Multivariate Inverse Problems

- imaging Cardiac or Brain electromagnetic functioning,
- fMRI,
- time series of inverse problems,
- problem of providing a regression matrix

For cardiac problem: quasi-static assumptions

$$y = X\beta + \eta, \tag{1}$$

where

- β is vector of e.g. epicardial potentials,
- y is measurement vector, e.g. BSP
- X is the transfer matrix,
- η is noise.

Same as determining a regression vector in univariate multiple linear regression.

$$\mathsf{y} = x'\beta + \mathsf{n}, \tag{2}$$

where

- y is an "endogenous" variable,
- x is a column vector of "exogenous" variables,
- β is the regression vector,
- "noise" n reflects imprecision of the linear dependence assumption.
- estimate regression vector β via design matrix X;
  a row of X corresponds to a component of y

Ill-conditioned X (multicollinearity) $\longrightarrow$ maximum likelihood estimate unstable: maximize $p(y|\beta)$ - requires only knowledge of $p(\eta)$.

- MAP is useful: maximize $p(\beta|y) \propto p(y|\beta)p(\beta)$ - also requiring knowledge of prior $p(\beta)$.
- If β and η are zero mean Gaussian and independent, MAP = Wiener Filter estimate,
$$\beta_{\text{map}} = \left[C_\beta X'(XC_\beta X' + C_\eta)^{-1}\right]y, \tag{3}$$
- Typically given C_β up to a scalar magnitude, and C_η up to a scalar magnitude.

Univariate RR-Axiom : β, η zero mean Gaussian, and $C_\beta \propto I$, $C_\eta \propto I$.

Multivariate Generalization

Above needs to be solved for many (image) vectors (a time series of them), or for many regression vectors, i.e., have

$$Y = XB + N, \tag{4}$$

where the above are matrices, e.g.,

- estimate spatiotemporal epicaridal potentials
  - *X has more columns than rows*,
- find a regression matrix B
  - *X has fewer columns than rows*,

Multivariate linear prediction:

$$\mathsf{y} = x'B + \mathsf{n} \tag{5}$$

y is a row vector of endogenous variables, that (roughly) linearly depend on a row vector of exogenous variables.

How do we solve $Y = XB + N$ in the case of Multicollinearity (ill-conditioned tx matrix)?????

Possibilities:

- Rx as a collection of independent univariate problems
  - *not a justified assumption*
- impose autoregressive assumptions
  - *not a justified assumption*
- Rx as a very large univariate problem
  - *RR or Tikhonov is wasteful! - the problem is ill-posed in space but not time (or, within a regression vector, not across regression vectors)*

$Y = XB + N$ is equivalent to

$$\mathrm{vec}(Y) = (I \otimes X)\mathrm{vec}(B) + \mathrm{vec}(N).$$

B and N independent and zero mean Gaussian random matrices, the MAP estimate is

$$\mathrm{vec}(B)_{\mathrm{map}} = \left\{C_B[I \otimes X]'([I \otimes X]C_B[I \otimes X]' + C_N)^{-1}\right\}\mathrm{vec}(Y), \tag{6}$$

where

$$\begin{aligned} C_B &\equiv \mathcal{E}[\mathrm{vec}(B)\mathrm{vec}(B)'] = \left[\left[\mathcal{E}[B_{:i}(B_{:j})']\right]\right] \\ C_N &\equiv \mathcal{E}[\mathrm{vec}(N)\mathrm{vec}(N)'] \end{aligned} \tag{7}$$

Generalize Univariate-RR Axiom in a **non-informative** way:

> *prior specification of features of the autocovariance matrices must be invariant under a time transformation*

in particular, prior specified features of C_B and $C_{BQ} = [[\mathcal{E}[BQ_{:i}(BQ_{:j})']]]$ must be identical and derive only from the Univariate RR-Axiom (or equivalent), e.g.,

- the derived univariate equation

$$\mathsf{y}q = x'(Bq) + \mathsf{n}q. \tag{8}$$

  must satisfy the Univariate RR-Axiom for any vector q of proper dimension
- $\mathcal{E}[B_{:i}B'_{:j}]$ is symmetic.

