Canonical Correlation Analysis in high dimensions with structured regularization (supplementary materials)

Elena Tuzhilina [1], Leonardo Tozzi [2], and Trevor Hastie [3]

[1] Department of Statistics, Stanford University, Stanford, CA, USA

[2] Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA

[3] Department of Statistics, Stanford University, Stanford, CA, USA

S1 Proof of RCCA kernel trick lemma

Lemma (RCCA kernel trick) The original RCCA problem stated for $\mathbf{X}$ and $\mathbf{Y}$ can be reduced to solving the RCCA problem for $\mathbf{R}$ and $\mathbf{Y}$. The resulting canonical correlations and variates for these two problems coincide. The canonical coefficients for the original problem can be recovered via the linear transformation $\alpha_X = V\alpha_R$.

Proof. Denote $V^\perp$ an orthogonal complement of matrix V, i.e. the matrix $V^\perp \in \mathbb{R}^{p\times(p-n)}$ such that $\widetilde{V} = \left(V, V^\perp\right) \in \mathbb{R}^{p\times p}$ is a full-rank orthogonal matrix. Then $V^\top\widetilde{V} = (I_n, 0)$. Denote also

$$\gamma = \widetilde{V}^\top\alpha = \begin{pmatrix} V^\top\alpha \\ \left(V^\perp\right)^\top\alpha \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}.$$

Note that since there is a one-to-one correspondence betwineen α and γ, the optimization of $\rho_{RCCA}(\alpha, \beta; \lambda_1, \lambda_2)$ w.r.t. to α is equivalent to optimization w.r.t. γ. Further, the following relation is true

$$\alpha^\top \frac{\mathbf{X}^\top \mathbf{Y}}{n} \beta = \gamma_1^\top \frac{\mathbf{R}^\top \mathbf{Y}}{n} \beta$$
$$\alpha^\top \left(\frac{\mathbf{X}^\top \mathbf{X}}{n} + \lambda_1 I \right) \alpha = \gamma_1^\top \frac{\mathbf{R}^\top \mathbf{R}}{n} \gamma_1 + \lambda_1 \gamma_1^\top \gamma_1 + \lambda_1 \gamma_2^\top \gamma_2$$

Therefore, the correlation coefficient (2.4) with $\lambda_2 = 0$ can be rewritten in terms of R and Y as

$$\rho_{RCCA}(\gamma, \beta; \lambda_1) = \frac{\gamma_1^\top \widehat{\Sigma}_{RY} \beta}{\sqrt{\gamma_1^\top (\widehat{\Sigma}_{RR} + \lambda_1 I)\gamma_1 + \lambda_1 \|\gamma_2\|^2} \sqrt{\beta^\top \widehat{\Sigma}_{YY} \beta}}.$$

It is easy to show that the maximum value of $\rho_{RCCA}(\gamma, \beta; \lambda_1)$ is attained when $\gamma_2 = 0$, so the above correlation coefficient is nothing but the RCCA correlation coefficient computed for $\mathbf{R}$ and $\mathbf{Y}$. Furthermore, the optimal value of $\alpha_X = \alpha$ can be recovered from $\alpha_R = \gamma_1$ by $\alpha = V\gamma_1$ and, since $\mathbf{X}\alpha = \mathbf{R}\gamma_1$, the canonical variates computed for $\mathbf{X}$ coincide with the ones computed for $\mathbf{R}$. $\square$

S2 Proof of PRCCA Kernel trick

LEMMA (PRCCA Kernel Trick) The original PRCCA problem stated for $\mathbf{X}$ and $\mathbf{Y}$ can be reduced to solving the PRCCA problem for $\mathbf{R} = \binom{\mathbf{R}_1}{\mathbf{X}_2} \in \mathbb{R}^{n+p_2}$ and $\mathbf{Y}$. The resulting canonical correlations and variates for these two problems coincide. The canonical coefficients for the original problem can be recovered via the linear transformation $\alpha_X = A\begin{pmatrix} V_1 & 0 \\ 0 & I \end{pmatrix}\alpha_R$.

