Supplementary material:
Online Adaptive Principal Component Analysis and Its extensions

The supplementary material contains proofs of the main results of the paper along with supporting results.

Before presenting the proofs, we need the following lemma from previous literature:

**Lemma 4.** (Freund & Schapire, 1997) Suppose \( 0 \leq L \leq \tilde{L} \) and \( 0 < R \leq \tilde{R} \). Let \( \beta = g(\tilde{L}/R) \) where \( g(z) = 1/(1 + \sqrt{2/z}) \). Then

\[
- L \ln \beta + R \leq L + \sqrt{2LR} + R
\]

Additionally, we need the following classic bound on traces for positive semidefinite matrices. See, e.g. (Tsuda et al., 2005).

**Lemma 5.** For any positive semi-definite matrix \( A \) and any symmetric matrices \( B \) and \( C \), \( B \preceq C \) implies \( \text{Tr}(AB) \leq \text{Tr}(AC) \).

### A. Proof of Theorem 1

**Proof.** Fix \( 1 \leq r \leq s \leq T \). We set \( q_t = q \in B_{n-k}^n \) for \( t = r, \ldots, s \) and \( 0 \) elsewhere. Thus, we have that \( \|q_t\|_1 \) is either \( 0 \) or \( 1 \).

According to Lemma 1, for both cases of \( q_t \), we have

\[
\|q_t\|_1 \|w_t\| T \ell_t (1 - \exp(-\eta)) - \eta q_t^T \ell_t \leq \sum_{i=1}^n q_{t,i} \ln \left( \frac{v_{t,i+1}}{v_{t,i}} \right)
\]

The analysis for \( \sum_{i=1}^n q_{t,i} \ln \left( \frac{v_{t,i+1}}{v_{t,i}} \right) \) follows the Proof of Proposition 2 in (Cesa-Bianchi et al., 2012b). We describe the steps for completeness, since it is helpful for understanding the effect of the fixed-share step, Eq.(4b). This analysis will be crucial for understanding how the fixed-share step can be applied to PCA problems.

\[
\sum_{i=1}^n q_{t,i} \ln \left( \frac{v_{t,i+1}}{v_{t,i}} \right) = \sum_{i=1}^n \left( q_{t,i} \ln \left( \frac{1}{v_{t,i}} \right) - q_{t-1,i} \ln \left( \frac{1}{v_{t-1,i}} \right) \right)
+ \sum_{i=1}^n \left( q_{t-1,i} \ln \left( \frac{1}{v_{t-1,i}} \right) - q_{t,i} \ln \left( \frac{1}{v_{t,i}} \right) \right)
\]

For the expression of \( A \), we have

\[
A = \sum_{i: q_{t,i} \geq q_{t-1,i}} (q_{t,i} - q_{t-1,i}) \ln \left( \frac{1}{v_{t,i}} \right) + q_{t-1,i} \ln \left( \frac{v_{t,i}}{v_{t-1,i}} \right)
+ \sum_{i: q_{t,i} < q_{t-1,i}} (q_{t,i} - q_{t-1,i}) \ln \left( \frac{1}{v_{t,i}} \right) + q_{t,i} \ln \left( \frac{v_{t,i}}{v_{t-1,i}} \right)
\]

Based on the update in Eq.(4), we have \( 1/\tilde{w}_{t,i} \leq n/\alpha \) and \( v_{t,i}/\tilde{w}_{t,i} \leq 1/(1 - \alpha) \). Plugging the bounds into the above equation, we have

\[
A \leq \sum_{i: q_{t,i} \geq q_{t-1,i}} (q_{t,i} - q_{t-1,i}) \ln \left( \frac{n}{\alpha} \right)
+ \left( \sum_{i: q_{t,i} \geq q_{t-1,i}} q_{t-1,i} + \sum_{i: q_{t,i} < q_{t-1,i}} q_{t,i} \right) \ln \left( \frac{1}{1 - \alpha} \right)
\]