Definition 1 *A random matrix is*

- column-simple *if*

  1. *the cross-covariance matrix of any two columns is symmetric*
  2. *there is a unit trace matrix such that the autocovariance matrix of every linear combination of columns of the random matrix is proportional to this unit trace matrix*

- row-simple *if its transpose is column-simple,*
- simple *if it is both row-simple and column-simple.*

Lemma 1 *If a random matrix B is column-simple then its column cross-covariance matrices are proportional to its column autocovariance matrices.*

PROOF: Given some column-simple random matrix B, let A be the unit trace matrix to which the autocovariance matrix of any linear combinations of its columns is proportional. Then the sum of the i-th and j-th columns of B will have autocovariance proportional to A. Since the cross-covariance matrices are symmetric,

$$\mathcal{E}[(B_{:i}+B_{:j})(B_{:i}+B_{:j})'] = \mathcal{E}[B_{:i}(B_{:i})'] + \mathcal{E}[B_{:j}(B_{:j})'] + 2\mathcal{E}[B_{:i}(B_{:j})'] \propto A.$$

Since the first and second matrices after the first equality above are proportional to A, it follows that the third matrix after the first equality is also proportional to A. **QED**

Theorem 1 *A random matrix is simple if and only if it is either row-simple or column-simple.*

PROOF: By Lemma 1, column-simple B has column autocovariance and cross-covariance matrices proportional to the same unit trace matrix. It is sufficient to show that a random matrix B has proportional column autocovariance and column cross-covariance matrices if and only if its row autocovariance and row cross-covariance matrices are proportional to each other. Thus, suppose the row autocovariance and row cross-covariance matrices of a random matrix B are proportional to each other. Let C be a matrix to which the row autocovariance and cross-covariance matrices of B are proportional. Let R be the matrix whose (i, j) entry is the proportionality scalar relating C to the cross-covariance matrix of the i-th and j-th rows of B (it is understood that the (i, i) entry of R is the proportionality constant between C and the autocovariance matrix of the i-th row of B). Then

$$(\mathcal{E}[(B_{i:})'B_{j:}])' = \mathcal{E}[(B_{j:})'B_{i:}] = (R_{ij}C)' = R_{ij}C = R_{ji}C,$$

since C is necessarily symmetric (being proportional to an autocovariance matrix). Thus, R is seen to be symmetric. By definition of C and R, we have

$$\mathcal{E}[(B_{i:})'B_{j:}] = R_{ij}C = \mathcal{E}[(B_{iu}B_{jv})_{u,v}] = (R_{ij}C_{uv})_{u,v}.$$

Thus, $\mathcal{E}[(B_{iu}B_{jv})] = R_{ij}C_{uv}$. Hence,

$$(\mathcal{E}[(B_{iu}B_{jv})])_{i,j} = \mathcal{E}[B_{:u}(B_{:v})'] = (R_{ij}C_{uv})_{i,j} = RC_{uv}.$$

Thus, the cross-covariance matrices of any two columns of B, as well as the column autocovariance matrices, are all proportional to R. Hence, the column autocovariance and column cross-covariance matrices of B are proportional to each other. A similar argument demonstrates that if B has proportional column autocovariance and column cross-covariance matrices, then its row autocovariance and row cross-covariance matrices are proportional to each other. The theorem then follows. **QED**

Generalization of Univariate Axioms in a "non-informative" way leads to the specification that B is simple.

Multivariate Transition Axiom (MTA): *B and N are independent simple random matrices.*

> *MTA is sufficient to provide a complete treatment of the Univariate to Multivariate inverse problem transition*

To deal with the multivariate setting, there are two very desirable properties that B and N could satisfy. The first is

Definition 2 *An $(m \times k)$ random matrix M is* separable *if there exist $(m \times m)$ matrix F and $(k \times k)$ matrix G such that*

$$\mathcal{E}[\text{vec}(M) \otimes \text{vec}(M)'] = F \otimes G.$$

Separability

- implies splitting of the autocovariance matrix into a portion defining the "column prior" G and a portion defining the "row prior" F
- without loss, G can be assumed to have unit trace (the unknown scalar amplitude can be transfered to matrix F in the decomposition) $\longrightarrow$,
- thus, the column prior would be provided fully by the univariate axiom

Secondly,

Definition 3 *An $(m \times k)$ random matrix M is* isotropic *if for any symmetric positive-definite $(k \times k)$ matrix Z there is a scalar γ_z such that*

$$\mathcal{E}[\beta_1' Z \beta_2] = \gamma_z \mathcal{E}[\beta_1' \beta_2],$$

where β_1, β_2 are any two linear combinations of columns of M.

i.e., all directions look the same in the space spanned by the columns of M

Isotropy implies that expressions like $\mathcal{E}[B'B]$ (the row prior of C_B - assuming B is separable) can be derived from expressions like $\mathcal{E}[B'(X'X)B]$, which can be estimated from the data in an obvious way.

