Math 5760/6890: Introduction to Mathematical Finance
Review: linear algebra and differential equation

Akil Narayan[1]

[1] Department of Mathematics, and Scientific Computing and Imaging (SCI) Institute
University of Utah

Fall 2024

We've discussed the basics of finance and investing – concepts of interest and present value.

A more advanced understanding of pricing and policies requires some math:

- **linear algebra**
- **differential equations**
- probability

These topics are prerequisites for this course!

Vectors and matrices, I

Let $m, n \in \mathbb{N}$. ($m > n$, $m = n$, $m < n$ are allowed.)

We'll typically use lowercase boldface letters, e.g., $\boldsymbol{v}$, to denote *vectors*, elements of $\mathbb{R}^n$.
Vectors can be described by their components:

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \sum_{j=1}^{n} v_j \boldsymbol{e}_j \in \mathbb{R}^n, \qquad \boldsymbol{e}_j = \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}.$$

I.e., the components v_j are the *coordinates* of $\boldsymbol{v}$ in an expansion of the canonical vectors $\{\boldsymbol{e}_j\}_{j=1}^n$.

Vectors and matrices, I

Let $m, n \in \mathbb{N}$. ($m > n$, $m = n$, $m < n$ are allowed.)

We'll typically use lowercase boldface letters, e.g., $\boldsymbol{v}$, to denote *vectors*, elements of $\mathbb{R}^n$.
Vectors can be described by their components:

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \sum_{j=1}^{n} v_j \boldsymbol{e}_j \in \mathbb{R}^n, \qquad \boldsymbol{e}_j = \begin{pmatrix} \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}.$$

I.e., the components v_j are the *coordinates* of $\boldsymbol{v}$ in an expansion of the canonical vectors $\{\boldsymbol{e}_j\}_{j=1}^n$.

We'll use uppercase boldface letters, e.g., $\boldsymbol{A}$, to denote *matrices*, elements of $\mathbb{R}^{m \times n}$ that are also described by their components:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

Matrices are *linear maps* (functions) taking $\mathbb{R}^n$ to $\mathbb{R}^m$.

It is sometimes useful to consider vectors as specializations of matrices:

- If $n = 1$ and $m > 1$, then $\boldsymbol{A} \in \mathbb{R}^{m \times 1}$ is a *column vector*
- If $m = 1$ and $n > 1$, then $\boldsymbol{A} \in \mathbb{R}^{1 \times n}$ is a *row vector*

When considering vectors as specializations of matrices, we will assume that vectors are column vectors, unless otherwise indicated.

Example (Portfolio parameterization)

Suppose we have some initial amount of money, $V(0)$, that we wish to invest.

Suppose there are $N \in \mathbb{N}$ securities, which are financial products of which we can purchase a quantity.

The price (per unit) of security i at time t is given by $S_i(t)$.

The number of units we purchase of security i is n_i (can be non-integer).

The *weight* of our portfolio for the ith security is $w_i = n_i S_i(0)/V(0)$, which is the relative amount of worth we invest in security i.

We represent all these things as vectors:

$$\boldsymbol{S}(t) = \begin{pmatrix} S_1(t) \\ \vdots \\ S_N(t) \end{pmatrix} \in \mathbb{R}^N, \qquad \boldsymbol{n} = \begin{pmatrix} n_1 \\ \vdots \\ n_N \end{pmatrix} \in \mathbb{R}^N, \qquad \boldsymbol{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix} \in \mathbb{R}^N.$$

The vector $\boldsymbol{n}$ is the "trading strategy", and $\boldsymbol{w}$ is the (portfolio) "weight" vector.

The space of vectors $\mathbb{R}^n$ has *Euclidean* structure. One source of this structure comes from the notion of inner products: With $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, then the **inner product** of these vectors is

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \sum_{j=1}^{n} v_j w_j.$$

The inner product allows us to define lengths of vectors:

$$\|\boldsymbol{v}\| := \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \geqslant 0,$$

with $\|\boldsymbol{v}\| = 0$ iff $\boldsymbol{v} = 0$.

The space of vectors $\mathbb{R}^n$ has *Euclidean* structure. One source of this structure comes from the notion of inner products: With $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, then the **inner product** of these vectors is

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \sum_{j=1}^{n} v_j w_j.$$

The inner product allows us to define lengths of vectors:

$$\|\boldsymbol{v}\| := \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \geqslant 0,$$

with $\|\boldsymbol{v}\| = 0$ iff $\boldsymbol{v} = 0$.

From the definition, we observe that the inner product satisfies some key properties:

- *Symmetry*: $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{v} \rangle$.
- *Bilinearity*: $\langle a\boldsymbol{u} + b\boldsymbol{v}, \boldsymbol{w} \rangle = a \langle \boldsymbol{u}, \boldsymbol{w} \rangle + b \langle \boldsymbol{v}, \boldsymbol{w} \rangle$ for any $a, b \in \mathbb{R}$.

