What Is a Generalized Inverse?

Nicholas J. Higham*

April 3, 2020

The matrix inverse is defined only for square nonsingular matrices. A generalized inverse is an extension of the concept of inverse that applies to square singular matrices and rectangular matrices. There are many definitions of generalized inverses, all of which reduce to the usual inverse when the matrix is square and nonsingular.

A large class of generalized inverses of an $m \times n$ matrix A can be defined in terms of the Moore–Penrose conditions, in which X is $n \times m$:

$$\begin{array}{llll}
(1) & AXA = A, & (2) & XAX = X, \\
(3) & (AX)^* = AX, & (4) & (XA)^* = XA.
\end{array}$$

Here, the superscript $*$ denotes the conjugate transpose. A 1-inverse is any X satisfying condition (1), a (1,3)-inverse is any X satisfying conditions (1) and (3), and so on for any subset of the four conditions.

Condition (1) implies that if $Ax = b$ then $A(Xb) = A(XAx) = AXAx = Ax = b$, so Xb solves the equation, meaning that any 1-inverse is an equation-solving inverse. Condition (2) implies that $X = 0$ if $A = 0$.

A (1,3) inverse can be shown to provide a least squares solution to an inconsistent linear system. A (1,4) inverse can be shown to provide the minimum 2-norm solution of a consistent linear system (where the 2-norm is defined by $\|x\|_2 = (x^*x)^{1/2}$).

There is not a unique matrix satisfying any one, two, or three of the Moore–Penrose conditions. But there is a unique matrix satisfying all four of the conditions, and it is called the *Moore-Penrose pseudoinverse*, denoted by A^+ or $A^\dagger$. For any system of linear equations $Ax = b$, $x = A^+b$ minimizes $\|Ax - b\|_2$ and has the minimum 2–norm over all minimizers.

The pseudoinverse can be expressed in terms of the singular value decomposition (SVD). If $A = U\Sigma V^*$ is an SVD, where the $m \times m$ matrix U and $n \times n$ matrix V are orthogonal, and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$ (so that $\mathrm{rank}(A) = r$), then

$$A^+ = V\mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0)U^*.$$

In MATLAB, the function `pinv` computes A^+ using this formula. If $\mathrm{rank}(A) = n$ then the concise formula $A^+ = (A^*A)^{-1}A^*$ holds.

For square matrices, *the Drazin inverse* is the unique matrix A^D such that

$$A^D A A^D = A^D, \quad AA^D = A^D A, \quad A^{k+1}A^D = A^k,$$

---

*Department of Mathematics, University of Manchester, Manchester, M13 9PL, UK (nick.higham@manchester.ac.uk).

where $k = \mathrm{index}(A)$. The first condition is the same as the second of the Moore–Penrose conditions, but the second and third have a different flavour. The index of a matrix of A is the smallest nonnegative integer k such that $\mathrm{rank}(A^k) = \mathrm{rank}(A^{k+1})$; it is characterized as the dimension of the largest Jordan block of A with eigenvalue zero.

If $\mathrm{index}(A) = 1$ then A^D is also known as the *group inverse* of A and is denoted by $A^{\#}$. The Drazin inverse is an equation-solving inverse precisely when $\mathrm{index}(A) \leq 1$, for then $AA^D A = A$, which is the first of the Moore–Penrose conditions.

The Drazin inverse can be represented explicitly as follows. If

$$A = P \begin{bmatrix} B & 0 \\ 0 & N \end{bmatrix} P^{-1},$$

where P and B are nonsingular and N has only zero eigenvalues, then

$$A^D = P \begin{bmatrix} B^{-1} & 0 \\ 0 & 0 \end{bmatrix} P^{-1}.$$

Here is the pseudoinverse and the Drazin inverse for a particular matrix with index 2:

$$A = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \quad A^+ = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & -1 & 0 \end{bmatrix}, \quad A^D = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Applications

The Moore–Penrose pseudoinverse is intimately connected with orthogonality, whereas the Drazin inverse has spectral properties related to those of the original matrix. The pseudoinverse occurs in all kinds of least squares problems. Applications of the Drazin inverse include population modelling, Markov chains, and singular systems of linear differential equations. It is not usually necessary to compute generalized inverses, but they are valuable theoretical tools.

References

This is a minimal set of references, which contain further useful references within.

- Adi Ben-Israel, The Moore of the Moore–Penrose Inverse, Electron. J. Linear Algebra 9, 150–157, 2002.
- Adi Ben-Israel and Thomas N. E. Greville, Generalized Inverses: Theory and Applications, second edition, Springer-Verlag, New York, 2003
- Stephen Campbell and Carl Meyer, Generalized Inverses of Linear Transformations, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2009. published (Originally published by Pitman in 1979.)
- Stephen J. Kirkland and Michael Neumann, Group Inverses and M-Matrices and their Applications, Chapman and Hall/CRC, 2013
- Guorong Wang, Yimin Wei and Sanzheng Qiao, Generalized Inverses: Theory and Computations, second edition, Springer-Verlag, Singapore, 2018.

Related Blog Posts

- What is a Matrix? (2020)

This article is part of the "What Is" series, available from `https://nhigham.com/category/what-is` and in PDF form from the GitHub repository `https://github.com/higham/what-is`.