Linear Algebra Short Course

Lecture 4

Matthew J. Holland

matthew-h@is.naist.jp

Mathematical Informatics Lab
Graduate School of Information Science, NAIST

Some useful references

- Finite dimensional inner-product spaces, normal operators: Axler (1997, Ch. 6-7)
- Projection theorem on infinite-dimensional Hilbert spaces: Luenberger (1968, Ch. 3)
- Unitary matrices: Horn and Johnson (1985, Ch. 2)

Lecture contents

1. **Inner products: motivations, terms, and basic properties**
2. **Projections, orthogonal complements, and related problems**
3. **Linear functionals and the adjoint**
4. **Normal operators and the spectral theorem**
5. **Positive operators and isometries**
6. **Some famous decompositions**

Lecture contents

1. **Inner products: motivations, terms, and basic properties**
2. **Projections, orthogonal complements, and related problems**
3. **Linear functionals and the adjoint**
4. **Normal operators and the spectral theorem**
5. **Positive operators and isometries**
6. **Some famous decompositions**

Key idea: generalizing geometric notions

Our progress thus far:

- Built a framework for *sets* with a linearity property
- Built a framework for *functions* with a linearity property
- Looked at some deeper results based on this framework

Note our framework was very general (operations on linear spaces of functions, etc.).

Can we add length and angle to our general framework?

Yes, and the key notion is that of an “inner product” between vectors.

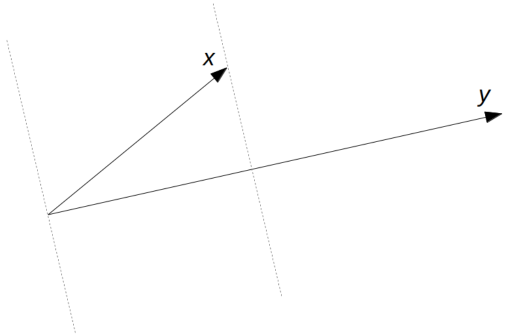
Geometric motivations from vector analysis on $\mathbb{R}^3$ 1

Typically define "projection" by its length, to start.

That is, if $\mathrm{proj}(x; y) \in \mathbb{R}^3$ denotes projection of x onto direction of y, we require $\mathrm{proj}(x; y)$ satisfy

$$\|\mathrm{proj}(x; y)\| = \|x\| |\cos(\angle xy)|,$$

natural considering the right-triangle of hypotenuse length $\|x\|$.



Geometric motivations from vector analysis on $\mathbb{R}^3$ 2

To define the actual projection, just scale y. That is,

$$\operatorname{proj}(x; y) := \frac{\|x\| \cos(\angle xy)}{\|y\|} y.$$

This naturally depends on “what goes where” (i.e., asymmetric in arguments).

A convenient quantity for examining the direction of a vector pair $x, y \in \mathbb{R}^3$ is

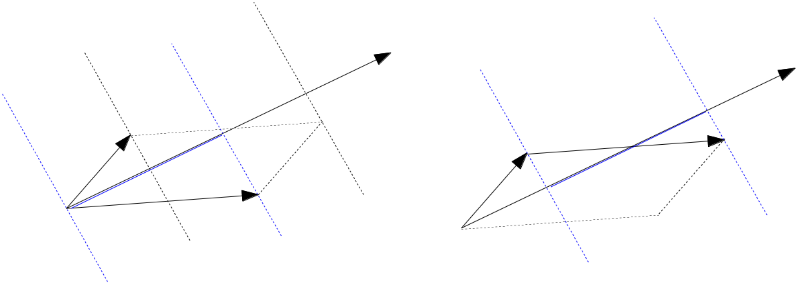
$$x \cdot y := \|x\| \|y\| \cos(\angle xy).$$

Clearly $x \cdot y = y \cdot x$ and

$$\begin{aligned} x \perp y &\iff x \cdot y = 0 \\ \angle xy \text{ acute} &\iff x \cdot y > 0 \\ \angle xy \text{ obtuse} &\iff x \cdot y < 0. \end{aligned}$$

Geometric motivations from vector analysis on $\mathbb{R}^3$ 3

It is easy to geometrically motivate the validity of scalar product being linear in both terms, namely $(x+z)\cdot y = (x\cdot y) + (y\cdot y)$.



With this, and perpendicular unit coordinate vectors $e_1, e_2, e_3 \in \mathbb{R}^3$ note

$$\begin{aligned} x\cdot y &= (x_1e_1 + x_2e_2 + x_3e_3)\cdot(y_1e_1 + y_2e_2 + y_3e_3) \\ &= x_1y_1 + x_2y_2 + x_3y_3 \end{aligned}$$

where $x_i := \|x\| \cos(\angle x e_i)$, same for y_j.

The inner product as a generalized scalar product 1

Enough geometry, now we do algebra.

Scalar product $x \cdot y$ captures both length and angle. Let's generalize.

For $x, y \in \mathbb{R}^n$, natural to extend length and angle quantifiers via

$$\|x\| := \sqrt{x_1^2 + \cdots + x_n^2}$$

$$x \cdot y := x_1 y_1 + \cdots + x_n y_n.$$

What about complex case? For $u = a + ib \in \mathbb{C}$, just like $\mathbb{R}^2$, i.e.,

$$\|u\| := \sqrt{a^2 + b^2} = (u\bar{u})^{1/2} = \sqrt{|u|^2}.$$

Extending this length to $u = (u_1, \ldots, u_n) \in \mathbb{C}^n$, naturally try

$$\|u\| := \sqrt{|u_1|^2 + \cdots + |u_n|^2}.$$

As $\|u\|^2 = u_1\bar{u}_1 + \cdots + u_n\bar{u}_n$, intuitively we'd like to consider defining

$$u \cdot v := u_1\bar{v}_1 + \cdots + u_n\bar{v}_n.$$

The inner product as a generalized scalar product 2

(*) Note that the following properties of our extended scalar products on both $\mathbb{R}^n$ and $\mathbb{C}^n$ hold:

- conjugate symmetry, $\overline{v \cdot u} = u \cdot v$
- definiteness, $u \cdot u = 0 \iff u = 0$
- linearity in first argument, $(\alpha u + \beta u') \cdot v = \alpha(u \cdot v) + \beta(u' \cdot v)$

Recall that these properties are shared by the original scalar product.