Telescoping the expression of \( B \), substituting the above inequality in Eq.(33), and summing over \( t = 2, \ldots, T \), we have

\[
\sum_{i=1}^n \sum_{t=2}^T q_{t,i} \ln \left( \frac{v_{t,i+1}}{v_{t,i}} \right) \leq m(q_1) T \ln \left( \frac{n}{\alpha} \right)
+ \left( \sum_{i=1}^n q_{t,i} \ln \left( \frac{1}{1 - \alpha} \right) + \sum_{i=1}^n q_{t,i} \ln \left( \frac{1}{1 - \alpha} \right) \right)
\]

Adding the \( t = 1 \) term to the above inequality, we have

\[
\sum_{i=1}^n \sum_{t=1}^T q_{t,i} \ln \left( \frac{v_{t,i+1}}{v_{t,i}} \right) \leq \|q_1\|_1 \ln (n) + m(q_1) T \ln \left( \frac{n}{\alpha} \right)
+ \left( \sum_{i=1}^n q_{t,i} \ln \left( \frac{1}{1 - \alpha} \right) \right)
\]

Now we bound the right side, using the choices for \( q_0 \) described at the beginning of the proof. If \( r \geq 2, m(q_1) T = 1 \), and \( \|q_1\|_1 = 0 \). If \( r = 1, m(q_1) T = 0 \), and \( \|q_1\|_1 = 1 \). Thus, \( m(q_1) T + \|q_1\|_1 = 1 \), and the right part can be upper bounded by \( \ln \frac{n}{\alpha} + T \ln \frac{1}{1 - \alpha} \).

Combine the above inequality with Eq.(32), set \( q_t = q \in B_{n-k}^n \), and multiply both
sides by \( n - k \), we have
\[
(1 - \exp(-\eta)) \sum_{t=r}^{s} (n-k)w_t^T \ell_t - \eta \sum_{t=r}^{s} (n-k)q_t^T \ell_t \\
\leq (n-k) \ln \frac{n}{\alpha} + (n-k)T \ln \frac{1}{1-\alpha} 
\]
If we set \( \alpha = 1/(1 + (n-k)T) \), then the right part can be upper bounded by \((n-k) \ln \left( (1 + (n-k)T) \right) + 1\), which equals to \( D \) as defined in the Theorem 1. Thus, the above inequality can be reformulated as
\[
\sum_{t=r}^{s} (n-k)w_t^T \ell_t \leq \frac{\eta \sum_{t=r}^{s} (n-k)q_t^T \ell_t + D}{1 - \exp(-\eta)} 
\]
Since the above inequality holds for arbitrary \( q \in B_{n-k} \), we have
\[
\sum_{t=r}^{s} (n-k)w_t^T \ell_t \leq \frac{\eta \min_{q \in B_{n-k}} \sum_{t=r}^{s} (n-k)q_t^T \ell_t + D}{1 - \exp(-\eta)} 
\]
We will apply the inequality in Lemma 4 to upper bound the right part in Eq.(40). With \( \min_{q \in B_{n-k}} \sum_{t=r}^{s} (n-k)q_t^T \ell_t \leq L \) and \( \eta = \ln(1 + \sqrt{2D/L}) \), we have
\[
\sum_{t=r}^{s} (n-k)w_t^T \ell_t - \min_{q \in B_{n-k}} \sum_{t=r}^{s} (n-k)q_t^T \ell_t \leq 2LD + D 
\]
Since the above inequality always holds for all intervals, \([r, s]\), the result is proved by maximizing the left side over \([r, s]\).

**B. Proof of Theorem 3**

Proof. In the proof, we will use two cases of \( Q_t \): \( Q_t \in B_{n-k} \), and \( Q_t = 0 \).

We first apply the eigendecomposition to \( Q_t \) as \( Q_t = \hat{D} \text{diag}(q_t) \hat{D}^T \), where \( \hat{D} = [\hat{d}_1, \ldots, \hat{d}_n] \). Since in the adaptive setting, \( Q_{t-1} \) is either equal to \( Q_t \) or 0, they share the same eigenvectors and \( Q_{t-1} \) can be expressed as \( Q_{t-1} = \hat{D} \text{diag}(q_{t-1}) \hat{D}^T \).

