Supplementary Material: Finding Linear Structure in Large Datasets with Scalable Canonical Correlation Analysis

1. Appendix

A brief review of the notations in the main paper:

$$\mathbf{S_x} = \mathbf{X}^\top\mathbf{X}/n, \mathbf{S_{xy}} = \mathbf{X}^\top\mathbf{Y}/n, \mathbf{S_y} = \mathbf{Y}^\top\mathbf{Y}/n, \|u\|_x = (u^\top\mathbf{S_x}u)^{\frac{1}{2}}, \|v\|_y = (v^\top\mathbf{S_y}v)^{\frac{1}{2}}$$

$$\Delta\widetilde{\phi}^t = \widetilde{\phi}^t - \widetilde{\phi}_1, \Delta\widetilde{\psi}^t = \widetilde{\psi}^t - \widetilde{\psi}_1, \Delta\phi^t = \phi^t - \phi_1, \Delta\psi^t = \psi^t - \psi_1$$

Further, we define $cos_x(u,v) = \frac{u^\top\mathbf{S_x}v}{\|u\|_x\|v\|_x}$, the cosine of the angle between two vectors induced by the inner product $\langle u,v\rangle = u^\top\mathbf{S_x}v$. Similarly, we define $cos_y(u,v) = \frac{u^\top\mathbf{S_y}v}{\|u\|_y\|v\|_y}$.
To prove the theorem, we will repeatedly use the following two lemmas.

Lemma 1. $\mathbf{S_{xy}} = \mathbf{S_x}\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Psi}^\top\mathbf{S_y}$

Proof of Lemma 1. The proof is in the main paper.

Lemma 2. $\|\Delta\phi^t\|_x \leq \frac{1}{\lambda_1}\sqrt{\frac{2}{1+cos_x(\phi^t,\phi_1)}}\|\Delta\widetilde{\phi}^t\|_x$ *and* $\|\Delta\psi^t\|_y \leq \frac{1}{\lambda_1}\sqrt{\frac{2}{1+cos_y(\psi^t,\psi_1)}}\|\Delta\widetilde{\psi}^t\|_y$

Proof of Lemma 2. Notice that $cos_x(\widetilde{\phi}^t, \widetilde{\phi}_1) = cos_x(\phi^t, \phi_1)$, then

$$\|\Delta\widetilde{\phi}^t\|_x^2 = \|\widetilde{\phi}^t - \widetilde{\phi}_1\|_x^2 \geq \|\widetilde{\phi}_1\|^2 sin_x^2(\widetilde{\phi}^t, \widetilde{\phi}_1) = \lambda_1^2 sin_x^2(\phi^t, \phi_1)$$

Also notice that $\|\phi^t\|_x = \|\phi_1\|_x = 1$, which implies $cos_x(\phi^t,\phi_1) = 1 - \|\phi^t - \phi_1\|_x^2/2 = 1 - \|\Delta\phi^t\|_x^2/2$. Further

$$\|\Delta\widetilde{\phi}^t\|_x^2 \geq \lambda_1^2 sin_x^2(\phi^t,\phi_1) = \lambda_1^2(1 - cos_x^2(\phi^t,\phi_1)) = \frac{\lambda_1^2}{2}\|\Delta\phi^t\|_x^2(1 + cos_x(\phi^t,\phi_1))$$

Square root both sides,

$$\|\Delta\phi^t\|_x \leq \frac{1}{\lambda_1}\sqrt{\frac{2}{1+cos_x(\phi^t,\phi_1)}}\|\Delta\widetilde{\phi}^t\|_x$$

Similar argument will show that

$$\|\Delta\psi^t\|_y \leq \frac{1}{\lambda_1}\sqrt{\frac{2}{1+cos_y(\psi^t,\psi_1)}}\|\Delta\widetilde{\psi}^t\|_y$$

1.1. Proof of Theorem 2.1

Without loss of generality, we can always assume $cos_x(\widetilde{\phi}^t, \widetilde{\phi}_1), cos_y(\widetilde{\psi}^t, \widetilde{\psi}_1) \geq 0$ because the canonical vectors are only identifiable up to a flip in sign and we can always choose $\widetilde{\phi}_1, \widetilde{\psi}_1$ such that the cosines are nonnegative. Apply simple algebra to the gradient step $\widetilde{\phi}^{t+1} = \widetilde{\phi}^t - \eta(\mathbf{S_x}\widetilde{\phi}^t - \mathbf{S_{xy}}\psi^t)$

$$\widetilde{\phi}^{t+1} - \widetilde{\phi}_1 = \widetilde{\phi}^t - \widetilde{\phi}_1 - \eta(\mathbf{S_x}(\widetilde{\phi}^t - \widetilde{\phi}_1) + \mathbf{S_x}\widetilde{\phi}_1 - \mathbf{S_{xy}}(\psi^t - \psi_1) - \mathbf{S_{xy}}\psi_1)$$
$$\Delta\widetilde{\phi}^{t+1} = \Delta\widetilde{\phi}^t - \eta(\mathbf{S_x}\Delta\widetilde{\phi}^t - \mathbf{S_{xy}}\Delta\phi^t) - \eta(\mathbf{S_x}\widetilde{\phi}_1 - \mathbf{S_{xy}}\psi_1)$$