In linear algebra, we *start* with inner product properties as axioms.

We abandon the clunky "dot" for the more standard $\langle \cdot, \cdot \rangle$ notation henceforth.

Inner product

Consider vector space V on field $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Defn. Call $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$ an **inner product** on V if $\forall\, u, v, w \in V, \alpha \in \mathbb{F}$,

IP.1 $\langle u, v\rangle = \overline{\langle v, u\rangle}$

IP.2 $\langle \alpha u, v\rangle = \alpha\langle u, v\rangle$

IP.3 $\langle u + w, v\rangle = \langle u, v\rangle + \langle w, v\rangle$

IP.4 $\langle u, u\rangle \geq 0$, and $\langle u, u\rangle = 0 \iff u = 0$.

(*) Additivity actually holds in both arguments. Also, $\langle u, \alpha v\rangle = \bar{\alpha}\langle u, v\rangle$, and $\langle 0, v\rangle = \langle v, 0\rangle = 0$ for any $v \in V$.

Defn. Call $(V, \langle\cdot,\cdot\rangle)$ an **inner product space**. A complete IP space is called **Hilbert space**.

Inner product examples

(*) Note $\mathbb{R}^n$ and $\mathbb{C}^n$ with $\langle \cdot, \cdot \rangle$ defined by our generalized dot product are IP spaces.

(*) Note $\mathcal{P}_m()$ equipped with

$$\langle p, q \rangle := \int_0^1 p(x)q(x)\, dx$$

is a valid inner product space.

(**) Recalling the space of real sequences

$$\ell_p := \left\{ (x_1, x_2, \ldots) \in \mathbb{R}^\infty : \sum_{i=1}^{\infty} |x_i|^p < \infty \right\},$$

if use the Hölder inequality to show finiteness, can show the natural IP

$$\langle x, y \rangle := \sum_{i=1}^{\infty} x_i y_i$$

makes ℓ_2 an inner product space.

Inner product properties 1

Defn. On IP space V, call $\|u\| := \sqrt{\langle u, u\rangle}$ the **norm** on V.

Let's verify this naming is valid (considering general norm definition).

(*) [Cauchy-Schwartz] On IP space V,

$$|\langle u, v\rangle| \leq \|u\|\|v\|, \quad \forall\, u, v \in V.$$

Expand $0 \leq \langle u - \alpha v, u - \alpha v\rangle$ and cleverly pick α.

(*) We just need the triangle inequality. Expand $\|u+v\|^2$ and use C-S to verify

$$\|u+v\| \leq \|u\| + \|v\|, \quad \forall\, u, v \in V.$$

Be sure to check the other axioms to conclude that inner products induce valid norms.

Inner product properties 2

(*) The generalized "Parallelogram Law" clearly follows:

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Defn. If $\langle u, v\rangle = 0$, we say u and v are **orthogonal**, often denoted $u \perp v$. For any $W \subset V$, say $u \perp W$ iff $u \perp w, \forall\, w \in W$.

(*) Clearly the Pythagorean theorem extends nicely,

$$u \perp v \implies \|u+v\|^2 = \|u\|^2 + \|v\|^2.$$

(*) If $u \perp v$, $\forall\, v \in V$, then $u = 0$.

(*) A superb fact: the IP is continuous. That is, if sequences $(u_n), (v_n)$ in V converge to $u_n \to u$ and $v_n \to v$, then

$$\langle u_n, v_n\rangle \to \langle u, v\rangle.$$

Lecture contents

1. Inner products: motivations, terms, and basic properties
2. Projections, orthogonal complements, and related problems
3. Linear functionals and the adjoint
4. Normal operators and the spectral theorem
5. Positive operators and isometries
6. Some famous decompositions

Lecture contents

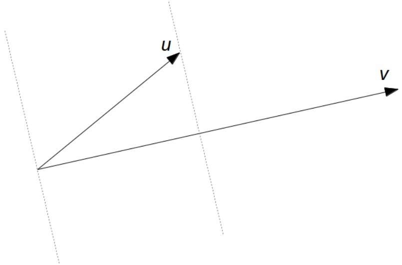
1. Inner products: motivations, terms, and basic properties
2. **Projections, orthogonal complements, and related problems**
3. **Linear functionals and the adjoint**
4. **Normal operators and the spectral theorem**
5. **Positive operators and isometries**
6. **Some famous decompositions**

Considering the notion of a projection again

We previously considered $\mathrm{proj}(u; v)$ only geometrically.
Now it is quite easy. Intuitively, we seek $\alpha v \in [\{v\}]$ such that

$$u = \alpha v + w, \text{ where } w \perp v.$$

(*) Check that the scalar then must be $\alpha = \langle u, v \rangle / \|v\|^2$, a nostalgic form indeed.



This “orthogonal projection” (formalized shortly) will play an important role moving ahead.

An optimization problem

Consider the following problem.
Let V be an IP space, and $X \subset V$ a subspace. Fix $u_0 \in V$, and

$$\text{find } \widehat{x} \in X \text{ which minimizes } \|x - u_0\| \text{ in } x.$$

Natural questions:

Does a solution exist? Is it unique? What is it?

The answers to these questions are given by the “Projection Theorem,” a truly classic result.

Note: no requirement that $\dim V < \infty$ thus far.

The Projection Theorem

(*) Say $\widehat{x} \in X$ is s.t. $\|\widehat{x} - u_0\| \leq \|x - u_0\|$ for all $x \in X$. Then $\widehat{x}$ (the minimizing vector in X) is unique.

(*) Element $\widehat{x}$ (uniquely) minimizes $\|x - u_0\| \iff \widehat{x} - u_0 \perp X$.

We have not shown that such an element need exist; to do this we need slightly stronger conditions:

(**) Let V be a *Hilbert* space, and $X \subset V$ a *closed* subspace. Then, for any $u_0 \in V$,

$$\exists \widehat{x} \in X, \ \|\widehat{x} - u_0\| \leq \|x - u_0\|, \forall x \in X.$$

This result is typically called the classical Projection theorem.
Let's develop these ideas further.

Orthogonal complements 1

Let's develop these ideas further. Let V be IP space.

Defn. Take any subset $U \subset V$, and denote by

$$U^\perp := \{v \in V : u \perp v\},$$

called the **orthogonal complement** of U.