Happily, a simple random matrix is both separable and isotropic.

Theorem 2 *If B is simple, then*

$$C_B = \mathcal{E}[\mathrm{vec}(B) \otimes \mathrm{vec}(B)'] = \frac{\mathcal{E}[B'B] \otimes \mathcal{E}[BB']}{\mathcal{E}[\|B\|^2]}. \tag{9}$$

PROOF: Note that the vector outer product in (7) is a Kronecker product, so that the first equation in (9) follows. Since $\mathcal{E}[BB']$ is the sum of the autocovariance matrices of each of the columns of simple B, there is a scalar $\gamma_{i,j}$ such that

$$\mathcal{E}[B_{:i}(B_{:j})'] = \gamma_{i,j}\mathcal{E}[BB']. \tag{10}$$

Taking the trace of both sides, it follows that $\mathcal{E}[(B_{:i})'B_{:j}] = \gamma_{i,j}\mathcal{E}[\|B\|^2]$. Thus,

$$\mathcal{E}[B_{:i}(B_{:j})'] = \frac{\mathcal{E}[(B_{:i})'B_{:j}]}{\mathcal{E}[\|B\|^2]}\mathcal{E}[BB']. \tag{11}$$

Now note that C_B can be written in block matrix form as

$$C_B = \left[\left[\mathcal{E}[B_{:i}(B_{:j})']\right]\right]. \tag{12}$$

Equations (11) and (12) imply (9). **QED**

Theorem 3 *For any nonzero $(p \times m)$ matrix X, and simple $(m \times k)$ random matrix B, the autocovariance matrix of* $\text{vec}(B)$ *is given by*

$$C_B = \frac{\mathcal{E}[B'B] \otimes \mathcal{E}[BB']}{I \odot \mathcal{E}[BB']} = \frac{\mathcal{E}[B'(X'X)B] \otimes \mathcal{E}[BB']}{(X'X) \odot \mathcal{E}[BB']}. \tag{13}$$

PROOF: We have

$$\begin{aligned}
\left(\mathcal{E}[B'(X'X)B]\right)_{ij} &= \sum_{u,v} (X'X)_{uv} \mathcal{E}[B_{ui}B_{vj}] & (14)\\
&= \frac{\mathcal{E}[(B_{:i})'B_{:j}]}{\mathcal{E}[\|B\|^2]} \sum_{u,v} (X'X)_{uv} \left(\mathcal{E}[BB']\right)_{uv} & (15)\\
&= \frac{\mathcal{E}[(B_{:i})'B_{:j}]}{I \odot \mathcal{E}[BB']} \left((X'X) \odot \mathcal{E}[BB']\right), & (16)
\end{aligned}$$

where (14) follows from simple matrix multiplication and rearrangement of terms, and (15) follows from substituting into (14) the (u, v) entry of the matrix on the left-hand-side of (11) as given by the right-hand-side of (11). Equations (16) and (9) imply (13). **QED**

Corollary 1 *A simple random matrix is separable and isotropic.*

PROOF: Equation (9) implies separability, and (16) implies isotropy. **QED**

The above allow one to express C_B in terms of the "measurement" expectations. That is,

Corollary 2 *In the context of (4), a Univariate Axiom and MTA imply that the autocovariance matrix of* $\text{vec}(B)$ *is given by*

$$C_B = \mathcal{E}[B'B] \otimes R_\beta = \frac{\left(\mathcal{E}[Y'Y] - \mathcal{E}[N'N]\right) \otimes R_\beta}{(X'X) \odot R_\beta}, \tag{17}$$

where R_β is the unit trace symmetric matrix given by the Univariate Axiom.

PROOF: Since $\mathcal{E}[BB']$ is the sum of the autocovariance matrices of the columns of B, it follows that $R_\beta = \mathcal{E}[BB']/\text{trace}(\mathcal{E}[BB'])$. The first equality in (17) then follows from the first equality in (13) and the identity

$$\text{trace}(\mathcal{E}[BB']) = I \odot \mathcal{E}[BB'].$$

Dividing the numerator and denominator on the right-hand-side of the second equality in (13) by $\text{trace}(\mathcal{E}[BB'])$, we obtain

$$C_B = \mathcal{E}[B'B] \otimes R_\beta = \frac{\mathcal{E}[(XB)'XB] \otimes R_\beta}{(X'X) \odot R_\beta}. \tag{18}$$

Equation (17) then follows from substituting $(Y - N)$ for XB on the right-hand-side of the above, and noting that $\mathcal{E}[N'(Y-N)]$ and $\mathcal{E}[(Y-N)'N]$ are zero matrices since the entries of N are uncorrelated with the entries of B. **QED**

Since MTA specifies that N is simple, we also have

$$C_N = \mathcal{E}[N'N] \otimes R_\eta, \tag{19}$$

where R_η is the unit trace matrix supplied by the given univariate axiom (e.g., proportional to the identity matrix for the Univariate RR-Axiom or Univariate TR-Axiom).