By Lemma 1, $\eta(\mathbf{S_x}\widetilde{\phi}_1 - \mathbf{S_{xy}}\psi_1) = \eta(\mathbf{S_x}\widetilde{\phi}_1 - \lambda_1\mathbf{S_x}\phi_1) = 0$, which implies

$$\Delta\widetilde{\phi}^{t+1} = \Delta\widetilde{\phi}^t - \eta(\mathbf{S_x}\Delta\widetilde{\phi}^t - \mathbf{S_{xy}}\Delta\psi^t)$$

Square both sizes,

$$\|\Delta\widetilde{\phi}^{t+1}\|^2 = \|\Delta\widetilde{\phi}^t\|^2 + \eta^2\|\mathbf{S_x}\Delta\widetilde{\phi}^t - \mathbf{S_{xy}}\Delta\psi^t\|^2 - 2\eta(\Delta\widetilde{\phi}^t)^\top(\mathbf{S_x}\Delta\widetilde{\phi}^t - \mathbf{S_{xy}}\Delta\psi^t) \tag{1}$$

Apply Lemma 1,

$$\|\mathbf{S_{xy}}\Delta\psi^t\| = \|\mathbf{S_x}\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Psi}^T\mathbf{S_y}\Delta\psi^t\| \le \|\mathbf{S_x^{\frac{1}{2}}}\|\|\mathbf{S_x^{\frac{1}{2}}}\mathbf{\Phi}\|\|\mathbf{\Lambda}\|\|\mathbf{\Psi}^\top\mathbf{S_y^{\frac{1}{2}}}\|\|\mathbf{S_y^{\frac{1}{2}}}\Delta\psi^t\| \le \lambda_1 L_1^{\frac{1}{2}}\|\Delta\psi^t\|_y$$

The last inequality uses the assumption that $\lambda_{max}(\mathbf{S_x}), \lambda_{max}(\mathbf{S_y}) \le L_1$. By Lemma2, $\|\Delta\psi^t\|_y \le \frac{\sqrt{2}}{\lambda_1}\|\Delta\widetilde{\psi}^t\|_y$. Hence, $\|\mathbf{S_{xy}}\Delta\psi^t\| \le \sqrt{2L_1}\|\Delta\widetilde{\psi}^t\|_y$. Also notice that $\|\mathbf{S_x}\Delta\widetilde{\phi}^t\| \le \|\mathbf{S_x^{\frac{1}{2}}}\|\|\mathbf{S_x^{\frac{1}{2}}}\Delta\widetilde{\phi}^t\| \le L_1^{\frac{1}{2}}\|\Delta\widetilde{\phi}^t\|_x$, then

$$\|\mathbf{S_x}\Delta\widetilde{\phi}^t - \mathbf{S_{xy}}\Delta\psi^t\|^2 \le (L_1^{\frac{1}{2}}\|\Delta\widetilde{\phi}^t\|_x + \sqrt{2}L_1^{\frac{1}{2}}\|\Delta\widetilde{\psi}^t\|_y)^2 \le 2L_1(\|\Delta\widetilde{\phi}^t\|_x^2 + 2\|\Delta\widetilde{\psi}^t\|_y^2)$$

Substitute into (1),

$$\|\Delta\widetilde{\phi}^{t+1}\|^2 \le \|\Delta\widetilde{\phi}^t\|^2 - 2\eta\|\Delta\widetilde{\phi}^t\|_x^2 + 2L_1\eta^2(\|\Delta\widetilde{\phi}^t\|_x^2 + 2\|\Delta\widetilde{\psi}^t\|_y^2) + 2\eta(\Delta\widetilde{\phi}^t)^\top\mathbf{S_{xy}}\Delta\psi^t \tag{2}$$