(*) Note $\{0\}^\perp = V$ and $V^\perp = \{0\}$. Also, for any U, have that $U^\perp \subset V$ is a closed subspace.

(*) Some additional properties (still allowing $\dim V = \infty$):

- $U \subset U^{\perp\perp}$
- $U \subset W \implies W^\perp \subset U^\perp$
- $U^{\perp\perp\perp} = U^\perp$
- $U^{\perp\perp} = \overline{[U]}$

Orthogonal complements 2

The use of the term "complement" will now be justified.

(*) Let V be a Hilbert space and $X \subset V$ a closed subspace. Then,

$$V = X \oplus X^{\perp}, \text{ and } X^{\perp\perp} = X.$$

This result may be proved using the Projection Theorem.

Thus, orthogonal complements furnish a nice direct sum decomposition. Uniquely have $v = x + x'$ with $x \in X, x' \in X^{\perp}$.

(*) If specialize to $\dim V < \infty$ everything simplifies further:

- In this case, V an IP space $\implies$ V is Hilbert (recall Lec 1).
- Thus, above result and Proj Theorem hold for any subspace.
- Similarly, for subspace $U \subset V$ have $U^{\perp\perp} = U$.
- Naturally have $\dim U^{\perp} = \dim V - \dim U$.

Orthogonal projection 1

With these terms down, we provide a general projection notion.

Defn. Let $X \subset V$ be a subset, and take $u \in V$. Uniquely, have

$$u = x + x'$$

where $x \in X, x' \in X^{\perp}$. Define the **orthogonal projection** of u onto X by $\mathrm{proj}(u; X) \coloneqq x' = u - x$.

(*) This pops up naturally in the Proj Theorem, since

$$\widehat{x} \in X \text{ minimizes } \|x - u_0\| \iff \widehat{x} - u_0 \in X^{\perp},$$

and as $u_0 = (u_0 - \widehat{x}) + \widehat{x}$, have $\widehat{x} = \mathrm{proj}(u; X)$.

Orthogonal projection 2

(*) Projecting some $x \in V$ in the direction of $y \in V$ is tantamount to acquiring $\mathrm{proj}(x; [\{y\}])$. It takes a familiar form. Decompose

$$x = \alpha y + w,\ w \perp y. \text{ Thus, } \mathrm{proj}(x; [\{y\}]) = \langle x, y\rangle / \|y\|^2$$

as we would hope.

Properties of orthogonal projection

Let $U \subset V$ be a subspace of V. Assume $\dim V < \infty$. Denote $P_U(v) := \operatorname{proj}(v; U)$ here.

(*) The following may readily be checked:

- $P_U \in \mathcal{L}(V)$, a linear operator.
- $\operatorname{range} P_U = U$, $\operatorname{null} P_U = U^{\perp}$.
- $P_U^2 = P_U$ (idempotent map)
- $\|P_U(v)\| \leq \|v\|, \forall\, v \in V$, a contraction.

(*) Interestingly, the latter two properties characterize the orthogonal projections. That is, taking some $S \in \mathcal{L}(V)$,

$$S^2 = S \text{ and } \|S(v)\| \leq \|v\|, \forall\, v \in V \implies S = P_U \text{ for some subspace } U.$$

Orthogonal sets 1

Let $S \subset V$ be a subset of IP space V.

Defn. We call S an **orthogonal set** if for any $u, v \in S$, we have $u \perp v$. We call S **orthonormal** if it is orthogonal and each $u \in S$ is $\|u\| = 1$.

In the following useful way, here orthogonality connects with the more fundamental notion of independence seen earlier:

(*) If $S \subset V$ is orthogonal, it is (linearly) independent.

Orthogonal sets 2

Conversely, given independent sets, we can always "orthonormalize," in the following sense.

(*) Given independent sequence $v_1, v_2, \ldots$, exists orthonormal sequence $e_1, e_2, \ldots$ such that

$$[\{v_1, \ldots, v_n\}] = [\{e_1, \ldots, e_n\}], \text{ for any } n > 0.$$

Proving this is straightforward, and can be done constructively. Initialize $e_1 := v_1 / \|v_1\|$. The rest are induced by

$$e_n = v_n - \sum_{i=1}^{n-1} \langle v_n, e_i \rangle e_i.$$

This is often called the "Graham-Schmidt procedure."

(*) Thus every inner product space has an orthonormal *basis*.

Orthogonal bases

Orthonormal sets play an important role as convenient bases.

Let V be $\dim V = n$, with orthonormal basis $\{e_1, \ldots, e_n\}$. Then for any $v \in V$, have

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$
$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2.$$

A classical result is nice to check here.

(*) (Schur's theorem, 1909). For any finite-dim V on $\mathbb{C}$ and $T \in \mathcal{L}(V)$, there exists basis B such that

$$M(T; B) \text{ is upper-tri and } B \text{ is orthonormal.}$$

Show this with our "portmanteau theorem" for upper-tri representations (Lec 3), and the previous result via Graham-Schmidt.

Optimization example 1

First example: find the optimal approximation of sin(x) on [-pi,pi] by a 5th-degree polynomial.

Optimization example 2

Second example: Find the closest element in the subspace generated by m vectors to an arbitrary vector.

Optimization example 3

Third example: Find the element of an affine set which has the smallest norm (this is of course the distance from any element in that affine set to the associated hyperplane through the origin).

Optimization example 4

Fourth example: Minimum distance from an arbitrary element to a convex set.

Lecture contents

Lecture contents

Linear functionals 1

Linear maps which return scalars play an important role in linear algebra (and other related fields).

Defn. Let V be a vector space on $\mathbb{F}$. Any $f \in \mathcal{L}(V, \mathbb{F})$ is called a **linear functional**.

(*) The following are linear functionals:

- On $\mathbb{R}^n$ over $\mathbb{F}$, $f(x) := \sum_{i=1}^n \alpha_i x_i$, any fixed $\alpha \in \mathbb{F}^n$. In fact, every $g \in \mathcal{L}(\mathbb{R}^n, \mathbb{F})$ takes this form.
- On real $\mathcal{P}_6(\mathbb{R})$, $f(x) := \int_0^1 x(t) \cos(t) \, dt$.
- On $\mathcal{C}[0,1]$, $f(x) := x(0.5)$.
- On Hilbert space H, $f(x) := \langle x, \bar{h} \rangle$ for fixed $\bar{h} \in H$.

