# Module 9: Linear Algebra III

## Operational math bootcamp

Statistical Sciences
UNIVERSITY OF TORONTO

Ichiro Hashimoto

University of Toronto

July 23, 2025

# Outline

- Adjoints, unitaries and orthogonal matrices
- Orthogonal decomposition
- Spectral theory
  - Eigenvalues and eigenvectors
  - Algebraic and geometric multiplicity of eigenvalues
  - Matrix diagonalization

# Recall

## Definition

Let $V$ be an $\mathbb{F}$-vector space. A function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ is called *inner product* on $V$ if the following holds:

1. (Conjugate) symmetry: $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ for all $\mathbf{x}, \mathbf{y} \in V$, where $\bar{a}$ denotes the complex conjugate for $a \in \mathbb{C}$
2. Linearity in the first argument: $\langle \alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\alpha, \beta \in \mathbb{F}$
3. Positive definiteness: $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$

A vector space equipped with an inner product is called an *inner product space*.

# Recall

## Example

$= x^T y$

- Standard inner product on $\mathbb{R}^n$: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- Standard inner product on $\mathbb{C}^n$: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i \overline{y}_i$ for $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$

$= x^T \bar{y}$

- On the space of polynomials $\mathbb{P}_n(\mathbb{R})$: $\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_{-1}^{1} p(x)q(x)\mathrm{d}x$ for $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}_n(\mathbb{R})$

## Proposition

Let $V$ be an inner product space. Then

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle}$$

for all $\mathbf{x}, \mathbf{y} \in V$.

## Proposition

Let $V$ be an inner product space. Then $\langle \cdot, \cdot \rangle$ induces a norm on $V$ via $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ for all $\mathbf{x} \in V$.

*Proof.*

Note: With this identification the Cauchy-Schwarz inequality can be restated as: $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in V$.

**Example**

The norm introduced by the standard inner product on $\mathbb{R}^n$ is the Euclidean distance.

# Adjoint

**Definition**

Let $U, V$ be inner product spaces and $S\colon U \to V$ be a linear map. The *adjoint* $S^*$ of $S$ is the linear map $S^*\colon V \to U$ defined such that

$$\langle S\mathbf{u}, \mathbf{v}\rangle_V = \langle \mathbf{u}, S^*\mathbf{v}\rangle_U \qquad \text{for all } \mathbf{u} \in U, \mathbf{v} \in V.$$

Q Existence?

Uniqueness?

## Proposition

Let $U, V$ be inner product spaces and $S\colon U \to V$ be a linear map. Then $S^*$ is unique and linear.

*Proof.* Let $T_1$ and $T_2$ satisfies the definition of $S^*$.

For any $u \in U$, $v \in V$,

since $T_1$ and $T_2$ are both adjoint of $S$.

$$\langle u, T_1 v\rangle_U = \langle Su, v\rangle_V = \langle u, T_2 v\rangle_U$$

Then $\langle u, T_1 v - T_2 v\rangle_U = 0$ for $\forall u \in U$, $\forall v \in V$.

Let $u = T_1 v - T_2 v$, we have

$$\| T_1 v - T_2 v\|^2 = 0 \implies T_1 v = T_2 v. \quad \text{(uniqueness)}$$

Statistical Sciences
UNIVERSITY OF TORONTO

$$\langle u, \underline{\xi^*(\alpha v_1 + \beta v_2)} \rangle_U = \langle \xi u, \underline{\alpha v_1 + \beta v_2} \rangle_V$$

conjugate linearity

$$= \bar{\alpha} \langle \xi u, v_1 \rangle_V + \bar{\beta} \langle \xi u, v_2 \rangle_V$$

$$= \bar{\alpha} \langle u, \xi^* v_1 \rangle_U + \bar{\beta} \langle u, \xi^* v_2 \rangle_U$$

$$= \langle u, \alpha \xi^* v_1 \rangle_U + \langle u, \beta \xi^* v_2 \rangle_U$$

$$= \langle u, \underline{\alpha \xi^* v_1 + \beta \xi^* v_2} \rangle_U$$

By the uniqueness we've proved earlier, we have

$$\xi^*(\alpha v_1 + \alpha v_2) = \alpha \xi^* v_1 + \beta \xi^* v_2.$$

## Example

Define $S: \mathbb{R}^3 \to \mathbb{R}^2$ by $S\mathbf{x} = (2x_1 + x_3, -x_2)$. What is the adjoint operator $S^*$?

