Multiple Perspectives on the Moore-Penrose Pseudoinverse

Theory, Computation, and Application

Galyna Kriukova

Department of Mathematics
National University of Kyiv-Mohyla Academy

September 18, 2025

Inverse Problem

Find x

$$Ax = y$$

Well-posedness (aka Hadamard's conditions)

- The problem has a solution
- The solution is unique
- The solution's behavior changes continuously with the initial conditions

Well-posed problem $\implies$ use inverse $x = A^{-1}y$
Ill-posed problem $\implies$ use pseudo-inverse!

Moore–Penrose inverse

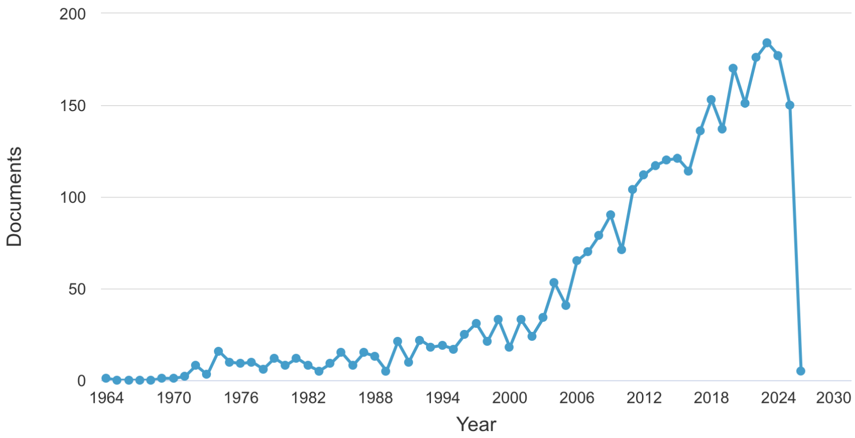


Figure 1: Number of papers published, in whose titles, abstracts or keynotes occurs the phrase "Moore–Penrose inverse" according to Elsevier Scopus (as of 17.09.2025)

Definition[1,2]

Pseudo-inverse $A^\dagger$ as generalized inverse (Moore–Penrose conditions)

- $AA^\dagger A = A$ ($AA^\dagger$ is an orthogonal projector onto column spaces)
- $A^\dagger AA^\dagger = A^\dagger$ (weak inverse)
- $(AA^\dagger)^* = AA^\dagger$ (is Hermitian)
- $(A^\dagger A)^* = A^\dagger A$ (is Hermitian)

Every matrix A has its Moore–Penrose pseudo-inverse $A^\dagger$;
the pseudo-inverse is unique;
if A is square and non-singular, then $A^\dagger = A^{-1}$.
$(A^\dagger)^\dagger = A$, $(A^*)^\dagger = (A^\dagger)^*$, $(A^T)^\dagger = (A^\dagger)^T$, $A^\dagger = (A^*A)^\dagger A^* = A^*(AA^*)^\dagger$

---

[1] [Moore, 1920]
[2] [Penrose, 1955]

Limit definition[3]

Limit definition

For any $A \in \mathbb{C}^{m \times n}$, as $\lambda \to 0$ through any neighborhood of 0 in $\mathbb{C}$, the following limits exist and

$$A_L^\dagger = \lim_{\lambda \to 0} (A^* A + \lambda I)^{-1} A^* = A^\dagger$$

$$A_R^\dagger = \lim_{\lambda \to 0} A^* (A A^* + \lambda I)^{-1} = A^\dagger$$

$A^* A$ and $A A^*$ are semi-positive definite and symmetric,
therefore $A^* A + \lambda I$ and $A A^* + \lambda I$ are invertible

- If A has linearly independent columns ($A^* A$ is invertible), $A^\dagger = (A^* A)^{-1} A^*$
- If A has linearly independent rows ($A A^*$ is invertible), $A^\dagger = A^* (A A^*)^{-1}$

[3] [Ben-Israel and Greville, 2003]

Explicit formula

Theorem (MacDuffee, 1959)

If $A \in \mathbb{C}_r^{m,n}$, $r > 0$, has a full-rank factorization

$$A = FG,$$

then

$$A^\dagger = G^*(F^*AG^*)^{-1}F^*.$$

Moreover, $A^\dagger = G^\dagger F^\dagger$.

THEOREM 2 (The Singular Value Decomposition). *Let* $O \neq A \in \mathbb{C}_r^{m \times n}$ *and let*

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \tag{0.32}$$

be the singular values of A.

Then there exist unitary matrices $U \in U^{m \times m}$ *and* $V \in U^{n \times n}$ *such that the matrix*

$$\Sigma = U^* A V = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & O \\ & & \sigma_r & \\ \hline & O & & O \end{array}\right] \tag{1}$$

is diagonal.

