormal set*

We might want to avoid the extra computational cost of normalization

Given: basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$

Find: orthogonal basis $\mathbf{b}_i$

Solution: set $\mathbf{b}_1 = \mathbf{v}_1$ then

$$\mathbf{b}_k = \mathbf{v}_k - \frac{\langle \mathbf{v}_k, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle}\mathbf{b}_1 - \ldots - \frac{\langle \mathbf{v}_k, \mathbf{b}_{k-1} \rangle}{\langle \mathbf{b}_{k-1}, \mathbf{b}_{k-1} \rangle}\mathbf{b}_{k-1} \qquad k = 2, \ldots, n$$

QR Decomposition

Matrix computation is fundamental to linear algebra

Apply this concept again to the Gram-Schmidt method

The QR decomposition will emerge

Immediate benefit is a new perspective on methods for solving least squares approximation

QR Decomposition

Given: n linearly independent vectors $\mathbf{a}_i$ in $\mathbb{R}^n$ (stored in A)

Find: n orthonormal vectors $\mathbf{q}_i$ in $\mathbb{R}^n$ (stored in Q)

Develop method with a example from the Gram-Schmidt section

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Step through the Gram-Schmidt process using a matrix representation

Map $\mathbf{a}_1$ to a unit vector $\mathbf{q}_1$

$$A_1 = AR_1 = [\mathbf{q}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_4] \qquad \text{where } R_1 = I$$

QR Decomposition

Next map $\mathbf{a}_2 \rightarrow \mathbf{q}_2$

$$\mathbf{q}_2 = \frac{1}{\sqrt{3}}\mathbf{a}_2 - \frac{1}{\sqrt{3}}\mathbf{q}_1$$

represented as an elementary matrix

$$R_2 = \begin{bmatrix} 1 & -\frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{then} \quad A_2 = AR_1R_2 = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{a}_3 \quad \mathbf{a}_4]$$

By right-multiplying by R_2
the elementary matrix is acting on the second column of A only

QR Decomposition

Continue ... (see text for details) then final step:

$$A_4 = AR_1R_2R_3R_4 = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3 \quad \mathbf{q}_4] = Q$$

Let $R^{-1} = R_1R_2R_3R_4$ then

$$A = QR \quad \text{is the QR decomposition of } A$$

To complete the example:

$$R = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & \sqrt{3} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & 0 & \sqrt{\frac{3}{2} - \frac{1}{\sqrt{6}}} & -\frac{1}{\sqrt{6}} \\ 0 & 0 & 0 & -\frac{1}{3\sqrt{2}} + \frac{2\sqrt{2}}{3} \end{bmatrix}$$

and the columns of Q are given on slide 28 (called $\mathbf{b}_i$)

QR Decomposition

Upper triangular matrix R describes the transformation of $\mathbf{q}_i \rightarrow \mathbf{a}_i$

$$R = \begin{bmatrix} \mathbf{q}_1^{\mathrm{T}} \cdot \mathbf{a}_1 & \mathbf{q}_1^{\mathrm{T}} \cdot \mathbf{a}_2 & \mathbf{q}_1^{\mathrm{T}} \cdot \mathbf{a}_3 & \mathbf{q}_1^{\mathrm{T}} \cdot \mathbf{a}_4 \\ 0 & \mathbf{q}_2^{\mathrm{T}} \cdot \mathbf{a}_2 & \mathbf{q}_2^{\mathrm{T}} \cdot \mathbf{a}_3 & \mathbf{q}_2^{\mathrm{T}} \cdot \mathbf{a}_4 \\ 0 & 0 & \mathbf{q}_3^{\mathrm{T}} \cdot \mathbf{a}_3 & \mathbf{q}_3^{\mathrm{T}} \cdot \mathbf{a}_4 \\ 0 & 0 & 0 & \mathbf{q}_4^{\mathrm{T}} \cdot \mathbf{a}_4 \end{bmatrix}$$

