MATH 20700 (Honors Analysis in $\mathbb{R}^n$ I) Notes

Steven Labalme

October 7, 2021

Contents

List of Figures

Part I

Linear Algebra

Chapter 1

Basic Notions

9/27:
- **Vector space**: Basically, a set for which you have an addition and multiplication.
- $\mathbb{F}^d$ is used for $\mathbb{R}^d$ or $\mathbb{C}^d$ in Treil (2017).
- $\mathbb{P}_n$ is the vector space of polynomials up to degree n.
- $C([0,1])$ is the set of continuous functions defined on $[0,1]$, an infinite-dimensional vector space.
- **Generating set**: A subset of a vector space, all linear combinations of which generate the vector space. *Also known as* **spanning set**.
  - Any element of VS is a linear comb. of elements of the generating set.
- **Linearly independent** (list): A list of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ such that $\sum_{i=1}^k \alpha_i \mathbf{v}_i = 0$ implies $\alpha_i = 0$ for all i.
- **Base**: A generating set consisting of linearly independent vectors.
- Any element of a VS can be written as a *unique* linear combination of the vectors in a base.
  - If $\mathbf{x} = \sum_{i=1}^k \alpha_i \mathbf{v}_i = \sum_{i=1}^k \beta_i \mathbf{v}_i$, then $\alpha_i = \beta_i$ for all i.
- **Linear transformation**: A function $T : X \to Y$, where X, Y are VSs, such that

  $$T(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha T\mathbf{x} + \beta T\mathbf{y}$$

  for all $\mathbf{x} \in X$, $\mathbf{y} \in Y$.
- Examples of linear transformations:
  - Consider $\mathbb{P}_n$. Let $Tp_n = p_n'$. This T is linear.
  - Rotation in $\mathbb{R}^d$.
    - Think graphically about two vectors $\mathbf{x}, \mathbf{y}$.
    - Rotating and summing them is the same as summing and rotating. Same for scaling.
    - Thus, rotation is actually linear!
  - Reflection as well.
- Consider $T : \mathbb{R} \to \mathbb{R}$.
  - Any linear map on the line is a line.
  - We must have $Tx = \alpha x$: $Tx = T(1x) = xT(1) = x\alpha$.
- Consider $T : \mathbb{R}^n \to \mathbb{R}^m$ linear.

- Any linear map between $\mathbb{R}^n$ and $\mathbb{R}^m$ is linear.
- Thus, $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, where A is an $m \times n$ matrix.

- To find A, do the same calculation as for $Tx = \alpha x$ but more carefully:
  - Let $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis.
  - So $\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i$.
  - Thus, $T\mathbf{x} = \sum_{i=1}^{n} \alpha_i T(\mathbf{e}_i)$.
  - Each $T(\mathbf{e}_i)$ is part of the matrix that we multiply by the column vector representing $\mathbf{x}$.
- Multiplication of matrices is equivalent to composition of linear maps.
- Consider $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $T_2 : \mathbb{R}^m \to \mathbb{F}^r$.
  - $T_2 \circ T_1$ is equivalent to BA, if A represents T_1 and B represents T_2. In other words, $(T_2 \circ T_1)(\mathbf{x}) = BA\mathbf{x}$ for all $\mathbf{x}$.
- Recall that if $A = (\alpha_{ij})$ and $B = (\beta_{ij})$, then $(BA)_{ij} = (\sum \beta_{ik}\alpha_{kj})$.
- Properties of multiplication:

$$
\begin{aligned}
(AB)C &= A(BC) \\
A(B + C) &= AB + AC \\
(A + B)C &= AC + BC
\end{aligned}
$$

