On the Relative Gain Array (RGA) with Singular and Rectangular Matrices

Jeffrey Uhlmann
University of Missouri-Columbia
201 Naka Hall, Columbia, MO 65211
573.884.2129, uhlmannj@missouri.edu

Abstract

This paper identifies a significant deficiency in the literature on the application of the Relative Gain Array (RGA) formalism in the case of singular matrices. Specifically, it is shown that the conventional use of the Moore-Penrose pseudoinverse is inappropriate because it fails to preserve critical properties that can be assumed in the nonsingular case. It is then shown that such properties can be rigorously preserved using an alternative generalized matrix inverse.

Keywords: Consistency Analysis, Control Systems, Generalized Matrix Inverse, Inverse Problems, Linear Estimation, Linear Systems, Matrix Analysis, Moore-Penrose Pseudoinverse, Relative Gain Array, RGA, Singular Value Decomposition, SVD, Stability of Linear Systems, System Design, UC inverse, Unit Consistency.

AMS Code: 15A09

I. Introduction

The relative gain array (RGA) provides a jointly-conditioned relative measure of input-output interactions among a multi-input multi-output (MIMO) system [3]. The RGA is a function of a matrix, which typically is interpreted as a plant/gain matrix $\mathbf{G}$ for a set of input parameters that control a set of output parameters, and is defined for nonsingular $\mathbf{G}$ as:

$$\mathcal{R}GA(\mathbf{G}) \doteq \mathbf{G} \circ (\mathbf{G}^{-1})^{\mathrm{T}} \tag{1}$$

where $\circ$ represents the elementwise Hadamard matrix product.

The RGA has been and continues to be used in a wide variety of important practical control applications. This is due in part to its convenient use and interpretation and in part to its established mathematical properties. The RGA is also applied widely in a form that is generalized for applications with singular $\mathbf{G}$ and is presumed to maintain at least some of the rigor and properties that hold in the nonsingular case [8], [4]. In this paper it is shown that this presumption is unjustified and therefore should undermine confidence in existing safety-critical systems in which it is applied. We then propose an alternative generalization that provably maintains a key property which does not hold for the convential formulation presented in the literature.

II. Properties of the RGA

When the gain matrix $\mathbf{G}$ is nonsingular the RGA has several important properties [7], [6], which include for diagonal matrices $\mathbf{D}$ and $\mathbf{E}$ and permutation matrices $\mathbf{P}$ and $\mathbf{Q}$:

$$\mathcal{R}GA(\mathbf{PGQ}) = \mathbf{P} \cdot \mathcal{R}GA(\mathbf{G}) \cdot \mathbf{Q} \tag{2}$$

$$\mathcal{R}GA(\mathbf{DGE}) = \mathcal{R}GA(\mathbf{G}). \tag{3}$$

The first property simply says that a permutation, i.e., reordering, of the elements of the input and output vectors leads to a conformant reordering of the rows and columns of the resulting RGA matrix. The second property says that the relative gain values are invariant to the units applied to elements of the input and output vectors. In other words, the relative gain is independent of the choice of units chosen for state variables, e.g., celsius versus farenheit for temperature variables or radians versus degrees for angle variables.

Both of the above properties represent intuitively natural sanity checks because there should certainly be grave concerns if the integrity of a system were impacted by the choice of how the parameters/variables are ordered or the choice of units used for those parameters/variables. The value of the RGA is that it provides a unit-invariant measure of the sensitivity of each output variable to each input parameter[1]. For example, if the output variables are redefined from imperial units to metric units the resulting controls obtained from the RGA will produce *identical* system behavior.