The last example here will be of particular interest to us.

Linear functionals 2

A very neat fact:

(*) On IP space $(V, \langle \cdot, \cdot \rangle)$ with $\dim V = n$, let f be a linear functional. Then, there exists a unique $\bar{v} \in V$ such that

$$f(u) = \langle u, \bar{v} \rangle, \quad \forall\, u \in V.$$

To see this, recall we can always find an orthonormal basis $\{e_1, \ldots, e_n\}$ of V. Expand arbitrary u wrt this basis, examine $f(u)$ using linearity of f.

This result is a special case of the Riesz-Fréchet theorem, which extends things to infinite-dim case. See for example Luenberger (1968, Ch. 4).

Adjoint of a linear map

A very important notion moving forward.

Defn. Let U, V be IP spaces on $\mathbb{F}$, with $\dim U, \dim V < \infty$. Take any $T \in \mathcal{L}(U, V)$, fixed. For any $v \in V$ fixed,

$$f(u) := \langle Tu, v \rangle$$

is clearly a linear functional $f \in \mathcal{L}(U, \mathbb{F})$. By Riesz-Fréchet, $\exists\, u^* \in U$, unique, s.t.

$$f(u) = \langle u, u^* \rangle, \quad u \in U.$$

The initial v was arbitrary, so, we may define a map $T^* : V \to U$ by

$$T^*(v) := u^* \text{ as above.}$$

We call T^* the **adjoint** of T. Somewhat subtle, but critical.
Critical to memorize: $\langle T(u), v \rangle = \langle u, T^*(v) \rangle$.

Properties of adjoints

(*) Define $T : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$T(x_1, x_2, x_3) := (x_2 + 3x_3, 2x_1),$$

and verify with usual inner products that $T^*(y) = (2y_2, y_1, 3y_1)$.
Simply note we must have for any $y \in \mathbb{R}^2$ that $\langle Tx, y\rangle = \langle x, T^*y\rangle$.

(*) For any $T \in \mathcal{L}(U, V)$ as in previous slide, have $T^* \in \mathcal{L}(V, U)$.

(*) Verify the following properties of $(\cdot)^*$ map of $T \mapsto T^*$. Take $T, T' \in \mathcal{L}(U, V)$ and $\alpha \in \mathbb{F}$.

- $(T + T')^* = (T)^* + (T')^*$
- $(\alpha T)^* = \bar{\alpha}(T)^*$
- $(T^*)^* = T$.
- For $T \in \mathcal{L}(U, V), S \in \mathcal{L}(V, W)$ have $(ST)^* = T^*S^*$, where W any IP space.

More properties of adjoints

(*) Let $T \in \mathcal{L}(V)$ and take $\alpha \in \mathbb{F}$. Then,

$$\alpha \in \sigma(T) \iff \bar{\alpha} \in \sigma(T^*).$$

(*) Let $U \subset V$ be a subspace, and $T \in \mathcal{L}(V)$. Then,

$$U \text{ is } T\text{-invariant} \iff U^\perp \text{ is } T^*\text{-invariant.}$$

(*) Take $T \in \mathcal{L}(V, W)$. Prove

- T is injective iff T^* is surjective.
- T is surjective iff T^* is injective.

(*) With this, take $T \in \mathcal{L}(V, W)$ and verify

$$\dim \operatorname{null} T^* = \dim \operatorname{null} T + \dim W - \dim V$$

as well as $\dim \operatorname{range} T^* = \dim \operatorname{range} T$.

(*) Note the above result completes the generalization of Strang's first fundamental theorem (i.e., row/colspaces have same dimension), mentioned in Lec 2.

Connections between a map and its adjoint

(*) Take any $T \in \mathcal{L}(U, V)$, both U, V IP spaces. Then,

- $\operatorname{null} T^* = (\operatorname{range} T)^\perp$
- $\operatorname{range} T^* = (\operatorname{null} T)^\perp$
- $\operatorname{null} T = (\operatorname{range} T^*)^\perp$
- $\operatorname{range} T = (\operatorname{null} T^*)^\perp$

Defn. Denote the **conjugate transpose** of matrix $A = [a_{ij}] \in \mathbb{F}^{m \times n}$ by $A^* := \overline{A^T} = [\overline{a_{ji}}]$.

Given a proper matrix representation of a linear map, we can easily find the representation of its adjoint:

(*) Take $T \in \mathcal{L}(U, V)$ for finite-dim IP spaces U, V. Let B_U, B_V be respectively orthonormal bases of U and V. Then,

$$(M(T; B_U, B_V))^* = M(T^*; B_V, B_U).$$

Lecture contents

Lecture contents

1. Inner products: motivations, terms, and basic properties
2. Projections, orthogonal complements, and related problems
3. Linear functionals and the adjoint
4. **Normal operators and the spectral theorem**
5. **Positive operators and isometries**
6. **Some famous decompositions**

Operators on inner product spaces

The flow through the first three lectures was:

- Linear spaces (sets with linearity)
- Linear maps (functions with linearity)
- Linear operators on general spaces

Now, considering what we've seen in this lecture, the next key point to tackle is

- **Linear operators on inner product spaces**

That is precisely what we look at now.

Self-adjoint operators

Let V be a finite-dim IP space, take $T \in \mathcal{L}(V)$.

Defn. Call operator T **self-adjoint** or **Hermitian** when $T = T^*$.

(*) Take $T \in \mathcal{L}(\mathbb{F}^2)$ defined to have matrix

$$M(T) = \begin{bmatrix} 19 & \gamma \\ 7 & 59 \end{bmatrix}$$

wrt standard basis of course. Note T self-adjoint $\iff \gamma = 7$.

(*) Similarly, we may confirm that for arbitrary orthonormal basis B,

$$T = T^* \iff M(T;B) = (M(T;B))^*.$$

A natural connection to the more familiar matrix territory.

Properties of self-adjoint operators

(*) If $T, S \in \mathcal{L}(V)$ are self-adjoint, then $T + S$ is self-adjoint.

(*) If $T \in \mathcal{L}(V)$ is self-adjoint, then for $\alpha \in \mathbb{R}$, αT is self-adjoint.

