Discovering Matrix Inversion Through Transport Theory

Mónica N. Burgos Chaparro and Dr. Rocío Díaz Martín

Florida State University (FSU), Department of Mathematics, Tallahassee, FL 32304

FSU | FLORIDA STATE UNIVERSITY

FSU UNDERGRADUATE RESEARCH OPPORTUNITY PROGRAM
CENTER FOR UNDERGRADUATE RESEARCH & ACADEMIC ENGAGEMENT

MOTIVATION

Matrices, what are they?

Matrices are mathematical objects that organize numbers into rows and columns, providing a structured way to represent and solve systems of linear equations. They are widely used across disciplines such as engineering, physics, machine learning, and neuroscience because they model relationships compactly and efficiently.

Beyond simply representing data, matrices follow their own algebraic rules. For real numbers, recovering a missing factor is straightforward:

if 3x=1, then x=1/3.

For matrices, however, this process is more involved. The analogous operation is computing the inverse of a matrix.

$$\begin{cases} \frac{13}{2}x + \frac{5}{2}y = 1 \\ \frac{5}{2}x + \frac{13}{2}y = 0 \end{cases} \longrightarrow \begin{bmatrix} \frac{13}{2} & \frac{5}{2} \\ \frac{5}{2} & \frac{13}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \longrightarrow \text{Solution: } \begin{bmatrix} x \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{13}{2} & \frac{5}{2} \\ \frac{5}{2} & \frac{13}{2} \end{bmatrix}^{-1}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Figure 1: Matrix representation of a system of 2 equations with 2 unknown variables.

Many methods exist to compute the inverse of a matrix. While these approaches are effective, we are curious whether alternative methods can emerge by connecting matrix inversion to areas beyond purely matrix algebra.

We propose focusing on symmetric positive definite matrices, which can be interpreted as covariance matrices of probability distributions, such as multivariate Gaussian distributions. This probabilistic viewpoint allows us to use Optimal Transport as a bridge between matrices and probability measures.

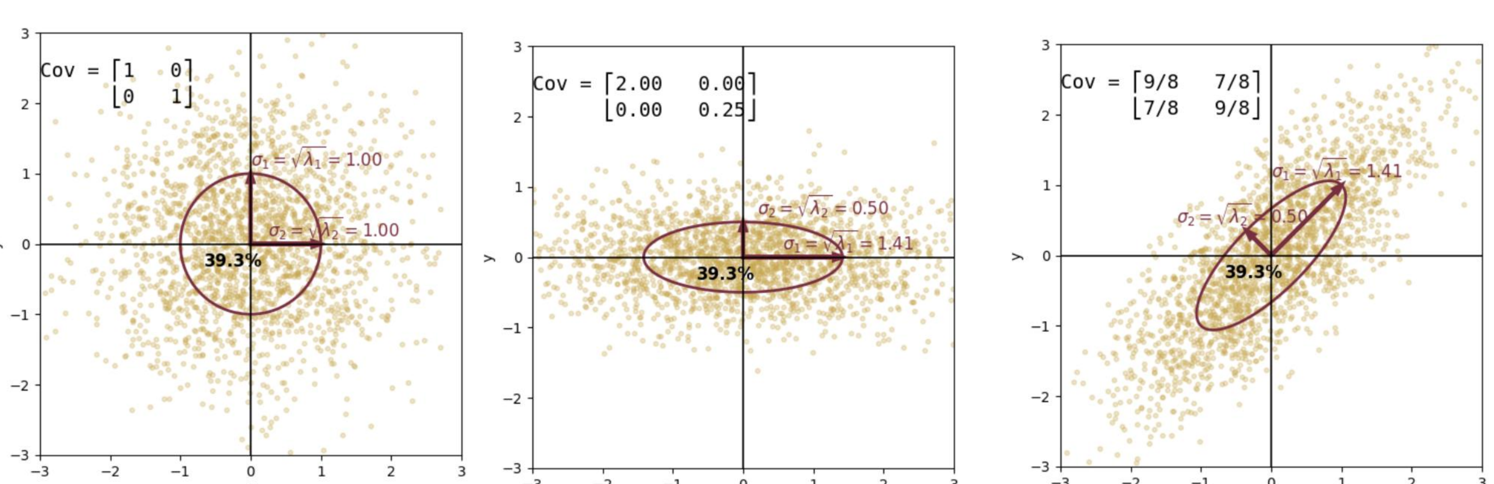


Figure 2: Samples from 3 different 2D Gaussians with different covariance matrices.

Optimal Transport provides a mathematical framework for comparing mass distributions by measuring the "cheapest cost" required to move, reshape, and redistribute mass so that one distribution matches another.

Figure 3: Optimal deformation of the FSU spear into the FSU logo.

Through this connection, we reformulate matrix inversion as multiple 1D transport problems, relying primarily on simple vector projections.