Of course the definition of the RGA as $\mathbf{G} \circ (\mathbf{G}^{-1})^{\mathrm{T}}$ requires $\mathbf{G}$ to be nonsingular to ensure the existence of its inverse. The most obvious approach for generalizing the definition to include the case of singular $\mathbf{G}$ is to replace the matrix inverse $\mathbf{G}^{-1}$ with a generalized inverse $\mathbf{G}^{\widetilde{-1}}$ such that, ideally, the result:

$$\mathcal{R}GA(\mathbf{G}) \doteq \mathbf{G} \circ (\mathbf{G}^{\widetilde{-1}})^{\mathrm{T}} \tag{4}$$

preserves the key properties satisfied for nonsingular $\mathbf{G}$. Universally in the RGA literature the generalized inverse is taken to be the Moore-Penrose (MP) pseudoinverse [8], [4], [1], which is denoted here as $\mathbf{G}^{-P}$. This choice seems to derive from a widely-held misperception that the MP inverse is somehow uniquely suited to be the default whenever a generalized inverse is required [10]. The fact is that the MP inverse is only applicable under certain sets of assumptions, and in the case of the RGA its use will not generally preserve the key property of unit invariance. This will be demonstrated in the following section, where it will also be shown that there exists an alternative generalized inverse that does preserve the unit-invariance property.

III. Generalized Matrix Inverses

Although the generalized matrix inverse of a singular matrix $\mathbf{A}$ as $\mathbf{A}^{\widetilde{-1}}$ cannot satisfy all properties of a true inverse, it is typically expected to satisfy the following algebraic identities:

$$\mathbf{A}\mathbf{A}^{\widetilde{-1}}\mathbf{A} = \mathbf{A}, \tag{5}$$

$$\mathbf{A}^{\widetilde{-1}}\mathbf{A}\mathbf{A}^{\widetilde{-1}} = \mathbf{A}^{\widetilde{-1}}. \tag{6}$$

Beyond these properties, the appropriate generalized inverse of choice for a particular application must be determined by the conditions it is expected to preserve. In the case of the MP inverse, it satisfies the following:

$$(\mathbf{UAV})^{-P} = \mathbf{V}^{*}\mathbf{A}^{-P}\mathbf{U}^{*} \tag{7}$$

where $\mathbf{U}$ and $\mathbf{V}$ are arbitrary rotation matrices (or, more generally, arbitrary orthonormal/unitary matrices) such that $\mathbf{U}\mathbf{U}^{*} = \mathbf{I}$, where $\mathbf{U}^{*}$ is the conjugate-transpose of $\mathbf{U}$. This consistency with respect to rigid rotations is necessary when Euclidean distances must be preserved by all state space transformations. This preservation of the Euclidean norm is also what makes it appropriate for determining the $n \times m$ matrix $\mathbf{T}$ representing the minimum mean-squared error (MMSE) optimal linear transformation, $\mathbf{TX} \approx \mathbf{Y}$, where the n-dimensional columns of rectangular matrix $\mathbf{X}$ represent a set of p input vectors and the m-dimensional columns of $\mathbf{Y}$ represent a set of p output vectors.

The MP inverse is intimately connected to "least-squares" optimization, and both are commonly used in scientific and engineering applications without regard to whether the MMSE criterion is appropriate. In the case of vectors with elements that represent state variables, e.g., that include a mix of quantities such as temperature, mass, velocity, etc., the minimization of squared deviations, or equivalently the preserving of Euclidean distances between vectors, has no physical meaning. The problem for least-squares is, of course, that the magnitude of squared deviations is strongly dependent on the choice of units [12]. In the case of a range-bearing estimate (r, θ), for example, the choice to represent r in centimeters rather than meters, or even kilometers, has many orders of magnitude implications when squared deviations are minimized. The bearing angle θ, by contrast, is likely to be constrained between $-\pi$ and π, or between 0 and 2π, so depending on the units chosen for r the deviations in angle may be relatively so small as to be negligible from a least-squares perspective.

