Block matrix pseudoinverse

In [mathematics](#), a **block matrix pseudoinverse** is a formula for the [pseudoinverse](#) of a [partitioned matrix](#). This is useful for decomposing or approximating many algorithms updating parameters in [signal processing](#), which are based on the [least squares](#) method.

Derivation

Consider a column-wise partitioned matrix:

$$[\mathbf{A} \quad \mathbf{B}], \quad \mathbf{A} \in \mathbb{R}^{m \times n}, \quad \mathbf{B} \in \mathbb{R}^{m \times p}, \quad m \geq n + p.$$

If the above matrix is full column rank, the [Moore–Penrose inverse](#) matrices of it and its transpose are

$$[\mathbf{A} \quad \mathbf{B}]^+ = \left([\mathbf{A} \quad \mathbf{B}]^\mathsf{T}[\mathbf{A} \quad \mathbf{B}]\right)^{-1}[\mathbf{A} \quad \mathbf{B}]^\mathsf{T},$$

$$\begin{bmatrix} \mathbf{A}^\mathsf{T} \\ \mathbf{B}^\mathsf{T} \end{bmatrix}^+ = [\mathbf{A} \quad \mathbf{B}]\left([\mathbf{A} \quad \mathbf{B}]^\mathsf{T}[\mathbf{A} \quad \mathbf{B}]\right)^{-1}.$$

This computation of the pseudoinverse requires $(n + p)$-square matrix inversion and does not take advantage of the block form.

To reduce computational costs to n- and p-square matrix inversions and to introduce parallelism, treating the blocks separately, one derives [1]

$$[\mathbf{A} \quad \mathbf{B}]^+ = \begin{bmatrix} \mathbf{P}_B^\perp \mathbf{A}(\mathbf{A}^\mathsf{T}\mathbf{P}_B^\perp \mathbf{A})^{-1} \\ \mathbf{P}_A^\perp \mathbf{B}(\mathbf{B}^\mathsf{T}\mathbf{P}_A^\perp \mathbf{B})^{-1} \end{bmatrix} = \begin{bmatrix} (\mathbf{P}_B^\perp \mathbf{A})^+ \\ (\mathbf{P}_A^\perp \mathbf{B})^+ \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{A}^\mathsf{T} \\ \mathbf{B}^\mathsf{T} \end{bmatrix}^+ = \begin{bmatrix} \mathbf{P}_B^\perp \mathbf{A}(\mathbf{A}^\mathsf{T}\mathbf{P}_B^\perp \mathbf{A})^{-1}, & \mathbf{P}_A^\perp \mathbf{B}(\mathbf{B}^\mathsf{T}\mathbf{P}_A^\perp \mathbf{B})^{-1} \end{bmatrix} = \begin{bmatrix} (\mathbf{A}^\mathsf{T}\mathbf{P}_B^\perp)^+ & (\mathbf{B}^\mathsf{T}\mathbf{P}_A^\perp)^+ \end{bmatrix},$$

where [orthogonal projection](#) matrices are defined by

$$\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{A}(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T},$$
$$\mathbf{P}_B^\perp = \mathbf{I} - \mathbf{B}(\mathbf{B}^\mathsf{T}\mathbf{B})^{-1}\mathbf{B}^\mathsf{T}.$$

The above formulas are not necessarily valid if $[\mathbf{A} \quad \mathbf{B}]$ does not have full rank – for example, if $\mathbf{A} \neq 0$, then

$$[\mathbf{A} \quad \mathbf{A}]^+ = \frac{1}{2}\begin{bmatrix} \mathbf{A}^+ \\ \mathbf{A}^+ \end{bmatrix} \neq \begin{bmatrix} (\mathbf{P}_A^\perp \mathbf{A})^+ \\ (\mathbf{P}_A^\perp \mathbf{A})^+ \end{bmatrix} = 0$$

Application to least squares problems

Given the same matrices as above, we consider the following least squares problems, which appear as multiple objective optimizations or constrained problems in signal processing. Eventually, we can implement a parallel algorithm for least squares based on the following results.