Proof. To find the required linear transformation A we first regress $\mathbf{X}_2$ from $\mathbf{X}_1$. We denote the matrix of regression coefficients by $B = (\mathbf{X}_2^\top \mathbf{X}_2)^{-1}\mathbf{X}_2^\top \mathbf{X}_1 \in \mathbb{R}^{p_2 \times p_1}$ and set $A = \left(\begin{smallmatrix} I & 0 \\ -B & I \end{smallmatrix}\right)$. This transformation leads to the following transformed matrix

$$\widetilde{\mathbf{X}} = \mathbf{X}A = (\widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2) = (\mathbf{X}_1 - \mathbf{X}_2 B, \mathbf{X}_2)\,.$$

It is easy to check that matrix A is invertible and that the following relations hold

$$A^{-1} = \left(\begin{smallmatrix} I & 0 \\ B & I \end{smallmatrix}\right) \text{ and } A^{-T}\left(\begin{smallmatrix} I_{p_1} & 0 \\ 0 & 0 \end{smallmatrix}\right)A^{-1} = \left(\begin{smallmatrix} I_{p_1} & 0 \\ 0 & 0 \end{smallmatrix}\right).$$

Thus, if we denote $\widetilde{\alpha} = A^{-1}\alpha$, then

$$\begin{aligned} &\alpha^\top \frac{\mathbf{X}^\top \mathbf{Y}}{n}\beta = \widetilde{\alpha}^\top \frac{\widetilde{\mathbf{X}}^\top \mathbf{Y}}{n}\beta \\ &\alpha^\top \left(\frac{\mathbf{X}^\top \mathbf{X}}{n} + \lambda_1 \left(\begin{smallmatrix} I_{p_1} & 0 \\ 0 & 0 \end{smallmatrix}\right)\right)\alpha = \widetilde{\alpha}^\top \left(\frac{\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}}}{n} + \lambda_1 \left(\begin{smallmatrix} I_{p_1} & 0 \\ 0 & 0 \end{smallmatrix}\right)\right)\widetilde{\alpha} \end{aligned}$$

The above equations imply that the PRCCA correlation coefficient (4.2) with $\lambda_2 = 0$ can be rewritten in terms of $\widetilde{\mathbf{X}}$ and $\mathbf{Y}$

$$\rho_{PRCCA}(\widetilde{\alpha}, \beta; \lambda_1) = \frac{\widetilde{\alpha}^\top \widehat{\Sigma}_{\widetilde{X}Y}\beta}{\sqrt{\widetilde{\alpha}^\top \left(\widehat{\Sigma}_{\widetilde{X}\widetilde{X}} + \lambda_1 \left(\begin{smallmatrix} I_{p_1} & 0 \\ 0 & 0 \end{smallmatrix}\right)\right)\widetilde{\alpha}}\;\sqrt{\beta^\top \widehat{\Sigma}_{YY}\beta}}.$$

Next, let $\widetilde{\alpha} = \left(\begin{smallmatrix} \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \end{smallmatrix}\right)$, where $\widetilde{\alpha}_1 \in \mathbb{R}^{p_1}$ and $\widetilde{\alpha}_2 \in \mathbb{R}^{p_2}$ correspond to blocks $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{X}}_2$, respectively. Denote the orthogonal complement of V_1 by $V_1^\perp \in \mathbb{R}^{p_1 \times (p_1 - n)}$ and consider the following transformation of the PRCCA coefficients

$$\gamma_1 = V_1^\top \widetilde{\alpha}_1 \text{ and } \gamma_2 = \left(V_1^\perp\right)^\top \widetilde{\alpha}_1$$

as well as the concatenation $\gamma = \left(\begin{smallmatrix}\gamma_1 \\ \tilde{\alpha}_2\end{smallmatrix}\right)$. Then, by analogy one can show that