Thus, given that B and N are simple and independent (MTA),

- the requisite ingredients for estimation of B are supplied by $\mathcal{E}[Y'Y]$, $\mathcal{E}[N'N]$, and the univariate regularization matrices R_β and R_η (see (17)).
- use available data to estimate $\mathcal{E}[Y'Y]$, perhaps up to some set of regularization parameters (whose values are governed by cross-validation principles, since use of a data-generated autocovariance matrix would imply a data reuse method).

$$\text{vec}(B)_{\text{map}} = \left\{C_B[I \otimes X]'([I \otimes X]C_B[I \otimes X]' + C_N)^{-1}\right\}\text{vec}(Y)$$

$$C_B = \mathcal{E}[B'B] \otimes R_\beta = \frac{\left(\mathcal{E}[Y'Y] - \mathcal{E}[N'N]\right) \otimes R_\beta}{(X'X) \odot R_\beta}$$

Complexity reduction

$Y = XB + N$ is equivalent to

$$Y_{:i} = XB_{:i} + N_{:i}, \tag{20}$$

$i = 1, \ldots, n$.

- The Standard Method: assume these equations are independent - low complexity
- this implies that C_B and C_N are block diagonal, i.e. $\mathcal{E}[Y'Y] \propto I$
- In our approach, we do not impose this assumption.

Suppose matrix Q simultaneously diagonalizes both $\mathcal{E}[(XB)'XB]$ and $\mathcal{E}[N'N]$ (and hence $\mathcal{E}[Y'Y]$). Then we have

$$YQ = X(BQ) + NQ$$

with block diagonal C_{BQ} and C_{NQ}, i.e., a the sequence of independent equations

$$YQ_{:i} = X(BQ)_{:i} + NQ_{:i}, \tag{21}$$

$i = 1. \ldots, n$.

In this case,

- $(BQ_{:k})_{\text{map}}$ is obtained from isolated treatment of the k-th equation of sequence (21), becoming the k-th column of $(BQ)_{\text{map}}$.
- The requisite Q is not orthogonal in general - which means that application of Q^{-1} to provide
$$B_{\text{reg}} = (BQ)_{\text{map}} Q^{-1}, \tag{22}$$
can entail an unstable amplification of error.
- For example, suppose $(BQ)_{\text{map}} - BQ = \epsilon$, with $\mathcal{E}[\epsilon] = 0$ and $\mathcal{E}[\|\epsilon\|^2] = \alpha^2 I$. Then
$$\mathcal{E}[\|(BQ)_{\text{map}} Q^{-1} - B\|^2] = \mathcal{E}[\|\epsilon Q^{-1}\|^2] = \alpha^2 \|Q^{-1}\|^2$$

However,

> *the requisite Q is orthogonal if $\mathcal{E}[N'N] \propto I$ (e.g., if the time series noise is white Gaussian.*

In this setting,

- we can write
$$\mathcal{E}[N'N] = \sigma^2 I.$$
- If Q is the orthogonal matrix diagonalizing $\mathcal{E}[Y'Y]$, then by (17) we have that C_{BQ} and C_{NQ} are block diagonal,
- thus, the equations of sequence $YQ_{:i} = X(BQ)_{:i} + NQ_{:i}$ are independent. Hence, they can be individually regularized to find estimates of $BQ_{:i}$, $i = 1, \ldots, n$.
- These estimates are collected as the columns of the regularized solution estimate $(BQ)_{\text{map}}$, and the estimate of B then becomes $(BQ)_{\text{map}} Q'$.

Note: assuming X is noiseless, the component of any column of Y that is not in the range of X cannot be fit by any column of any candidate for the regression matrix B. Evidently, the latter component must represent noise. Thus we can replace (4) by

$$X(X'X)^{\dagger}X'Y = XB + N. \tag{23}$$

(if X is surjective then $X(X'X)^{\dagger}X'Y = Y$).