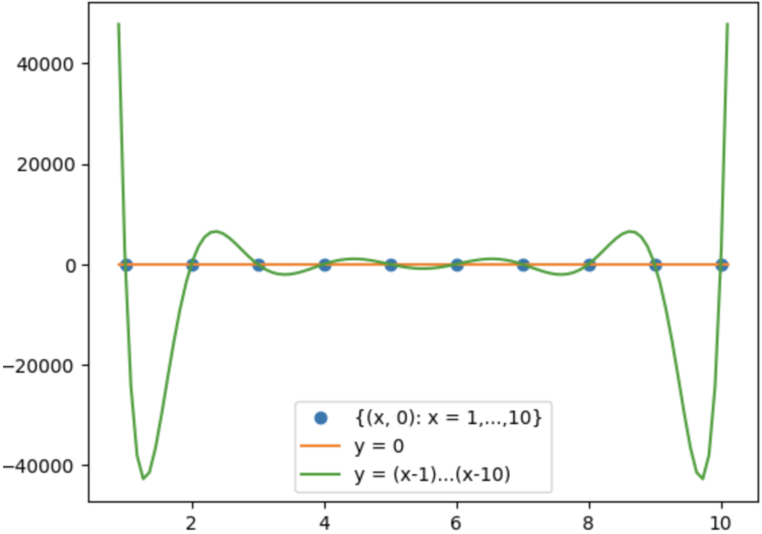
COROLLARY 1 (Penrose [**635**]). *Let* A, Σ, U, *and* V *be as in Theorem* 2. *Then*

$$A^\dagger = V\Sigma^\dagger U^* \tag{27}$$

where

$$\Sigma^\dagger = \operatorname{diag}\left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_r}, 0, \dots, 0\right) \in \mathbb{R}^{n\times m}. \tag{28}$$

For given dataset $\{(x, y)\}$ find the best (?) function f, such that $y = f(x)$



Machine learning

The Bayesian approach, or maximum a posteriori probability (MAP) estimate, finds an x such that maximizes the conditional probability $p(x|y)$. According to the Bayes rule

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx} \propto p(y|x)p(x),$$

therefore maximisation of $p(x|y)$ corresponds to the following problem:

$$\arg\min_x(-\log p(y|x) - \log p(x)).$$

Real probability distribution functions are unknown, therefore some heuristics are used

$$\hat{x} = \arg\min_x\{L(f,x,y) + \alpha\rho(x)\},$$

where $l(x,y)$ is a loss function and $\rho(x)$ is a regularization term.

Hypothesis space

Let's consider f from Hypothesis space $\mathcal{H}$.
Expected risk for given loss-function L:

$$\mathcal{E}(f) = \int_{X \times Y} L(f(x), y) d\rho(x, y) \to \min$$

Empirical risk (Tikhonov-Phillips regularization):

$$\sum_{(x_i, y_i)} L(f(x_i), y_i) + \lambda \|f\|^2_{\mathcal{H}}$$

If we consider Hypothesis space reproducing kernel Hilbert space (RKHS) generated by a kernel $K : X \times X \to \mathbb{R}$, by the representer theorem for RKHS, the minimizer of empirical risk is equal to

$$f^{\lambda}_{\{(x_i, y_i)\}} = \sum_{x_i} c_i K(\cdot, x_i).$$

Least squares method

$$Ax = y$$

$$L(x) = \|Ax - y\|^2 \to \min$$

$$\nabla L(x) = 2A^T Ax - 2A^T y = 0$$
$$A^T Ax = A^T y$$

$$x = A^\dagger y$$

Linear least squares

$$Ax = y$$

$$L(x) = \|Ax - y\| \to \min$$

$$x = A^{\dagger} y + (I - A^{\dagger} A) v,$$

for any vector v.

Applications

- Sparse and Redundant Representations
- Artificial Neural Networks for Computer Vision, Natural Language Processing
- Physics research ([Baksalary and Trenkler, 2021])

and many more...

Sir Roger Penrose was awarded the Nobel Prize in Physics in 2020

Resistance Distance

The Moore–Penrose inverse of Laplacian matrix of a graph can be applied to study the resistance distance between vertices of the graph.

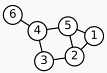
The resistance distance between two vertices of a simple, connected graph, G, is equal to the resistance between two equivalent points on an electrical network, constructed so as to correspond to G, with each edge being replaced by a resistance of one ohm. It is a metric on graphs.

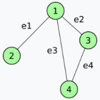
The resistance distance between two vertices u and v of G can be obtained via the formula

$$r(u, v) = L^{\dagger}_{u,u} + L^{\dagger}_{v,v} - 2L^{\dagger}_{u,v}.$$

Laplacian matrix

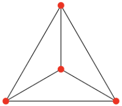
The Laplacian matrix $L(G)$ is an $n \times n$ matrix whose rows and columns are indexed by vertices of G. The (i, j)-entry of $L(G)$ is equal to $\deg_G(v_i)$, the degree of the vertex v_i, if $i = j$, and it is -1 or 0 if the vertices v_i and v_j are adjacent or non-adjacent, respectively.