$$\tilde{\alpha}^\top \frac{\widetilde{\mathbf{X}}^\top \mathbf{Y}}{n}\beta = \gamma_1^\top \frac{\mathbf{R}_1^\top \mathbf{Y}}{n}\beta + \tilde{\alpha}_2^\top \frac{\widetilde{\mathbf{X}}_2^\top \mathbf{Y}}{n}\beta = \gamma^\top \frac{\mathbf{R}^\top \mathbf{Y}}{n}\beta$$
$$\tilde{\alpha}^\top \left(\frac{\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}}}{n} + \lambda_1 \left(\begin{smallmatrix} I_{p_1} & 0 \\ 0 & 0\end{smallmatrix}\right)\right)\tilde{\alpha} = \gamma_1^\top \left(\frac{\mathbf{R}_1^\top \mathbf{R}_1}{n} + \lambda_1 I\right)\gamma_1 + \lambda_1 \gamma_2^\top \gamma_2 + \tilde{\alpha}_2^\top \frac{\widetilde{\mathbf{X}}_2^\top \widetilde{\mathbf{X}}_2}{n}\tilde{\alpha}_2 =$$
$$= \gamma^\top \left(\frac{\mathbf{R}^\top \mathbf{R}}{n} + \lambda_1 \left(\begin{smallmatrix} I_{p_1} & 0 \\ 0 & 0\end{smallmatrix}\right)\right)\gamma + \lambda_1 \gamma_2^\top \gamma_2$$

where the last equation holds since $\mathbf{R}_1^\top \widetilde{\mathbf{X}}_2 = V_1^\top \widetilde{\mathbf{X}}_1^\top \widetilde{\mathbf{X}}_2 = 0$.

Again, we can ignore $\lambda_1 \gamma_2^\top \gamma_2$ term as it is present only in the denominator of PRCCA correlation coefficient, which, therefore, can be rewritten in terms of $\mathbf{R}$ and $\mathbf{Y}$ as

$$\rho_{PRCCA}(\gamma, \beta; \lambda_1) = \frac{\gamma^\top \widehat{\Sigma}_{RY}\beta}{\sqrt{\gamma^\top \left(\widehat{\Sigma}_{RR} + \lambda_1 \left(\begin{smallmatrix} I_{p_1} & 0 \\ 0 & 0\end{smallmatrix}\right)\right)\gamma}\ \sqrt{\beta^\top \widehat{\Sigma}_{YY}\beta}},$$

which is exactly the PRCCA correlation coefficient computed for $\mathbf{R}$ and $\mathbf{Y}$. The optimal value of $\alpha_X = \alpha$ for the original problem can be recovered from $\alpha_R = \gamma$ by

$$\alpha = A\tilde{\alpha} = A\left(\begin{smallmatrix} V_1\gamma_1 \\ \tilde{\alpha}_2\end{smallmatrix}\right) = A\left(\begin{smallmatrix} V_1 & 0 \\ 0 & I\end{smallmatrix}\right)\gamma.$$

Moreover,

$$\mathbf{X}\alpha = \widetilde{\mathbf{X}}\tilde{\alpha} = (\mathbf{R}_1, \widetilde{\mathbf{X}}_2)\left(\begin{smallmatrix} V_1^\top \tilde{\alpha}_1 \\ \tilde{\alpha}_2\end{smallmatrix}\right) = \mathbf{R}\gamma,$$

so the canonical variates computed for $\mathbf{X}$ coincide with the ones computed for $\mathbf{R}$. □

S3 Proof of General RCCA to RCCA/PRCCA lemma

Lemma (General RCCA to RCCA/PRCCA) If both K_X and K_Y are positive definite then, by some proper change of basis, the general RCCA problem can be reduced to the RCCA one. Alternatively, if one of K_X and K_Y has zero eigenvalues then general RCCA boils down to solving the PRCCA problem with number of unpenalized coefficients equal to the multiplicity of zero eigenvalue.

