Supplementary Materials

A Step size choices.

In this appendix we show that the step sizes chosen for BRAINZOOM leads to monotonic decrease of the cost function of (3.7). The idea is to simply show that the selected step sizes are smaller than the reciprocal of the respective Liptchitz constants of the conditional functions while fixing the other variables. We show that for the step size of \( Z \), and that of \( W \) follow similar arguments.

The Liptchitz constant is the supremum of the spectral norm of the Hessian matrix of a function—for quadratic functions, it boils down to the spectral norm of Hessians. To derive it, let us first write the \( Z \)-subproblem \([3.11]\) in the vectorized form

\[
\tilde{f}(z; w, \tau) = \|x_t - \tau \vec{T}_t z\|^2 + \|x_s - \tau \vec{T}_s \text{Diag}(w) z\|^2 + \rho \left( \|\text{Diag}(d_s) \vec{H}_s z\|^2 + \|	ext{Diag}(d_t) \vec{H}_t z\|^2\right) + \mu \|z - w\|^2,
\]

where the lowercase letters are vectorized versions of the matrices denoted by the uppercase letters, and

\[
\vec{T}_t = I \otimes T_t, \quad \vec{T}_s = T_s \otimes I, \\
\vec{H}_s = I \otimes H_s, \quad \vec{H}_t = H_t \otimes I.
\]

The Hessian matrix can easily be calculated as

\[
\nabla^2\tilde{f} = \tau^2 \vec{T}_t^T \vec{T}_t + \text{Diag}(w) \vec{T}_s^T \vec{T}_s \text{Diag}(w) + \rho \vec{H}_s^T \text{Diag}(d_s^2) \vec{H}_s + \rho \vec{H}_t^T \text{Diag}(d_t^2) \vec{H}_t + \mu I
\]

a summation of several individual matrices. An upper-bound on the spectral norm of \( \nabla^2 \tilde{f} \) is the sum of the individual spectral norms. Furthermore, we invoke the following properties on the matrix spectral norm:

\[
\|AB\| \leq \|A\|\|B\|,
\]

\[
\|A \otimes B\| = \|A\|\|B\|.
\]

Then we have that

\[
\|\nabla^2\tilde{f}\| \leq \tau^2 \lambda_{\max}(\vec{T}_t^T \vec{T}_t) + \max(w^2) \lambda_{\max}(\vec{T}_s^T \vec{T}_s) + \rho \left( \max(d_s^2)\|H_s\|^2 + \max(d_t^2)\|H_t\|^2\right) + \mu.
\]

Finally, we show that \( \|H_s\| \) and \( \|H_t\| \) are upperbounded by \( c_H \) defined in \([3.15]\), by taking the total variation regularization as an example, assuming its size is \((n-1) \times n\).

The definition of matrix spectral norm is

\[
\|H\| = \max_{\|u\|=1}\|Hu\|.
\]

Assume \( \hat{H} \) is obtained by adding one more row into \( H \), then we trivially have \( \|\hat{H}\| \leq \|H\| \). Consider \( \hat{H} \) to be the following circulant matrix

\[
\hat{H} = \Phi \text{Diag}(\hat{h}) \Phi^*,
\]

where \( \hat{h} \) is the DFT of first row of \( \hat{H} \). Because \( \Phi \) has orthogonal columns, by rotating the elements of \( \hat{h} \) to be non-negative real, this becomes the singular value decomposition of \( \hat{H} \), and the largest absolute value of \( \hat{h} \) is the spectral norm of \( \|\hat{H}\| \). By the definition of DFT,

\[
\hat{h} = 1 + \left[ 1 e^{-i\pi/n} e^{-i2\pi/n} \ldots e^{-i(n-1)\pi/n} \right]^T,
\]

therefore we trivially have \( \max(\|\hat{h}\|) \leq 2 \), thus

\[
\|\hat{H}\| \leq 2
\]

Similarly, for the smoothness regularization, \( \|H\| \leq 4 \).

B Proof of Theorem[1]

Two claims were made in Theorem [1]. Here we separate them into two propositions, and prove them individually.

**Proposition 1.** Let \( \{(Z^{(r)}, W^{(r)}, \tau^{(r)})\} \) be the solution sequence produced by the proposed BRAINZOOM \([3.16]\), then every limit point of \( \{(Z^{(r)}, W^{(r)}, \tau^{(r)})\} \) is a stationary point of Problem \([3.7] \).

