A History of Computational Linear Algebra

The Theory of Tables of Numbers through Time

Jesse He

OSU Reading Classics

16 February, 2020

Sheaves of wheat

Systems of simultaneous linear equations are known to have been computed as far back as ancient Mesopotamia, but we will begin with the first matrices.

Sheaves of wheat

Systems of simultaneous linear equations are known to have been computed as far back as ancient Mesopotamia, but we will begin with the first matrices. Consider the following problem:

Problem

"The yield of 3 sheaves of superior grain, 2 sheaves of medium grain, and 1 sheaf of inferior grain is 39 pieces of bread.
The yield of 2 sheaves of superior grain, 3 sheaves of medium grain, and 1 sheaf of inferior grain is 34 pieces of bread.
The yield of 1 sheaf of superior grain, 2 sheaves of medium grain, and 3 sheaves of inferior grain is 26 pieces of bread.
What is the yield of superior, inferior, and medium grains?"

Sheaves of wheat

With x representing superior grain, y representing medium grain, and z representing inferior grain, we have a system of linear equations:

$$\begin{aligned} x + 2y + \ \ z &= 39 \\ 2x + 3y + \ \ z &= 34 \\ 1x + 2y + 3z &= 26 \end{aligned}$$

which we can solve with Gaussian elimination.

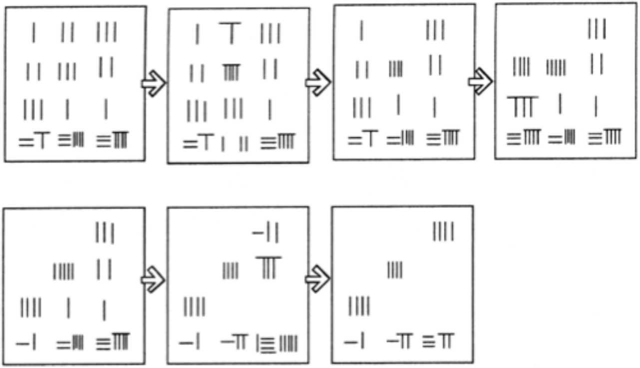
However, the Chinese mathematicians who wrote this problem did not have the benefit of Gauss's work, since this was published in the *Nine Chapters on the Art of Calculation* about 600 years before Gauss's birth.[6]

Sheaves of wheat

To solve, we simply represent this as a matrix and perform Gaussian elimination the usual way:

Sheaves of wheat

To solve, we simply represent this as a matrix and perform Gaussian elimination the usual way:



Adapted from Horng Wann-Sheng by Randy K. Schwartz [5]

Sheaves of wheat

In more familiar notation,

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 39 \\ 2 & 3 & 1 & 34 \\ 1 & 2 & 3 & 26 \end{array}\right] \to \left[\begin{array}{ccc|c} 3 & 2 & 1 & 39 \\ 0 & 5 & 1 & 26 \\ 1 & 2 & 3 & 26 \end{array}\right] \to \left[\begin{array}{ccc|c} 3 & 2 & 1 & 39 \\ 0 & 5 & 1 & 26 \\ 0 & 4 & 8 & 39 \end{array}\right] \to \left[\begin{array}{ccc|c} 3 & 2 & 1 & 39 \\ 0 & 5 & 1 & 26 \\ 0 & 0 & 36 & 99 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 3 & 2 & 1 & 39 \\ 0 & 5 & 1 & 26 \\ 0 & 0 & 4 & 11 \end{array}\right] \to \left[\begin{array}{ccc|c} 12 & 8 & 0 & 145 \\ 0 & 4 & 0 & 17 \\ 0 & 0 & 4 & 11 \end{array}\right] \to \left[\begin{array}{ccc|c} 4 & 0 & 0 & 37 \\ 0 & 4 & 0 & 17 \\ 0 & 0 & 4 & 11 \end{array}\right]$$

so $x = \frac{37}{4}$, $y = \frac{17}{4}$, $z = \frac{11}{4}$.