Proof. As usual, we do not penalize Y part assuming $K_Y = 0$. Consider the eigendecomposition of matrix K_X, i.e. $K_X = UDU^\top$ with orthogonal $U \in \mathbb{R}^{p\times p}$ and diagonal $D \in \mathbb{R}^{p\times p}$. Since K_X is supposed to be a positive semi-definite matrix, D has non-negative diagonal elements. Applying transformation $\widetilde{\mathbf{X}} = \mathbf{X}U$ and $\widetilde{\alpha} = U^\top\alpha$, and using the same reasoning as in S1 and S2 we obtain

$$\alpha^\top \frac{\mathbf{X}^\top\mathbf{Y}}{n}\beta = \widetilde{\alpha}^\top \frac{\widetilde{\mathbf{X}}^\top\mathbf{Y}}{n}\beta$$

$$\alpha^\top \left(\frac{\mathbf{X}^\top\mathbf{X}}{n} + K_X\right)\alpha = \widetilde{\alpha}^\top \left(\frac{\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}}{n} + D\right)\widetilde{\alpha}$$

thus the equivalent modified correlation coefficient in the new basis is

$$\rho(\widetilde{\alpha}, \beta; D) = \frac{\widetilde{\alpha}^\top \widehat{\Sigma}_{\widetilde{X}Y}\beta}{\sqrt{\widetilde{\alpha}^\top\left(\widehat{\Sigma}_{\widetilde{X}\widetilde{X}} + D\right)\widetilde{\alpha}}\ \sqrt{\beta^\top\widehat{\Sigma}_{YY}\beta}}.$$

Further, we decompose remaining diagonal matrix as $D = SLS$ as follows. We have two cases:

1. If all diagonal elements of D are positive then we put $S = D^{\frac{1}{2}}$ and $L = I$.

2. Suppose the first p_1 elements of D are positive and the rest $p_2 = p - p_1$ elements are zero. Let $D = \left(\begin{smallmatrix} D_{11} & 0 \\ 0 & 0 \end{smallmatrix}\right)$, where $D_{11} \in \mathbb{R}^{p_1 \times p_1}$ is the block containing all positive diagonal elements of matrix D. Then we can set

$$S = \begin{pmatrix} D_{11}^{\frac{1}{2}} & 0 \\ 0 & I \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} I_{p_1} & 0 \\ 0 & 0 \end{pmatrix}$$

Note that, unlike D, matrix S is non-singular, so the change of basis $\widetilde{\widetilde{\mathbf{X}}} = \widetilde{\mathbf{X}} S^{-1}$ and $\widetilde{\widetilde{\alpha}} = S\widetilde{\alpha}$ is well-defined and leads to the following equalities

$$\widetilde{\alpha}^\top \frac{\widetilde{\mathbf{X}}^\top \mathbf{Y}}{n} \beta = \widetilde{\widetilde{\alpha}}^\top \frac{\widetilde{\widetilde{\mathbf{X}}}^\top \mathbf{Y}}{n} \beta$$
$$\widetilde{\alpha}^\top \left(\frac{\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}}}{n} + D \right) \widetilde{\alpha} = \widetilde{\widetilde{\alpha}}^\top \left(\frac{\widetilde{\widetilde{\mathbf{X}}}^\top \widetilde{\widetilde{\mathbf{X}}}}{n} + L \right) \widetilde{\widetilde{\alpha}}.$$

Therefore, the equivalent modified correlation coefficients

$$\rho(\widetilde{\widetilde{\alpha}}, \beta; L) = \frac{\widetilde{\widetilde{\alpha}}^\top \widehat{\Sigma}_{\widetilde{\widetilde{X}}\widetilde{\widetilde{X}}} \widetilde{\widetilde{\alpha}}}{\sqrt{\widetilde{\widetilde{\alpha}}^\top \left(\widehat{\Sigma}_{\widetilde{\widetilde{X}}\widetilde{\widetilde{X}}} + L \right) \widetilde{\widetilde{\alpha}}} \sqrt{\beta^\top \widehat{\Sigma}_{YY} \beta}}.$$

