A. Proofs of Lemma

Lemma A.1. *Let* $\mathbf{M}$ *be a matrix with rank* K*. Let* $\mathbf{B}$ *be a matrix with* $\|\mathbf{B}-\mathbf{M}\|_{\mathrm{F}} \leq \epsilon$*. And* $p(\cdot)$ *is the projection onto the top-*K *spectral spaces of* $\mathbf{M}$*, then*

$$\|p(\mathbf{B})-\mathbf{B}\|_{\mathrm{F}} \leq \epsilon.$$

Proof. Note that

$$\mathbf{B} \in \underset{\mathbf{X}}{\operatorname{argmin}} \left\|\mathbf{B}-\left(\mathbf{U}\mathbf{U}^{\top}\right)\mathbf{X}\left(\mathbf{V}\mathbf{V}^{\top}\right)\right\|_{\mathrm{F}}^{2},$$

where $\mathbf{M}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\top}$. So we have

$$\begin{aligned}\|p(\mathbf{B})-\mathbf{B}\|_{\mathrm{F}} =&\left\|\mathbf{B}-\left(\mathbf{U}\mathbf{U}^{\top}\right)\mathbf{B}\left(\mathbf{V}\mathbf{V}^{\top}\right)\right\|_{\mathrm{F}} \\ \leq&\left\|\mathbf{B}-\left(\mathbf{U}\mathbf{U}^{\top}\right)\mathbf{M}\left(\mathbf{V}\mathbf{V}^{\top}\right)\right\|_{\mathrm{F}} \\ =&\|\mathbf{B}-\mathbf{M}\|_{\mathrm{F}} \\ =&\epsilon.\end{aligned}$$

□

Lemma A.2. *Let* $\mathbf{M}$ *be a matrix with* $\|\mathbf{M}\|_{\mathrm{F}} \leq C$*. Let* $\mathbf{B}$ *be a matrix with* $\|\mathbf{B}-\mathbf{M}\|_{\mathrm{F}} \leq \epsilon$*. We have that* $p_m(\cdot)$ *and* $p_b(\cdot)$ *are the projection onto the top-*K *spectral spaces of* $\mathbf{M}$ *and* $\mathbf{B}$ *respectively, then*

$$\|p_m(\mathbf{B})-p_b(\mathbf{B})\|_{\mathrm{F}} \leq 2\epsilon.$$

Proof. It follows the Lemma A.1 and the triangular inequality of Frobenius norm. □

Without the loss of generality, we prove the results for 3-way tensors.

Lemma A.3. *Given a tensor* $\mathbf{X} \in \mathbb{R}^{I\times J\times K}$ *and a target rank* r*, then for* $\mathbf{X}' = TSM(X, r)$*, we have that its mode* i *rank is no greater than* r *for any* i*.*

Proof. Assume that the Tucker form of $\mathbf{X}$ is as follows:

$$\mathbf{X}=\mathbf{S}_1 \times_1 \mathbf{U}_{1,1} \times_2 \mathbf{U}_{1,2} \times_3 \mathbf{U}_{1,3},$$

where $\mathbf{S}_1 \in \mathbb{R}^{I\times J\times K}$ is the core tensor, $\mathbf{U}_{1,1} \in \mathbb{R}^{I\times I}$, $\mathbf{U}_{1,2} \in \mathbb{R}^{J\times J}$ and $\mathbf{U}_{1,3} \in \mathbb{R}^{K\times K}$ are unitary matrices containing the singular vector for each mode.

By the connection between tensor Tucker form and the SVD of its mode (De Lathauwer et al., 2000), we have that the operation $\mathbf{X}_{(1)} \leftarrow p\left(\mathbf{X}_{(1)}, r\right)$ is equivalent to

$$\mathbf{X} \leftarrow \mathbf{S}_1' \times_1 \mathbf{U}_{1,1}' \times_2 \mathbf{U}_{1,2} \times_3 \mathbf{U}_{1,3},$$

where $\mathbf{S}_1' \in \mathbb{R}^{r\times J\times K}$ is the first r slabs in $\mathbf{S}$ and $\mathbf{U}_{1,1}' \in \mathbb{R}^{I\times r}$ is the first r column of $\mathbf{U}_{1,1}$.

