5 Matrix Inversion

To summarize the developments so far, matrix techniques, including inversion, are used in systems of linear algebraic equations and co-ordinate transformations, ranging from eigenvalue problems, changes of integration variables (Jacobians), and vector analysis.

For solutions of large systems of equations, matrix inversion by Cramer's rule is extraordinarily inefficient. We seek efficient computation algorithms and highlight two here. Throughout, we can check results using `Mathematica`.

5.1 Gauss-Jordan Elimination

This technique is to simultaneously perform a sequence of linear operations on the rows and columns of the matrix $\mathcal{A}$ to be inverted, and the identity matrix $\mathcal{I}$, so as to sequentially reduce $\mathcal{A}$ to $\mathcal{I}$. In the process, $\mathcal{I}$ becomes $\mathcal{A}^{-1}$. Each of these linear operations (e.g. adding and subtracting rows or columns, multiplication by scalars) is tantamount to multiplication by a matrix $\mathcal{O}_j$, so that we have the effective protocol

A&W pp. 170-2

$$\mathcal{I} \;=\; \left(\prod_{j=1}^{m} \mathcal{O}_j \right) \mathcal{A} \quad \Leftrightarrow \quad \mathcal{A}^{-1} \;=\; \left(\prod_{j=1}^{m} \mathcal{O}_j \right) \mathcal{I} \quad . \tag{36}$$

The equivalence of these two forms is obvious. The elimination aspect of the method is the objective of generating zeros for off-diagonal elements of the matrix. We illustrate the Gauss-Jordan technique by example.

• This approach generates of the <u>order of a few times n^3 computations</u> for an $n \times n$ matrix, contrasting Cramer's rule that involves the <u>order of $n!$ operations</u>. Accordingly, for $n \geq 4$, the Gauss-Jordan process is clearly significantly more efficient.

• **Example 5:** To illustrate the Gauss-Jordan protocol we invert a 3×3 matrix, starting with the following pair:

$$\mathcal{A} \equiv \begin{pmatrix} \pi & e & \pi \\ -e & \pi & e \\ \pi & -e & \pi \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad .$$

First subtract the first row from the third and add e times the first row to π times the second:

$$\begin{pmatrix} \pi & e & \pi \\ 0 & \pi^2+e^2 & 2e\pi \\ 0 & -2e & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 \\ e & \pi & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad .$$

Now multiply the first row by $2e$ and subtract from this the second row:

$$\begin{pmatrix} 2e\pi & e^2-\pi^2 & 0 \\ 0 & \pi^2+e^2 & 2e\pi \\ 0 & -2e & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} e & -\pi & 0 \\ e & \pi & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad .$$

Multiply the first row by $2e$ and then add to it $e^2-\pi^2$ times the third row:

$$\begin{pmatrix} 4e^2\pi & 0 & 0 \\ 0 & \pi^2+e^2 & 2e\pi \\ 0 & -2e & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} e^2+\pi^2 & -2e\pi & e^2-\pi^2 \\ e & \pi & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad .$$

Multiply the third row by π^2+e^2 and add to this $2e$ times the second row:

$$\begin{pmatrix} 4e^2\pi & 0 & 0 \\ 0 & \pi^2+e^2 & 2e\pi \\ 0 & 0 & 4e^2\pi \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} e^2+\pi^2 & -2e\pi & e^2-\pi^2 \\ e & \pi & 0 \\ e^2-\pi^2 & 2e\pi & e^2+\pi^2 \end{pmatrix} \quad .$$

Finally, multiply the second row by $2e$ and subtract the third row, to yield a diagonal matrix on the left. Follow this by dividing rows 1 and 3 by $4e^2\pi$ and row 2 by $2e(e^2+\pi^2)$ to get

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathcal{A}^{-1} = \frac{1}{2e} \begin{pmatrix} \dfrac{e^2+\pi^2}{2e\pi} & -1 & \dfrac{e^2-\pi^2}{2e\pi} \\ 1 & 0 & -1 \\ \dfrac{e^2-\pi^2}{2e\pi} & 1 & \dfrac{e^2+\pi^2}{2e\pi} \end{pmatrix}$$

as the inverse of $\mathcal{A}$.