The requisite Q is orthogonal even under the additional complication that there is white Gaussian noise in the transfer matrix X, as well as white Gaussian measurement (electronic) noise in the data matrix Y.

[[[Thus, suppose

$$Y = (X + N_2)B + N_1. \tag{24}$$

The total noise $N = N_1 + N_2 B$ is not independent of the signal B. Suppose both N_1 and N_2 are composed of zero mean independent and identically distributed Gaussian random variables and B, N_1, N_2 are independent of each other. Then

$$\mathcal{E}[N'N] = \mathcal{E}[N_1'N_1] + \mathcal{E}[B'N_2'N_2B] = \sigma_1^2 I + \sigma_2^2 \mathcal{E}[B'B]. \tag{25}$$

But

$$\mathcal{E}[Y'Y] = \mathcal{E}[(XB)'XB] + \sigma_1^2 I + \sigma_2^2 \mathcal{E}[B'B]. \tag{26}$$

From isotropy, $\mathcal{E}[(XB)'XB]$ is proportional to $\mathcal{E}[B'B]$. Thus, (26) implies that $\mathcal{E}[B'B]$ is a linear combination of $\mathcal{E}[Y'Y]$ and I. This and (25) imply that $\mathcal{E}[N'N]$ is also linear combination of $\mathcal{E}[Y'Y]$ and I. Thus, if Q is the orthogonal matrix diagonalizing $\mathcal{E}[Y'Y]$, then (as in the prior paragraph) the individual equations of (21) are independent of each other, and we again have the complexity reduction.]]]

But

- if there is transfer matrix noise it is no longer true that the component of Y outside the range of the noisy transfer matrix is necessarily entirely noise. In this case, it does not immediately follow that it is desirable to substitute (23) in place of (4).

Thus, in the case of Gaussian measurement noise, and possibly Gaussian transfer matrix noise, Corollary 2 implies that three ingredients are used to efficiently fashion the estimate of B from $Y = XB + N$:

- An orthogonal transform Q which provides the sequence of independent equations $\{YQ_{:i} = X(BQ)_{:i} + NQ_{:i}\}$,
- a matrix R_β (of unit trace) proportional to the autocovariance matrix of any linear combination of columns of B - thereby constituting a regularization matrix for any equation of the sequence in the context of Tikhonov regularization,
- a set of regularization parameters for the above equation sequence.

These three ingredients constitute a kind of "stochastic" singular value decomposition relevant to simple random matrices.

Corollary 3 *For $Y+N_y = (X+N_x)B$, with Gaussian N_y such that $\mathcal{E}[N_y'N_y] = \sigma_y^2 I$, and Gaussian N_x such that $\mathcal{E}[N_x'N_x] = \sigma_x^2 I$, we have*

$$B = PDQ', \tag{27}$$

where Q is the orthogonal matrix diagonalizing $\mathcal{E}[Y'Y]$, random matrix P is such that $\mathcal{E}[\text{vec}(P)(\text{vec}(P))'] = I \otimes R_\beta$ with R_β of unit trace, and diagonal D has diagonal elements $\{d_i\}$ such that

$$d_i^2 = \frac{\psi_i^2 - \sigma_y^2}{X'X \odot R_\beta + \sigma_x^2}, \tag{28}$$

where $\{\psi_i^2\}$ is the set of eigenvalues of $\mathcal{E}[Y'Y]$.

PROOF: Let Q be the orthogonal matrix diagonalizing $\mathcal{E}[Y'Y]$. As indicated in the prior discussion, BQ is an isotropic random matrix whose column autocovariance matrices are proportional to unit trace R_β, and whose column cross-covariance matrices are zero. Thus, $BQ = PD$, where $\mathcal{E}[\text{vec}(P)(\text{vec}(P))'] = I \otimes R_\beta$, and D is diagonal. From (13), and (26),

$$\mathcal{E}[B'B] = \frac{\mathcal{E}[Y'Y] - \sigma_y^2 I}{X'X \odot R_\beta + \sigma_x^2},$$

so the squares of the proportionality constants (diagonal entries of D^2) are given by (28). **QED**

$$YQ_{:i} = X(BQ)_{:i} + NQ_{:i},$$

$i = 1, 2, \dots, n$

- For the transfer matrix noise case, the individual equations (21) can be approached by regularized total least square methodology, however, for the case of zero-order Tikhonov regularization this adds nothing [Golub et al, 1999].