Let the Tucker form of $\mathbf{S}_1'$ be

$$\mathbf{S}_1' = \mathbf{S}_2 \times_1 \mathbf{U}_{2,1} \times_2 \mathbf{U}_{2,2} \times_3 \mathbf{U}_{2,3},$$

where the core tensor $\mathbf{S}_2 \in \mathbb{R}^{r\times J\times K}$, $\mathbf{U}_{1,1} \in \mathbb{R}^{r\times r}$, $\mathbf{U}_{1,2} \in \mathbb{R}^{J\times J}$, $\mathbf{U}_{1,3} \in \mathbb{R}^{K\times K}$. And after the update, we have

$$\mathbf{X}=\mathbf{S}_2 \times_1 \left(\mathbf{U}_{1,1}'\mathbf{U}_{2,1}\right) \times_2 \left(\mathbf{U}_{1,2}\mathbf{U}_{2,2}\right) \times_3 \left(\mathbf{U}_{1,3}\mathbf{U}_{2,3}\right),$$

as the new Tucker form.

Similarly, after cutting $\mathbf{S}_2$ to keep its top r slabs in the second mode, updating the Tucker form and doing so for the third mode, we have

$$\mathbf{X}' = \mathbf{S}' \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)},$$

where the core tensor $\mathbf{S}' \in \mathbb{R}^{r\times r\times r}$, $\mathbf{U}^{(1)} \in \mathbb{R}^{I\times r}$, $\mathbf{U}^{(2)} \in \mathbb{R}^{J\times r}$ and $\mathbf{U}^{(3)} \in \mathbb{R}^{K\times r}$. Again, by the connection between tensor Tucker form and the SVD of its mode, we reach the conclusion. □

Lemma A.4. *Given a tensor* $\mathbf{Y} \in \mathbb{R}^{I\times J\times K}$ *where its mode* i *rank is no greater than* r *for all* i*. If tensor* $\mathbf{X} \in \mathbb{R}^{I\times J\times K}$ *satisfies* $\|\mathbf{Y}-\mathbf{X}\|_{\mathrm{F}} \leq \epsilon$ *and* $\mathbf{X}' = TSM(X, r)$*, then* $\|\mathbf{Y}-\mathbf{X}'\|_{\mathrm{F}} \leq 8\epsilon$*.*

Proof. For any tensor $\mathbf{T}$ and any mode i, we have that $\|\mathbf{T}\|_{\mathrm{F}} = \left\|\mathbf{T}_{(i)}\right\|_{\mathrm{F}}$. Therefore, we can arrive at the conclusion by applying Lemma A.2 three times, i.e., $\|\mathbf{Y}-\mathbf{X}'\|_{\mathrm{F}} \leq 2^3\epsilon$. □

Lemma A.5. *Let* $\mathbf{W}$ *be an* $N\times N$ *matrix with (1) rank*$(\mathbf{W}) = R$*, (2)*$\|\mathbf{W}\|_{\mathrm{F}} < C_w$*, (3)*$\sigma_k(\mathbf{W}) > \sigma_w$*,* $\mathbf{W}'$ *be an* $N\times N$ *matrix such that* $\|\mathbf{W}'-\mathbf{W}\|_{\mathrm{F}} \leq \epsilon$*,* $\mathbf{E}$ *be a random matrix with (3) zero mean, (4)* $\sigma_N(\mathbf{E}) \geq \sigma_e$*, (5)*$\|\mathbf{E}\|_{\mathrm{F}} \leq \epsilon_e$*, then we have that*

$$\|p(\mathbf{W}'+\mathbf{E})-\mathbf{W}\|_{\mathrm{F}}^2 \leq \|\mathbf{W}'+\mathbf{E}-\mathbf{W}\|_{\mathrm{F}}^2,$$

when

$$(N-2R) \geq \frac{8(\epsilon+\epsilon_e)^4 C_w^2}{\sigma_w^4\sigma_e^2} \quad \text{AND} \quad \sigma_w \geq 4(\epsilon_e+\epsilon).$$