Angles

A useful concept that inner products provide is a measure of angles between vectors:

$$\theta := \angle(\boldsymbol{v}, \boldsymbol{w}), \qquad \cos\theta = \frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{\|\boldsymbol{v}\| \|\boldsymbol{w}\|}, \qquad \boldsymbol{v}, \boldsymbol{w} \neq \boldsymbol{0}.$$

In particular this allows us to define *orthogonal* vectors: $\boldsymbol{v}$ is orthogonal to $\boldsymbol{w}$ if $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$.

A useful concept that inner products provide is a measure of angles between vectors:

$$\theta := \angle(\boldsymbol{v}, \boldsymbol{w}), \qquad \cos\theta = \frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}, \qquad \boldsymbol{v}, \boldsymbol{w} \neq \boldsymbol{0}.$$

In particular this allows us to define *orthogonal* vectors: $\boldsymbol{v}$ is orthogonal to $\boldsymbol{w}$ if $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = 0$.

Why should $\frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}$ be a number between -1 and 1? Recall:

$$\left\langle \boldsymbol{v}, \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \right\rangle = \text{“Amount” of } \boldsymbol{v} \text{ pointing in the direction of } \boldsymbol{w}.$$

$$\left\langle \boldsymbol{v}, \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \right\rangle \boldsymbol{w} = \text{The projection of } \boldsymbol{v} \text{ onto } \boldsymbol{w}$$

A useful concept that inner products provide is a measure of angles between vectors:

$$\theta := \angle(\boldsymbol{v}, \boldsymbol{w}), \qquad \cos\theta = \frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}, \qquad \boldsymbol{v}, \boldsymbol{w} \neq \boldsymbol{0}.$$

In particular this allows us to define *orthogonal* vectors: $\boldsymbol{v}$ is orthogonal to $\boldsymbol{w}$ if $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = 0$.

Why should $\frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}$ be a number between -1 and 1? Recall:

$$\left\langle \boldsymbol{v}, \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right\rangle = \text{“Amount” of } \boldsymbol{v} \text{ pointing in the direction of } \boldsymbol{w}.$$

$$\left\langle \boldsymbol{v}, \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right\rangle \boldsymbol{w} = \text{The projection of } \boldsymbol{v} \text{ onto } \boldsymbol{w}$$

If the first expression is the “amount” of $\boldsymbol{v}$ pointing in a direction, then this “amount” shouldn't be larger than $\|\boldsymbol{v}\|$:

$$\left|\left\langle \boldsymbol{v}, \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right\rangle\right| \leqslant \|\boldsymbol{v}\| \quad \Longrightarrow \quad |\langle \boldsymbol{v}, \boldsymbol{w}\rangle| \leqslant \|\boldsymbol{v}\|\|\boldsymbol{w}\|$$

This is the **Cauchy-Schwarz** inequality. (Equality iff $\boldsymbol{v}$ is a scalar multiple of $\boldsymbol{w}$.)

Because of Cauchy-Schwarz, the quantity $\frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|} \in [-1, 1]$, so that it can be the cosine of some angle.

Example

With a portfolio weight vector $\boldsymbol{w}$, the trading strategy $\boldsymbol{n}$, the per-unit security price $\boldsymbol{S}(t)$, and the initial capital $V(0)$, we have the following relations:

$$\langle \boldsymbol{w}, \mathbf{1} \rangle = \sum_{j=1}^{N} w_j = 1.$$

$$\langle \boldsymbol{n}, \boldsymbol{S}(0) \rangle = \sum_{j=1}^{N} n_j S_j(0) = V(0)$$

Example

With a portfolio weight vector $\boldsymbol{w}$, the trading strategy $\boldsymbol{n}$, the per-unit security price $\boldsymbol{S}(t)$, and the initial capital $V(0)$, we have the following relations:

$$\langle \boldsymbol{w}, \boldsymbol{1} \rangle = \sum_{j=1}^{N} w_j = 1.$$

$$\langle \boldsymbol{n}, \boldsymbol{S}(0) \rangle = \sum_{j=1}^{N} n_j S_j(0) = V(0)$$

There is no restriction on the values of the weights w_i: they can be negative or greater than 1.

- $w_i > 0$ corresponds to purchasing units, with the intention to sell later (a long position)
- $w_i < 0$ corresponds to borrowing units and selling them now, with the intention to buy them back later ("short selling", a short position)

If there is no short selling, then $w_i \geqslant 0$, and hence $0 \leqslant w_i \leqslant 1$ for all i.

A core concept we'll need involves algebra on matrices, specifically *matrix multiplication*:

Given matrices $\boldsymbol{A} \in \mathbb{R}^{m \times \ell}$ and $\boldsymbol{B} \in \mathbb{R}^{\ell \times n}$, then the product $\boldsymbol{AB}$ is given by,

$$\boldsymbol{AB} \in \mathbb{R}^{m \times n}, \qquad (\boldsymbol{AB})_{j,k} = \sum_{q=1}^{\ell} A_{j,q} B_{q,k}$$

I.e., $(\boldsymbol{AB})_{j,k}$ is the inner product between the jth row of $\boldsymbol{A}$ and the kth row of $\boldsymbol{B}$.