  - However, it is not true in general that $AB = BA$.
- **Trace** (of an $n \times n$ matrix A): The sum of the diagonal entries of A. *Denoted by* $\textbf{tr}\,\boldsymbol{(A)}$. *Given by*

$$\operatorname{tr}(A) = \sum \alpha_{ii}$$

- It is true that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.
  - Indeed, on the diagonals, multiplication is commutative; it's the other terms that mess you up in general.
- Invertibility of matrices.
  - In general, matrices are not invertible: Not every system of equations is solveable; $Ax = b$ does not always have a solution $x = A^{-1}b$.
- C is the inverse from the left: $CA = I$. B is the inverse from the right: $AB = I$. A matrix can have a left and a right inverse and still not be invertible. A matrix is invertible iff $C = B$.
- Any time we write "inverse," we do so under the assumption that it exists.
- $(AB)^{-1} = B^{-1}A^{-1}$ — easy proof by multiplication.
- If $A = (a_{ij})$, $A^T = (a_{ji})$.
  - $(A^{-1})^T = (A^T)^{-1}$.
  - $(AB)^T = B^T A^T$.
- Let X, Y VS.
  - $X \cong Y$[1] if there exists a linear $T : X \to Y$ that is one-to-one and onto.
  - Check: $A(\text{basis of } X) = \text{basis of } Y$. Prove by definition and expression of elements as linear combinations.
- **Subspace**: A subset of a vector space which happens to be a vector space, itself.

[1] "X is isomorphic to Y."

Chapter 2

Systems of Linear Equations

9/29:

- Row elimination:
  - Let
    $$A = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 3 & 1 & 2 & 7 \\ 2 & 1 & 2 & 1 \end{pmatrix}$$
  - Then the **eschelon form** matrix
    $$A_e = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 2 & -4 \end{pmatrix}$$
  - Lastly, the **reduced eschelon form** matrix
    $$A_{re} = \begin{pmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{pmatrix}$$
- **Eschelon form**:
  - All zero rows are below nonzero rows.
  - For any nonzero row, its leading element is strictly to the left of the nonzero entry of the next row.
- **Reduced eschelon form**:
  - All pivots are 1.
  - Used to solve systems of the form $Ax = b$.
- **Inconsistent** (system of equations): A system with no solution.
  - If the last row is of the form $(0, \ldots, 0, b)$ where $b \neq 0$, then there is no solution.
- Unique solution if A_e has a pivot in every column.
- There exists a solution for every b if there is a pivot in every row?
- Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a matrix. Then $\ker A = \{x \in \mathbb{R}^n : Ax = 0\}$ (subspace of $\mathbb{R}^n$) and $\operatorname{range} A = \{Ax : x \in \mathbb{R}^n\}$ (subspace of $\mathbb{R}^m$).
- Also consider $\ker(A^T)$ and $\operatorname{range}(A^T)$, the basis of the kernel and range, and dimension.
- Finite-dimensional vector spaces:

  - A basis is a generating set (so every element of V can be written uniquely as a linear combination of the basis) the length of which is equal to the dimension of V.
  - All bases of finite-dimensional vector spaces have the same number of elements.
    - Let v_1, v_2, v_3 and w_1, w_2 be two generating sets of V.
    - Then

$$\begin{aligned} v_1 &= \lambda_{11} w_1 + \lambda_{12} w_2 \\ v_2 &= \lambda_{21} w_1 + \lambda_{22} w_2 \\ v_3 &= \lambda_{31} w_1 + \lambda_{32} w_2 \end{aligned}$$

    - Suppose the only solution to $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = 0$ is $\alpha_1 = \alpha_2 = \alpha_3 = 0$.
    - But this is not true, as we can find another one in terms of the λs.
  - If you have a list of linearly independent vectors, you can complete it into a basis.
    - If there exists a vector that can't be written as a linear combination of the list, add it to the list.
  - If you find any particular solution to a system $Ax = b$, and you add to it any element of $\ker A$, you will obtain another solution.
    - $Ax_1 = b$ and $Ax_h = 0$ implies that $A(x_1 + x_h) = b$.
    - $Ax_1 = b$ and $Ax_2 = b$ imply that $A(x_1 - x_2) = 0$, i.e., that $x_1 - x_2 \in \ker A$.
  - If $A : \mathbb{R}^n \to \mathbb{R}^m$ and $\dim \operatorname{range} A = m$, then $Ax = b$ is solveable for all $b \in \mathbb{R}^m$.
  - Let $\operatorname{rank} A = \dim \operatorname{range} A$.
  - Rank theorem:
    - $\operatorname{rank} A = \operatorname{rank} A^T$.
    - Let $A : \mathbb{R}^n \to \mathbb{R}^m$. We know that $\dim \ker A + \dim \operatorname{range} A = n$.
    - $\dim \ker A^T + \operatorname{rank} A^T = m$.
    - This theorem survives linear algebra and enters functional analysis under the name **Fredholm's alternative**.