Proof. Assume that the full SVD of $\mathbf{W}'+\mathbf{E}$ is $[\mathbf{U}_1, \mathbf{U}_2]\operatorname{diag}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)[\mathbf{V}_1, \mathbf{V}_2]^{\top}$, then we have

$$\begin{aligned}\|\mathbf{W}'+\mathbf{E}-\mathbf{W}\|_{\mathrm{F}}^2 - &\|p(\mathbf{W}'+\mathbf{E})-\mathbf{W}\|_{\mathrm{F}}^2 \\ &=\left\|\boldsymbol{\Sigma}_2-\mathbf{U}_2^{\top}\mathbf{W}\mathbf{V}_2\right\|_{\mathrm{F}}^2 - \left\|\mathbf{U}_2^{\top}\mathbf{W}\mathbf{V}_2\right\|_{\mathrm{F}}^2 \\ &\geq\|\boldsymbol{\Sigma}_2\|_{\mathrm{F}}\left(\|\boldsymbol{\Sigma}_2\|_{\mathrm{F}} - 2\left\|\mathbf{U}_2^{\top}\mathbf{W}\mathbf{V}_2\right\|_{\mathrm{F}}\right)\end{aligned}$$

By Wedin $\sin\theta$ theorem (Wedin, 1972), when $\sigma_w \geq 4(\epsilon_e+\epsilon)$, we have that $\left\|\mathbf{U}_2^{\top}\mathbf{W}\mathbf{V}_2\right\|_{\mathrm{F}}^2 \leq \frac{4(\epsilon+\epsilon_e)^4 C_w^2}{\sigma_w^4}$.

Together with $\|\boldsymbol{\Sigma}_2\|_{\mathrm{F}}^2 \geq (N-2R)\sigma_e^2$, we arrive at the condition $(N-2R) \geq \frac{8(\epsilon+\epsilon_e)^4 C_w^2}{\sigma_w^4\sigma_e^2}$. □

B. Additional Appendix

B.1. Matrix Update

Problem Given $(Y_1, X_1) \in (\mathbb{R}^{q\times n_1}, \mathbb{R}^{p\times n_1})$, we have estimated $\widehat{A}_1$ that minimizes $\|Y_1 - AX_1\|_F$. Now we receive some more data: $(Y_2, X_2) \in (\mathbb{R}^{q\times n_2}, \mathbb{R}^{p\times n_2})$, where $n_2 \ll n_1$. Denote $Y = [Y_1, Y_2]$ and $X = [X_1, X_2]$. The goal is to find $\widehat{A}$ such that it minimizes $\|Y - AX\|_F$.

Solution: We have $\widehat{A}_1 = (Y_1X_1^\top)(X_1X_1^\top)^{-1}$. The goal is to compute the following quantity efficiently:

$$\widehat{A} = (YX^\top)(XX^\top)^{-1} = (Y_1X_1^\top + Y_2X_2^\top)(X_1X_1^\top + X_2X_2^\top)^{-1}$$