Matrix multiplication is defined for matrices of *conforming* sizes, i.e., when the inner dimensions match.

Matrix multiplication is in general *not* commutative.

Matrix multiplication

A core concept we'll need involves algebra on matrices, specifically *matrix multiplication*:

Given matrices $\boldsymbol{A} \in \mathbb{R}^{m \times \ell}$ and $\boldsymbol{B} \in \mathbb{R}^{\ell \times n}$, then the product $\boldsymbol{AB}$ is given by,

$$\boldsymbol{AB} \in \mathbb{R}^{m \times n}, \qquad \qquad (\boldsymbol{AB})_{j,k} = \sum_{q=1}^{\ell} A_{j,q} B_{q,k}$$

I.e., $(\boldsymbol{AB})_{j,k}$ is the inner product between the jth row of $\boldsymbol{A}$ and the kth row of $\boldsymbol{B}$.

Matrix multiplication is defined for matrices of *conforming* sizes, i.e., when the inner dimensions match.

Matrix multiplication is in general *not* commutative.

Given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the *transpose* of $\boldsymbol{A}$ is the matrix $\boldsymbol{A}^T \in \mathbb{R}^{n \times m}$, formed by reflecting the entries of $\boldsymbol{A}$ across its main diagonal.

An inner product can be viewed as matrix multiplication:

$$\boldsymbol{v}^T \boldsymbol{w} = \langle \boldsymbol{v}, \boldsymbol{w} \rangle, \qquad \qquad \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n.$$

(Recall that when interpreting vectors $\boldsymbol{v} \in \mathbb{R}^n$ as matrices, we consider them as column vectors $\boldsymbol{v} \in \mathbb{R}^{n \times 1}$).

An *outer product* is another matrix multiplication between vectors, but this time when the inner dimension is 1:

$$\boldsymbol{v} = (v_1, \ldots, v_n)^T \in \mathbb{R}^n, \qquad \boldsymbol{w} = (w_1, \ldots, w_n)^T \in \mathbb{R}^n.$$

$$\boldsymbol{v}\boldsymbol{w}^T = \begin{pmatrix} | & | & & | \\ w_1\boldsymbol{v} & w_2\boldsymbol{v} & \cdots & w_n\boldsymbol{v} \\ | & | & & | \end{pmatrix} = \begin{pmatrix} — & v_1\boldsymbol{w}^T & — \\ — & v_2\boldsymbol{w}^T & — \\ & \vdots & \\ — & v_n\boldsymbol{w}^T & — \end{pmatrix} \in \mathbb{R}^{n\times n}$$

Linear independence, span, and basis, I

Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ be any collection of vectors, and let $\boldsymbol{V} \in \mathbb{R}^{n \times k}$ be the matrix whose columns are these vectors:

$$\boldsymbol{V} = \begin{pmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k \\ | & | & & | \end{pmatrix}$$

These vectors are **linearly dependent** if there exists a(ny) vector $\boldsymbol{c} \in \mathbb{R}^k$, $\boldsymbol{c} \neq \boldsymbol{0}$, such that,

$$\boldsymbol{V}\boldsymbol{c} = c_1 \boldsymbol{v}_1 + \ldots + c_k \boldsymbol{v}_k = \boldsymbol{0}.$$

Vectors that are *not* linearly dependent are **linearly independent**.

Vectors that are linearly dependent have a nontrivial linear relationship.
(If $\boldsymbol{0}$ is in the collection of vectors, the definition above implies they are linearly dependent.)

Linear independence, span, and basis, II

Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ be any collection of vectors, and let $\boldsymbol{V} \in \mathbb{R}^{n \times k}$ be the matrix whose columns are these vectors:

$$\boldsymbol{V} = \begin{pmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k \\ | & | & & | \end{pmatrix}$$

The **span** of these vectors is the collection of all linear combinations of these vectors:

$$\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} := \left\{ \boldsymbol{V}\boldsymbol{c} \;\middle|\; \boldsymbol{c} \in \mathbb{R}^k \right\}.$$

The span of vectors is a *linear/vector subspace*: it is a collection of vectors closed under addition and scalar multiplication.

Linear independence, span, and basis, III

Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ be any collection of vectors, and let $\boldsymbol{V} \in \mathbb{R}^{n \times k}$ be the matrix whose columns are these vectors:

$$\boldsymbol{V} = \begin{pmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k \\ | & | & & | \end{pmatrix}$$

Let S be some given vector subspace.

The vectors form a **basis** for S if the span of these vectors is S and they are linearly independent.