- **Fredholm's alternative**: $Ax = b$ has a solution for all $b \in \mathbb{R}^n$ iff $\dim \ker A^T = 0$.
  - $\dim \ker A^T = 0$ implies $\operatorname{rank} A^T = m$ implies $\operatorname{rank} A = m$ implies $\dim \operatorname{range} A = m$, as desired.
- **Pivot column** (of A): A column of A where A_e has pivots.
- The **pivot columns** of A give a basis for $\operatorname{range} A$.
- The pivot rows of A_e give a basis for $\operatorname{range} A^T$.
- A basis for the kernel is enough to solve $Ax = 0$.
- If you take these three things as givens, you can prove the rank theorem.

Chapter 3

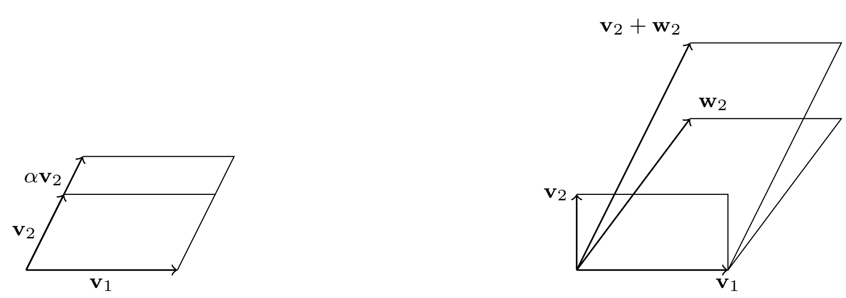
Determinants

9/29:

- The determinant, geometrically, is the volume of the object (in $\mathbb{R}^3$) you get when you take linear combinations of the vectors.
- In 2D:
  - Let v_1, v_2 be two vectors. Put tail to tail and forming a parallelogram, the determinant of the matrix (v_1, v_2) is the area of said parallelogram.
  - Linearity 1: $D(av_1, v_2, \dots, v_n) = aD(v_1, \dots, v_n)$ is the same as saying that if you stretch one vector by a, you scale up the area by that much, too.
  - Linearity 2: $D(v_1, \dots, v_{k+} + v_{k-}, \dots, v_n) = D(-) + D(+)$.
  - Antisymmetry: $D(v_1, \dots, v_k, \dots, v_j, \dots, v_n) = -D(v_1, \dots, v_j, \dots, v_k, \dots, v_n)$. Interchanging columns flips the sign of the determinant.
  - Basis: $D(e_1, \dots, e_n) = 1$.
- Determinant: Denoted by $D(v_1, \dots, v_n)$, where $(v_1, \dots, v_n)$ is an $n \times n$ matrix.

10/1:

- Consider an $n \times n$ matrix A consisting of n columns containing vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$.
  - $D(A)$ is the volume of the solid $V = \sum_{i=1}^{n} \alpha_i v_i$.
  - $D(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1$.



(a) $D(\mathbf{v}_1, \alpha\mathbf{v}_2) = \alpha D(\mathbf{v}_1, \mathbf{v}_2)$.

(b) $D(\mathbf{v}_1, \mathbf{v}_2 + \mathbf{w}_2) = D(\mathbf{v}_1, \mathbf{v}_2) + D(\mathbf{v}_1, \mathbf{w}_2)$.

Figure 3.1: Visualizing properties of determinants.