REFERENCES & ACKNOWLEDGMENTS

BACKGROUND

Linear Algebra:

Finding the inverse of a matrix isn't possible for all matrices; certain rules and properties have to be met for a matrix to be invertible. Common matrices that are simple to invert are:

- Diagonal Matrices with non-zero diagonal → Their inverse is equal to the diagonal matrix with reciprocal entries in the diagonal
- Rotations & Reflections (respectively) → Their inverse is equal to their transpose.

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} , \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

Singular Value Decomposition (SVD): Allows a matrix to be made up of two rotation or reflection matrices and a diagonal matrix. It looks like: $A = U\,D\,V^T$

If our chosen matrix A is invertible, then D does not have zero entries in the diagonal (σ_i) and

$$A = U\,D\,V^T \rightarrow (A)^{-1} = (U\,D\,V^T)^{-1} \rightarrow A^{-1} = V\,D^{-1}U^T$$

In this project, we consider **symmetric positive definite matrices (SPD)**, which are of the form

$$\Sigma = AA^T = (U\,D\,V^T)(V\,D\,U^T) = U\,D^2U^T \quad \text{for } A \text{ invertible.}$$

- (Positive Definite) All the eigenvalues of S are positive, $\sigma^2 = \lambda > 0$.
- (Symmetric) Column vectors of Σ can be changed to row vectors and it won't change; $\Sigma = \Sigma^T$.

Probability Theory:

Consider 2D Gaussian distributions, with Covariances Σ (SPD matrix) and the Identity matrix

Figure 4: 2D Gaussians with covariance matrices Σ and the identity matrix. Goal: math cross-section by cross-section.

$$F(\boldsymbol{x}) = \frac{1}{2\pi(det\Sigma)^{1/2}} e^{\left(-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}\right)} \quad \text{both with total mass =1} \quad G(\boldsymbol{x}) = \frac{1}{2\pi} e^{\left(-\frac{1}{2}||\boldsymbol{x}||^2\right)}.$$

Projecting F and G onto 1D eigendirection gives us two 1D Gaussian distributions

$$f(x) = \frac{1}{\sqrt{2\pi\lambda}} e^{\left(\frac{-x^2}{2\lambda}\right)} \quad \text{and} \quad g(x) = \frac{1}{\sqrt{2\pi}} e^{\left(\frac{-x^2}{2}\right)}$$

Optimal Transport:

Optimization Problem:

Find $\hbar$ that minimizes the total cost of moving and preserving mass from f to g.

Theorem: There exists one solution for $\hbar$ characterized by:

$$\int_{-\infty}^{x} f(s)ds = \int_{-\infty}^{h(x)} g(s)ds \qquad \text{for all points x}$$

Now, we must solve for $\hbar$(x) in our case to find the optimal transport map:

$$\int_{-\infty}^{x} \frac{1}{(\sqrt{2\pi})\sqrt{\lambda}} e^{-\frac{1}{2}\left(\frac{s}{\sqrt{\lambda}}\right)^2} ds = \int_{-\infty}^{h(x)} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}s^2\right)} ds$$

We need a change of variables, we propose $y = \frac{s}{\sqrt{\lambda}}$; $dy = \frac{1}{\sqrt{\lambda}} ds$

$$\int_{-\infty}^{x} \frac{1}{(\sqrt{2\pi})\sqrt{\lambda}} e^{-\frac{1}{2}\left(\frac{s}{\sqrt{\lambda}}\right)^2} ds = \int_{-\infty}^{x/\sqrt{\lambda}} \frac{1}{(\sqrt{2\pi})} e^{\left(-\frac{1}{2}y^2\right)} dy$$

Thus, $\hbar$ becomes that change of variables: $\hbar(x) = \frac{x}{\sqrt{\lambda}}$

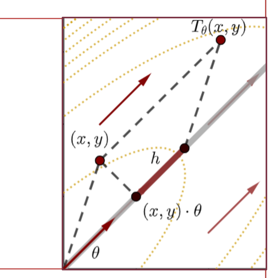
This is, ultimately, the foundation for our algorithm.

METHODOLOGY: THE ALGORITHM

Case 1: Eigenvectors (θ) and eigenvalues (λ) of the d x d SPD matrix Σ are known

1. Project every **x** = (x, y) onto the direction of θ_j.

2. Match the Gaussian spread of $\sigma = \sqrt{\lambda_j}$ with the Gaussian spread of σ = 1.
The optimal transport map is $\hbar$ from previous section with $\lambda = \lambda_j$.
The optimal displacement is $\left(\frac{(x,y)}{\sqrt{\lambda_j}} - (x,y)\right) \cdot \boldsymbol{\theta_j}$ in the $\boldsymbol{\theta_j}$ direction.