Linear independence, span, and basis, III

Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ be any collection of vectors, and let $\boldsymbol{V} \in \mathbb{R}^{n \times k}$ be the matrix whose columns are these vectors:

$$\boldsymbol{V} = \begin{pmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k \\ | & | & & | \end{pmatrix}$$

Let S be some given vector subspace.

The vectors form a **basis** for S if the span of these vectors is S and they are linearly independent.

In math: these vectors are a basis for S if

$$\forall\, \boldsymbol{w} \in S, \;\; \exists!\, \boldsymbol{c} \in \mathbb{R}^k \;\text{ such that }\; \boldsymbol{V}\boldsymbol{c} = \boldsymbol{w}.$$

(If $\boldsymbol{c}$ did not exist, the vectors wouldn't span S. If $\boldsymbol{c}$ weren't unique, then there would exist a nontrivial solution to $\boldsymbol{V}\boldsymbol{d} = \boldsymbol{0}$.)

Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ be any collection of vectors, and let $\boldsymbol{V} \in \mathbb{R}^{n \times k}$ be the matrix whose columns are these vectors:

$$\boldsymbol{V} = \begin{pmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k \\ | & | & & | \end{pmatrix}$$

Let S be some given vector subspace.

The vectors form a **basis** for S if the span of these vectors is S and they are linearly independent.

In math: these vectors are a basis for S if

$$\forall\, \boldsymbol{w} \in S, \quad \exists\,!\; \boldsymbol{c} \in \mathbb{R}^k \text{ such that } \boldsymbol{V}\boldsymbol{c} = \boldsymbol{w}.$$

(If $\boldsymbol{c}$ did not exist, the vectors wouldn't span S. If $\boldsymbol{c}$ weren't unique, then there would exist a nontrivial solution to $\boldsymbol{V}\boldsymbol{d} = \boldsymbol{0}$.)

A basis for S is not unique, but the size of a basis for S is unique.

This unique size of a basis for S is its **dimension**, $\dim S$.

If S contains m-dimensional vectors, then $\dim S \leqslant m$.

One particularly important application of linear algebra is as the theoretical and practical underpinning for solving linear equations for an unknown vector $\boldsymbol{x} \in \mathbb{R}^n$:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \qquad \boldsymbol{A} = \begin{pmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{m \times n}, \qquad \boldsymbol{b} \in \mathbb{R}^m.$$

To characterize solutions to such linear equations, consider the *range* or "column space" of $\boldsymbol{A}$, which is a subspace:

$$\operatorname{range}(\boldsymbol{A}) := \operatorname{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} \implies n \geqslant \dim \operatorname{range}(\boldsymbol{A})$$

One particularly important application of linear algebra is as the theoretical and practical underpinning for solving linear equations for an unknown vector $\boldsymbol{x} \in \mathbb{R}^n$:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \qquad \boldsymbol{A} = \begin{pmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{m \times n}, \qquad \boldsymbol{b} \in \mathbb{R}^m.$$

To characterize solutions to such linear equations, consider the *range* or "column space" of $\boldsymbol{A}$, which is a subspace:

$$\operatorname{range}(\boldsymbol{A}) := \operatorname{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} \implies n \geqslant \dim \operatorname{range}(\boldsymbol{A})$$

We can make very strong characterizations about solutions to linear systems:

1. If $\boldsymbol{b} \notin \operatorname{range}(\boldsymbol{A})$, then there is no solution $\boldsymbol{x}$.
2. If $\boldsymbol{b} \in \operatorname{range}(\boldsymbol{A})$ and $n > \dim \operatorname{range}(\boldsymbol{A})$ then there are infinitely many solutions $\boldsymbol{x}$, and the collection of these solutions form an *affine* space[1] of dimension $(n - \dim \operatorname{range}(\boldsymbol{A}))$.
3. If $\boldsymbol{b} \in \operatorname{range}(\boldsymbol{A})$ and $n = \dim \operatorname{range}(\boldsymbol{A})$, then there exists exactly one solution $\boldsymbol{x}$.

[1] An affine space is a subspace shifted by a fixed vector.

One particularly important application of linear algebra is as the theoretical and practical underpinning for solving linear equations for an unknown vector $\boldsymbol{x} \in \mathbb{R}^n$:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \qquad \boldsymbol{A} = \begin{pmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{m \times n}, \qquad \boldsymbol{b} \in \mathbb{R}^m.$$

To characterize solutions to such linear equations, consider the *range* or "column space" of $\boldsymbol{A}$, which is a subspace:

$$\mathrm{range}(\boldsymbol{A}) := \mathrm{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} \implies n \geqslant \dim \mathrm{range}(\boldsymbol{A})$$

We can make very strong characterizations about solutions to linear systems:

1. If $\boldsymbol{b} \notin \mathrm{range}(\boldsymbol{A})$, then there is no solution $\boldsymbol{x}$.
2. If $\boldsymbol{b} \in \mathrm{range}(\boldsymbol{A})$ and $n > \dim \mathrm{range}(\boldsymbol{A})$ then there are infinitely many solutions $\boldsymbol{x}$, and the collection of these solutions form an *affine* space[1] of dimension $(n - \dim \mathrm{range}(\boldsymbol{A}))$.
3. If $\boldsymbol{b} \in \mathrm{range}(\boldsymbol{A})$ and $n = \dim \mathrm{range}(\boldsymbol{A})$, then there exists exactly one solution $\boldsymbol{x}$.