$$\langle Sx, y \rangle = \left\langle \begin{pmatrix} 2x_1 + x_3 \\ -x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle.$$

$$= (2x_1 + x_3)\, y_1 + (-x_2) \cdot y_2.$$

$$= x_1 \cdot 2y_1 + x_2 \cdot (-y_2) + x_3 \cdot y_1$$

$$= \left\langle \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \boxed{\begin{pmatrix} 2y_1 \\ -y_2 \\ y_1 \end{pmatrix}} \right\rangle$$

$$= S^* y$$

$$S^* = \begin{pmatrix} 2 & 0 \\ 0 & -1 \\ 1 & 0 \end{pmatrix}$$

## Proposition

Let $A \in M_{m\times n}(\mathbb{F})$ be a matrix and $T_A \colon \mathbb{F}^n \to F^m \colon \mathbf{x} \mapsto A\mathbf{x}$. Then, $T_A^*(\mathbf{x}) = A^*\mathbf{x}$, where $A^* \in M_{n\times m}(\mathbb{F})$ with $(A^*)_{ij} = \overline{A_{ji}}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

In particular, if $\mathbb{F} = \mathbb{R}$, the adjoint of the matrix is given by its transpose,denoted $A^T$, and if $\mathbb{F} = \mathbb{C}$, it is given by its conjugate transpose, denoted $A^*$. $= \bar{A}^T$

$\longrightarrow$ ensures the existence of the adjoint operator for linear maps between finite dimensional vector spaces.

*Proof* for $\mathbb{R}$:

$$\langle Ax, y \rangle = (Ax)^T y$$

$$= x^T (A^T y).$$

$$= \langle x, \underbrace{A^T y}_{= T_A^* y} \rangle$$

## Definition

A matrix $O \in M_n(\mathbb{R})$ is called *orthogonal* if its inverse is given by its transpose, i.e. $O^T O = OO^T = I$.

A matrix $U \in M_n(\mathbb{C})$ is called *unitary* if the inverse is given by the conjugate transpose, i.e. $U^* U = UU^* = I$.

Let $U$ be unitary

$$\langle Ux, Uy \rangle = \langle x, \underbrace{U^* U}_{= Id} y \rangle = \langle x, y \rangle$$

$U$ does not change inner product, norm, angle, etc.

## Example

- Let $\varphi \in [0, 2\pi]$. Then

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$$

Rotation

is an orthogonal matrix. What does it describe geometrically?

- The following is a unitary matrix:

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = A$$

$$A A^{*} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathrm{Id}.$$

## Definition

Let $A \in M_n(\mathbb{F})$. We call $A$ *self-adjoint* if $A^* = A$. In the case $\mathbb{F} = \mathbb{R}$, such an $A$ is called *symmetric* and if $\mathbb{F} = \mathbb{C}$, such an $A$ is called *Hermitian*.

# Orthogonality and Gram-Schmidt

**Definition**

Two vectors $\mathbf{x}, \mathbf{y} \in V$ are called *orthogonal* if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, denoted $\mathbf{x} \perp \mathbf{y}$. We call them *orthonormal* if additionally the vectors are normalized, i.e. $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$. A basis $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of $V$ is called *orthonormal basis (ONB)*, if the vectors are pairwise orthogonal and normalized.

In $\mathbb{R}^n$ or $\mathbb{C}^n$, $\quad e_i = (0, \cdots, \underset{i\text{th}}{1}, \cdots, 0)^T$ forms ONB

## Proposition

Let $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$ be orthonormal. Then the system of vectors is linearly independent.

*Proof.* Let $\sum_{i=1}^{k} \alpha_i x_i = 0$.