QR Decomposition and Least Squares

Revisit finding the best fit line to seven time and temperature data pairs

— Overdetermined linear system

— Find best approximation with respect to the least squares error

— Normal equations formed with QR decomposition of A

$$\begin{aligned}
(QR)^{\mathrm{T}}(QR)\mathbf{u} &= (QR)^{\mathrm{T}}\mathbf{b} \\
R^{\mathrm{T}}Q^{\mathrm{T}}QR\mathbf{u} &= R^{\mathrm{T}}Q^{\mathrm{T}}\mathbf{b} \\
R^{\mathrm{T}}R\mathbf{u} &= R^{\mathrm{T}}Q^{\mathrm{T}}\mathbf{b} \\
R\mathbf{u} &= Q^{\mathrm{T}}\mathbf{b}.
\end{aligned}$$

QR decomposition provides a new approach to the normal equations

QR Decomposition

Householder method is numerically more stable than the possibly ill-conditioned normal equations

Transforms the linear system via orthogonal reflection matrices H_i

$$H_{n-1} \ldots H_1 A\mathbf{u} = H_{n-1} \ldots H_1 \mathbf{b}$$

Let $Q^{\mathrm{T}} = H_{n-1} \ldots H_1$ then

$$R\mathbf{u} = Q^{\mathrm{T}}\mathbf{b}$$

Householder can be used to construct the QR decomposition instead of the Gram-Schmidt method

— Probably the better choice due to potential rounding error problems in Gram-Schmidt

Chapter 15: another application of the QR decomposition – the QR algorithm for finding eigenvalues

Let's highlight some special linear spaces—but there are many more!

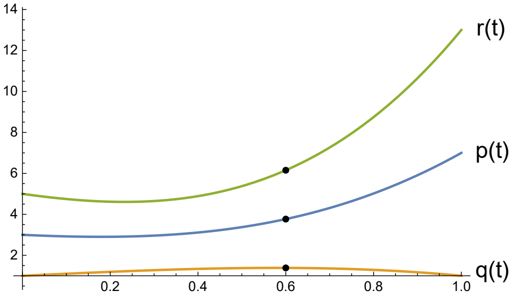
— Polynomials, continuous functions, matrices, linear maps

Polynomials: linear space $\mathcal{P}_n$ whose elements are all polynomials of degree $\leq n$

$$p(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n$$

Addition: coefficient by coefficient

Multiplication: polynomial times a real number



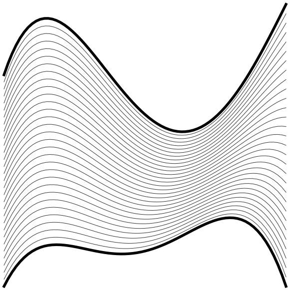
Check linearity property with an example:

$$p(t) = 3 - 2t + 3t^2 \qquad q(t) = -1 + t + 2t^2$$

$$2p(t) + 3q(t) = 3 - t + 12t^2$$

Yet another polynomial of the same degree

A Gallery of Spaces



Application of the linear space properties of $\mathcal{P}_n$ in *shape design*

Feature curves (polynomials) designed over a common domain interval

Shape formed from convex combinations of the feature curves

Idea can be used for 3D surface design using the techniques in Chapter 20

A Gallery of Spaces

Linear map: derivative p' of a degree n polynomial p

$$p'(t) = a_1 + 2a_2 t + \ldots + n a_n t^{n-1}$$

Rank of this map is $n - 1$

Example: Two cubic polynomials

$$p(t) = 3 - t + 2t^2 + 3t^3 \quad \text{and} \quad q(t) = 1 + t - t^3$$

in the linear space of cubic polynomials $\mathcal{P}_3$

$$\text{Let} \quad r(t) = 2p(t) - q(t) = 5 - 3t + 4t^2 + 7t^3$$

(See Figure on slide 40)