NB: Situations 1 and 2 can happen for any relationship between n and m. Situation 3 can happen only if $m \geqslant n$.
The canonical algorithm to compute solutions to linear equations is Gaussian elimination.

[1] An affine space is a subspace shifted by a fixed vector.

Example

Recall that portfolio weights satisfy,

$$\langle \boldsymbol{w}, \mathbf{1} \rangle = 1.$$

This is equivalent to:

$$\boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}, \qquad \boldsymbol{A} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{1 \times N}, \qquad \boldsymbol{b} = 1 \in \mathbb{R}^1.$$

Example

Recall that portfolio weights satisfy,

$$\langle \boldsymbol{w}, \boldsymbol{1} \rangle = 1.$$

This is equivalent to:

$$\boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}, \qquad \boldsymbol{A} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{1 \times N}, \qquad \boldsymbol{b} = 1 \in \mathbb{R}^1.$$

In this case, the dimension of the range is $\dim \operatorname{range}(\boldsymbol{A}) = 1$ (and clearly $\boldsymbol{b} \in \operatorname{range}(\boldsymbol{A})$).

Hence, there are infinitely many valid portfolio weight vectors $\boldsymbol{w}$, and they form an affine space of dimension $N - \dim \operatorname{range}(\boldsymbol{A}) = N - 1$.

When $m = n$, consider the "square" linear system,

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \qquad \qquad \boldsymbol{A}, \boldsymbol{b} \text{ given}$$

There are some equivalent statements about a unique solution:

- There is a unique solution $\boldsymbol{x}$.
- The **rank** of $\boldsymbol{A}$, that is $\dim \operatorname{range}(\boldsymbol{A})$, has maximal value n.
- The *determinant* of $\boldsymbol{A}$ does not vanish: $\det \boldsymbol{A} \neq 0$.
- The matrix $\boldsymbol{A}$ has an *inverse* $\boldsymbol{A}^{-1}$, satisfying $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$.

When $m = n$, consider the "square" linear system,

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \qquad \boldsymbol{A}, \boldsymbol{b} \text{ given}$$

There are some equivalent statements about a unique solution:

- There is a unique solution $\boldsymbol{x}$.
- The **rank** of $\boldsymbol{A}$, that is $\dim \mathrm{range}(\boldsymbol{A})$, has maximal value n.
- The *determinant* of $\boldsymbol{A}$ does not vanish: $\det \boldsymbol{A} \neq 0$.
- The matrix $\boldsymbol{A}$ has an *inverse* $\boldsymbol{A}^{-1}$, satisfying $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$.

When any (hence all) of the above statements is true, then

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

is the unique solution.

Orthogonal matrices

Matrix inverses are generally "hard" to compute (analytically or numerically).

There is one class of matrices for which matrix inversion is rather simple:

A matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is **orthogonal** if its columns are (pairwise) orthgonal and unit norm:

$$\boldsymbol{A} = \begin{pmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{pmatrix}, \qquad \langle \boldsymbol{a}_j, \boldsymbol{a}_k \rangle = \delta_{j,k} := \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

Orthogonal matrices

Matrix inverses are generally "hard" to compute (analytically or numerically).

There is one class of matrices for which matrix inversion is rather simple:

A matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is **orthogonal** if its columns are (pairwise) orthgonal and unit norm:

$$\boldsymbol{A} = \begin{pmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{pmatrix}, \qquad \langle \boldsymbol{a}_j, \boldsymbol{a}_k \rangle = \delta_{j,k} := \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

A straightforward computation using matrix multiplication reveals:

$$\boldsymbol{A} \text{ orthogonal} \implies \boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{I} \implies \boldsymbol{A}^{-1} = \boldsymbol{A}^T.$$

Hence, orthogonality is a particularly useful practical property. (And $\boldsymbol{A}$ orthogonal implies $\boldsymbol{A}^{-1} = \boldsymbol{A}^T$ is also orthogonal.)

Orthogonal matrices

Matrix inverses are generally "hard" to compute (analytically or numerically).

There is one class of matrices for which matrix inversion is rather simple:

A matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is **orthogonal** if its columns are (pairwise) orthgonal and unit norm:

$$\boldsymbol{A} = \begin{pmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{pmatrix}, \qquad \langle \boldsymbol{a}_j, \boldsymbol{a}_k \rangle = \delta_{j,k} := \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

A straightforward computation using matrix multiplication reveals:

$$\boldsymbol{A} \text{ orthogonal} \implies \boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{I} \implies \boldsymbol{A}^{-1} = \boldsymbol{A}^T.$$

Hence, orthogonality is a particularly useful practical property. (And $\boldsymbol{A}$ orthogonal implies $\boldsymbol{A}^{-1} = \boldsymbol{A}^T$ is also orthogonal.)