Gaussian elimination, 600 years before Gauss

The Chinese matrix (*fāngchěng*) method described in the *Nine Chapters* was used to solve equations in several unknowns, including a procedure for handling negative numbers and an exercise with 6 unknowns in 5 equations.

Gaussian elimination, 600 years before Gauss

The Chinese matrix (*fāngchěng*) method described in the *Nine Chapters* was used to solve equations in several unknowns, including a procedure for handling negative numbers and an exercise with 6 unknowns in 5 equations.

The method appeared later in Newton's notes in 1670, and methods of elimination and matrix reduction were developed in Europe throughout the 18th and 19th centuries including Gauss's contribution in 1811[2][1].

Determinate and indeterminate systems

Definition

A system of linear equations is said to be underspecified if there are fewer equations than unknowns.
If there are more equations than unknowns, the system is said to be overspecified.

Definition

A system of equations is determinate if there is a unique solution.
If there is more than one solution, it is indeterminate.

Determinate and indeterminate systems

Remark

An underspecified linear system must necessarily be indeterminate, but the converse may not hold.

Determinate and indeterminate systems

Remark

An underspecified linear system must necessarily be indeterminate, but the converse may not hold.

Under what conditions is a linear system of equations determinate?

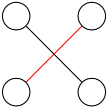
Discovery

The determinant of a matrix was used as a tool for solving simultaneous polynomial equations by Japanese mathematician Seki Kōwa in 1683[4], and introduced independently by Leibnitz in 1693[3]

Discovery

The determinant of a matrix was used as a tool for solving simultaneous polynomial equations by Japanese mathematician Seki Kōwa in 1683[4], and introduced independently by Leibnitz in 1693[3]

Seki illustrates the 2×2 case with the following picture:



where the red line corresponds to a positive ("constructive") product and the black to a negative ("destructive") product.

Laplace exapansion, half a century before Laplace

Seki describes the method for calculating determinants up to the 5×5 case in his *Fukudai* in 1683, and by 1710 Seki and his contemporaries had described a full, correct description of the Laplace expansion.

Laplace exapansion, half a century before Laplace

Seki describes the method for calculating determinants up to the 5×5 case in his *Fukudai* in 1683, and by 1710 Seki and his contemporaries had described a full, correct description of the Laplace expansion.

In Europe, Gabriel Cramer used the determinant in 1750 to describe closed-form solutions to systems of linear equations in terms of their coefficients (what we now call Cramer's rule), and other European mathematicians would continue to study determinants throughout the 18th and 19th centuries.

Singular and nonsingular matrices

Cramer noted that system whose determinant was zero did not have a unique solution, and moreover used his closed-form solutions to characterize when a system had zero, one, or infinitely many solutions.

Euler noticed in the same year that a system of n equations in n unknowns need not have a unique solution:

$$3x - 2y = 5$$
$$6x - 4y = 10$$

which led to the idea of "inclusive dependence" (what we now refer to as *linear dependence*)[2].

Algebra with tables of numbers

James Joseph Sylvester coined the term "matrix" in 1850, and his collegue Arthur Cayley published his *Treatise on the Theory of Matrices* where he developed a theory of matrix addition, multiplication, and inversion.

Algebra with tables of numbers

James Joseph Sylvester coined the term "matrix" in 1850, and his collegue Arthur Cayley published his *Treatise on the Theory of Matrices* where he developed a theory of matrix addition, multiplication, and inversion.

Let $\mathbb{F}$ be a field. For now, a matrix is a table of numbers, which we write $A = (a_{ij})$; a_{ij} is the entry in the i-th row and j-th column of A.

Algebra with tables of numbers

James Joseph Sylvester coined the term “matrix” in 1850, and his collegue Arthur Cayley published his *Treatise on the Theory of Matrices* where he developed a theory of matrix addition, multiplication, and inversion.

Let $\mathbb{F}$ be a field. For now, a matrix is a table of numbers, which we write $A = (a_{ij})$; a_{ij} is the entry in the i-th row and j-th column of A.

Given two $n \times m$ matrices $A = (a_{ij})$ and $B = (b_{ij})$, we write their sum $A + B = (a_{ij} + b_{ij})$.