- Basic properties of the determinant.
  - If A has a zero column, then $\det A = 0$: Scalar property.
  - If A has two equal columns, then $\det A = 0$: Multiply one by minus and add.

  - If A has a column which is a multiple of another, then $\det A = 0$: Pull out the multiple and then you have the previous one.
  - If columns are linearly dependent, then $\det A = 0$: Decompose it into sums, split, add back up with previous properties.
  - The determinant is preserved under column reduction.
  - $\det A^T = \det A$: Put everything in rref.
  - If A is not invertible, then $\det A = 0$ (not invertible implies linearly dependent columns, implies $\det A = 0$).
  - $\det(AB) = \det A \det B$.

- Determinant of...
  - A diagonal matrix: The product of the diagonal entries (pull out the terms, and then note that the remaining identity matrix has determinant 1).
  - An upper triangular matrix: The product of the diagonal entries (column reduction to make it into a diagonal matrix, and then the property above).

Chapter 4

Introduction to Spectral Theory

10/1:

- **Difference equation**: Like a differential equation, but instead of writing a differentials, you write differences.
- Suppose we want to solve $x_{n+1} = Ax_n$ with x_0 given.
  - You will find that $x_n = A^n x_0$.
  - This gets hard to compute, so we want to find a way to simplify the computation.
- Thus, we want to diagonalize the matrix, and this concept is inherently linked to eigenvalues and eigenvectors.
  - If you can decompose the x_0 into a linear combination of eigenvectors, then you can simplify the computation a lot:
    $$x_n = \sum \alpha_i A^n v_i = \sum \alpha_i \lambda_i^n v_i$$
  - An $n \times n$ matrix will have n eigenvalues. You want n linearly independent eigenvectors, creating an eigenbasis.
- To find eigenvalues and eigenvectors, we need to solve $Ax = \lambda x$, i.e., $(A - \lambda I)x = 0$. Thus, $\ker(A - \lambda I) \neq \{0\}$, so $\det(A - \lambda I) = 0$.
- The eigenvalues of A are independent of the choice of basis of the domain of A or the range.

10/4:

- We need to know everything in Treil (2017).
  - We don't need to know the applications sections, but you should be interested.
- **Spectral theory**: Decomposing a linear operator.
- Let $A : V \to V$ be a linear operator. $\lambda \in \mathbb{C}$ is an eigenvalue if there exists $x \in V$ nonzero such that $Ax = \lambda x$.
  - Let A be an $n \times n$ matrix over $\mathbb{C}$ or $\mathbb{R}$.
  - The eigenvalues are the roots of the polynomial $\det(A - \lambda I) = 0$ in λ.
- Things we want to do:
  - Given A, find the eigenvalues and eigenvectors (solve $(A - \lambda I)x = 0$).
  - In order to simplify A, make it a diagonal matrix:
    $$A = S \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} S^{-1}$$