Now, we are going to bound $(\Delta\widetilde{\phi}^t)^T\mathbf{S_{xy}}\Delta\psi^t$. Because $\mathbf{S_y^{\frac{1}{2}}}\mathbf{\Psi}$ is an orthonormal matrix (orthogonal if $p = p_1$) and $\mathbf{S_y^{\frac{1}{2}}}\psi_t$ is a unit vector, there exisit coefficients $\alpha_1, \cdots, \alpha_p, \alpha_\perp$ and unit vector $\psi_\perp \in ColSpan(\mathbf{S_y^{\frac{1}{2}}}\mathbf{\Psi})^\perp$ such that $\mathbf{S_y^{\frac{1}{2}}}\psi_t = \sum_{i=1}^p \alpha_i\mathbf{S_y^{\frac{1}{2}}}\psi_i + \alpha_\perp\mathbf{S_y^{\frac{1}{2}}}\psi_\perp, \sum_{i=1}^p \alpha_i^2 + \alpha_\perp^2 = 1$. Therefore,

$$\begin{aligned}(\Delta\widetilde{\phi}^t)^\top\mathbf{S_x}\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Psi}^\top\mathbf{S_y}\Delta\psi^t &= \Delta\widetilde{\phi}^t\mathbf{S_x}\mathbf{\Phi}\mathbf{\Lambda}(\mathbf{S_y^{\frac{1}{2}}}\mathbf{\Psi})^\top\{(\alpha_1 - 1)\mathbf{S_y^{\frac{1}{2}}}\psi_1 + \sum_{i=2}^p \alpha_i\mathbf{S_y^{\frac{1}{2}}}\psi_i + \alpha_\perp\mathbf{S_y^{\frac{1}{2}}}\psi_\perp\} \\ &= \lambda_1(\alpha_1 - 1)(\Delta\widetilde{\phi}^t)^\top\mathbf{S_x}\phi_1 + \sum_{i=2}^p \alpha_i\lambda_i(\Delta\widetilde{\phi}^t)^\top\mathbf{S_x}\phi_i\end{aligned}$$

By Cauthy-Schwarz inequality,

$$\begin{aligned}(\Delta\widetilde{\phi}^t)^\top\mathbf{S_x}\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Psi}^\top\mathbf{S_y}\Delta\psi^t &\le \Big(\lambda_1^2(1-\alpha_1)^2 + \sum_{i=2}^p \alpha_i^2\lambda_i^2\Big)^{\frac{1}{2}}\Big(\sum_{i=1}^p \big((\Delta\widetilde{\phi}^t)^\top\mathbf{S_x}\phi_1\big)^2\Big)^{\frac{1}{2}} \\ &\le \Big(\lambda_1^2(1-\alpha_1)^2 + \lambda_2^2(1-\alpha_1^2)\Big)^{\frac{1}{2}}\|\Delta\widetilde{\phi}^t\|_x \\ &= \Big(\lambda_1^2\frac{1-\alpha_1}{1+\alpha_1} + \lambda_2^2\Big)^{\frac{1}{2}}(1-\alpha_1^2)^{\frac{1}{2}}\|\Delta\widetilde{\phi}^t\|_x\end{aligned}$$

By definition, $1 - \alpha_1 = 1 - cos_y(\psi^t, \psi_1) = \frac{\|\Delta\psi^t\|_y^2}{2}$. Further by Lemma 2,

$$1 - \alpha_1 \le \frac{1}{\lambda_1^2(1+\alpha_1)}\|\Delta\widetilde{\psi}^t\|_y^2$$

Therefore,

$$\begin{aligned}(\Delta\widetilde{\phi}^t)^\top\mathbf{S_x}\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Psi}^\top\mathbf{S_y}\Delta\psi^t &\le \Big(\frac{1-\alpha_1}{1+\alpha_1} + \frac{\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}}\|\Delta\widetilde{\phi}^t\|_x\|\Delta\widetilde{\psi}^t\|_y \\ &\le \Big(\frac{\|\Delta\widetilde{\psi}^t\|_y^2}{\lambda_1^2(1+\alpha_1)^2} + \frac{\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}}\|\Delta\widetilde{\phi}^t\|_x\|\Delta\widetilde{\psi}^t\|_y \\ &\le \frac{1}{2}\Big(\frac{\|\Delta\widetilde{\psi}^t\|_y^2}{\lambda_1^2} + \frac{\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}}\Big(\|\Delta\widetilde{\phi}^t\|_x^2 + \|\Delta\widetilde{\psi}^t\|_y^2\Big)\end{aligned}$$