(*) Let $T \in \mathcal{L}(V)$ be self-adjoint. Then every eigenvalue is real (recalling $\mathbb{F}$ may be either $\mathbb{C}$ or $\mathbb{R}$).

(*) Of course, for $T \in \mathcal{L}(\mathbb{F}^n)$ specified by $A \in \mathbb{F}^{m \times n}$, this is already a matrix WRT the standard basis, so just look at A.

A nice analogy:
Think of the self-adjoint operators among all operators like $\mathbb{R}$ as a subset of $\mathbb{C}$ (adjoint operation $(\cdot)^*$ analogous to complex conjugate operation $\overline{(\cdot)}$).

Characterizing the self-adjoint operators

A characterization of the self-adjoint operators is given at the start of section 5, but technically is used for upcoming results.

The normal operators

The next very important class of operators.

Defn. Take $T \in \mathcal{L}(V)$. When T commutes with its adjoint, that is, if

$$TT^* = T^*T,$$

we call T a **normal** operator.

(*) Every self-adjoint operator is normal.

(*) Let B be an orthonormal basis. Then, T is normal iff $M(T; B)$ and $M(T^*; B)$ commute.

(*) Consider $T \in \mathcal{L}(\mathbb{F}^2)$ with matrix (wrt standard basis)

$$M(T) = \begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix},$$

clearly normal, not self-adjoint. Thus normals are *larger class*.

Properties of normal operators

Normal operators *need not* equal their adjoints, yet they have *a lot in common* with them:

(**) Norms of maps are common:

$$T \text{ is normal} \iff \|T(v)\| = \|T^*(v)\|, \quad v \in V.$$

(*) This implies for normal $T \in \mathcal{L}(V)$,

$$\operatorname{null} T = \operatorname{null} T^*.$$

(*) Their eigenvectors are closely related. Let T is normal and $\alpha \in \sigma(T)$. We have that

$$\text{if } Tv = \alpha v, \text{ then } T^* v = \bar{\alpha} v.$$

(*) This gives us a *critical* property. Let $\alpha_1, \ldots, \alpha_m$ be the distinct eigenvalues of $T \in \mathcal{L}(V)$, with eigenvectors $v_1, \ldots, v_m$. Then,

$$\{v_1, \ldots, v_m\} \text{ is orthogonal.}$$

This clearly strengthens previous results (only had independence).

The spectral theorem, intuitively

Recall we know that for $T \in \mathcal{L}(V)$, $\dim V = n$,

T is "diagonalizable" $\iff$ $\exists$ basis $\{v_1, \ldots, v_n\}, v_i$ eigenvectors.

While such a T is nice, in general *we have no guarantee that basis* $\{v_1, \ldots, v_n\}$ *is orthogonal*, which is really the "nicest" setup.

The spectral theorem characterizes the very nicest operators:

$\mathbb{C}$ *version:*

The nicest operators are the normal operators.

$\mathbb{R}$ *version:*

The nicest operators are the self-adjoint operators.

Why is this useful?

It gives us easy access to an orthonormal basis!

(in general, we have only existence guarantees)

The spectral theorem

Work on V, $\dim V = n$. Take $T \in \mathcal{L}(V)$.

(**) Complex spectral theorem. Assume V on $\mathbb{F} = \mathbb{C}$.

$$T \text{ is normal} \iff \exists \text{ ortho basis } \{v_1, \ldots, v_n\}, \text{ all eigenvecs}$$

(**) Real spectral theorem. Assume V on $\mathbb{F} = \mathbb{R}$.

$$T \text{ is self-adjoint} \iff \exists \text{ ortho basis } \{v_1, \ldots, v_n\}, \text{ all eigenvecs}$$

Proving these results is somewhat involved (though we have the tools required), but absolutely worth doing.

Key take-aways:

For the "nicest" operators (and only the nicest operators), the eigenvectors furnish an *orthogonal* basis.

All the self-adjoint operators (on general $\mathbb{F}$) can be diagonalized via an orthogonal basis.

Illustrative examples 1

Example. (*) Define $T \in \mathcal{L}(\mathbb{C}^2)$ by

$$M(T) = \begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix}$$

wrt standard basis of $\mathbb{C}^2$. Confirm

$$B = \left\{ \frac{(i,1)}{\sqrt{2}}, \frac{(-i,1)}{\sqrt{2}} \right\}$$

is a orthonormal basis, both are eigenvectors of T, and indeed that $M(T;B)$ is diagonal.

Illustrative examples 2

Example. (*) Similar deal, this time $T \in \mathcal{L}(\mathbb{R}^3)$, with matrix (wrt std. basis)

$$M(T) = \begin{bmatrix} 14 & -13 & 8 \\ -13 & 14 & 8 \\ 8 & 8 & -7 \end{bmatrix}.$$

Check the same properties as in previous slide, this time for the basis

$$B' = \left\{ \frac{(1,-1,0)}{\sqrt{2}}, \frac{(1,1,1)}{\sqrt{3}}, \frac{(1,1,-2)}{\sqrt{6}} \right\}.$$

Some comments on the $\mathbb{R}$ case

Even on $\mathbb{R}$, when restricted to self-adjoint operators, things simplify.

(**) Let $T \in \mathcal{L}(V)$ on $\mathbb{R}$ be self-adjoint. Take $a, b \in \mathbb{R}$ s.t. $a^2 < 4b$. Then,

$$T^2 + aT + bI \in \mathcal{L}(V) \text{ is invertible.}$$

(*) This implies T has *no* "eigenpairs," and thus (recall Lec 3),

$$\implies \sigma(T) \neq \emptyset.$$

Of course, we know this last fact must hold, since we've already presented the real spectral theorem.

Note: we haven't characterized the normal operators in the real case. For this, see Axler (1997, Ch. 7).

Specialized structural results

With the extra assumptions of the "nice" operators, the structural results specialize nicely, proving us with respectively *orthogonal* subspaces.

(*) Let $T \in \mathcal{L}(V)$ be self-adjoint if $\mathbb{F} = \mathbb{R}$ (normal if $\mathbb{F} = \mathbb{C}$), with distinct eigenvalues $\alpha_1, \ldots, \alpha_m$. Then,

$$V = \mathrm{null}(T - \alpha_1 I) \oplus \cdots \oplus \mathrm{null}(T - \alpha_m I)$$

and $\mathrm{null}(T - \alpha_i I) \perp \mathrm{null}(T - \alpha_j I)$, all $i \neq j$.