- Eigenvalues are independent of the choice of basis.
  - From the book, we have that
    $$[A]_{\mathcal{A}\mathcal{A}} = [S]_{\mathcal{A}\mathcal{B}}[B]_{\mathcal{B}\mathcal{B}}[S]_{\mathcal{A}\mathcal{B}}^{-1}$$
  - It follows that
    $$A - \lambda I = [S]_{\mathcal{A}\mathcal{B}}(B - \lambda I)[S]_{\mathcal{A}\mathcal{B}}^{-1}$$
    so
    $$\det(A - \lambda I) = \det\big([S]_{\mathcal{A}\mathcal{B}}(B - \lambda I)[S]_{\mathcal{A}\mathcal{B}}^{-1}\big) = \det\big([S]_{\mathcal{A}\mathcal{B}}[S]_{\mathcal{A}\mathcal{B}}^{-1}(B - \lambda I)\big) = \det(B - \lambda I)$$
- If $p(z) = (z - \lambda)^k q(z)$, then k is the **algebraic multiplicity** of λ. The **geometric multiplicity** of λ is $\dim\ker(A - \lambda I)$.
  - These terms are not always the same, but they are related.
- Diagonalization:
  - Given A that corresponds to $T : V \to V$, can we find a basis of V in which the operator is a diagonal matrix?
  - $A = SDS^{-1}$ iff there exists a basis of V consisting of the eigenvectors of A.
  - Proves $A^N = SD^N S^{-1}$ via $A^2 = SDS^{-1}SDS^{-1} = SDIDS^{-1} = SD^2S^{-1}$.
- Let A be an $n \times n$ matrix over $\mathbb{F}$. If $\lambda_1, \dots, \lambda_r$ are distinct eigenvalues, then their eigenvectors are linearly independent.
  - Prove with induction contradiction argument. Assume true for $\mathbf{v}_{r-1}$. Then
    $$0 = (A - \lambda_r I)[\mathbf{v}_1 + \cdots + \mathbf{v}_r] = (\lambda_1 - \lambda_r)\mathbf{v}_1 + \cdots + (\lambda_{r-1} - \lambda_r)\mathbf{v}_{r-1}$$
  - Implies $\lambda_r = \lambda_i$ for all $i \in [r-1]$, a contradiction.
  - If A has n distinct eigenvalues, then A is diagonalizable.
- If $A : V \to V$ has n complex eigenvalues, then A is diagonalizable iff the algebraic multiplicity equals the geometric multiplicity for each eigenvalue.
- Goes through a sample diagonalization with $\left(\begin{smallmatrix} 1 & 2 \\ 8 & 1 \end{smallmatrix}\right)$.
  - We have
    $$A - \lambda I = \begin{pmatrix} 1 - \lambda & 2 \\ 8 & 1 - \lambda \end{pmatrix}$$
    so
    $$0 = \det(A - \lambda I) = (1 - \lambda)^2 - 16$$
  - It follows that $\lambda = 5, -3$.
  - This yields
    $$\begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix}^{-1}$$
    by inspection.
- As another example, consider $\left(\begin{smallmatrix} 1 & 2 \\ -2 & 1 \end{smallmatrix}\right)$.
  - Here, we have $\lambda = 1 \pm 2i$.

Chapter 5

Inner Product Spaces

10/6:

- We define

$$\ell^2(\mathbb{R}) = \left\{ \{a_n\}_{n\geq 1} \subset \mathbb{R} : \sum_1^\infty |a_n|^2 < \infty \right\}$$

- **Inner product**: A map $V \times V \to \mathbb{F}$ that takes $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} \cdot \mathbf{y}$. *Denoted by* $\cdot, (\cdot, \cdot), \langle \cdot, \cdot \rangle$.
- Properties of the inner product:
  - $(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{y}, \mathbf{x})}$ (symmetry).
  - $(\alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z}) = \alpha(\mathbf{x}, \mathbf{z}) + \beta(\mathbf{y}, \mathbf{z})$ (linearity).
  - $(\mathbf{x}, \mathbf{x}) \geq 0$.
  - $(\mathbf{x}, \mathbf{x}) = 0$ iff $\mathbf{x} = 0$.
- If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i y_i$$

- If $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, then

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i \bar{y}_i$$

- If $f, g \in \mathbb{P}_n(t)$, then

$$(f, g) = \int_{-1}^1 f \bar{g} \, \mathrm{d}t$$

  - The conjugate of a polynomial is the polynomial with the conjugate of the coefficients of the original polynomial. Symbolically, if $f = \sum_{i=0}^n \alpha_i t^i$ is a polynomial, then $\bar{f} = \sum_{i=0}^n \bar{\alpha}_i t^i$.
- It is a fact that

$$\left| \sum^\infty a_n \bar{b}_n \right| \leq \|(a_n)_{n\geq 1}\| \|(b_n)_{n\geq 1}\|$$

- Suppose we want to define the inner product between two matrices.
  - A common one is

$$(A, B) = \mathrm{tr}(B^* A)$$

    where $B^* = \bar{B}^T = \overline{B^T}$ is the conjugate transpose.