5.2 LU Decomposition

This approach is to render the matrix $\mathcal{A}$ as a product of **lower triangular** ($\mathcal{L}$, where all elements above the diagonal are zero) and **upper triangular** ($\mathcal{U}$, where all elements below the diagonal are zero) matrices. This decomposition then can be solved as a two sequential systems of simultaneous equations involving the elements of $\mathcal{L}$ and $\mathcal{U}$. We illustrate by example, for a system of linear simultaneous equations.

- This technique involves of the order of $n^3/2$ computations when employing *Crout's algorithm* for defining the elements of each matrix, which is of the same order, but a factor of a few faster than Gauss-Jordan elimination.

- **Example 6:** For a simple example of LU decomposition, we consider

$$\begin{aligned} \mathcal{A} &= \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \alpha_{21} & 1 \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} \\ 0 & \beta_{22} \end{pmatrix} \\ &= \begin{pmatrix} \beta_{11} & \beta_{12} \\ \alpha_{21}\beta_{11} & \alpha_{21}\beta_{12} + \beta_{22} \end{pmatrix} . \end{aligned} \tag{37}$$

From this one can simply determine that $\beta_{11} = 1$ and $\beta_{12} = 2$, and then that $\alpha_{21} = 3/\beta_{11} = 3$ and $\beta_{22} = 4 - \alpha_{21}\beta_{12} = -2$, in an order where all needed elements are <u>determined prior to use</u>. It then follows that

$$\mathcal{L} = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \quad \text{and} \quad \mathcal{U} = \begin{pmatrix} 1 & 2 \\ 0 & -2 \end{pmatrix} . \tag{38}$$

Observe that the LU decomposition is such that $Det(\mathcal{L}) = 1$. The inversion is now routinely achieved, *e.g. using Gauss-Jordan elimination*:

$$\mathcal{A}^{-1} = \mathcal{U}^{-1}\mathcal{L}^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} . \tag{39}$$

This result can be checked with `Mathematica`. Observe that $\mathcal{U}^{-1}$ and $\mathcal{L}^{-1}$ are upper and lower triangular matrices, respectively, like their inverses.

6 Four-Vectors

Tensors define a sophisticated mathematical formalism for dealing with coordinate transformations and their impact on physical quantities. As prefatory material, we consider their simplest version, **4-vectors**, which appear naturally in the spacetime manipulations of Einstein's theory of special relativity.

6.1 Four-Vectors in Spacetime

Start with the spacetime coordinate position **four-vector** x^μ, defined by

$$x^\mu \;=\; (x^0 = ct,\, x^1 = x,\, x^2 = y,\, x^3 = z) \quad . \tag{40}$$

Here the use of Greek indices denotes four-vectors. The **magnitude** of this 4D four-vector is given by the line-element length squared

$$s^2 \;\equiv\; x_\mu x^\mu \;=\; (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \quad . \tag{41}$$

This is an invariant under Lorentz transformations, and we therefore term it a **scalar**. The product notation resembles the dot product in 4D if we define

$$x_\mu \;=\; (x_0 = ct,\, x_1 = -x,\, x_2 = -y,\, x_3 = -z) \quad . \tag{42}$$

Both x^μ and x_μ are two forms representing the same quantity. We call x^μ the **contravariant** form and x_μ the **covariant** version. The dot product form then obeys the **Einstein summation convention**, namely that

$$x_\mu x^\mu \;\equiv\; x^\mu x_\mu \;\rightarrow\; \sum_{\mu=0}^{3} x^\mu x_\mu \quad ; \tag{43}$$

all indices appearing twice (μ here) are summed over. This is also termed the **scalar product** of the contravariant and covariant spacetime four-vectors.