Thus, the spectral information of any "nice" T yields an orthogonal decomposition of V.

Lecture contents

1. Inner products: motivations, terms, and basic properties
2. Projections, orthogonal complements, and related problems
3. Linear functionals and the adjoint
4. **Normal operators and the spectral theorem**
5. **Positive operators and isometries**
6. **Some famous decompositions**

Lecture contents

Characterizing the self-adjoint operators

Real case:
The real spectral theorem characterizes self-adjoint operators.

Complex case:
We haven't discussed this yet.

(**) For any $T \in \mathcal{L}(V)$ on $\mathbb{C}$, say

$$\langle Tv, v \rangle = 0, \quad \forall\, v \in V.$$

Then, $T = 0$ (for self-adjoint T, holds for $\mathbb{R}$ case).

(*) It follows that for $T \in \mathcal{L}(V)$ on $\mathbb{C}$,

$$T \text{ self-adjoint} \iff \langle Tv, v \rangle \in \mathbb{R},\ v \in V.$$

So, complex self-adjoint operators are precisely those for which any v and its map $T(v)$ have a real inner product.

Important special case: when $\langle Tv, v \rangle \geq 0$, all $v \in V$.

Positive operators and square roots

Let V be $\dim V < \infty$.

Defn. Focus on *self-adjoint* $T \in \mathcal{L}(V)$. Call T a **positive (semi-definite)** operator if

$$\langle Tv, v\rangle \geq 0, \quad \forall\, v \in V.$$

(*) Of course, for $\mathbb{C}$ case, self-adjoint requirement is superfluous.

(*) For any subspace $U \subset V$, the projection operator $\operatorname{proj}(\cdot; U)$ is positive.

Defn. Take any operator $T \in \mathcal{L}(V)$. If exists $S \in \mathcal{L}(V)$ such that

$$S^2 = T,$$

then call S a **square root** of operator T.

(*) Find a square root of $T \in \mathcal{L}(\mathbb{F}^3)$ defined $T(z_1, z_2, z_3) := (z_3, 0, 0)$.

Portmanteau theorem for positive operators

(**) Take any $T \in \mathcal{L}(V)$ on finite-dim V. The following are equivalent.

- A T is positive; i.e., $T = T^*$ and $\langle Tv, v\rangle \geq 0$, all v.
- B $T = T^*$ and eigenvalues of T are non-negative.
- C Exists positive $Q \in \mathcal{L}(V)$ such that $Q^2 = T$.
- D Exists self-adjoint $R \in \mathcal{L}(V)$ such that $R^2 = T$.
- E Exists $S \in \mathcal{L}(V)$ such that $S^*S = T$.

(**) When $T \in \mathcal{L}(V)$ is positive, there exists a *unique* $Q \in \mathcal{L}(V)$ s.t. $Q^2 = T$. That is, T has a unique square root. Denote $\sqrt{T} := Q$.

Great. We'll sort out the implications in the next slide.

Key properties of positive operators

(*) From the results of the previous slide:

- Only positive operators have positive square roots
- If an operator has a positive square root, this root is unique.
- Positive operators form a subset of self-adjoint operators.
- Not only are eigenvalues real, they're positive.
- For any $S \in \mathcal{L}(V)$, S^*S is positive.
- If S is positive or self-adjoint, S^2 is positive.

(*) Let $T \in \mathcal{L}(V)$ be positive. Show

$$T \text{ invertible} \iff \langle Tv, v\rangle > 0, \forall\, v \neq 0.$$

Isometries

Norm-preserving operators are also naturally of interest.

Defn. Call $T \in \mathcal{L}(V)$ an **isometry** if

$$\|Tv\| = \|v\|, \, \forall \, v \in V.$$

This is a general term. For specific cases, other names are used:

- If $\mathbb{F} = \mathbb{C}$, call T a **unitary operator**.
- If $\mathbb{F} = \mathbb{R}$, call T an **orthogonal operator**.

(*) Let $\beta \in \mathbb{F}$ be $|\beta| = 1$. Note $T := \beta I$ is an isometry.

(*) Let $\{v_1, \ldots, v_n\}$ be orthonormal basis of V. Define $T \in \mathcal{L}(V)$ by

$$T(v_i) := \beta_i v_i,$$

with $|\beta_i| = 1$ for each $i = 1, \ldots, n$. Then T is positive.

(*) Counter-clockwise rotation on $V = \mathbb{R}^2$ is an isometry.

Many useful properties of isometries

(*) If $T \in \mathcal{L}(V)$ an isometry, T^{-1} exists.

(**) Let $T \in \mathcal{L}(V)$. The following are equivalent.

A T is an isometry.

B $\langle Tu, Tv \rangle = \langle u, v \rangle$ for all $u, v \in V$ (preserves IP)

C $T^*T = I$

D For any orthonormal set $\{e_1, \ldots, e_m\}$, the mapped $\{Te_1, \ldots, Te_m\}$ also orthonormal ($0 \leq m \leq n$).

E Exists basis $\{v_1, \ldots, v_n\}$ such that $\{Tv_1, \ldots, Tv_n\}$ is orthonormal.

F T^* is an isometry.

This is a nice collection of characterizations for the special case of isometries, which yields some critical implications.

Implications of isometry equivalences

Some *key* implications:

(*) Clearly, if T an isometry, have $T^{-1} = T^*$.

(*) T preserves norms $\iff$ T preserves inner products.

(*) Every isometry is normal.

(*) Now a great equivalence. Let $E := \{e_1, \ldots, e_n\}$ be any orthonormal basis of V on $\mathbb{F}$. Then,

$$T \text{ an isometry} \iff \text{columns of } M(T; E) \text{ orthonormal}$$

To see this, use $\mathbf{A} \implies \mathbf{D}$ for the $\implies$ direction, and $\mathbf{E} \implies \mathbf{A}$ for the $\impliedby$ direction.

(*) Using $\mathbf{A} \iff \mathbf{F}$, show that analogous condition holds using the rows of $M(T; E)$ above.