Substitute into (2),

$$\|\Delta\widetilde{\phi}^{t+1}\|^2 \le \|\Delta\widetilde{\phi}^t\|^2 - 2\eta\|\Delta\widetilde{\phi}^t\|_x^2 + 2L_1\eta^2\Big(\|\Delta\widetilde{\phi}^t\|_x^2 + 2\|\Delta\widetilde{\psi}^t\|_y^2\Big) + \eta\Big(\frac{\|\Delta\widetilde{\psi}^t\|_y^2}{\lambda_1^2} + \frac{\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}}\Big(\|\Delta\widetilde{\phi}^t\|_x^2 + \|\Delta\widetilde{\psi}^t\|_y^2\Big)$$

Similar analysis implies that,

$$\|\Delta\widetilde{\psi}^{t+1}\|^2 \le \|\Delta\widetilde{\psi}^t\|^2 - 2\eta\|\Delta\widetilde{\psi}^t\|_y^2 + 2L_1\eta^2\Big(\|\Delta\widetilde{\psi}^t\|_y^2 + 2\|\Delta\widetilde{\phi}^t\|_x^2\Big) + \eta\Big(\frac{\|\Delta\widetilde{\phi}^t\|_x^2}{\lambda_1^2} + \frac{\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}}\Big(\|\Delta\widetilde{\phi}^t\|_x^2 + \|\Delta\widetilde{\psi}^t\|_y^2\Big)$$

Add these two inequalities,

$$\begin{aligned}\|\Delta\widetilde{\phi}^{t+1}\|^2 + \|\Delta\widetilde{\psi}^{t+1}\|^2 \le \Big(\|\Delta\widetilde{\phi}^t\|^2 + \|\Delta\widetilde{\psi}^t\|^2\Big) - 2\eta\Big\{1 - \frac{1}{2}\Big(\frac{\|\Delta\widetilde{\psi}^t\|_y^2}{\lambda_1^2} + \frac{\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}} - \frac{1}{2}\Big(\frac{\|\Delta\widetilde{\phi}^t\|_x^2}{\lambda_1^2} + \frac{\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}} \\ - 3L_1\eta\Big\}\Big(\|\Delta\widetilde{\phi}^t\|_x^2 + \|\Delta\widetilde{\psi}^t\|_y^2\Big)\end{aligned}$$

Notice that $\sqrt{a} + \sqrt{b} \le \sqrt{2(a+b)}$, we have

$$\begin{aligned}\Big(\frac{\|\Delta\widetilde{\psi}^t\|_y^2}{\lambda_1^2} + \frac{\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}} + \Big(\frac{\|\Delta\widetilde{\phi}^t\|_x^2}{\lambda_1^2} + \frac{\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}} &\le \Big(\frac{2\|\Delta\widetilde{\psi}^t\|_y^2}{\lambda_1^2} + \frac{2\|\Delta\widetilde{\phi}^t\|_x^2}{\lambda_1^2} + \frac{4\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}} \\ &\le \Big(\frac{2L_1\|\Delta\widetilde{\psi}^t\|^2}{\lambda_1^2} + \frac{2L_1\|\Delta\widetilde{\phi}^t\|^2}{\lambda_1^2} + \frac{4\lambda_2^2}{\lambda_1^2}\Big)^{\frac{1}{2}} \\ &= \frac{1}{2\lambda_1}\Big(\frac{L_1}{2}\|\Delta\widetilde{\psi}^t\|^2 + \frac{L_1}{2}\|\Delta\widetilde{\phi}^t\|^2 + \lambda_2^2\Big)^{\frac{1}{2}}\end{aligned}$$

Then,

$$\begin{aligned}\|\Delta\widetilde{\phi}^{t+1}\|^2 + \|\Delta\widetilde{\psi}^{t+1}\|^2 \le \Big(\|\Delta\widetilde{\phi}^t\|^2 + \|\Delta\widetilde{\psi}^t\|^2\Big) - 2\eta\Big\{1 - \frac{1}{\lambda_1}\Big(\frac{L_1}{2}\|\Delta\widetilde{\psi}^t\|^2 + \frac{L_1}{2}\|\Delta\widetilde{\phi}^t\|^2 + \lambda_2^2\Big)^{\frac{1}{2}} \\ - 3L_1\eta\Big\}\Big(\|\Delta\widetilde{\phi}^t\|_x^2 + \|\Delta\widetilde{\psi}^t\|_y^2\Big)\end{aligned} \tag{3}$$