Algebra with tables of numbers

James Joseph Sylvester coined the term "matrix" in 1850, and his collegue Arthur Cayley published his *Treatise on the Theory of Matrices* where he developed a theory of matrix addition, multiplication, and inversion.

Let $\mathbb{F}$ be a field. For now, a matrix is a table of numbers, which we write $A = (a_{ij})$; a_{ij} is the entry in the i-th row and j-th column of A.

Given two $n \times m$ matrices $A = (a_{ij})$ and $B = (b_{ij})$, we write their sum $A + B = (a_{ij} + b_{ij})$.

Given an $n \times m$ matrix A and an $m \times n$ matrix B, we write their product $AB = (\sum_{k=1}^{n} a_{ik} b_{kj})$.

Algebra with tables of numbers

James Joseph Sylvester coined the term “matrix” in 1850, and his collegue Arthur Cayley published his *Treatise on the Theory of Matrices* where he developed a theory of matrix addition, multiplication, and inversion.

Let $\mathbb{F}$ be a field. For now, a matrix is a table of numbers, which we write $A = (a_{ij})$; a_{ij} is the entry in the i-th row and j-th column of A.

Given two $n \times m$ matrices $A = (a_{ij})$ and $B = (b_{ij})$, we write their sum $A + B = (a_{ij} + b_{ij})$.

Given an $n \times m$ matrix A and an $m \times n$ matrix B, we write their product $AB = (\sum_{k=1}^{n} a_{ik} b_{kj})$.

Given a matrix A and a scalar λ, the scalar product is $\lambda A = (\lambda a_{ij})$.

Algebra with tables of numbers

The $n \times n$ identity matrix is the matrix

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

with the property that $AI = IA = A$ for any matrix A (where I is understood to be of the proper size).

The inverse A^{-1} of an $n \times n$ matrix A is the matrix such that $AA^{-1} = A^{-1} = A$.

Abstract lists of numbers

Frobenius identified a linear equation in n variables with its n-tuple of coefficients, and formulated a definition of linear dependence that applied to both linear equations and n-tuples.

Abstract lists of numbers

Frobenius identified a linear equation in n variables with its n-tuple of coefficients, and formulated a definition of linear dependence that applied to both linear equations and n-tuples.

Hamilton used the term "vector" (which he borrowed from physics) to describe the non-real parts of his quaternions.

Abstract lists of numbers

Frobenius identified a linear equation in n variables with its n-tuple of coefficients, and formulated a definition of linear dependence that applied to both linear equations and n-tuples.

Hamilton used the term "vector" (which he borrowed from physics) to describe the non-real parts of his quaternions.

Weierstrass and Kronecker treated the determinant rigorously around 1860 (as a "normed, linear, homogeneous function")[3].

Abstract lists of numbers

Frobenius identified a linear equation in n variables with its n-tuple of coefficients, and formulated a definition of linear dependence that applied to both linear equations and n-tuples.

Hamilton used the term "vector" (which he borrowed from physics) to describe the non-real parts of his quaternions.

Weierstrass and Kronecker treated the determinant rigorously around 1860 (as a "normed, linear, homogeneous function")[3].

For completeness's sake, I will include some basic definitions and results from linear algebra.

Vectors

Peano gave a formal definition of a vector space in 1880[3].

Vectors

Peano gave a formal definition of a vector space in 1880[3].

Definition

A vector space over a field $\mathbb{F}$ is a set V with operations addition $+ : V \times V \longrightarrow V$ and scalar multiplication $\bullet : \mathbb{F} \times V \longrightarrow V$ such that

- For all $u, v, w \in V$, $u + (v + w) = (u + v) + w$
- For all $u, v \in V$, $u + v = v + u$
- There exists $0_V \in V$ such that for all $v \in V$, $v + 0_V = v$
- For all $v \in V$, there exists $-v \in V$ so $v + (-v) = 0_V$
- For all $\lambda, \mu \in \mathbb{F}$, $v \in V$, $\lambda(\mu v) = (\lambda\mu)v$
- For all $v \in V$, $1_{\mathbb{F}} v = v$
- For all $\lambda \in \mathbb{F}$, $u, v \in V$, $\lambda(u + v) = \lambda u + \lambda v$
- For all $\lambda, \mu \in \mathbb{F}$, $v \in V$, $(\lambda + \mu)v = \lambda v + \mu v$

Vectors

Definition

A set $U \subseteq V$ is a linear subspace of V if U is a vector space under the operations on V.