The Lorentz transformation has the property that it preserves the magnitude of differential spacetime elements, dx^μ:

$$dx_\mu dx^\mu \;=\; dx'_\nu dx'^\nu \quad \text{with} \quad dx^\mu \;=\; \frac{\partial x^\mu}{\partial x'^\nu}\, dx'^\nu \quad . \tag{44}$$

This constraint is imposed by the fixed value of the speed of light in all inertial reference frames, the *fundamental premise of special relativity*.

For simplicity, consider velocity **boosts** of magnitude βc in the x^1 direction, with $|\beta| < 1$. Given the form of Eq. (41), a single parameter description of boost invariance is provided by hyperbolic functions, with $x^0 \propto \cosh\eta$ and $x^1 \propto \sinh\eta$, and $\beta = \tanh\eta$. Here η is called the **rapidity** of the Lorentz transformation, and we can define the **Lorentz factor** of the boost to be $\gamma \equiv 1/\sqrt{1-\beta^2} = \cosh\eta$. One way to facilitate the Lorentz transformation algebra is via a matrix construction. Representing the 4-vectors as column matrices, then the boost equations can be compactly expressed via

$$\begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{bmatrix} (x^0)' \\ (x^1)' \\ (x^2)' \\ (x^3)' \end{bmatrix} \quad . \tag{45}$$

The covariant form of this can be quickly cast in terms of row-vectors. It naturally must involve a matrix describing the inverse Lorentz transformation matrix, which sends $\beta \to -\beta$, and this path can be used to establish the Lorentz invariance $x_\mu x^\mu = (x_\mu)' \, (x^\mu)'$.

• We can extend this structure to other four-vectors and arrive at a Lorentz invariance scalar product if we ascribe to the four-vector the same Lorentz transformation properties that are satisfied by the spacetime coordinates under boosts. This now *becomes a matter of definition*: we <u>define a four-vector</u> A^μ to be a 4D vector that obeys (*contravariant form*)

$$\begin{gathered} A^0 \;=\; \gamma\Big[(A^0)' + \beta(A^1)'\Big] \quad , \quad A^1 \;=\; \gamma\Big[(A^1)' + \beta(A^0)'\Big] \\ A^2 \;=\; (A^2)' \quad , \quad A^3 \;=\; (A^3)' \quad . \end{gathered} \tag{46}$$

This applies specifically for boosts in the x^1 direction, but can be routinely generalized to arbitrary boosts. The *covariant* form is obtained from

$$A_0 \;=\; A^0 \quad , \quad A_1 \;=\; -A^1 \quad , \quad A_2 \;=\; -A^2 \quad , \quad A_3 \;=\; -A^3 \quad , \tag{47}$$

and automatically obeys the Lorentz transformations. Thus,

$$A_\mu A^\mu \;\equiv\; A^\mu A_\mu \;=\; \sum_{\mu=0}^{3} A^\mu A_\mu \;=\; (A_\mu)' \, (A^\mu)' \tag{48}$$

is a Lorentz invariant, as is easily shown algebraically from Eq. (46).

Eq. (48) expresses the square of the magnitude, or **length** of the spacetime 4-vector A^μ. It does not have to be real. If it is real, i.e. $A_\mu A^\mu$ is positive, then this 4-vector is called *timelike*, contrasting $A_\mu A^\mu < 0$ cases where A^μ is termed *spacelike*. A 4-vector that is of zero length is called a **null vector**.

The **inner product** or dot product of two four vectors A^μ and B^μ is simply defined by the summation

$$A_\mu B^\mu \equiv \sum_{\mu=0}^{3} A_\mu B^\mu \quad , \tag{49}$$

and can be positive or negative.