Is it easy to see that in the first case when $L = I$ the above correlation coefficient coincides with the RCCA correlation coefficient with $\lambda_1 = 1$. Alternatively, if $L = \left(\begin{smallmatrix} I_{p_1} & 0 \\ 0 & 0 \end{smallmatrix}\right)$ it is equal to the PRCCA correlation coefficient with $\lambda_1 = 1$ and p_2 unpenalized coefficients. Thus, we conclude that the General RCCA solution computed for X and Y can be found by means of either RCCA or PRCCA method applied to $\widetilde{\widetilde{X}}$ and Y. The canonical variates stay the same regardless the basis as $\widetilde{\widetilde{\mathbf{X}}}\widetilde{\widetilde{\alpha}} = \widetilde{\mathbf{X}}\widetilde{\alpha} = \mathbf{X}\alpha$. The corresponding inverse transform for the coefficients is $\alpha = SU\widetilde{\widetilde{\alpha}}$. □

S4 Link between GRCCA and RCCA/PRCCA via the SVD of the penalty matrix

Since GRCCA is just a special case of general RCCA, one can use previous lemma to map the GRCCA problem to either RCCA or PRCCA and, subsequesntly, find the canonical coefficients via the kernel trick. However, to find this transformation it is required to do an extra step: the eigendecomposition of the kernel matrices $K_X(\lambda_1, \mu_1)$ and $K_Y(\lambda_2, \mu_2)$. Although this can be infeasible in high dimensions for general penalty matrices it turns out that one can use the specific structure of the GRCCA penalty matrix to get around this eigendecomposition.

We again assume that the regularization was imposed on the X part only (i.e. $\lambda_2 = \mu_2 = 0$); however, it is not difficult to derive similar results for the general case. Recall that matrix $C_m = \frac{\mathbb{1}\mathbb{1}^\top}{m} \in \mathbb{R}^{m\times m}$ has

- unit eigenvalue with corresponding eigenvector $\frac{\mathbb{1}}{\sqrt{m}} \in \mathbb{R}^m$,
- zero eigenvalue with corresponding eigenspace $\left[\frac{\mathbb{1}}{\sqrt{m}}\right]^\perp \in \mathbb{R}^{m\times(m-1)}$.

Here $[A]^\perp$ refers to the orthogonal complement of matrix A. The resulting eigendecomposition is therefore

$$C_m = U_m \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) U_m^\top \quad \text{with} \quad U_m = \left(\tfrac{\mathbb{1}}{\sqrt{m}} \; \left[\tfrac{\mathbb{1}}{\sqrt{m}}\right]^\perp \right).$$

It is easy to show the following eigendecomposition as well

$$\lambda_1(I - C_m) + \mu_1 C_m = U_m D_m U_m^\top \quad \text{with} \quad D_m = \left(\begin{smallmatrix} \mu_1 & 0 \\ 0 & \lambda_1 I_{m-1} \end{smallmatrix}\right).$$

Thus the penalty matrix $K_X(\lambda_1, \mu_1)$ can be decomposed as

$$K_X(\lambda_1, \mu_1) = UDU^\top \quad \text{with} \quad U = U_{p_1} \oplus \ldots \oplus U_{p_K} \quad \text{and} \quad D = D_{p_1} \oplus \ldots \oplus D_{p_K}.$$

Using the lemma from Section S3 we conclude that if $\lambda_1, \mu_1 > 0$ then the GRCCA problem can be solved via the RCCA approach. Alternatively, if $\lambda_1 = 0$ or $\mu_1 = 0$ it can be reduced to the PRCCA problem with $n - K$ and K unpenalized coefficients, respectively. Hereafter we will assume $\lambda_1 > 0$, i.e. the presence of group homogeneity.

Note that due to the specific structure of $K_X(\lambda_1, \mu_1)$ the eigendecomposition can be calculated block-wise. Moreover, the resulting computational cost is equal to the cost of computing the orthogonal complements $\left[\frac{\mathbb{1}}{\sqrt{p_1}}\right]^\perp, \ldots, \left[\frac{\mathbb{1}}{\sqrt{p_K}}\right]^\perp$, which can be efficiently found via, for example, Helmert contrasts (c.f. `contr.helmert()` function in R) and without computing any eigendecomposition.