$$r'(t) = -3 + 8t + 21t^2$$
$$p'(t) = -1 + 4t + 9t^2$$
$$q'(t) = 1 - 3t^2$$

Linearity of the derivative map $\quad \Rightarrow \quad r'(t) = 2p'(t) - q'(t)$

A Gallery of Spaces

The usual inner product for $\mathcal{P}_n$

$$\langle p(t), q(t) \rangle = \int_a^b p(t)q(t)dt$$

Example: For $t \in [-1, 1]$

$$p_1(t) = 1 \qquad p_2(t) = t \qquad p_3(t) = t^2$$

Calculate the inner products:

$$\langle 1, t \rangle = \int_{-1}^{1} (1 \times t)dt = \frac{1}{2}t^2 \Big|_{-1}^{1} = 0$$

$$\langle 1, t^2 \rangle = \int_{-1}^{1} (1 \times t^2)dt = \frac{1}{3}t^3 \Big|_{-1}^{1} = \frac{2}{3}$$

$$\langle t, t^2 \rangle = \int_{-1}^{1} (t \times t^2)dt = \frac{1}{4}t^4 \Big|_{-1}^{1} = 0$$

(These polynomials are not an orthogonal)

Inner product spaces offer the concept of length

$$||p(t)|| = \langle p(t), p(t) \rangle = \sqrt{\int_a^b p(t)^2 dt}$$

Example: For $t \in [-1, 1]$

$$||p_1(t)|| = \sqrt{\int_{-1}^{1} 1 dt} = \sqrt{2}$$

Build an orthogonal set of polynomials with the Gram-Schmidt method

Example:

For $t \in [-1, 1]$ transform $p_i(t) = \{1, t, t^2\}$

to an orthogonal set of polynomials $\{q_1(t), q_2(t), q_3(t)\}$

— use the inner product definition from previous slide

$$q_1 = p_1 = 1$$

$$q_2 = t - \frac{\langle t, 1\rangle}{\langle 1, 1\rangle}1 = t$$

$$q_3 = t^2 - \frac{\langle t^2, t\rangle}{\langle t, t\rangle}t - \frac{\langle t^2, 1\rangle}{\langle 1, 1\rangle}1 = t^2 - \frac{1}{3}$$

The q_i are the quadratic Legendre polynomials

Orthogonal polynomials provide for more computationally efficient and better conditioned solutions to least squares approximation

Continuous functions:

A linear space given by the set of all real-valued continuous functions over the interval $[0, 1]$
— This space is typically named $C[0, 1]$
— The linearity condition is met:
If f and g are elements of $C[0, 1]$ then $\alpha f + \beta g$ is also in $C[0, 1]$
— This is an *infinite-dimensional* linear space
No finite set of functions forms a basis for $C[0, 1]$

Matrices:

The set of all 3×3 matrices form a linear space
— This space consists of matrices
— Linear combinations formed using standard matrix addition and multiplication with a scalar

A Gallery of Spaces

Linear Maps: (A more abstract example)

The linear space formed from
the set of all linear maps from a linear space $\mathcal{L}_n$ into the reals
— Called the dual space $\mathcal{L}_n^*$ of $\mathcal{L}_n$
— Its dimension equals that of $\mathcal{L}_n$
— The linear maps in $\mathcal{L}_n^*$ are known as linear functionals

Let a fixed vector $\mathbf{v}$ and an variable vector $\mathbf{u}$ be in $\mathcal{L}_n$

The linear functionals defined by $\Phi_{\mathbf{v}}(\mathbf{u}) = \langle \mathbf{u}, \mathbf{v} \rangle$ are in $\mathcal{L}_n^*$

For any basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ of $\mathcal{L}_n$ define linear functionals

$$\Phi_{\mathbf{b}_i}(\mathbf{u}) = \langle \mathbf{u}, \mathbf{b}_i \rangle \quad \text{for } i = 1, \ldots, n$$

These functionals form a basis for $\mathcal{L}_n^*$

Example: In $\mathbb{R}^2$ consider the fixed vector

$$\mathbf{v} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \text{then} \quad \Phi_{\mathbf{v}}(\mathbf{u}) = \langle \mathbf{u}, \mathbf{v} \rangle = u_1 - 2u_2$$

for all vectors $\mathbf{u}$ where $\langle \cdot, \cdot \rangle$ is the dot product