**Proof.** We prove that BRAINZOOM described in \([3.16]\) falls into the framework of successive upper-bound minimization (BSUM) [18]. As a result, every limit point is a stationary point, according to [18] Theorem 2. To do so, we re-write the algorithm as

\[
\tau^{(r+1)} \leftarrow \arg \min_{\tau} f(Z^{(r)}, W^{(r)}, \tau),
\]

\[
W^{(r+1)} \leftarrow \arg \min_{W} u_w(W; Z^{(r)}, W^{(r)}, \tau^{(r+1)}),
\]

\[
Z^{(r+1)} \leftarrow \arg \min_{Z} u_z(Z; W^{(r+1)}, Z^{(r)}, \tau^{(r+1)}),
\]

where the arg min’s are uniquely defined, function \( u_z \) is an auxiliary function that satisfies that \( \forall Z, W, \tilde{\tau},
\]

\[
u_z(Z; Z, W, \tilde{\tau}) \geq f(Z, W, \tilde{\tau}), \forall Z
\]

\[
\nabla Z u_z(Z; Z, W, \tilde{\tau}) = \nabla Z f(Z, W, \tilde{\tau}),
\]

and
and similarly for \( u_w \).

For the update of \( \tau \), it is a scalar least-squares problem, and the minimizer is unique as long as \( \|T_iZ\|_F^2 \neq 0 \), which can be guaranteed as long as the regularization parameter \( \rho \) is not too big.

As for the auxiliary function with respect to \( Z \), we define it as

\[
u_z(Z, \tilde{Z}, \tilde{W}, \hat{\tau}) := \bar{f}(\tilde{Z}, \tilde{W}, \hat{\tau}) + \left\langle \nabla Z \bar{f}(\tilde{Z}, \tilde{W}, \hat{\tau}), Z - \bar{Z} \right\rangle + \frac{1}{2\zeta} \|Z - \bar{Z}\|^2_2.
\]

As we have shown in Appendix \( \ref{app:proof} \), \( 1/\zeta \) is larger than the Lipschitz constant of \( \bar{f} \) when fixing \( W \) and \( \tau \), therefore

\[u_z(Z, \tilde{Z}, \tilde{W}, \hat{\tau}) \geq \bar{f}(Z, \tilde{W}, \hat{\tau}) \geq f(Z, \tilde{W}, \hat{\tau}).\]

It is also easy to see that the function value and gradient also coincides with that of \( f \) with respect to \( Z \). Furthermore, \( u_z \) is strongly convex, implying the minimizer is unique.

Similarly, and with simpler derivations, we have that \( u_w \) also satisfies the sharp upperbound requirements and the minimizer is unique. Invoking \cite[Theorem 2]{18}, every limit point of BRAINZOOM is a stationary point.

**Proposition 2**  In addition to Proposition 1, the optimality gap between \( \{(Z^{(r)}, W^{(r)}, \tau^{(r)})\}_r \) and a stationary point is at most \( \mathcal{O}(1/r) \); i.e., the algorithm approaches a stationary point at least sub-linearly.

**Proof.** For the \( \tau \)-subproblem, because

\[
\tau^{(r+1)} = \text{tr}(X_i^T T_i Z^{(r)}) / \|T_iZ^{(r)}\|_F^2,
\]

we have that

\[
f(Z^{(r)}, W^{(r)}, \tau^{(r)}) - f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}) = \|X_i - \tau^{(r)} T_i Z^{(r)}\|_F^2 - \|X_i - \tau^{(r+1)} T_i Z^{(r)}\|_F^2
\]

\[
= \frac{1}{\|T_i Z^{(r)}\|_F^2} \left( \partial_\tau f(Z^{(r)}, W^{(r)}, \tau^{(r)}) \right)^2
\]

\[
= l^{(r)} \left( \partial_\tau f(Z^{(r)}, W^{(r)}, \tau^{(r)}) \right)^2,
\]

where we define \( l^{(r)} = 1/\|T_i Z^{(r)}\|_F^2 \).

For the \( W \)-subproblem, we have the following inequality:

\[
f(Z^{(r)}, W^{(r+1)}, \tau^{(r+1)})
\]

\[
\leq f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}) + \frac{L_w^{(r)}}{2} \|W^{(r)} - W^{(r+1)}\|_F^2 + \left\langle \nabla Z f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}), W^{(r+1)} - W^{(r)} \right\rangle,
\]

We also notice that

\[
W^{(r+1)} = \arg \min_W \frac{1}{2u_w^{(r)}} \|W - W^{(r)}\|_F^2 + \left\langle \nabla W f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}), W - W^{(r)} \right\rangle,
\]

(B.5)

\[
\geq \left\langle \nabla W f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}), W^{(r+1)} - W^{(r)} \right\rangle + \frac{1}{2u_w^{(r)}} \|W^{(r+1)} - W^{(r)}\|_F^2 \leq 0,
\]

therefore

\[
f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}) \leq f(Z^{(r)}, W^{(r+1)}, \tau^{(r+1)})
\]

(B.6)

\[
\geq \left( \frac{1}{2u_w^{(r)}} - \frac{L_w^{(r)}}{2} \right) \|W^{(r+1)} - W^{(r)}\|_F^2.
\]

On the other hand, since \( W^{(r+1)} \) is the minimizer of \( \mathcal{B}_3 \), by the first order optimality condition, we have

\[
\nabla W f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}) - f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}) = \frac{1}{u_w^{(r)}} (W^{(r+1)} - W^{(r)}) = 0.
\]