Proposition

$U \subseteq V$ is a linear subspace of V if and only if U is closed under vector addition and scalar multiplication.

Linear combinations

Definition

A linear combination of a set $S \subseteq V$ is a finite sum

$$\sum_{v \in S} \alpha_v v$$

where $\alpha_v \in \mathbb{F}$ and only finitely many are nonzero.
The set of all linear combinations of S is the span of S.

Linear dependence and independence

Definition

A set $S \subseteq V$ is linearly dependent if there exist $(\alpha_v)_{v \in V}$, finitely many nonzero, not all zero, such that

$$0 = \sum_{v \in S} \alpha_v v$$

If S is not linearly dependent we say it is linearly independent.

Basis and dimension

Definition

A basis for a vector space V is a maximal linearly independent set $S \subseteq V$.
Equivalently, S is linearly independent and $\operatorname{span}(S) = V$.

Definition

The dimension of V is the cardinality of a basis for V.

Tables of numbers acting on lists of numbers

From this new perspective of vector spaces, the theory of matrices was recontextualized as the theory of linear maps between vector spaces.

Tables of numbers acting on lists of numbers

From this new perspective of vector spaces, the theory of matrices was recontextualized as the theory of linear maps between vector spaces.

Definition

A function $T : V \longrightarrow W$ between vector spaces is linear if for all $u, v \in V$, $\lambda, \mu \in \mathbb{F}$,

$$T(\lambda v + \mu u) = \lambda T(v) + \mu T(u).$$

Frequently we write Tv to mean $T(v)$.

Tables of numbers acting on lists of numbers

From this new perspective of vector spaces, the theory of matrices was recontextualized as the theory of linear maps between vector spaces.

Definition

A function $T : V \longrightarrow W$ between vector spaces is linear if for all $u, v \in V$, $\lambda, \mu \in \mathbb{F}$,

$$T(\lambda v + \mu u) = \lambda T(v) + \mu T(u).$$

Frequently we write Tv to mean $T(v)$. A linear map can be encoded as a matrix by choosing a basis for V and then writing the images of each basis vector in terms of the basis.

Kernels and images

Definition

Let U, V be vector spaces and let $T : V \longrightarrow W$ be a linear map. Then

$$\ker(T) = \{v \in V \mid Tv = 0_W\}, \ \operatorname{img}(T) = \{Tv | v \in V\}.$$

Proposition

$\ker(M)$ and $\operatorname{img}(M)$ are linear subspaces.

The characteristic polynomial

Cayley's work defining matrix operations, as well as the identity matrix I, allowed for a more algebraic treatment of matrices.

The characteristic polynomial

Cayley's work defining matrix operations, as well as the identity matrix I, allowed for a more algebraic treatment of matrices.

Definition

The characteristic polynomial of a square matrix M is $p(\lambda) = \det(M - \lambda I)$

The characteristic polynomial

Cayley's work defining matrix operations, as well as the identity matrix I, allowed for a more algebraic treatment of matrices.

Definition

The characteristic polynomial of a square matrix M is $p(\lambda) = \det(M - \lambda I)$

Cayley gave a proof of the Cayley-Hamilton theorem in 1858 (Frobenius gave a full proof in 1878):

Theorem (Cayley-Hamilton, 1858)

Any square matrix M satisfies its characteristic polynomial.

Eigenvalues, eigenvectors, and eigenspaces

Euler first observed "principal axes" when studying rotational motion, which Lagrange later noticed were eigenvectors:

Eigenvalues, eigenvectors, and eigenspaces

Euler first observed “principal axes” when studying rotational motion, which Lagrange later noticed were eigenvectors:

Definition

Let $T : V \longrightarrow V$ be a linear map. We say $\lambda \in \mathbb{F}$ is an eigenvalue of T if there is some $v \in V$, $v \neq 0_V$ such that

$$Tv = \lambda v.$$

The vector v is said to be an eigenvector of T associated with λ, and the set of all such vectors is called the eigenspace associated with λ.