S5 Link between GRCCA and RCCA/PRCCA via the feature matrix extension

In Section S4 we already proposed the way to solve GRCCA problem. If the number of groups K is rather small then penalty matrix $K_X(\lambda_1, \mu_1)$ will consist of a few large blocks. In this case it would be quite expensive to compute the orthogonal complements and to further apply the linear transformation mapping General RCCA to RCCA/PRCCA. It turns out that there is an alternative linear transformation that leads to equivalent RCCA/PRCCA problem while being less expensive.

We need to establish some notation first. Suppose that matrix $\mathbf{X}$ is divided into blocks according to groups, i.e. $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_K)$. Denote the column average matrix for group k by $\bar{\mathbf{X}}_k = \mathbf{X}_k \frac{\mathbb{1}}{p_k}$ and consider an extended matrix $\widetilde{\mathbf{X}}$ that consists of (scaled) mean centered blocks and extra K columns corresponding to (scaled) group means

$$\widetilde{\mathbf{X}}(a,b) = \left(\sqrt{\frac{1}{a}}\left(\mathbf{X}_1 - \bar{\mathbf{X}}_1\right), \sqrt{\frac{p_1}{b}}\bar{\mathbf{X}}_1, \dots, \sqrt{\frac{1}{a}}\left(\mathbf{X}_K - \bar{\mathbf{X}}_K\right), \sqrt{\frac{p_K}{b}}\bar{\mathbf{X}}_K\right) \in \mathbb{R}^{n \times (p+K)}.$$

Lemma (GRCCA to RCCA/PRCCA) If $\mu_1 > 0$ the GRCCA problem for $\mathbf{X}$ and $\mathbf{Y}$ can be reduced to solving the RCCA problem for $\widetilde{\mathbf{X}}(\lambda_1, \mu_1)$ and $\mathbf{Y}$. If $\mu_1 = 0$, then GRCCA boils down to the PRCCA problem for $\widetilde{\mathbf{X}}(\lambda_1, 1)$ and $\mathbf{Y}$ with K unpenalized coefficients.

Proof. Let us prove the statement for $K = 1$, one can easily extend the proof for arbitrary K. For $K = 1$ the penalty matrix is

$$K_X(\lambda_1, \mu_1) = \lambda_1 \left(I - \frac{\mathbb{1}\mathbb{1}^\top}{n}\right) + \mu_1 \frac{\mathbb{1}\mathbb{1}^\top}{n}$$

which can be decomposed as

$$K_X(\lambda_1, \mu_1) = \widetilde{U}\widetilde{D}\widetilde{U}^\top \text{ with } \widetilde{U} = \left(I - \tfrac{\mathbb{1}\mathbb{1}^\top}{n} \; \tfrac{\mathbb{1}}{\sqrt{n}}\right) \text{ and } \widetilde{D} = \left(\begin{smallmatrix} \lambda_1 I_n & 0 \\ 0 & \mu_1 \end{smallmatrix}\right).$$

This decomposition can be considered as an alternative to the eigendecomposition $K_X = UDU^\top$ discussed in the previous section. However, unlike matrix U which was square and orthogonal, $\widetilde{U}$ is a rectangular $n \times (n+1)$ matrix with orthogonal rows, i.e. $\widetilde{U}\widetilde{U}^\top = I$.

Further, similar to the previous section, we do decomposition $\widetilde{D} = \widetilde{S}\widetilde{L}\widetilde{S}$. We again have two cases here. If $\mu_1 > 0$ then we have

$$\widetilde{S} = \begin{pmatrix} \sqrt{\lambda_1} I_n & 0 \\ 0 & \sqrt{\mu_1} \end{pmatrix} \quad \text{and} \quad \widetilde{L} = I$$

and the following relation is true

$$\mathbf{X}\widetilde{U}\widetilde{S}^{-1} = \left(\sqrt{\tfrac{1}{\lambda_1}} \left(\mathbf{X} - \bar{\mathbf{X}} \right), \sqrt{\tfrac{n}{\mu_1}} \bar{\mathbf{X}} \right) = \widetilde{\mathbf{X}}(\lambda_1, \mu_1).$$