The unit-consistent (UC) generalized inverse [11], denoted as $\mathbf{A}^{-U}$, satisfies a different consistency condition:

$$(\mathbf{DAE})^{-U} = \mathbf{E}^{-1}\mathbf{A}^{-U}\mathbf{D}^{-1} \tag{8}$$

where $\mathbf{D}$ and $\mathbf{E}$ are arbitrary nonsingular diagonal matrices. Thus it satisfies

$$\mathbf{E} \cdot (\mathbf{DAE})^{-U} \cdot \mathbf{D} = \mathbf{A}^{-U} \tag{9}$$

[1]The RGA actually provides more than just a measure of sensitivity, e.g., the ratio of open-loop gain to closed-loop gain is a very useful interpretation for controller pairing.

while the MP inverse does not. As suggested in [9] this is precisely the condition required to ensure that a generalized definition of the RGA:

$$\mathcal{UC}\text{-}\mathcal{RGA}(\mathbf{G}) \doteq \mathbf{G} \circ (\mathbf{G}^{\text{-U}})^{\mathrm{T}} \tag{10}$$

preserves the unit-invariance property. Specifically, it is straightforward to show that the UC-RGA is invariant with respect to the scaling of its argument $\mathbf{G}$ by nonsingular diagonal matrices $\mathbf{D}$ and $\mathbf{E}$:

$$\begin{aligned}
\mathcal{UC}\text{-}\mathcal{RGA}(\mathbf{DGE}) &= (\mathbf{DGE}) \circ ((\mathbf{DGE})^{\text{-U}})^{\mathrm{T}} && (11)\\
&= (\mathbf{DGE}) \circ (\mathbf{E}^{\text{-1}}\mathbf{G}^{\text{-U}}\mathbf{D}^{\text{-1}})^{\mathrm{T}} && (12)\\
&= (\mathbf{DGE}) \circ (\mathbf{D}^{\text{-1}}(\mathbf{G}^{\text{-U}})^{\mathrm{T}}\mathbf{E}^{\text{-1}}) && (13)\\
&= (\mathbf{DD}^{\text{-1}}) \left(\mathbf{G} \circ (\mathbf{G}^{\text{-U}})^{\mathrm{T}}\right) (\mathbf{EE}^{\text{-1}}) && (14)\\
&= \mathbf{G} \circ (\mathbf{G}^{\text{-U}})^{\mathrm{T}} && (15)\\
&= \mathcal{UC}\text{-}\mathcal{RGA}(\mathbf{G}) && (16)
\end{aligned}$$

where the step of Eq. (12) exploits the unit-consistency property of the UC inverse, which is not satisfied by the MP inverse.

IV. Properties of the UC-RGA

The UC-RGA and the RGA are equivalent when $\mathbf{G}$ is nonsingular because $\mathbf{G}^{\text{-U}} = \mathbf{G}^{\text{-1}}$. Therefore, it is the case of singular $\mathbf{G}$ that remains to be examined. In the 2×2 case singular $\mathbf{G}$ is either the zero matrix or has rank 1, i.e., the system is constrained to one dimension and therefore interactions among the two input-output pairings are scalar-equivalent. The rank-1 case is degenerate but illustrative because the UC-RGA is always

$$\mathcal{UC}\text{-}\mathcal{RGA}(\mathbf{G}) = \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}, \tag{17}$$

where the row and column sums are the same in analogy to the general case for the RGA of a nonsingular matrix. This is not generally the case for singular $\mathbf{G}$, as can be inferred from

$$\mathbf{G} = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}, \tag{18}$$

for which the second row of $\mathcal{RGA}(\mathbf{G})$ must also be zero and therefore cannot have the same sum as the columns (assuming a and b are nonzero). What can be said in general is that the sum of the elements of $\mathcal{RGA}(\mathbf{G})$ will equal the rank of $\mathbf{G}$, and the fact that the sum of every row and column equals 1 when $\mathbf{G}$ is nonsingular is due to the fact that its rows and columns are linearly independent.