Cauchy coined the term “characteristic root” for the idea of eigenvalues.

Eigenvalues and the characteristic polynomial

Proposition

Consider a matrix M as a linear transformation $M : V \longrightarrow V$. Then the eigenvalues of M are the roots of its characteristic polynomial.

Eigenvalues and the characteristic polynomial

Proposition

Consider a matrix M as a linear transformation $M : V \longrightarrow V$. Then the eigenvalues of M are the roots of its characteristic polynomial.

Definition

The set of eigenvalues of M is the spectrum of M, $\text{Spec}(M)$. For convenience, allow $\text{Spec}(M)$ to have multiple copies of the same element, accounting for multiplicity.

Eigenvalues and the characteristic polynomial

Proposition

Consider a matrix M as a linear transformation $M : V \longrightarrow V$. Then the eigenvalues of M are the roots of its characteristic polynomial.

Definition

The set of eigenvalues of M is the spectrum of M, $\mathrm{Spec}(M)$. For convenience, allow $\mathrm{Spec}(M)$ to have multiple copies of the same element, accounting for multiplicity.

Proposition

$$\det(M) = \prod_{\lambda \in \mathrm{Spec}(M)} \lambda$$

Eigenvalue algorithms

There are a number of algorithms to find the spectrum of a matrix:

Eigenvalue algorithms

There are a number of algorithms to find the spectrum of a matrix:

The naive algorithm is to simply expand the characteristic polynomial $p(\lambda) = \det(T - \lambda I)$ and solve for λ.

Eigenvalue algorithms

There are a number of algorithms to find the spectrum of a matrix:

The naive algorithm is to simply expand the characteristic polynomial $p(\lambda) = \det(T - \lambda I)$ and solve for λ.

Some algorithms are iterative, producing a sequence of values that converges to an eigenvalue, like Jacobi's algorithm, power iteration, or the QR algorithm.

Eigenvalue algorithms

There are a number of algorithms to find the spectrum of a matrix:

The naive algorithm is to simply expand the characteristic polynomial $p(\lambda) = \det(T - \lambda I)$ and solve for λ.

Some algorithms are iterative, producing a sequence of values that converges to an eigenvalue, like Jacobi's algorithm, power iteration, or the QR algorithm.

Others involve direct calculation, but there are no general (non-naive) algorithms to directly calculate the spectrum of an arbitrary matrix.

Power Iteration

One of the earliest computational algorithms for identifying eigenvalues is power iteration (sometimes von Mises algorithm).

Power Iteration

One of the earliest computational algorithms for identifying eigenvalues is power iteration (sometimes von Mises algorithm).

Given a matrix $M : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, generate a random vector $b_0 = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and repeatedly apply M, then renormalize:

$$b_{k+1} = \frac{Mb_k}{\|Mb_k\|}$$

Power Iteration

One of the earliest computational algorithms for identifying eigenvalues is power iteration (sometimes von Mises algorithm).

Given a matrix $M : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, generate a random vector $b_0 = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and repeatedly apply M, then renormalize:

$$b_{k+1} = \frac{Mb_k}{\|Mb_k\|}$$

Then if M has a largest eigenvalue λ with (unit) eigenvector v and $\langle b_0, v \rangle > 0$, the sequence (b_k) will converge to v, and the sequence of Rayleigh quotients

$$\mu_k = \frac{\langle b_k, Mb_k \rangle}{\|b_k\|} \xrightarrow[k \to \infty]{} \lambda.$$

Generalized eigenspaces

Definition

Let V be a vector space, let $T : V \longrightarrow V$ be a linear map, and let λ be an eigenvalue of T. The generalized eigenspace associated with λ is the subspace
$E_\lambda = \{v \in V \mid (T - \lambda I)^n v = 0_V \text{ for some } n \in \mathbb{N}\}$.