By definition, $\delta = 1 - \frac{1}{\lambda_1}\Big(\frac{L_1}{2}\|\Delta\widetilde{\psi}^0\|^2 + \frac{L_1}{2}\|\Delta\widetilde{\phi}^0\|^2 + \lambda_2^2\Big)^{\frac{1}{2}}$ and $\eta = \frac{\delta}{6L_1}$. Substitute in (3) with $t = 0$,

$$\begin{aligned}\|\Delta\widetilde{\phi}^1\|^2 + \|\Delta\widetilde{\psi}^1\|^2 &= \Big(\|\Delta\widetilde{\phi}^0\|^2 + \|\Delta\widetilde{\psi}^0\|^2\Big) - \frac{\delta^2}{6L_1}\Big(\|\Delta\widetilde{\phi}^0\|_x^2 + \|\Delta\widetilde{\psi}^0\|_y^2\Big) \\ &\le \Big(\|\Delta\widetilde{\phi}^0\|^2 + \|\Delta\widetilde{\psi}^t\|^2\Big) - \frac{\delta^2}{6L_1L_2}\Big(\|\Delta\widetilde{\phi}^0\|^2 + \|\Delta\widetilde{\psi}^0\|^2\Big) \\ &\le \Big(1 - \frac{\delta^2}{6L_1L_2}\Big)\Big(\|\Delta\widetilde{\phi}^0\|^2 + \|\Delta\widetilde{\psi}^0\|^2\Big)\end{aligned}$$

It follows by induction that $\forall\, t \in \mathbb{N}_+$

$$\|\Delta\widetilde{\phi}^{t+1}\|^2 + \|\Delta\widetilde{\psi}^{t+1}\|^2 \le \Big(1 - \frac{\delta^2}{6L_1L_2}\Big)\Big(\|\Delta\widetilde{\phi}^t\|^2 + \|\Delta\widetilde{\psi}^t\|^2\Big)$$

1.2. Proof of Proposition 2.3

Substitute $(\mathbf{\Phi}^t, \mathbf{\Psi}^t, \widetilde{\mathbf{\Phi}}^t, \widetilde{\mathbf{\Psi}}^t) = (\mathbf{\Phi}_k, \mathbf{\Psi}_k, \mathbf{\Phi}_k\mathbf{\Lambda}_k, \mathbf{\Psi}_k\mathbf{\Lambda}_k)\mathbf{Q}$ into the iterative formula in Algorithm 4.

$$\begin{aligned}\widetilde{\mathbf{\Phi}}^{t+1} &= \mathbf{\Phi}_k\mathbf{\Lambda}_k\mathbf{Q} - \eta_1(\mathbf{S_x}\mathbf{\Phi}_k\mathbf{\Lambda}_k - \mathbf{S_{xy}}\mathbf{\Psi}_k)\mathbf{Q} \\ &= \mathbf{\Phi}_k\mathbf{\Lambda}_k\mathbf{Q} - \eta_1(\mathbf{S_x}\mathbf{\Phi}_k\mathbf{\Lambda}_k - \mathbf{S_x}\mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Psi}^\top\mathbf{S_y}\mathbf{\Psi}_k)\mathbf{Q} \\ &= \mathbf{\Phi}_k\mathbf{\Lambda}_k\mathbf{Q} - \eta_1(\mathbf{S_x}\mathbf{\Phi}_k\mathbf{\Lambda}_k - \mathbf{S_x}\mathbf{\Phi}_k\mathbf{\Lambda}_k)\mathbf{Q} \\ &= \mathbf{\Phi}_k\mathbf{\Lambda}_k\mathbf{Q}\end{aligned}$$

The second equality is direct application of Lemma 1. The third equality is due to the fact that $\mathbf{\Psi}^\top\mathbf{S_y}\mathbf{\Psi} = I_p$. Then,

$$(\widetilde{\mathbf{\Phi}}^{t+1})^\top\mathbf{S_x}\widetilde{\mathbf{\Phi}}^{t+1} = \mathbf{Q}^\top\mathbf{\Lambda}_k^2\mathbf{Q}$$

and

$$\mathbf{\Phi}^{t+1} = \widetilde{\mathbf{\Phi}}^{t+1}\mathbf{Q}^{\top}\mathbf{\Lambda}_k^{-1}\mathbf{Q} = \mathbf{\Phi}_k\mathbf{Q}$$

Therefore $(\mathbf{\Phi}^{t+1}, \widetilde{\mathbf{\Phi}}^{t+1}) = (\mathbf{\Phi}^{t}, \widetilde{\mathbf{\Phi}}^{t}) = (\mathbf{\Phi}_k, \mathbf{\Phi}_k\mathbf{\Lambda}_k)\mathbf{Q}$. A symmetric argument will show that $(\mathbf{\Psi}^{t+1}, \widetilde{\mathbf{\Psi}}^{t+1}) = (\mathbf{\Psi}^{t}, \widetilde{\mathbf{\Psi}}^{t}) = (\mathbf{\Psi}_k, \mathbf{\Psi}_k\mathbf{\Lambda}_k)\mathbf{Q}$, which completes the proof.