- We define the norm as a function $V \to [0, \infty)$ given by

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$$

- Properties of the norm.
  - $\|\alpha \mathbf{x}\| = |\alpha|\|\mathbf{x}\|$.
  - $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.
  - $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$.
- In $\mathbb{R}^n$,

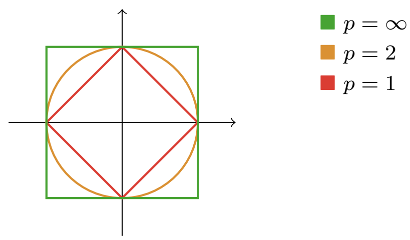


Figure 5.1: The unit ball of norms corresponding to $p = 1, 2, \infty$.

  - The standard norm is
$$\|\mathbf{x}\| = \sqrt{\sum |x_i|^2}$$
  - We can also define
$$\|\mathbf{x}\|_p = \sqrt[p]{\sum |x_i|^p}$$
  - We can even define
$$\|\mathbf{x}\|_\infty = \max |x_i|$$
  - And we can prove that all of these are valid norms.
  - Only the norm corresponding to ℓ^2 is given by an inner product, but all the other quantities are still norms as defined by the properties (see Treil (2017)).
  - Figure 5.1 shows the unit ball of each norm, i.e., the set of all points which have norm 1.
- The parallelogram rule:
$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)$$
- Orthogonality: Given $\mathbf{v}, \mathbf{w}$, if $\mathbf{v} \perp \mathbf{w}$, then $(\mathbf{v}, \mathbf{w}) = 0$.
- In particular, if $\mathbf{v} \perp \mathbf{w}$, then
$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$$
- Let E be a subspace of V. If $\mathbf{v} \perp E$, then $\mathbf{v} \perp \mathbf{e}$ for all $\mathbf{e} \in E$, i.e., $\mathbf{v} \perp$ a set of vectors spanning E.
- Any set of orthogonal vectors is linearly independent. Thus, if V is n dimensional, then $\mathbf{v}_1, \dots, \mathbf{v}_n$ orthogonal is a basis.
- Let E be a subspace of V. Take $\mathbf{v} \in V$. We want to define the projection $P_E\mathbf{v}$ of $\mathbf{v}$ onto E.
  - We have that $P_E\mathbf{v} \in E$ and $v - P_E\mathbf{v} \perp E$.
  - Additionally, we have that
$$\|\mathbf{v} - P_E\mathbf{v}\| \leq \|\mathbf{v} - \mathbf{e}\|$$
for all $\mathbf{e} \in E$.

  - Lastly, we have that $P_E\mathbf{v}$ is unique.

- If we receive a basis of a vector space, how do we create out of that a basis that is orthogonal? The process of doing this is called **Gram-Schmidt orthogonalization**.
  - We keep $\mathbf{v}_1$, subtract $P_{\mathbf{v}_1}\mathbf{v}_2$ from $\mathbf{v}_2$, subtract $P_{\{\mathbf{v}_1,\mathbf{v}_2\}}\mathbf{v}_3$ from $\mathbf{v}_3$, and on and on.
- If we are given a set of orthogonal vectors, we can normalize them by dividing each by its norm. This creates an orthonormal list. The standard basis is orthonormal.
- Let
$$E^\perp = \{v \in V : v \perp E\}$$
  - It follows that $V = E \oplus E^\perp$.
- How close can we come to solving $A\mathbf{x} = \mathbf{b}$ if we cannot solve it exactly (i.e., if the columns are not linearly independent)?
  - Let A be an $m \times n$ matrix, and let $\mathbf{b} \in \mathbb{R}^m$.
  - Then the best solution is given by minimizing $\|A\mathbf{x} - \mathbf{b}\|$. We minimize this with projections. A special case of this is least squares regression! More details in Treil (2017).

References

Treil, S. (2017). *Linear algebra done wrong* [http://www.math.brown.edu/streil/papers/LADW/LADW_2017-09-04.pdf].