Theorem (Jordan Decomposition, 1870)

If V is finite-dimensional and $T : V \longrightarrow V$ has all its eigenvalues in $\mathbb{F}$, (i.e. its characteristic polynomial splits), then we can decompose V into its generalized eigenspaces

$$V = \bigoplus_{\lambda \in \mathrm{Spec}(T)} E_\lambda$$

Jordan canonical form

The Jordan canonical form expresses the spectrum and generalized eigenspaces of a transformation:

$$\mathrm{JCF}(T) = \begin{pmatrix} J_0 & 0 & \dots & 0 \\ 0 & J_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_n \end{pmatrix}$$

where $\mathrm{Spec}(T) = \{\lambda_0, \dots, \lambda_n\}$ and each Jordan block J_i is of the form

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i \end{pmatrix}$$

Finding Jordan normal form

The algorithm to find the Jordan normal form of a linear transformation T (assuming the characteristic polynomial splits) is, at a high level, as follows:

Finding Jordan normal form

The algorithm to find the Jordan normal form of a linear transformation T (assuming the characteristic polynomial splits) is, at a high level, as follows:

Calculate $\mathrm{Spec}(T)$.

Finding Jordan normal form

The algorithm to find the Jordan normal form of a linear transformation T (assuming the characteristic polynomial splits) is, at a high level, as follows:

Calculate $\mathrm{Spec}(T)$.

For each $\lambda \in \mathrm{Spec}(T)$, calculate the associated generalized eigenspace E_λ.

Finding Jordan normal form

The algorithm to find the Jordan normal form of a linear transformation T (assuming the characteristic polynomial splits) is, at a high level, as follows:

Calculate $\mathrm{Spec}(T)$.

For each $\lambda \in \mathrm{Spec}(T)$, calculate the associated generalized eigenspace E_λ.

Start with a basis for E_λ and iteratively find a Jordan basis for E_λ.

Finding Jordan normal form

The algorithm to find the Jordan normal form of a linear transformation T (assuming the characteristic polynomial splits) is, at a high level, as follows:

Calculate $\mathrm{Spec}(T)$.

For each $\lambda \in \mathrm{Spec}(T)$, calculate the associated generalized eigenspace E_λ.

Start with a basis for E_λ and iteratively find a Jordan basis for E_λ.

Gather all the Jordan bases for each E_λ into a Jordan basis for T.

The $J_0 \ker(S)$ trick

Computing the Jordan canonical form of a linear map can be computationally expensive, and the Jordan canonical form is very sensitive to small numerical changes.

The $J_0 \ker(S)$ trick

Computing the Jordan canonical form of a linear map can be computationally expensive, and the Jordan canonical form is very sensitive to small numerical changes.

To verify that the result of such a computation is correct, we can test each Jordan block, beginning with J_0:

The $J_0 \ker(S)$ trick

Computing the Jordan canonical form of a linear map can be computationally expensive, and the Jordan canonical form is very sensitive to small numerical changes.

To verify that the result of such a computation is correct, we can test each Jordan block, beginning with J_0:

Let $S = T - \lambda_0 I$. Then $E_{\lambda_0} = \ker(S^{n-1})$, so we apply the restriction of $T_{E_{\lambda_0}} = J_0$ to the kernel of S to verify that

$$J_0 \ker(S) = (J_0)^2 \ker(S^2) = \cdots = (J_0)^n \ker(S^n) = \{0_V\}$$

and then we can repeat this "$J_0 \ker(S)$ trick" for each Jordan block.

- Intern et al.
- Christine Andrews-Larson. "Roots of Linear Algebra: An Historical Exploration of Linear Systems". In: *PRIMUS* (Apr. 2015).
- Israel Kleiner. *A History of Abstract Algebra*. Birkhäuser Boston, 2007.
- Yoshio Mikami. "On the Japanese Theory of Determinants". In: *Isis* Vol. 2.1 (1914).
- Randy K. Schwartz. "A Classic from China: The Nine Chapters - Matrices". In: *Convergence* (Dec. 2018).
- B. L. van der Waerden. *Geometry and Algebra in Ancient Civilizations*. New York, New York: Springer-Verlag, 1983.