A guide to the Moore-Penrose generalized inverse operators for data assimilation

Ross Bannister, National Centre for Earth Observation, University of Reading, RG6 6BB

August 1, 2015

There are two types of operator that appear frequently in data assimilation which do not have defined inverse counterparts. In this note we derive the Moore-Penrose generalized inverse (MPGI) counterparts and interpret their meanings.

1 Type I operator (low-to-high dimensional operator)

1.1 Definitions

The first type of operator that we shall consider the MPGI for is one whose input space is smaller in dimension that the output space:

$$\mathbf{x} = \mathbf{X}\mathbf{w}, \tag{1}$$

$$\mathbf{x} \in \mathbb{R}^n, \quad \mathbf{w} \in \mathbb{R}^N, \quad \mathbf{X} \in \mathbb{R}^{n \times N}, \quad n > N.$$

An example of a type I operator is that which gives a model space output ($\mathbf{x}$) as a linear combination ($\mathbf{w}$) of ensemble members (columns of $\mathbf{X}$). Important: note that (1) works only when $\mathbf{x}$ lies in the column space of $\mathbf{w}$. The MPGI for (1) is $\mathbf{X}^+$:

$$\mathbf{w} = \mathbf{X}^+\mathbf{x}, \tag{2}$$

$$\text{where } \mathbf{X}^+ = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}. \tag{3}$$

1.2 Explanation

Eliminating $\mathbf{x}$ between (1) and (2), with definition (3) gives:

$$\mathbf{w} = \mathbf{X}^+\mathbf{X}\mathbf{w} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w} = \mathbf{w}.$$

This means that we can go from $\mathbf{w}$ to $\mathbf{x}$ and then back to $\mathbf{w}$ exactly. This does not work the other way round: eliminating $\mathbf{w}$ between (1) and (2), with definition (3) in general gives:

$$\mathbf{x} \neq \mathbf{X}\mathbf{X}^+\mathbf{x} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{x}.$$

There is a special case though: if $\mathbf{x}$ lies in the column space of $\mathbf{X}$ then $\mathbf{x}$ can be represented exactly with (1) and the above then reduces to (1) itself so, in this case, the above inequality becomes an equality.

1.3 Derivation

To derive (2) and (3), we pose the question, "Given a data vector $\mathbf{x}_{\mathrm{d}}$, what $\mathbf{w}$ yields a corresponding $\mathbf{x}$ using (1) that is as close as possible to $\mathbf{x}_{\mathrm{d}}$?" This is the $\mathbf{w}$ that is the minimum of the following cost function:

$$J_I(\mathbf{w}) = \frac{1}{2}(\mathbf{x}_{\mathrm{d}} - \mathbf{X}\mathbf{w})^{\mathrm{T}}(\mathbf{x}_{\mathrm{d}} - \mathbf{X}\mathbf{w}). \tag{4}$$

Minimizing this cost function with the standard method of differentiating and setting to zero gives:

$$\nabla_{\mathbf{w}}(J_I) = -\mathbf{X}^{\mathrm{T}}(\mathbf{x}_{\mathrm{d}} - \mathbf{X}\mathbf{w}) = 0, \tag{5}$$

which can be rearranged to give (2) and (3).

2 Type II operator (high-to-low dimensional operator)

2.1 Definitions

The second type of operator that we shall consider the MPGI for is one whose input space is larger in dimension that the output space:

$$\mathbf{y} = \mathbf{H}\mathbf{x}, \tag{6}$$

$$\mathbf{x} \in \mathbb{R}^n, \quad \mathbf{y} \in \mathbb{R}^m, \quad \mathbf{H} \in \mathbb{R}^{m \times n}, \quad n > m.$$

An example of a type II operator is an observation operator that has fewer model observations ($\mathbf{y}$) than state vector elements ($\mathbf{x}$). The MPGI for (6) is $\mathbf{H}^+$:

$$\mathbf{x} = \mathbf{H}^+\mathbf{y}, \tag{7}$$

$$\text{where } \mathbf{H}^+ = \mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}. \tag{8}$$

2.2 Explanation

Eliminating $\mathbf{x}$ between (6) and (7) with definition (8) gives:

$$\mathbf{y} = \mathbf{H}\mathbf{H}^+\mathbf{y} = \mathbf{H}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{y} = \mathbf{y}.$$

This means that we can go from $\mathbf{y}$ to $\mathbf{x}$ and then back to $\mathbf{y}$ exactly. This does not work the other way round: eliminating $\mathbf{y}$ between (6) and (7) in general gives:

$$\mathbf{x} \neq \mathbf{H}^+\mathbf{H}\mathbf{x} = \mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{H}\mathbf{x}.$$

There is a special case though: if $\mathbf{x}$ lies in the row space of $\mathbf{H}$ then $\mathbf{x}$ can be represented exactly with (7) and the above then reduces to (6) itself so, in this case, the above inequality becomes an equality.

2.3 Derivation

To derive (7), we note that there is no unique $\mathbf{x}$ that is consistent with (6). We pose the question, "Given a data vector $\mathbf{y}$, what is the *smallest* $\mathbf{x}$ that satisfies (6)?". The cost function to minimize is:

$$J_{II}(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{x} + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{y} - \mathbf{H}\mathbf{x}), \tag{9}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$ is the vector of Lagrange multipliers. The first term in (9) specifies that the smallest $\mathbf{x}$ is to be found and the second term describes the constraint. Minimizing this cost function with the standard method of differentiating and setting to zero gives:

$$\nabla_{\mathbf{x}}(J_{II}) = \mathbf{x} - \mathbf{H}^{\mathrm{T}}\boldsymbol{\lambda} = 0, \tag{10}$$

$$\nabla_{\boldsymbol{\lambda}}(J_{II}) = \mathbf{y} - \mathbf{H}\mathbf{x} = 0. \tag{11}$$

Eliminating $\mathbf{x}$ gives $\mathbf{y} = \mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\lambda}$, giving $\boldsymbol{\lambda} = (\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{y}$. Now eliminating $\boldsymbol{\lambda}$ between this and (10) leads to (7) and (8).