A revealing example of the difference between the UC-RGA and the MP-RGA is the case of

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \tag{19}$$

for which the interactions among inputs and outputs clearly should be the same. In this case both the UC-RGA and MP-RGA can be verified to yield the same result

$$\frac{1}{9} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{20}$$

where, as expected, all of the interaction values are the same (and sum to 1, the rank of $\mathbf{G}$). However, if the first row and column are scaled by 2:

$$\mathbf{G} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}, \tag{21}$$

the UC-RGA result is unaffected whereas the MP-RGA gives

$$\frac{1}{9} \cdot \begin{bmatrix} 4 & 1 & 1 \\ 1 & 1/4 & 1/4 \\ 1 & 1/4 & 1/4 \end{bmatrix}, \tag{22}$$

which demonstrates clear unit-scale dependency. The more general case of rectangular $\mathbf{G}$ can be sanity-checked using an arbitrary nonsingular 3×3 matrix

$$\mathbf{A} \;=\; \begin{bmatrix} 7 & 4 & 8 \\ 7 & 2 & 5 \\ 3 & 8 & 8 \end{bmatrix} \tag{23}$$

and

$$\mathbf{B} \;=\; \begin{bmatrix} 21 & 16 & 16 \\ 21 & 8 & 10 \\ 9 & 32 & 16 \end{bmatrix}, \tag{24}$$

which are equivalent up to diagonal scaling of their columns and therefore have the same RGA, which can be verified to be[2]:

$$\mathcal{RGA}(\mathbf{A}) = \mathcal{RGA}(\mathbf{B}) = \begin{bmatrix} \text{-}2.47 & \text{-}2.41 & 5.88 \\ 3.29 & 0.94 & \text{-}3.24 \\ 0.18 & 2.47 & \text{-}1.65 \end{bmatrix}. \tag{25}$$

Thus, one should expect that UC-RGA applied to the block-rectangular matrix $\mathbf{M} = [\mathbf{A}\ \mathbf{B}]$ will produce a result in which its two 3×3 rectangular blocks are identical, and it does indeed satisfy this expectation:

$$\mathcal{UC}\text{-}\mathcal{RGA}(\mathbf{M}) \;=\; \frac{1}{2} \cdot \begin{bmatrix} \text{-}2.47 & \text{-}2.41 & 5.88 & \text{-}2.47 & \text{-}2.41 & 5.88 \\ 3.29 & 0.94 & \text{-}3.24 & 3.29 & 0.94 & \text{-}3.24 \\ 0.18 & 2.47 & \text{-}1.65 & 0.18 & 2.47 & \text{-}1.65 \end{bmatrix}. \tag{26}$$

The MP-RGA result, by contrast, fails this sanity test:

$$\mathcal{MP}\text{-}\mathcal{RGA}(\mathbf{M}) \;=\; \frac{1}{2} \cdot \begin{bmatrix} \text{-}4.47 & \text{-}4.54 & 9.41 & \text{-}0.49 & \text{-}0.28 & 2.35 \\ 5.93 & 1.77 & \text{-}5.18 & 0.66 & 0.11 & \text{-}1.29 \\ 0.32 & 4.65 & \text{-}2.64 & 0.04 & 0.29 & \text{-}0.66 \end{bmatrix} \tag{27}$$

as its first 3×3 block is completely different from the second 3×3 block despite the fact they are derived from matrices that are identical up to a scaling of columns and should both equal the matrix of Equation (25).

V. Discussion

This paper has provided evidence that the Moore-Penrose (MP) pseudoinverse is inappropriate for use to generalize the relative gain array (RGA) for applications involving singular matrices. Given the fact that it has been used in this way for decades, it is natural to ask how such an error could have gone unnoticed. It turns out that questions have been raised in the literature. In [1], for example, the authors express a caveat regarding use of the MP inverse for RGA “control in the least-square sense” and that its implications “should be investigated.”

It is also reasonable to speculate that self-selection has discouraged publication of reports of failed applications of the MP-RGA and that published reports of successful applications may have omitted discussion of engineering efforts applied to mitigate unexplained performance issues. For example, units may have been changed judiciously so that (implicitly) input-output variables receive comparable weight under the least-squares criterion[3].

[2] Numeric values in this example have been rounded to two decimal places. It should be noted that the values in the example matrices of this section are chosen to be integers purely for simplicity in replicating results using code provided in the appendix. The matrices are not intended to represent or be interpreted as meaningful plant matrices, though realistic examples can be generated and examined using code from the appendix.