If, on the contrary, $\mu_1 = 0$ then

$$\widetilde{S} = \begin{pmatrix} \sqrt{\lambda_1} I_n & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \widetilde{L} = \begin{pmatrix} I_n & 0 \\ 0 & 0 \end{pmatrix}$$

which leads us to

$$\mathbf{X}\widetilde{U}\widetilde{S}^{-1} = \left(\sqrt{\tfrac{1}{\lambda_1}} \left(\mathbf{X} - \bar{\mathbf{X}} \right), \sqrt{n} \bar{\mathbf{X}} \right) = \widetilde{\mathbf{X}}(\lambda_1, 1).$$

Note that the proof in Section S3 requires matrix $\widetilde{U}$ to have orthogonal rows only. Thus, following this proof, one can conclude that the original GRCCA problem is equivalent to the RCCA problem solved for $\widetilde{\mathbf{X}}(\lambda_1, \mu_1)$ and $\mathbf{Y}$ if $\mu_1 > 0$. Alternatively, if $\mu_1 = 0$ then GRCCA boils down to solving the PRCCA problem for $\widetilde{\mathbf{X}}(\lambda_1, 1)$ and $\mathbf{Y}$, and one unpenalized coefficient. Again, such change of basis does not influence the canonical variates, whereas the canonical coefficients are transformed according to $\alpha = \widetilde{U}\widetilde{S}^{-1}\widetilde{\alpha}$. □

Note that the corresponding data transformation boils down to computing the group means and adjusting the feature matrix by group means. Although this approach can be considered as convenient alternative to the one suggested in Section S4, there is a trade-off. On the one hand, we reduce the cost by getting around the eigendecomposition (computing the orthogonal complements). On the other hand, we increase the feature matrix dimension from p to $p + K$, which can be quite inefficient for large K.

S6 Neuroimaging analysis

Details about the protocol and measures collected by HCP-DES are outlined in Tozzi et al. (2020). Here, only the details relevant to this study are discussed.

6.1 Neuroimaging acquisition details

Images were acquired at the Stanford Center for Cognitive and Neurobiological Imaging (CNI) on a GE Discovery MR750 3 T scanner using a Nova Medical 32-channel head coil. Two spin-echo fieldmaps were acquired at the beginning of each session, one with a posterior-anterior phase encoding direction, the other with an anterior-posterior direction. All fMRI scans were conducted using a blipped-CAIPI simultaneous multislice "multiband" acquisition (Setsompop et al., 2012).

1. Spin-echo fieldmaps: TE = 55.5 ms, TR = 6 s, FA = 90°, acquisition time = 18 s, field of view = 220.8 × 220.8 mm, 3D matrix size = 92 × 92 × 60, slice orientation = axial, angulation to anterior commissure - posterior commissure (AC-PC) line, phase encoding = AP and PA, receiver bandwidth = 250 kHz, readout duration = 49.14 ms, echo spacing = 0.54 ms, voxel size = 2.4 mm isotropic.

2. Single-band calibration: TE = 30 ms, TR = 4.4 s, FA = 90°, acquisition time = 13 s, field of view = 220.8 × 220.8 mm, 3D matrix size = 92 × 92 × 60, slice orientation = axial, angulation to AC-PC line, phase encoding = PA,

receiver bandwidth = 250 kHz, readout duration = 49.14 ms, echo spacing = 0.54 ms, number of volumes = 4, voxel size = 2.4 mm isotropic.

3. Multiband fMRI: TE = 30 ms, TR = 0.71 s, FA = 54°, acquisition time = 3:44 (Gambling task), field of view = 220.8 × 220.8 mm, 3D matrix size = 92 × 92 × 60, slice orientation = axial, angulation to AC-PC line, phase encoding = PA, receiver bandwidth = 250 kHz, readout duration = 49.14 ms, echo spacing = 0.54 ms, number of volumes for Gambling task = 316, multiband factor = 6, calibration volumes = 2, voxel size = 2.4 mm isotropic.