Given the Woodbury identity $(D + CC^\top)^{-1} = D^{-1} - D^{-1}C(I + C^\top D^{-1}C)^{-1}C^\top D^{-1}$, if we define $B = D^{-1}C$, we have $(D + CC^\top)^{-1} = D^{-1} - B(I + C^\top B)^{-1}B^\top$. Thus, defining $B = (X_1X_1^\top)^{-1}X_2$, , $E = B(I + X_2^\top B)^{-1}B^\top$, we can write:

$$\begin{aligned}\widehat{A} &= (Y_1X_1^\top + Y_2X_2^\top)\left((X_1X_1^\top)^{-1} - B(I + X_2^\top B)^{-1}B^\top\right),\\ &= \widehat{A}_1 + Y_2X_2^\top(X_1X_1^\top)^{-1} - Y_1X_1^\top E - Y_2X_2^\top E.\end{aligned}$$

Thus, we have found the required update term and it only depends on efficient matrix multiplications:

$$\Delta_A = \qquad Y_2B^\top - Y_1X_1^\top - Y_2X_2^\top E$$

Thus, we can design an algorithm that whenever receives a new set of samples, updates its version of $(XX^\top)^{-1}$ and $\widehat{A}$ and becomes ready for the next update.

B.2. Reformulation

Define $\mathbf{H}$ as the Cholesky decomposition of $\mathbf{I}_P + \mu\mathbf{L}$ $\mathbf{H}\mathbf{H}^\top = \mathbf{I}_P + \mu\mathbf{L}$. Since $\mathbf{H}$ is full rank and the mapping defined by $\mathcal{W} \mapsto \tilde{\mathcal{W}} : \tilde{\mathcal{W}}_{:,:,m} = \mathbf{H}\mathcal{W}_{:,:,m}$ for $m = 1, \ldots, M$ preserves the tensor rank, i.e., $\mathrm{rank}(\mathcal{W}) = \mathrm{rank}(\tilde{\mathcal{W}})$. We can rewrite the loss function in Equation 3 and 3 as

$$\sum_{m=1}^{M}\sum_{t=K+1}^{T} \|\mathbf{H}\mathcal{W}_{:,:,m}\mathbf{X}_{t,m} - (\mathbf{H}^{-1})\mathcal{X}_{:,t,m}\|_F^2$$

This suggests that we can solve first solve the quadratic loss $\sum_{m=1}^{M} \|\mathcal{W}_{:,:,m}\mathcal{Y}_{:,:,m} - \mathcal{V}_{:,:,m}\|_F^2$ with $\mathcal{Y}_{:,:,m} = \mathbf{X}_{K+1:T,m}$ and $\mathcal{V}_{:,:,m} = \mathcal{X}_{:,:,m}$ and obtain its solution as $\tilde{\mathcal{W}}$; then compute $\mathcal{W}_{:,:,m} = \mathbf{H}^{-1}\tilde{\mathcal{W}}_{:,:,m}$.

B.3. Multi-model Data

Seven models are selected from http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php

1. BCCR-BCM2.0: Norway Bjerknes Centre for Climate Research
2. CGCM3.1(T47): Canadian Centre for Climate Modelling
3. INM-CM3.0: Russia Institute for Numerical Mathematics
4. MRI-CGCM2.3.2: Japan Meteorological Research Institute
5. MIROC3.2(hires): Japan Center for Climate System Research
6. MIROC3.2(medres): Japan Center for Climate System Research
7. FGOALS-g1.0: China ASG / Institute of Atmospheric Physics China

The variables are then downsampled into a 5' by 5' latitude-longitude grid. For flux-like variables, first order conservative mapping is used. For pressures and temperatures, bilinear interpolation is used. The 19 variables we use are 'lhtfl.sfc', 'shtfl.sfc', 'shum.2m', 'soilw.0-10cm', 'prate.sfc', 'cprat.sfc','pres.sfc', 'mslp', 'dlwrf.sfc', 'ulwrf.sfc','ulwrf.ntat','dswrf.sfc','dswrf.ntat' , 'uswrf.sfc', 'uswrf.ntat','air.2m', 'skt.sfc', 'vwnd.10m', 'runof.sfc'. The meaning of the variables is available at http://esg.llnl.gov:8080/about/ipccTables.do