Column-wise partitioning in over-determined least squares

Suppose a solution $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$ solves an over-determined system:

$$[\mathbf{A}, \quad \mathbf{B}] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \mathbf{d}, \quad \mathbf{d} \in \mathbb{R}^{m \times 1}.$$

Using the block matrix pseudoinverse, we have

$$\mathbf{x} = [\mathbf{A}, \quad \mathbf{B}]^+ \mathbf{d} = \begin{bmatrix} \left(\mathbf{P}_B^\perp \mathbf{A}\right)^+ \\ \left(\mathbf{P}_A^\perp \mathbf{B}\right)^+ \end{bmatrix} \mathbf{d}.$$

Therefore, we have a decomposed solution:

$$\mathbf{x}_1 = \left(\mathbf{P}_B^\perp \mathbf{A}\right)^+ \mathbf{d}, \quad \mathbf{x}_2 = \left(\mathbf{P}_A^\perp \mathbf{B}\right)^+ \mathbf{d}.$$

Row-wise partitioning in under-determined least squares

Suppose a solution $\mathbf{x}$ solves an under-determined system:

$$\begin{bmatrix} \mathbf{A}^\mathsf{T} \\ \mathbf{B}^\mathsf{T} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{e} \\ \mathbf{f} \end{bmatrix}, \quad \mathbf{e} \in \mathbb{R}^{n \times 1}, \quad \mathbf{f} \in \mathbb{R}^{p \times 1}.$$

The minimum-norm solution is given by

$$\mathbf{x} = \begin{bmatrix} \mathbf{A}^\mathsf{T} \\ \mathbf{B}^\mathsf{T} \end{bmatrix}^+ \begin{bmatrix} \mathbf{e} \\ \mathbf{f} \end{bmatrix}.$$

Using the block matrix pseudoinverse, we have

$$\mathbf{x} = \begin{bmatrix} \left(\mathbf{A}^\mathsf{T} \mathbf{P}_B^\perp\right)^+ & \left(\mathbf{B}^\mathsf{T} \mathbf{P}_A^\perp\right)^+ \end{bmatrix} \begin{bmatrix} \mathbf{e} \\ \mathbf{f} \end{bmatrix} = \left(\mathbf{A}^\mathsf{T} \mathbf{P}_B^\perp\right)^+ \mathbf{e} + \left(\mathbf{B}^\mathsf{T} \mathbf{P}_A^\perp\right)^+ \mathbf{f}.$$

Comments on matrix inversion

Instead of $\left([\mathbf{A} \quad \mathbf{B}]^\mathsf{T} [\mathbf{A} \quad \mathbf{B}]\right)^{-1}$, we need to calculate directly or indirectly

$$\left(\mathbf{A}^\mathsf{T} \mathbf{A}\right)^{-1}, \quad \left(\mathbf{B}^\mathsf{T} \mathbf{B}\right)^{-1}, \quad \left(\mathbf{A}^\mathsf{T} \mathbf{P}_B^\perp \mathbf{A}\right)^{-1}, \quad \left(\mathbf{B}^\mathsf{T} \mathbf{P}_A^\perp \mathbf{B}\right)^{-1}.$$

In a dense and small system, we can use singular value decomposition, QR decomposition, or Cholesky decomposition to replace the matrix inversions with numerical routines. In a large system, we may employ iterative methods such as Krylov subspace methods.

Considering [parallel algorithms](), we can compute $\left(\mathbf{A}^{\mathsf{T}}\mathbf{A}\right)^{-1}$ and $\left(\mathbf{B}^{\mathsf{T}}\mathbf{B}\right)^{-1}$ in parallel. Then, we finish to compute $\left(\mathbf{A}^{\mathsf{T}}\mathbf{P}_B^{\perp}\mathbf{A}\right)^{-1}$ and $\left(\mathbf{B}^{\mathsf{T}}\mathbf{P}_A^{\perp}\mathbf{B}\right)^{-1}$ also in parallel.

See also

- [Invertible matrix § Blockwise inversion]()

References

1. [J.K. Baksalary]() and O.M. Baksalary (2007). "Particular formulae for the Moore–Penrose inverse of a columnwise partitioned matrix". *Linear Algebra Appl*. **421**: 16–23. [doi]():[10.1016/j.laa.2006.03.031 (https://doi.org/10.1016%2Fj.laa.2006.03.031)]().

External links

- [The Matrix Reference Manual (http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html)]() by [Mike Brookes (http://www.ee.ic.ac.uk/hp/staff/dmb/dmb.html)]()
- [Linear Algebra Glossary (https://web.archive.org/web/20060414125709/http://www.csit.fsu.edu/~burkardt/papers/linear_glossary.html)]() by [John Burkardt (http://www.csit.fsu.edu/~burkardt/)]()
- [The Matrix Cookbook (http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf)]() by [Kaare Brandt Petersen (http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3274/)]()
- [Lecture 8: Least-norm solutions of undetermined equations (http://see.stanford.edu/materials/lsoeldsee263/08-min-norm.pdf)]() by [Stephen P. Boyd (https://stanford.edu/~boyd/)]()

-