Another useful property of orthogonal matrices: they correspond to *isometric* maps.

In particular, if $\boldsymbol{A}$ is orthogonal:

$$\langle \boldsymbol{A}\boldsymbol{v}, \boldsymbol{A}\boldsymbol{w} \rangle = \boldsymbol{v}^T \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{w} = \boldsymbol{v}^T \boldsymbol{I} \boldsymbol{w} = \boldsymbol{v}^T \boldsymbol{w} = \langle \boldsymbol{v}, \boldsymbol{w} \rangle.$$

I.e., the transformation $\boldsymbol{v} \mapsto \boldsymbol{A}\boldsymbol{v}$ *preserves angles and lengths*.
Orthogonal matrices are simple rotations and/or reflections.

Eigenvalues

For square matrices $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, an important concept is that of the *spectrum* of $\boldsymbol{A}$.

If there exists any (possibly complex-valued) scaled λ, and any non-zero vector $\boldsymbol{v}$ (possibly complex-valued) such that,

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v},$$

then

- λ is called an *eigenvalue* of $\boldsymbol{A}$
- $\boldsymbol{v}$ is called an *eigenvector* of $\boldsymbol{A}$.

For square matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, an important concept is that of the *spectrum* of $\boldsymbol{A}$.

If there exists any (possibly complex-valued) scaled λ, and any non-zero vector $\boldsymbol{v}$ (possibly complex-valued) such that,

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v},$$

then

- λ is called an *eigenvalue* of $\boldsymbol{A}$
- $\boldsymbol{v}$ is called an *eigenvector* of $\boldsymbol{A}$.

Fixing λ, note that the condition for being an eigenvector is invariant under addition of vectors and scalar multiplication: The set of eigenvectors associated to an eigenvalue λ is a subspace.

For square matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, an important concept is that of the *spectrum* of $\boldsymbol{A}$.

If there exists any (possibly complex-valued) scaled λ, and any non-zero vector $\boldsymbol{v}$ (possibly complex-valued) such that,

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v},$$

then

- λ is called an *eigenvalue* of $\boldsymbol{A}$
- $\boldsymbol{v}$ is called an *eigenvector* of $\boldsymbol{A}$.

Fixing λ, note that the condition for being an eigenvector is invariant under addition of vectors and scalar multiplication: The set of eigenvectors associated to an eigenvalue λ is a subspace.

Eigenvalues λ satisfy the characteristic equation:

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0,$$

so that eigenvalues are roots of a degree-n polynomial.

Matrix diagonalization

Every $n \times n$ matrix has exactly n eigenvalues (possibly repeated according to roots of the characteristic equation).

Matrix diagonalization

Every $n \times n$ matrix has exactly n eigenvalues (possibly repeated according to roots of the characteristic equation).

For each eigenvalue (counting multiplicity), there *may* be an eigenvector that is linearly independent from all others.

Matrices for which each eigenvalue has a corresponding linearly independent eigenvector are called *diagonalizable*.

If $\boldsymbol{A}$ is diagonalizable, then the following decomposition holds,

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}, \qquad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \qquad \boldsymbol{V} = \begin{pmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \\ | & | & & | \end{pmatrix},$$

where $(\lambda_j, \boldsymbol{v}_j)$ are eigenvalue-eigenvector pairs for $j = 1, \ldots, n$.

A particularly nice property about diagonalizable matrices is that the eigenvectors span $\mathbb{R}^n$ (possibly using complex scalar multiplication).

The upshot: if $\boldsymbol{A}$ is diagonalizable, then there is a linear transformation (defined by $\boldsymbol{V}$) such that multiplication by $\boldsymbol{A}$ corresponds to a simple diagonal scaling:

$$\boldsymbol{w} = \boldsymbol{A}\boldsymbol{x} \quad \xrightarrow{\boldsymbol{y}=\boldsymbol{V}^{-1}\boldsymbol{x},\, \boldsymbol{z}=\boldsymbol{V}^{-1}\boldsymbol{w}} \quad \boldsymbol{z} = \boldsymbol{\Lambda}\boldsymbol{y}.$$

Hence diagonalizations are *very* useful.

Orthogonal diagonalization

Diagonalizable matrices are diagonal under some transformation defined by $\boldsymbol{V}^{-1}$. But $\boldsymbol{V}^{-1}$ can be painful to compute.

Some matrices are *orthogonally* diagonalizable, meaning that $\boldsymbol{V}$ is an orthogonal matrix, and hence $\boldsymbol{V}^{-1}$ is “easy” to compute.