Labelled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2&0&0&0&0&0\\ 0&3&0&0&0&0\\ 0&0&2&0&0&0\\ 0&0&0&3&0&0\\ 0&0&0&0&3&0\\ 0&0&0&0&0&1 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&0&1&0\\ 1&0&1&0&1&0\\ 0&1&0&1&0&0\\ 0&0&1&0&1&1\\ 1&1&0&1&0&0\\ 0&0&0&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 2&-1&0&0&-1&0\\ -1&3&-1&0&-1&0\\ 0&-1&2&-1&0&0\\ 0&0&-1&3&-1&-1\\ -1&-1&0&-1&3&0\\ 0&0&0&-1&0&1 \end{pmatrix}$

Undirected graph	Incidence matrix	Laplacian matrix
	$\begin{pmatrix} 1&1&1&0\\ -1&0&0&0\\ 0&-1&0&1\\ 0&0&-1&-1 \end{pmatrix}$	$\begin{pmatrix} 3&-1&-1&-1\\ -1&1&0&0\\ -1&0&2&-1\\ -1&0&-1&2 \end{pmatrix}$

Resistance distance: examples

Tetrahedral Graph K_4



$$L(K_4) = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}$$

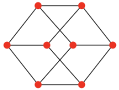
$$R(K_4) = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \qquad L^{\dagger}(K_4) = \begin{pmatrix} \frac{3}{16} & -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & \frac{3}{16} & -\frac{1}{16} & -\frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & \frac{3}{16} & -\frac{1}{16} \\ -\frac{1}{16} & -\frac{1}{16} & -\frac{1}{16} & \frac{3}{16} \end{pmatrix}$$

For a complete graph K_n, the resistance distance between any two distinct vertices i and j is given by the formula:

$$R_{ij}(K_4) = \frac{2}{n} \cdot \mathbf{1}_{i \neq j}$$

Resistance distance: examples

Cubical Graph Q_3



$$L(Q_3) = \begin{pmatrix} 3 & -1 & -1 & 0 & -1 & 0 & 0 & 0 \\ -1 & 3 & 0 & -1 & 0 & -1 & 0 & 0 \\ -1 & 0 & 3 & -1 & 0 & 0 & -1 & 0 \\ 0 & -1 & -1 & 3 & 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 & 3 & -1 & -1 & 0 \\ 0 & -1 & 0 & 0 & -1 & 3 & 0 & -1 \\ 0 & 0 & -1 & 0 & -1 & 0 & 3 & -1 \\ 0 & 0 & 0 & -1 & 0 & -1 & -1 & 3 \end{pmatrix}$$

$$R(Q_3) = \frac{1}{12}\begin{pmatrix} 0 & 5 & 5 & 7 & 5 & 7 & 7 & 8 \\ 5 & 0 & 7 & 5 & 7 & 5 & 8 & 7 \\ 5 & 7 & 0 & 5 & 7 & 8 & 5 & 7 \\ 7 & 5 & 5 & 0 & 8 & 7 & 7 & 5 \\ 5 & 7 & 7 & 8 & 0 & 5 & 5 & 7 \\ 7 & 5 & 8 & 7 & 5 & 0 & 7 & 5 \\ 7 & 8 & 5 & 7 & 5 & 7 & 0 & 5 \\ 8 & 7 & 7 & 5 & 7 & 5 & 5 & 0 \end{pmatrix}$$

$$L^{\dagger}(Q_3) = \frac{1}{96}\begin{pmatrix} 29 & 9 & 9 & 1 & 9 & 1 & 1 & -3 \\ 9 & 29 & 1 & 9 & 1 & 9 & -3 & 1 \\ 9 & 1 & 29 & 9 & 1 & -3 & 9 & 1 \\ 1 & 9 & 9 & 29 & -3 & 1 & 1 & 9 \\ 9 & 1 & 1 & -3 & 29 & 9 & 9 & 1 \\ 1 & 9 & -3 & 1 & 9 & 29 & 1 & 9 \\ 1 & -3 & 9 & 1 & 9 & 1 & 29 & 9 \\ -3 & 1 & 1 & 9 & 1 & 9 & 9 & 29 \end{pmatrix}$$

References

[Baksalary and Trenkler, 2021] Baksalary, O. and Trenkler, G. (2021).
The moore–penrose inverse: a hundred years on a frontline of physics research.
The European Physical Journal H, 46(1):9.

[Ben-Israel and Greville, 2003] Ben-Israel, A. and Greville, T. N. E. (2003).
Generalized Inverses: Theory and Applications.
Springer-Verlag, New York, NY, 2nd edition.

[Moore, 1920] Moore, E. (1920).
On the reciprocal of the general algebraic matrix.
Bulletin of the American Mathematical Society, 26:394–395.

[Penrose, 1955] Penrose, R. (1955).
A generalized inverse for matrices.
Mathematical Proceedings of the Cambridge Philosophical Society, 51:406–413.

Thank you!