- Since $\mathcal{E}[Y'Y]$ is not given, we take Q to be the matrix of eigenvectors of $Y'X(X'X)^{\dagger}X'Y$, unless there is transfer matrix noise (i.e., $N_x = 0$), in which case we take Q to be the matrix of eigenevectors of $Y'Y$.

- A sequence of "regularization parameters" derives from the diagonal entries of D. Thus, for the i-th equation of sequence (21), the signal autocovariance matrix is $d_i^2 R_\beta$, and by (25) the noise autocovariance matrix is $\sigma_y^2 I + \sigma_x^2 d_i^2 R_\beta$. In the setting where $\mathcal{E}[Y'Y]$ is estimated from the data as $Y'Y$, they are the regularization parameters associated with a maximum likelihood choice for the autocovariance matrix. However, this is a "data re-use" method, and thereby subject to modifications based on cross-validation (or other) principles.

- $B = PDQ$ can be considered to be a stochastic generalization of the SVD relevant to isotropic random matrices. As with other generalized SVDs (e.g., Product and Quotient SVDs, associated with equations of the form $Y = XB$), this "Stochastic SVD" is associated with its own equation, the noise-corrupted $(Y + N_y) = (X + N_x)B$, where N_x and N_y are white Gaussian.

- If there is some *a priori* temporal information available, so that one is supplied with a prior nontrivial estimate of a matrix proportional to $\mathcal{E}[B'B]$, then Q should be taken as the orthogonal matrix diagonalizing this latter matrix estimate, rather than the matrix diagonalizing the data-derived $Y'Y$ or $Y'\left(X(X'X)^{\dagger}X'\right)Y$. Recall that $Y'Y$ is only an estimate of $\mathcal{E}[Y'Y]$. Random matrix Y is ultimately dependent on random matrix B via $Y = XB + N$. In particular, note that (13) and (17) imply that $\mathcal{E}[B'B]$ is proportional to $(\mathcal{E}[Y'Y] - \mathcal{E}[N'N])$. Thus, a prior favored estimate of $\mathcal{E}[B'B]$ could more accurately reflect the relevant eigenvectors.

Appendix

$$W \odot Z = \text{trace}[W'Z]$$

$$C_B = \frac{\mathcal{E}[B'B] \otimes \mathcal{E}[BB']}{I \odot \mathcal{E}[BB']} = \frac{\mathcal{E}[B'(X'X)B] \otimes \mathcal{E}[BB']}{(X'X) \odot \mathcal{E}[BB']}$$

The matrix scalar product $W \odot Z \equiv \text{trace}(W'Z)$ has a useful role in making evident the cancellation property of simple random matrices (equation (13)). Note that $W \odot Z$ requires that W and Z have identical dimensions. The standard matrix product $W'Z$ requires only that the column dimension of W and Z be the same, while $W \otimes Z$ makes no demands whatever on the respective number of either rows or columns. Thus, we have a hierarchy of matrix products.

The following properties of $\odot$ are easily derived (we assume the matrices have compatible dimensions wherever required in the following expressions):

1. $W \odot Z = \text{vec}(W)'\text{vec}(Z)$,
2. $W \odot Z = Z \odot W$, (commutivity),
3. $(W + Z) \odot R = W \odot R + Z \odot R$, (distributivity)
4. $W \odot Z = W' \odot Z'$,
5. $W \odot W = \|W\|^2 = I \odot (WW')$,
6. $W \odot Z = (V^{-1})'WU' \odot VZU^{-1}$,
7. $(WZ) \odot R = Z \odot (W'R)$,
8. $(W \otimes Z) \odot (R \otimes S) = (W \odot R)(Z \odot S)$,
9. if W and Z are symmetric non-negative definite, $W \odot Z$ is the sum of the (generalized) product singular values of the pair (W, Z).

#1 through #7 are immediate. #9 derives from the existence of a matrix X such that $W = XD_wX'$, $Z = (X^{-1})'D_zX^{-1}$, where D_w, D_z are diagonal (guaranteed from the Product SVD). Demonstration of #8 derives from the definition of $\odot$ and three well-known identities concerning Kronecker products (transpose of a Kronecker product is the Kronecker product of transposes; the mixed product rule; the trace of a Kronecker product is the product of the traces). Thus,

$$\begin{aligned}(W \otimes Z) \odot (R \otimes S) &= \text{trace}\big((W' \otimes Z')(R \otimes S)\big) \\ &= \text{trace}\big((W'R) \otimes (Z' \otimes S)\big) \\ &= \text{trace}(W'R)\text{trace}(Z'S) \\ &= (W \odot R)(Z \odot S).\end{aligned}$$