More concrete characterization of isometries

The previous characterizations of isometries were quite general. Let's put forward a more concrete equivalence condition.

Complex case:
(*) Let $T \in \mathcal{L}(V)$, on $\mathbb{C}$. The following condition is both sufficient and necessary for $T \in \mathcal{L}(V)$ to be an isometry.

Exists $\{v_1, \ldots, v_n\}$, orthonormal basis of V, where the v_i are eigenvectors of T, with eigenvalues $|\alpha_i| = 1$.

Real case:
Similar to Lecture 3, a bit less elegant. See Axler (1997, Ch. 7).

Symmetric real matrices

In probability/statistics, *symmetric real matrices* appear frequently.

We've said a lot about how working on $\mathbb{R}$ is somewhat inconvenient. What's so special about symmetric matrices?

That's easy: Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then,

- $T \in \mathcal{L}(\mathbb{R}^n)$ def'd $T(x) := Ax$ is self-adjoint
- T is normal.
- T has eigenvalues, and $\mathbb{R}^{n \times n}$ has an orthonormal basis $\{v_1, \ldots, v_n\}$ of T's eigenvectors.
- T may be "diagonalized" by $\{v_1, \ldots, v_n\}$.
- Specifically, A may be diagonalized by COB matrix $[v_1 \cdots v_n]$.

Lecture contents

1. Inner products: motivations, terms, and basic properties
2. Projections, orthogonal complements, and related problems
3. Linear functionals and the adjoint
4. Normal operators and the spectral theorem
5. **Positive operators and isometries**
6. **Some famous decompositions**

Lecture contents

1. Inner products: motivations, terms, and basic properties
2. Projections, orthogonal complements, and related problems
3. Linear functionals and the adjoint
4. Normal operators and the spectral theorem
5. Positive operators and isometries
6. **Some famous decompositions**

Some famous decompositions

Here we look at conditions for some well-known decompositions:

- Schur
- Polar
- Singular value
- Spectral

Here we periodically switch over to “matrix language” to illustrate the generality of our results thus far.

Schur's decomposition

See for example Magnus and Neudecker (1999).

(*) Let A be complex, $n \times n$ matrix. Then, there exists unitary matrix R such that

$$R^*AR = \begin{bmatrix} \alpha_1 & & * \\ & \ddots & \\ 0 & & \alpha_n \end{bmatrix},$$

where the α_i are eigenvalues of A.

To see this:
Easy, just let $T(z) := Az$, which we know can always be upper-triangularized by an orthonormal basis $E = \{v_1, \ldots, v_n\}$. Construct COB matrix (here R) using these v_i. Done.

Polar decomposition: first, an analogy

A nice analogy exists between $\mathbb{C}$ and $\mathcal{L}(V)$:

$$\begin{aligned} z \in \mathbb{C} &\cdots T \in \mathcal{L}(V) \\ \bar{z} \in \mathbb{C} &\cdots T^* \in \mathcal{L}(V) \\ z = \bar{z}, \text{ i.e. } \mathbb{R} \subset \mathbb{C} &\cdots T = T^*, \text{ i.e. } \{\text{self-adjoint ops.}\} \subset \mathcal{L}(V) \\ x \in \mathbb{R}, x \geq 0 &\cdots \{\text{positive ops.}\} \subset \{\text{self-adjoint ops.}\} \\ \text{unit circle } \{z : z\bar{z} = 1\} &\cdots \text{isometries, } \{T : T^*T = I\} \end{aligned}$$

Note any $z \in \mathbb{C}$ can be written

$$z = \left(\frac{z}{|z|}\right)\sqrt{z\bar{z}}, \text{ of course noting } z/|z| \text{ on unit circle.}$$

Following the analogy, we wonder whether for any $T \in \mathcal{L}(V)$ we have an isometry S such that T breaks down into $S\sqrt{T^*T}$...

Polar decomposition

Indeed, the analogy proves to lead us in a fruitful direction.

(**) (Polar decomposition). Let $T \in \mathcal{L}(V)$ over $\mathbb{F}$. Then, exists isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

The naming refers to $z = e^{\theta i}r, \theta \in [0, 2\pi)$, where $r = |z|$.
Here S (like $e^{\theta i}$) only changes direction.
Magnitude is determined by $\sqrt{T^*T}$ (like r).

Why is this nice? T is totally general, but,

$$T = \text{isometry} \times \text{positive operator}$$

i.e., it breaks into two classes we know very well!

Polar decomposition, in matrix language

(*) Let $A \in \mathbb{F}^{n \times n}$. Then, exists unitary matrix Q and positive semi-definite matrix P such that

$$A = QP.$$

To see this:
Take matrix of $T \in \mathcal{L}(\mathbb{F}^n)$ defined by A with respect to usual basis B, so

$$A = M(T; B) = M(S; B)M(\sqrt{T^*T}; B),$$

where S an isometry, and $\sqrt{T^*T}$ is positive. Verify $M(S; B)$ is unitary and $M(\sqrt{T^*T}; B)$ is positive semi-definite. Done.

(*) Also note that in decomposing any T into an isometry/positive operator product, the *only* choice for the positive operator is $\sqrt{T^*T}$.

Singular values of operators

Clearly, for $T \in \mathcal{L}(V)$, the positive operator $\sqrt{T^*T}$ clearly plays an important role. It pops up in both theory and practice.

Take any $T \in \mathcal{L}(V)$ on general $\mathbb{F}$.

T need not have real eigenvalues, nor even any at all. However,

$\sqrt{T^*T}$ always has real, non-neg eigenvalues.

Defn. Call the eigenvalues $s_i \in \sigma(\sqrt{T^*T})$ the **singular values** of T.

Singular value decomposition (SVD) 1

(*) Let's see the interesting role that $\sigma(\sqrt{T^*T})$ plays.

Take $T \in \mathcal{L}(V)$, $\dim V = n$. Let $s_1, \ldots, s_n$ denote the eigenvalues of $\sqrt{T^*T}$, up to multiplicity.