[3] Various forms of “normalization” are used heuristically in many engineering domains to address unit-dependency issues for which rigorous solutions exist. Such heuristics are often regarded as ordinary aspects of implementation engineering and consequently may not be reported.

It is hoped that the unit-consistent (UC) solution described in this paper will promote renewed interest in applications of the generalized RGA and, more importantly, provide greater rigor and reliability to resulting systems.

APPENDIX A
CODE FOR PROPOSED RGA SOLUTION

The following is an Octave/Matlab implementation of the generalized RGA (adapted from an implementation of the UC-inverse from [11]). The function takes a matrix argument and returns the generalized RGA.

```
function R = ucrga(A)
    tol = 1e-15;
    [m, n] = size(A);
    L = zeros(m, n);    M = ones(m, n);
    S = sign(A);   AA = abs(A);
    idx = find(AA > 0.0);  L(idx) = log(AA(idx));
    idx = setdiff(1 : numel(AA), idx);
    L(idx) = 0;    M(idx) = 0;
    r = sum(M, 2);   c = sum(M, 1);
    u = zeros(m, 1); v = zeros(1, n);
    dx = 2*tol;
    while (dx > tol)
        idx = c > 0;
        p = sum(L(:, idx), 1) ./ c(idx);
        L(:, idx) = L(:, idx) - repmat(p, m, 1) .* M(:, idx);
        v(idx) = v(idx) - p;  dx = mean(abs(p));
        idx = r > 0;
        p = sum(L(idx, :), 2) ./ r(idx);
        L(idx, :) = L(idx, :) - repmat(p, 1, n) .* M(idx, :);
        u(idx) = u(idx) - p;  dx = dx + mean(abs(p));
    end
    dl = exp(u);   dr = exp(v);
    S = S.* exp(L);
    R = A .* transpose(pinv(S) .* (dl * dr)');
end
```

REFERENCES

[1] K.H. Baeka, T.F. Edgar, K. Song, G. Choi, H.K. Cho, and C. Han, "An effective procedure for sensor variable selection and utilization in plasma etching for semiconductor manufacturing," *Computers and Chemical Engineering*, 61, 20-29, 2013.

[2] Ben-Israel, Adi; Greville, Thomas N.E., "Generalized inverses: Theory and Applications," *New York: Springer*, 2003.

[3] E. Bristol, "On a new measure of interaction for multivariable process control," *IEEE Transactions on Automatic Control*, vol. 11, no. 1, pp. 133-134, 1966.

[4] J. Chang and C. Yu, "The relative gain for non-square multivariable systems," *Chemical Engineering Science*, vol. 45, no. 5, pp. 1309-1323, 1990.

[5] Robert Hanson, "Matrices Whose Inverse Contain Only Integers," *The Two-Year College of Mathematics Journal*, Vol 13, No. 1, 1982.

[6] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 2007.

[7] C.R. Johnson and H.M. Shapiro, "Mathematical Aspects of the Relative Gain Array $(A \circ A^{-T})$", *SIAM. J. on Algebraic and Discrete Methods*, 7(4), 627-644, 1985.

[8] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*, 2nd ed., New York: Wiley 2005.

[9] Jeffrey Uhlmann, "A Rank-Preserving Generalized Matrix Inverse for Consistency with Respect to Similarity," *IEEE Control Systems Letters*, ISSN: 2475-1456, 2018.

[10] Jeffrey Uhlmann, "Unit Consistency, Generalized Inverses, and Effective System Design Methods," arXiv:1604.08476v2 [cs.NA] 11 Jul 2017 (2015).

[11] Jeffrey Uhlmann, "A Generalized Matrix Inverse that is Consistent with Respect to Diagonal Transformations," *SIAM Journal on Matrix Analysis* (SIMAX), Vol. 39:3, 2018.

[12] Bo Zhang and Jeffrey Uhlmann, "Applying a New Generalized Matrix Inverse For Stable Control of Robotic Systems," arXiv:1806.01776v1 [eess.SP] 7 May, 2018.