4. T1-weighted: TE = 3.548 ms, MPRAGE TR = 2.84 s, FA = 8°, acquisition time = 8:33, field of view = 256 × 256 mm, 3D matrix size = 320 × 320 × 230, slice orientation = sagittal, angulation to AC-PC line, receiver bandwidth = 31.25 kHz, fat suppression = no, motion correction = PROMO, voxel size = 0.8 mm isotropic.

5. T2-weighted: TE = 74.4 ms, TR = 2.5 s, FA = 90°, acquisition time = 5:42, field of view = 240 × 240 mm, 3D matrix size = 320 × 320 × 216, slice orientation = sagittal, angulation to AC-PC line, receiver bandwidth = 125 kHz, fat suppression = no, motion correction = PROMO, voxel size = 0.8 mm isotropic.

6.2 Gambling task paradigm

The HCP-DES adopts a version of the HCP Gambling task modified to allow comparison of small and large gain and loss outcomes (Somerville et al., 2018; Tozzi et al., 2020). A question mark is displayed on the screen and the participant must guess whether a number is greater than or less than five (and indicate their answer via button presses). If the participant identifies correctly, they win money, and if they guess incorrectly, they lose money. At the end of the task, 5 trials are randomly selected and summed together to determine the participant's payment.

6.3 Preprocessing

Raw image files were converted to BIDS format and preprocessed using fMRIPrep (Esteban et al., 2019). Briefly, brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, Dale et al. (1999)). Susceptibility distortion for fMRI data were corrected using the two echo-planar imaging (EPI) references with opposing phase-encoding directions (Cox and Hyde, 1997). Surface data was registered to fsaverage space and subcortical data to MNI space. These were then merged to grayordinate CIFTI files. Automated labeling of noise components following ICA decomposition was performed using AROMA (Pruim et al., 2015). As final output, we down-sampled the preprocessed grey-ordinate functional CIFTI files to 32k FSLR space (Glasser et al., 2013). Then, we applied a 4 mm full-width half-maximum smoothing constrained to the grey matter boundaries. For the quantification of brain responses to the task, the following conditions were convolved with a canonical hemodynamic response function as implemented in FSL (Jenkinson et al., 2012): high win, low win, high loss, low loss, high cue, low cue. The regressors obtained were entered in a design matrix together with the confound regressors generated by AROMA. A GLM analysis was then performed using the HCP pipelines (Glasser et al., 2013) and the contrast win > loss was estimated for each participant (coefficients of high and low wins and losses were averaged). The z-scores corresponding to this contrast in each greyordinate were the features entered in the CCA analyses.

References

Cox, R. W. and Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, **10**(4-5), 171–178.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, **9**(2), 179–194.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., and Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, **16**(1), 111–116.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., and Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, **80**, 105–124.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *NeuroImage*, **62**(2), 782–790.

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., and Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, **112**, 267–277.

Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., and Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine*, **67**(5), 1210–1224.

Somerville, L. H., Bookheimer, S. Y., Buckner, R. L., Burgess, G. C., Curtiss, S. W., Dapretto, M., Elam, J. S., Gaffrey, M. S., Harms, M. P., Hodge, C., Kandala, S., Kastman, E. K., Nichols, T. E., Schlaggar, B. L., Smith, S. M., Thomas, K. M., Yacoub, E., Van Essen, D. C., and Barch, D. M. (2018). The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5-21 year olds. *NeuroImage*, **183**, 456–468.

Tozzi, L., Staveland, B., Holt-Gosselin, B., Chesnut, M., Chang, S. E., Choi, D., Shiner, M., Wu, H., Lerma-Usabiaga, G., Sporns, O., Barch, D. M., Gotlib, I. H., Hastie, T. J., Kerr, A. B., Poldrack, R. A., Wandell, B. A., Wintermark, M., and Williams, L. M. (2020). The human connectome project for disordered emotional states: Protocol and rationale for a research domain criteria study of brain connectivity in young adult anxiety and depression. *NeuroImage*, **214**, 116715.