Orthogonal diagonalization

Diagonalizable matrices are diagonal under some transformation defined by $\boldsymbol{V}^{-1}$. But $\boldsymbol{V}^{-1}$ can be painful to compute.

Some matrices are *orthogonally* diagonalizable, meaning that $\boldsymbol{V}$ is an orthogonal matrix, and hence $\boldsymbol{V}^{-1}$ is "easy" to compute.

One of the major results of linear algebra is the following identification of one class of orthogonal matrices:

Theorem (Spectral theorem for symmetric matrices)

Assume $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ *satisfies* $\boldsymbol{A} = \boldsymbol{A}^T$.
(Such matrices are called symmetric.)

Then:

- *All eigenvalues of* $\boldsymbol{A}$ *are real-valued.*
- $\boldsymbol{A}$ *is orthogonally diagonalizable.*
  (The eigenvectors can be chosen as orthogonal vectors.)

Diagonalizable matrices are diagonal under some transformation defined by $\boldsymbol{V}^{-1}$. But $\boldsymbol{V}^{-1}$ can be painful to compute.

Some matrices are *orthogonally* diagonalizable, meaning that $\boldsymbol{V}$ is an orthogonal matrix, and hence $\boldsymbol{V}^{-1}$ is "easy" to compute.

One of the major results of linear algebra is the following identification of one class of orthogonal matrices:

Theorem (Spectral theorem for symmetric matrices)

Assume $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ satisfies $\boldsymbol{A} = \boldsymbol{A}^T$.
(Such matrices are called symmetric.)

Then:

- *All eigenvalues of $\boldsymbol{A}$ are real-valued.*
- *$\boldsymbol{A}$ is orthogonally diagonalizable.*
  (The eigenvectors can be chosen as orthogonal vectors.)

I.e.,

$$\boldsymbol{A} = \boldsymbol{A}^T \implies \boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T.$$

Since symmetric matrices have real-valued eigenvalues, then one can make sensible definitions about where the eigenvalues lie on $\mathbb{R}$.

In particular, the following are well-defined for λ_j the eigenvalues of an $n \times n$ symmetric matrix:

- $\lambda_{\min} = \min_{j=1,\ldots,n} \lambda_j$
- $\lambda_{\max} = \max_{j=1,\ldots,n} \lambda_j$

Equality above occurs iff $\boldsymbol{x}$ is a multiple of the eigenvalue corresponding to the minimum/maximum eigenvalue of $\boldsymbol{A}$.

Since symmetric matrices have real-valued eigenvalues, then one can make sensible definitions about where the eigenvalues lie on $\mathbb{R}$.

In particular, the following are well-defined for λ_j the eigenvalues of an $n \times n$ symmetric matrix:

- $\lambda_{\min} = \min_{j=1,\ldots,n} \lambda_j$
- $\lambda_{\max} = \max_{j=1,\ldots,n} \lambda_j$

Equality above occurs iff $\boldsymbol{x}$ is a multiple of the eigenvalue corresponding to the minimum/maximum eigenvalue of $\boldsymbol{A}$.

The spectral theorem also implies the following extremely useful inequality: if $\boldsymbol{A}$ is symmetric, then,

$$\lambda_{\min} \|\boldsymbol{x}\|^2 \leqslant \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \leqslant \lambda_{\max} \|\boldsymbol{x}\|^2.$$

(The function $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}$ is an example of a *quadratic form*.)

Symmetric + definite matrices

Two final definitions are sub-classes of symmetric matrices:

- If $\boldsymbol{A}$ is symmetric and all its eigenvalues are strictly positive, then $\boldsymbol{A}$ is (symmetric) **positive definite**.
- If $\boldsymbol{A}$ is symmetric and all its eigenvalues are non-negative, then $\boldsymbol{A}$ is (symmetric) **positive semidefinite**.

One can equivalently define these matrix classes through their quadratic forms.

In particular, $\boldsymbol{A}$ is symmetric positive semi-definite iff $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geqslant 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

Differential equations govern how quantities change in time.

One class of general ordinary differential equations (DE) governing the unknown function $y(t)$ where t is a scalar (i.e., time) is,

$$F\left(t, y, y', y'', y''', \ldots\right) = 0, \qquad y(0) = y_0, \qquad y'(0) = y'_0 \qquad \cdots.$$

This is an **initial value problem**. The maximum derivative appearing in F is called the *order* of the equation.

Linear differential equations

Differential equations govern how quantities change in time.

One class of general ordinary differential equations (DE) governing the unknown function $y(t)$ where t is a scalar (i.e., time) is,

$$F\left(t, y, y', y'', y''', \ldots\right) = 0, \qquad y(0) = y_0, \qquad y'(0) = y'_0 \qquad \cdots .$$

This is an **initial value problem**. The maximum derivative appearing in F is called the *order* of the equation.