Then $0 = \left\langle \sum_{i=1}^{k} \alpha_i x_i, \sum_{i=1}^{k} \alpha_i x_i \right\rangle$

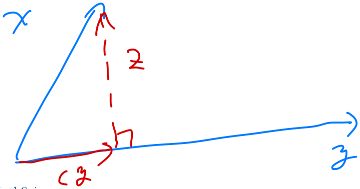
$$= \sum_{i=1}^{k} \underbrace{\alpha_i \overline{\alpha_i}}_{\|\alpha_i\|^2} \underbrace{\langle x_i, x_i \rangle}_{=1} + \sum_{i \neq j} \alpha_i \overline{\alpha_j} \underbrace{\langle x_i, x_j \rangle}_{=0}$$

$$= \sum_{i=1}^{k} \|\alpha_i\|^2.$$ Therefore, $\alpha_i = 0$ for all $i$.

### Proposition (Orthogonal Decomposition)

Let $\mathbf{x}, \mathbf{y} \in V$ with $\mathbf{y} \neq 0$. Then, there exist $c \in F$ and $\mathbf{z} \in V$ such that $\mathbf{x} = c\mathbf{y} + \mathbf{z}$ with $\mathbf{y} \perp \mathbf{z}$.

Given a basis, we can obtain an ONB from it using the Gram-Schmidt algorithm by repeating this orthogonal decomposition.




(pf) Let $c = \dfrac{\overline{\langle x, y \rangle}}{\|y\|^2}$.

Then,

$$\langle y, z \rangle = \langle y, x - cy \rangle$$
$$= \langle y, x \rangle - \bar{c} \langle y, y \rangle$$
$$= \langle y, x \rangle - \langle y, x \rangle = 0$$

## Proposition (Gram-Schmidt Algorithm)

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in V$ be a system of linearly independent vectors. Define $\mathbf{y}_1 = \mathbf{x}_1 / \|\mathbf{x}_1\|$. For $i = 2, \ldots, n$ define $\mathbf{y}_j$ inductively by

$$y_2 = \frac{x_2 - \langle x_1, y_1 \rangle y_1}{\| x_2 - \langle x_1, y_1 \rangle y_1 \|}$$

$$\mathbf{y}_i = \frac{\mathbf{x}_i - \sum_{k=1}^{i-1} \langle \mathbf{x}_i, \mathbf{y}_k \rangle \mathbf{y}_k}{\|\mathbf{x}_i - \sum_{k=1}^{i-1} \langle \mathbf{x}_i, \mathbf{y}_k \rangle \mathbf{y}_k\|}.$$

Then the $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are orthonormal and

$$\mathrm{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} = \mathrm{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}.$$

The proof is omitted but can be found in the book.

# Recall: connection between matrices and linear maps

## Multiplication by a matrix defines a linear map

Let $A \in M_{m\times n}$ be a fixed matrix. Then, we can define a linear map $T_A \colon \mathbb{F}^n \to \mathbb{F}^m$ via $T_A(\mathbf{v}) = A\mathbf{v}$, where we recall matrix vector multiplication $(A\mathbf{v})_i = \sum_{k=1}^n A_{ik} v_k$ for $i = 1, \ldots, m$.

## Given a bases for $U$ and $V$, $T : U \to V$ can be written as a matrix

Let $T \in \mathcal{L}(U, V)$ where $U$ and $V$ are vector spaces. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be bases for $U$ and $V$ respectively. The matrix of $T$ with respect to these bases is the $m \times n$ matrix $\mathcal{M}(T)$ with entries $A_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ defined by

$$T\mathbf{u}_k = A_{1k}\mathbf{v}_1 + \cdots + A_{mk}\mathbf{v}_m.$$

expand using $v_i$'s

# Eigenvalues

**Definition**

Given an operator $A\colon V \to V$ and $\lambda \in \mathbb{F}$, $\lambda$ is called an *eigenvalue* of $A$ if there exists a non-zero vector $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ such that

$$A\mathbf{v} = \lambda \mathbf{v}.$$

We call such $\mathbf{v}$ an *eigenvector* of $A$ with eigenvalue $\lambda$. We call the set of all eigenvalues of $A$ *spectrum* of $A$ and denote it by $\sigma(A)$.

Motivation in terms of linear maps: Let $T\colon V \to V$ be a linear map, where $V$ is a vector space. We would like to describe the action of this linear map in a particularly "nice" way: such that $T$ acts only by scaling, i.e. $T\mathbf{v}_i = \lambda_i \mathbf{v}_i$ where $\lambda_i \in \mathbb{F}$ for $i = 1, \ldots, n$.