- The Lorentz transformation matrix is just a Jacobian:

$$\mathbf{\Lambda}^\mu{}_\nu \equiv \frac{\partial x^\mu}{\partial x'^\nu} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{bmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 \\ \Lambda^2{}_0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ \Lambda^3{}_0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{bmatrix} \quad . \tag{50}$$

Observe that the first index, μ*, in* $\mathbf{\Lambda}^\mu{}_\nu$ *corresponds to the matrix* <u>*row*</u> *number, while the second,* ν*, marks the* <u>*column*</u> *number.* This is established via the construction in Eq. (45). One can then apply this matrix protocol to sequential Lorentz boosts since the chain rule for differentiation is operative for the two coordinate transformation functions $x^\mu(x'^\sigma)$ and $x'^\sigma(x''^\nu)$:

$$dx^\mu = \frac{\partial x^\mu}{\partial x'^\nu}\, dx'^\nu = \frac{\partial x^\mu}{\partial x'^\sigma}\, \frac{\partial x'^\sigma}{\partial x''^\nu}\, dx''^\nu \tag{51}$$

(remember that the Einstein summation convention applies here) which corresponds to a composite boost:

$$\mathbf{\Lambda}_c \;\to\; \mathbf{\Lambda}^\mu{}_\sigma \, \mathbf{\Lambda}^\sigma{}_\nu \quad . \tag{52}$$

This matrix manipulation technique appears extremely useful for 4-vectors: we will find it also so for tensors, our incipient focus.

- Observe that the Lorentz transformation matrix $\mathbf{\Lambda}^\mu_\nu$ is *symmetric* and *has a determinant equal to unity.* This unit determinant is a general property of boost matrices, because *all such boosts constitute a sequence of rotations in spacetime*, and <u>all rotations preserve volume elements and solid angles</u>.

6.2 Four-velocity

To serve as an immediate example of a four-vector, we generalize the ordinary 3D velocity $\mathbf{v}$ to define the **four-velocity** u^μ of a particle via

$$u^\mu \;=\; \frac{dx^\mu}{ds} \quad \text{for} \quad ds = c\,dt\,\sqrt{1-\frac{v^2}{c^2}} \;=\; cd\tau \quad . \tag{53}$$

Here $\tau \equiv t'$ is the proper time (a scalar), evaluated in the inertial frame in which the particle is at rest. Therefore the Lorentz factor of the particle is

$$\gamma \;=\; \frac{1}{\sqrt{1-v^2/c^2}} \;\equiv\; \cosh\eta \quad , \tag{54}$$

so that $ds = cdt/\gamma$. The x-component of the four velocity is

$$u^1 \;=\; \frac{dx}{c\,dt\,\sqrt{1-v^2/c^2}} \;=\; \gamma\,\frac{v_x}{c} \quad . \tag{55}$$

Here the v pertains to the *total* particle speed, not just v_x , since the time dilation factor is pertinent to the boost to the particle rest frame. Similar results ensue for the other velocity components. In compact form,

$$u^\mu \;=\; \gamma\Big(1,\, \beta\,\hat{v}\Big) \;\equiv\; \gamma\big(1,\, \boldsymbol{\beta}\big) \quad , \quad \beta \;=\; \frac{v}{c} \quad , \tag{56}$$

with $\hat{v} = \mathbf{v}/v$ as the unit vector in the direction of the velocity of the particle, as measured in the observer's frame.

- Since $dx_\alpha dx^\alpha = ds^2$, one simply has

$$u_\mu u^\mu \;\equiv\; \frac{dx_\mu dx^\mu}{ds^2} \;=\; 1 \quad , \tag{57}$$

i.e. the 4-velocity always has unit length that is therefore a Lorentz invariant. This just expresses the constancy of the speed of light in yet another form. Consequently, *the components of the four-velocity are not independent.*

∗ The "normalization condition" $u_\mu u^\mu = 1$ extends also to general relativity, where one generalizes the covariant form $u_\mu = g_{\mu\nu} u^\nu$ for $g_{\mu\nu} \neq \eta_{\mu\nu}$, where $g_{\mu\nu}$ and $\eta_{\mu\nu}$ are **metric tensors** that we will encounter shortly.