Understanding the theory (solvability, well-posedness) of these problems is generally quite difficult, but *linear* equations are quite flexible for modeling and are rather well-understood.

Linear DE's are those where y, y', etc., collectively appear in F in a *linear* fashion.

The initial value problem,

$$y'(t) = 3y, \qquad\qquad y(0) = 4,$$

is, with some experience, rather transparent to solve:

$$y(t) = 4e^{3t}.$$

The initial value problem,

$$y'(t) = 3y, \qquad y(0) = 4,$$

is, with some experience, rather transparent to solve:

$$y(t) = 4e^{3t}.$$

The general solution of a first-order "constant-coefficient" linear equation can be determined in a similar fashion:

$$y' + \lambda y = f(t), \qquad y(0) = y_0,$$

has the unique solution

$$y(t) = y_0 e^{-\lambda t} + e^{-\lambda t} \int_0^t f(s) e^{\lambda s} \mathrm{d}s$$

Example

Consider an annuity (say a loan) with interest continuously compounded at (annual) rate r.

Suppose we also consider a repayment cycle that is *continuous* instead of periodic. I.e., we pay money at a (continuous) rate of P dollars (per year).

Example

Consider an annuity (say a loan) with interest continuously compounded at (annual) rate r.

Suppose we also consider a repayment cycle that is *continuous* instead of periodic. I.e., we pay money at a (continuous) rate of P dollars (per year).

The equation modeling the time-t present value $V(t)$ of this annuity is given by,

$$V'(t) = rV(t) - P, \qquad V(0) = V_0,$$

where V_0 is the loan principal (annuity present value at time 0).

Example

Consider an annuity (say a loan) with interest continuously compounded at (annual) rate r.

Suppose we also consider a repayment cycle that is *continuous* instead of periodic. I.e., we pay money at a (continuous) rate of P dollars (per year).

The equation modeling the time-t present value $V(t)$ of this annuity is given by,

$$V'(t) = rV(t) - P, \qquad V(0) = V_0,$$

where V_0 is the loan principal (annuity present value at time 0).

The solution to this equation is given by,

$$\begin{aligned} V(t) &= V_0 e^{rt} - e^{rt} \int_0^t P e^{-rs} \mathrm{d}s \\ &= V_0 e^{rt} + \frac{P}{r} \left[1 - e^{rt} \right] \end{aligned}$$

Example

Consider an annuity (say a loan) with interest continuously compounded at (annual) rate r.

Suppose we also consider a repayment cycle that is *continuous* instead of periodic. I.e., we pay money at a (continuous) rate of P dollars (per year).

The equation modeling the time-t present value $V(t)$ of this annuity is given by,

$$V'(t) = rV(t) - P, \qquad V(0) = V_0,$$

where V_0 is the loan principal (annuity present value at time 0).

The solution to this equation is given by,

$$\begin{aligned} V(t) &= V_0 e^{rt} - e^{rt} \int_0^t P e^{-rs} \mathrm{d}s \\ &= V_0 e^{rt} + \frac{P}{r} \left[1 - e^{rt}\right] \end{aligned}$$

Note in particular that this implies $V_0 < P/r$ is required in order for the loan to eventually be repaid.

The rather interesting part of this comes with *systems* of linear constant-coefficient differential equations:

$$\boldsymbol{y}'(t) = \boldsymbol{A}\boldsymbol{y}, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0,$$

where $\boldsymbol{y}(t) = (y_1(t), y_2(t), \ldots, y_n(t))^T$.

The rather interesting part of this comes with *systems* of linear constant-coefficient differential equations:

$$\boldsymbol{y}'(t) = \boldsymbol{A}\boldsymbol{y}, \qquad \boldsymbol{y}(0) = \boldsymbol{y}_0,$$

where $\boldsymbol{y}(t) = (y_1(t), y_2(t), \ldots, y_n(t))^T$.

If $\boldsymbol{A}$ is diagonalizable, $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$, then this system is the same as

$$\boldsymbol{z}'(t) = \boldsymbol{\Lambda}\boldsymbol{z}, \qquad \boldsymbol{z}(0) = \boldsymbol{V}^{-1}\boldsymbol{y}_0,$$

where $\boldsymbol{z}(t) := \boldsymbol{V}^{-1}\boldsymbol{y}(t)$, and easily solvable:

$$\boldsymbol{z}(t) = e^{\boldsymbol{\Lambda}t}\boldsymbol{V}^{-1}\boldsymbol{y}_0 \quad \Longrightarrow \quad \boldsymbol{y}(t) = \boldsymbol{V}e^{\boldsymbol{\Lambda}t}\boldsymbol{V}^{-1}\boldsymbol{y}_0.$$

Above, $e^{\boldsymbol{\Lambda}t} = (e^{\lambda_1 t},\ e^{\lambda_2 t}, \ldots, e^{\lambda_n t})^T$.

(If $\boldsymbol{A}$ is orthogonally diagonalizable, this is computationally even easier.)

References I