3. Use the parallelogram rule to move all (x, y) in the direction of $\boldsymbol{\theta_j}$.
Equation: $(x,y) + \left(\left(\frac{(x,y)}{\sqrt{\lambda_j}} - (x,y)\right) \cdot \boldsymbol{\theta_j}\right)\boldsymbol{\theta_j}$.
The equation above yields our optimal transport map $\left(T_j(x,y)\right)$ in the direction of **θ.**

4. Repeat for j = 1, 2, … , d to get $T_1, T_2, \ldots, T_d$.

5. $\Sigma^{-1} = (T_d \ldots T_2T_1)^2$

Case 2: Eigenvectors and eigenvalues of the d x d SPD matrix Σ are not pre-computed

1. Pick n random directions.

2. Follow steps 1-4 from case 1 and substitute λ_j with $\theta_n^T\Sigma\theta_n$

3. Repeat for all n's, and the equation for Σ^{-1} should approximate to:

$$\Sigma^{-1} \approx (T_nT_{n-1}T_{n-2} \cdots T_2T_1)^2$$

PRELIMINARY RESULTS

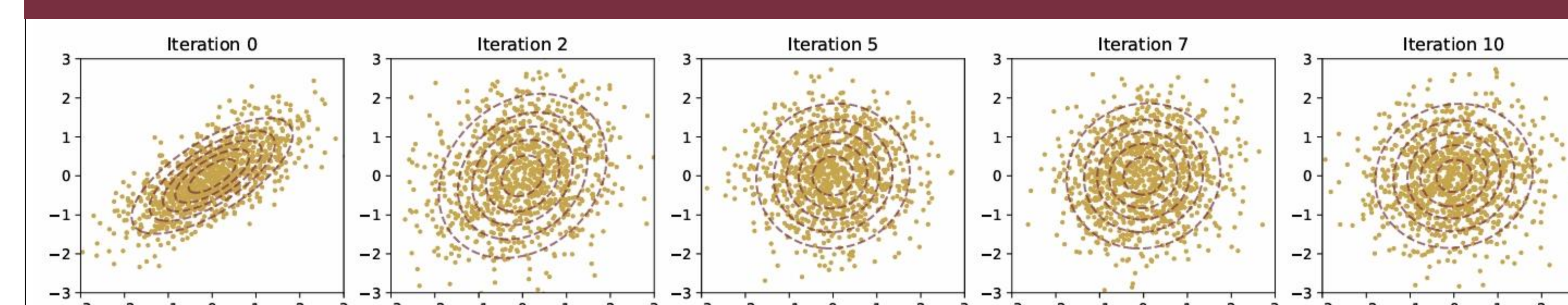


Figure 5

The final algorithm described in the methodology is currently under development. However, Figure 5 shows the expectation of a Gaussian with covariance matrix Σ at iteration 0 approaching a standard Gaussian with the identity matrix as the covariance matrix by iteration 10. Thus, we achieve inversion.

DISCUSSION/FUTURE DIRECTION

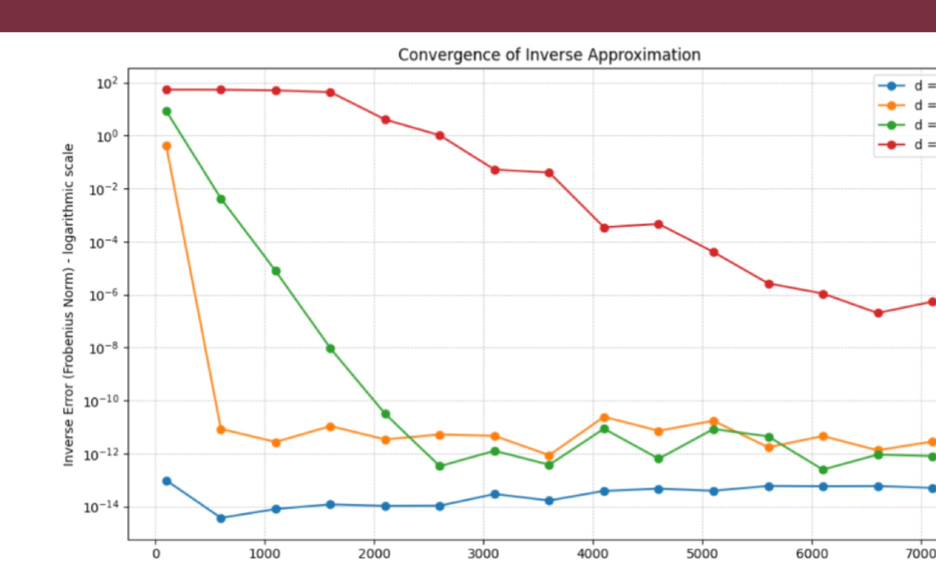


Figure 6: Preliminary depiction of the algorithm's error decay, for matrices of different sizes dxd, as the number of random angles N increases.

Question: What happens when we run the algorithms for matrices of different sizes (d x d)?

The algorithm is exact for **case 1** in all dimensions d.

For **case 2**, as a preliminary result, Figure 6 shows the error between the true inverse and the proposed algorithm. The plot illustrates that as the number of iterations increases, the error decreases. However, as the matrix size increases (higher dimensions), more iterations are needed and the error decreases slowly.