# Finding eigenvalues

Note: here we will assume $\mathbb{F} = \mathbb{C}$, so that we are working on an algebraically closed field.

- Rewrite $A\mathbf{v} = \lambda\mathbf{v}$ as $(A - \lambda I)v = 0 \iff v \in \ker(A - \lambda I)$
- Thus, if $\lambda$ is an eigenvalue, we can find the corresponding eigenvectors by finding the null space of $A - \lambda I$.
- The subspace null$(A - \lambda I)$ is called the *eigenspace*
- To find the eigenvalues of $A$, one must find the scalars $\lambda$ such that null$(A - \lambda I)$ contains non-trivial vectors (i.e. not $\mathbf{0}$)
- Recall: We saw that $T \in \mathcal{L}(U, V)$ is injective if and only if null $T = \{\mathbf{0}\}$.
- Thus $\lambda$ is an eigenvalue if and only if $A - \lambda I$ is not invertible. +
- Recall: $|A| \neq 0$ if and only if $A$ is invertible.
- Thus $\lambda$ is an eigenvalue if and only if $|A - \lambda I| = 0$

If $\dim U = \dim V = n$,
$T$ is invertible $\iff$ $T$ is injective

**Theorem**

*The following are equivalent*

1. $\lambda \in \mathbb{F}$ *is an eigenvalue of* $A$,
2. $(A - \lambda I)\mathbf{v} = 0$ *has a non-trivial solution,*
3. $|A - \lambda I| = 0$.

# Characteristic polynomial

**Definition**

If $A$ is an $n \times n$ matrix, $p_A(\lambda) = |A - \lambda I|$ is a polynomial of degree $n$ called the *characteristic polynomial* of $A$.

To find the eigenvectors of $A$, one needs to find the roots of the characteristic polynomial.

# Example

Find the eigenvalues of

$$A = \begin{bmatrix} 4 & -2 \\ 5 & -3 \end{bmatrix}.$$

$$|A - \lambda I| = \begin{vmatrix} 4-\lambda & -2 \\ 5 & -3-\lambda \end{vmatrix} = (4-\lambda)(-3-\lambda) - (-2)\cdot 5$$

$$= \lambda^2 - \lambda - 12 + 10 = \lambda^2 - \lambda - 2 = (\lambda-2)(\lambda+1)$$

$$\therefore \underbrace{\lambda = -1, 2}_{\text{eigen values}}$$

# Multiplicity

example)

$p(\lambda) = \lambda^{2}(\lambda-1)^{3}(\lambda-2)^{1}$

Algebraic multiplicity of $\lambda=0$ is 2,
$\lambda=1$ is 3,
$\lambda=2$ is 1,

## Definition

The multiplicity of the root $\lambda$ in the characteristic polynomial is called the *algebraic multiplicity* of the eigenvalue $\lambda$. The dimension of the eigenspace null$(A-\lambda I)$ is called the *geometric multiplicity* of the eigenvalue $\lambda$.

||

# of vectors consisting a basis of this subspace.

## Definition (Similar matrices)

Square matrices $A$ and $B$ are called *similar* if there exists an invertible matrix $S$ such that

$$A = SBS^{-1}.$$

Similar matrices have the same characteristic polynomials and hence the same eigenvalues (see exercise).

## Theorem

*Suppose $A$ is a square matrix with distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be eigenvectors corresponding to these eigenvalues. Then $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent.*

*Proof.* By induction on $n$,

Base case $n=1$ is trivial since there is only one vector.

Suppose $v_1, \dots, v_k$ are linearly independent.

Let $\sum_{i=1}^{k+1} \alpha_i v_i = 0$ --- (1)

Multiply $A$ to both sides gives

$$\sum_{i=1}^{k+1} \alpha_i \lambda_i v_i = 0 \quad \text{---} \; (2)$$

Multiply $\lambda_{k+1}$ to (1),

$$\sum_{i=1}^{k+1} \alpha_i \lambda_{k+1} v_i = 0 \quad \cdots\cdots (3)$$

By (2), (3), we have

$$\sum_{i=1}^{k} \alpha_i (\lambda_i - \lambda_{k+1}) v_i = 0$$

But $v_1, \cdots, v_k$ is linearly independent.