Example: Pick $\mathbf{e}_1, \mathbf{e}_2$ for a basis in $\mathbb{R}^2$

The associated linear functionals are

$$\Phi_{\mathbf{e}_1}(\mathbf{u}) = u_1 \qquad \Phi_{\mathbf{e}_2}(\mathbf{u}) = u_2$$

Any linear functional Φ can now be defined as

$$\Phi(\mathbf{u}) = r_1 \Phi_{\mathbf{e}_1}(\mathbf{u}) + r_2 \Phi_{\mathbf{e}_2}(\mathbf{u})$$

where r_1 and r_2 are scalars

Find the best approximation to the function $f(x)$ by another function $g(x)$ in a particular linear space of continuous functions over a fixed interval $[a, b]$

Example: given a cubic polynomial, find the best linear polynomial fit

Least Squares

Need a quantitative definition of "best"

Measure the difference between two functions f and g over the fixed interval

$$E = \int_a^b |f(x) - g(x)| dx$$

Easier:

$$E^2 = \int_a^b (f(x) - g(x))^2 dx$$

Approximation space: let's use the orthogonal trigonometric polynomials
— Well known in the context of the *Fourier series* of a function $f(x)$

$$f(x) = a_0 + a_1 \cos(x) + b_1 \sin x + a_2 \cos(2x) + b_2 \sin(2x) + \ldots$$

The a_i and b_i for $i = 1, \ldots, n \to \infty$ are called the *Fourier coefficients*
— For the approximation problem, choose a finite n

Let's choose $n = 2$ and the interval $[0, 2\pi]$ then
the least squares approximation to $f(x)$
is the orthogonal projection of f into space of trigonometric polynomials of degree less than or equal to 2

Least Squares

Compute the unknown coefficients a_0, a_1, b_1, a_2, b_2

Details for a_1:

$$\int_0^{2\pi} f(x)\cos(x)dx = a_0 \int_0^{2\pi} \cos(x)dx$$
$$+ a_1 \int_0^{2\pi} \cos^2(x)dx + b_1 \int_0^{2\pi} \sin(x)\cos(x)dx$$
$$+ a_2 \int_0^{2\pi} \cos^2(x)dx + b_2 \int_0^{2\pi} \sin(x)\cos(x)dx$$

Cancellation due to the interval and orthogonality of the basis functions

$$a_1 = \frac{\int_0^{2\pi} f(x)\cos(x)dx}{\int_0^{2\pi} \cos^2(x)dx} = \frac{\langle f, \cos(x)\rangle}{\langle \cos(x), \cos(x)\rangle}$$

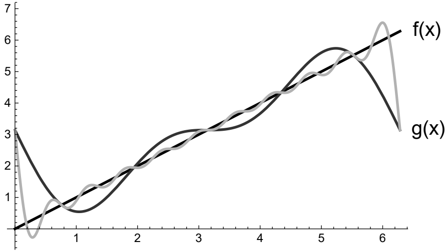
Example:

Given: $f(x) = x$ over $[0, 2\pi]$

Find: the least square approximation in the space of trigonometric polynomials of degree $n \leq 2$

$$g(x) = a_0 + a_1 \cos(x) + b_1 \sin(x) + a_2 \cos(2x) + b_2 \sin(2x)$$

$$\text{Solution:} \quad g(x) = \pi - 2\sin x - \sin 2x$$



Also illustrated in gray color: degree 10 solution

WYSK

- linear space
- vector space
- dimension
- linear combination
- linearity property
- linearly independent
- subspace
- span
- linear map
- image
- preimage
- domain
- range
- rank
- full rank
- rank deficient
- inverse
- determinant
- inner product
- inner product space
- distance in an inner product space
- length in an inner product space
- orthogonal
- Gram-Schmidt method
- projection
- basis
- orthonormal
- orthogonal decomposition
- best approximation
- dual space
- linear functional
- QR decomposition