In sum,

(B.8)

\[
f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}) \geq \left( \frac{u_w^{(r)}}{2} - \frac{L_w^{(r)}}{2u_w^{(r)}} \right) \|\nabla W f(Z^{(r)}, W^{(r)}, \tau^{(r+1)})\|_F^2.
\]

Similarly for \( Z \), we can show that

(B.9)

\[
f(Z^{(r)}, W^{(r+1)}, \tau^{(r+1)}) - f(Z^{(r+1)}, W^{(r+1)}, \tau^{(r+1)})
\]

\[
\geq \left( \frac{z^{(r)}}{2} - \frac{L_z^{(r)}}{2z^{(r)}} \right) \|\nabla Z f(Z^{(r)}, W^{(r)}, \tau^{(r+1)})\|_F^2.
\]

Combining \( \mathcal{B}_3, \mathcal{B}_8, \) and \( \mathcal{B}_9 \), we obtain

(B.10)

\[
f(Z^{(r)}, W^{(r)}, \tau^{(r)}) - f(Z^{(r+1)}, W^{(r+1)}, \tau^{(r+1)})
\]

\[
\geq l^{(r)} \left( \partial_\tau f(Z^{(r)}, W^{(r)}, \tau^{(r)}) \right)^2 + \left( \frac{u_w^{(r)}}{2} - \frac{L_w^{(r)}}{2u_w^{(r)}} \right) \|\nabla W f(Z^{(r)}, W^{(r)}, \tau^{(r+1)})\|_F^2
\]

\[
+ \left( \frac{z^{(r)}}{2} - \frac{L_z^{(r)}}{2z^{(r)}} \right) \|\nabla Z f(Z^{(r)}, W^{(r+1)}, \tau^{(r+1)})\|_F^2.
\]

To show the convergence rate, let us define

\[
\phi^{(r)} = \left( \partial_\tau f(Z^{(r)}, W^{(r)}, \tau^{(r)}) \right)^2 + \|\nabla W f(Z^{(r)}, W^{(r)}, \tau^{(r+1)})\|_F^2 + \|\nabla Z f(Z^{(r)}, W^{(r+1)}, \tau^{(r+1)})\|_F^2.
\]
One can see that
\[
\phi^{(r)} \to 0 \quad \implies \quad \left( \partial_\tau f(Z^{(r)}, W^{(r)}, \tau^{(r)}) \right)^2 \to 0
\]
\[
\| \nabla W f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}) \|_F^2 \to 0
\]
\[
\| \nabla Z f(Z^{(r)}, W^{(r+1)}, \tau^{(r+1)}) \|_F^2 \to 0
\]
which implies a stationary point is attained. Let us assume that the first time \( \phi^{(r)} < \varepsilon \) requires \( T \) iterations. Then, summing up (B.10) over \( r = 1, \ldots, T \), we have

\[
f(Z^{(1)}, W^{(1)}, \tau^{(1)}) - f(Z^{(T)}, W^{(T)}, \tau^{(T)})
\geq \sum_{r=1}^{T} \tau^{(r)} \left( \partial_\tau f(Z^{(r)}, W^{(r)}, \tau^{(r)}) \right)^2
\]
\[
+ \sum_{r=1}^{T} \left( \frac{w^{(r)}}{2} - \frac{L_{WW}^{(r)}}{2w^{(r)}r^2} \right) \| \nabla W f(Z^{(r)}, W^{(r)}, \tau^{(r+1)}) \|_F^2
\]
\[
+ \sum_{r=1}^{T} \left( \frac{z^{(r)}}{2} - \frac{L_{ZZ}^{(r)}}{2z^{(r)}r^2} \right) \| \nabla Z f(Z^{(r)}, W^{(r+1)}, \tau^{(r+1)}) \|_F^2.
\]
(B.11)
\[
\geq \sum_{r=1}^{T} c \phi^{(r)}
\]
where
\[
c = \min_r \left\{ \tau^{(r)} \left( \frac{w^{(r)}}{2} - \frac{L_{WW}^{(r)}}{2w^{(r)}r^2} \right), \left( \frac{z^{(r)}}{2} - \frac{L_{ZZ}^{(r)}}{2z^{(r)}r^2} \right) \right\}.
\]
The above implies
\[
f(Z^{(1)}, W^{(1)}, \tau^{(1)}) - f(Z^{(T)}, W^{(T)}, \tau^{(T)}) \geq c \phi^{(r)},
\]
and so
\[
\phi^{(r)} \leq \frac{1}{T} \left( \frac{f(Z^{(1)}, W^{(1)}, \tau^{(1)}) - f(Z^*, W^*, \tau^*)}{c} \right),
\]
where \( Z^*, W^*, \tau^* \) denote a global optimal solution of Problem (3.7). This completes the proof.