Thus $\alpha_i \underbrace{(\lambda_i - \lambda_{k+1})}_{\neq 0 \text{ by assumption.}} = 0, \ \forall i = 1, \cdots, k.$

Since $\lambda_i \neq \lambda_{k+1}$, we have $\alpha_i = 0, \ \forall i = 1, \cdots, k.$

Then, by (1), we also have $\alpha_{k+1} = 0.$ //

## Corollary

*If a $A \in M_n(\mathbb{C})$ has n distinct eigenvalues, then A is diagonalizable. That is there exists an invertible matrix $S \in M_n(\mathbb{C})$ such that $A = SDS^{-1}$, where D is a diagonal matrix with the eigenvalues of A in the diagonal.*

Let $v_i$ be eigenvectors corresponding to $\lambda_i$.

Let $S = (v_1 \cdots v_n)$, $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$.

S is invertible since $v_1, \ldots, v_n$ is linearly independent.

Then $A \cdot S = (Av_1 \cdots Av_n) = (\lambda_1 v_1 \cdots \lambda_n v_n) = SD$

$\therefore \quad S^{-1} A S = D$

## Theorem

*Let $A : V \to V$ be an operator with $n$ eigenvalues. $A$ is diagonalizable if and only if for each eigenvalue $\lambda$, the geometric multiplicity of $\lambda$ and the algebraic multiplicity of $\lambda$ are the same.*

# Example: a diagonalizable matrix

$$\begin{bmatrix} 1 & 2 \\ 8 & 1 \end{bmatrix} \text{ is diagonalizable.}$$

$$\begin{vmatrix} 1-\lambda & 2 \\ 8 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda + 1 - 16 = \lambda^2 - 2\lambda - 15$$

$$= \underbrace{(\lambda - 5)(\lambda + 3)}_{\text{distinct eigenvalues}}$$

$\Longrightarrow$ diagonalizable.

# Example continued

$\lambda = -3$

$$\underbrace{\begin{pmatrix} 4 & 2 \\ 8 & 4 \end{pmatrix}}_{A - \lambda I} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff 2x + y = 0$$

eigenvector : $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$

$\lambda = 5$

$$\begin{pmatrix} -4 & 2 \\ 8 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff 2x - y = 0$$

eigenvector : $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$

# Example continued

Let $S = \begin{pmatrix} 1 & 1 \\ -2 & 2 \end{pmatrix}$. $S^{-1} = \frac{1}{4}\begin{pmatrix} 2 & -1 \\ 2 & 1 \end{pmatrix}$.

$$S^{-1} A S = \frac{1}{4}\begin{pmatrix} 2 & -1 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -2 & 2 \end{pmatrix}$$

$$= \frac{1}{4}\begin{pmatrix} -6 & 3 \\ 10 & 5 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -2 & 2 \end{pmatrix}$$

eigenvalues

$$= \frac{1}{4}\begin{pmatrix} -12 & 0 \\ 0 & 20 \end{pmatrix} = \begin{pmatrix} -3 & 0 \\ 0 & 5 \end{pmatrix}.$$

# Example: a matrix that is not diagonalizable

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \text{ is } \textit{not} \text{ diagonalizable.}$$

$$\begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (\lambda-1)^{2}$$

algebraic multiplicity of $\lambda=1$ is $\underline{2}$

$$\begin{pmatrix} x \\ y \end{pmatrix} \in \ker(A - I)$$

$$\Leftrightarrow \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0 \Leftrightarrow y = 0 \quad \therefore \dim \ker(A-I) = 1$$

geometric multiplicity of $\lambda=1$ is $\underline{1}$

# References

Howard Anton and Chris Rorres. Elementary Linear Algebra. 11th ed. Wiley, 2014

Axler S. *Linear Algebra Done Right*. 3rd ed. Undergraduate Texts in Mathematics. Springer, 2015. Available from: https://link.springer.com/book/10.1007/978-3-319-11080-6

Treil S. *Linear Algebra Done Wrong*. 2017. Available from: https://www.math.brown.edu/streil/papers/LADW/LADW.html