By spectral theorem, exists $\{b_1, \ldots, b_n\}$, eigenvectors of $\sqrt{T^*T}$, forming an orthonormal basis of V. Taking $v \in V$, recall

$$v = \langle v, b_1 \rangle b_1 + \cdots + \langle v, b_n \rangle b_n.$$

Note that via polar decomp. $T = S\sqrt{T^*T}$,

$$\begin{aligned} Tv &= S\sqrt{T^*T}v \\ &= \langle v, b_1 \rangle s_1 S b_1 + \cdots + \langle v, b_n \rangle s_n S b_n, \end{aligned}$$

and as S is an isometry, $\{Sb_1, \ldots, Sb_n\}$ is an orthonormal basis of V.

Singular value decomposition (SVD) 2

(*) With this handy decomposition, one may easily check that for $B_1 := \{b_1, \ldots, b_n\}$ and $B_2 := \{Sb_1, \ldots, Sb_n\}$, have

$$M(T; B_1, B_2) = \begin{bmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{bmatrix},$$

another rare appearance of matrix reps with distinct bases.

To estimate singular values:
Finding $\sqrt{T^*T}$ explicitly may be hard. For fixed basis B, let G be the matrix that diagonalizes such that

$$GM(\sqrt{T^*T}; B)G^* = M(T; B_1, B_2),$$

clearly $M(T; B_1, B_2)^2 = GM(T^*T; B)G^*$. So

$$\sigma(T^*T) = \{s_1^2, \ldots, s_n^2\}.$$

Estimating the eigenvalues of positive T^*T is easier. Take roots.

SVD, in (square) matrix language

(**) (Matrix SVD). Consider $A \in \mathbb{F}^{n\times n}$. Then we may factorize A into

$$A = QDR^*,$$

where Q, R are unitary matrices, and D is diagonal, and whose diagonal entries are precisely the singular values of A.

To see this:

Defining $T(z) := Az$, via polar decomposition $T = S\sqrt{T^*T}$, letting B be usual basis,

$$\begin{aligned} A = M(T;B) &= M(S;B)M(\sqrt{T^*T};B) \\ &= M(S;B)GDG^*, \end{aligned}$$

where $G = M(I;E,B)$, and $E = \{v_1, \dots, v_n\}$ is an orthonormal basis diagonalizing $\sqrt{T^*T}$, where $\sqrt{T^*T}v_i = s_i$.

G has ortho cols, equivalent to G being unitary. Let $R^* := G^* = G^{-1}$. Set $Q := M(S;B)G$, unitary as both $M(S;B)$ and G are.

SVD-related additional properties

(*) Take $T \in \mathcal{L}(V)$, singular values $s_1, \ldots, s_n$. Then,

$$T \text{ invertible} \iff s_i \neq 0, i = 1, \ldots, n.$$

(*) Take $T \in \mathcal{L}(V)$. Then,

$$\dim \operatorname{range} T = |\{s \in \sigma(\sqrt{T^*T}) : s \neq 0\}|.$$

(*) Take $S \in \mathcal{L}(V)$, singular values $s_1, \ldots, s_n$. Then,

$$S \text{ an isometry} \iff s_i = 1, i = 1, \ldots, n.$$

(*) Let s_* and s^* denote the smallest and largest singular values of $T \in \mathcal{L}(V)$. Then,

$$s_* \|v\| \leq \|Tv\| \leq s^* \|v\|, \text{ any } v \in V.$$

Singular values of general linear maps

In fact, the "usual" singular value decomposition extends to the more general case of $A \in \mathbb{F}^{m\times n}$ quite easily.

(*) Note of course for finite-dim IP spaces U, V, taking $T \in \mathcal{L}(U, V)$,

$$T^*T \in \mathcal{L}(U), \quad (T^*T)^* = T^*T,$$

thus T^*T self-adjoint, and furthermore, for $u \in U$,

$$\langle T^*Tu, u\rangle = \langle T^*(Tu), u\rangle = \langle Tu, Tu\rangle \geq 0,$$

and so T^*T is in fact positive, as we would hope.

Thus may define singular values of $T \in \mathcal{L}(U, V)$ by $\sigma(\sqrt{T^*T})$.

For more on the general SVD, see Horn and Johnson (1985, Ch. 7).

Spectral (or eigen-) decomposition

Let $A \in \mathbb{F}^{n \times n}$ be self-adjoint. We may then express A as

$$A = \sum_{i=1}^{n} \alpha_i v_i v_i^T,$$

where α_i are eigenvalues of A, with respective orthonormal eigenvectors v_i.

To see this:
Letting $T(z) := Az$, as T is self-adjoint, have orthonormal basis of eigenvectors $E := \{v_1, \ldots, v_n\}$. Let B be usual basis. Then,

$$A = M(T; B) = M(I; E, B) D M(I; B, E),$$

where D is diagonal, populated by eigenvalues α_i, and $M(I; E, B) = [v_1 \cdots v_n]$. Matrix multiplication yields our result.

The rest of the decompositions

The rest of the basic famous decompositions can be shown using typically algorithmic approaches. For example:

QR factorization
Any $A \in \mathbb{F}^{m \times n}$, can get $A = QR$, $Q \in \mathbb{F}^{m \times n}$ with orthonormal columns, $R \in \mathbb{F}^{n \times n}$ upper-tri.

Cholesky factorization
Any positive definite $A \in \mathbb{F}^{n \times n}$ may be factorized as $A = LL^*$, with L lower-tri with non-neg diagonal elements. Note $A = S^*S$ for some square S. Applying the QR result to S yields the result.

For QR and Cholesky, see Horn and Johnson (1985, Ch. 2).

For LU decomposition (a lot of technical details), see Horn and Johnson (1985, Ch. 3).

Lecture contents

1. Inner products: motivations, terms, and basic properties
2. Projections, orthogonal complements, and related problems
3. Linear functionals and the adjoint
4. Normal operators and the spectral theorem
5. Positive operators and isometries
6. **Some famous decompositions**

Lecture contents

1. **Inner products: motivations, terms, and basic properties**
2. **Projections, orthogonal complements, and related problems**
3. **Linear functionals and the adjoint**
4. **Normal operators and the spectral theorem**
5. **Positive operators and isometries**
6. **Some famous decompositions**

References

Axler, S. (1997). *Linear Algebra Done Right*. Springer, 2nd edition.

Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis*. Cambridge University Press, 1st edition.

Luenberger, D. G. (1968). *Optimization by Vector Space Methods*. Wiley.

Magnus, J. R. and Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. Wiley, 3rd edition.