According to Lemma 2, the following inequality is true for both cases of \( Q_t \):
\[
\|q_t\|_1 \text{Tr}(W_t x_t x_t^T) (1 - \exp(-\eta)) - \eta \text{Tr}(Q_t x_t x_t^T) \leq -\text{Tr}(Q_t \ln \hat{W}_t) + \text{Tr}(Q_t \ln V_{t+1}) 
\]
The next steps extend proof of Proposition 2 in (Cesa-Bianchi et al., 2012b) to the matrix case.

We analyze the right part of the above inequality, which can be expressed as:
\[
-\text{Tr}(Q_t \ln \hat{W}_t) + \text{Tr}(Q_t \ln V_{t+1}) = \hat{A} + \hat{B} 
\]
where \( \hat{A} = -\text{Tr}(Q_t \ln \hat{W}_t) + \text{Tr}(Q_{t-1} \ln V_t) \), and \( \hat{B} = -\text{Tr}(Q_{t-1} \ln V_t) + \text{Tr}(Q_t \ln V_{t+1}) \).

We will first upper bound the \( \hat{A} \) term, and then telescope the \( \hat{B} \) term.

\( \hat{A} \) can be expressed as:
\[
\hat{A} = \sum_{i : q_{t,i} \geq q_{t-1,i}} \left( \text{Tr} \left( (q_{t,i} \hat{d}_i \hat{d}_i^T - q_{t-1,i} \hat{d}_i \hat{d}_i^T) \ln \hat{W}_t \right) \right) + \text{Tr}(q_{t,i} \hat{d}_i \hat{d}_i^T \ln V_t) - \text{Tr}(q_{t,i} \hat{d}_i \hat{d}_i^T \ln \hat{W}_t) 
\]

For (1), it can be expressed as:
\[
1 = \text{Tr} \left( (q_{t,i} \hat{d}_i \hat{d}_i^T - q_{t-1,i} \hat{d}_i \hat{d}_i^T) \ln \hat{W}_t^{-1} \right) \leq \text{Tr} \left( (q_{t,i} \hat{d}_i \hat{d}_i^T - q_{t-1,i} \hat{d}_i \hat{d}_i^T) \ln \frac{\alpha}{\alpha} \right) = (q_{t,i} - q_{t-1,i}) \ln \frac{\alpha}{\alpha} \]
For (4), we have $4 \leq q_{t, i} \ln \frac{1}{1-\alpha}$, which follows the same argument used to bound the term (2).

Thus, $\bar{A}$ can be upper bounded as follows:

$$
\bar{A} \leq \sum_{i, q_{i} \geq q_{t-1, i}} (q_{t, i} - q_{t-1, i}) \ln \frac{n}{\alpha} + \left( \sum_{i, q_{i} \geq q_{t-1, i}} \frac{q_{t, i}}{T} \ln \frac{1}{1-\alpha} \right)
$$

(48)

Then we telescope the $B$ term, substitute the above inequality for $\bar{A}$ into Eq.(43), and sum over $t = 2, \ldots, T$ to give:

$$
\sum_{t=2}^{T} \left( -\text{Tr}(Q_{t} \ln W_{t}) + \text{Tr}(Q_{t} \ln V_{t+1}) \right) \leq m(q_{t, 1, \mathbf{T}}) \ln \frac{n}{\alpha} + \left( \sum_{t=2}^{T} \|q_{t, 1, \mathbf{T}}\|_{1} - m(q_{t, 1, \mathbf{T}}) \right) \ln \frac{n}{\alpha} - \text{Tr}(Q_{1} \ln V_{2})
$$

(49)

Adding the $t=1$ term to the above inequality, we have

$$
\sum_{t=1}^{T} \left( -\text{Tr}(Q_{t} \ln W_{t}) + \text{Tr}(Q_{t} \ln V_{t+1}) \right) \leq \|q_{1, 1, \mathbf{T}}\|_{1} \ln n + m(q_{1, 1, \mathbf{T}}) \ln \frac{n}{\alpha} + \left( \sum_{t=1}^{T} \|q_{t, 1, \mathbf{T}}\|_{1} - m(q_{1, 1, \mathbf{T}}) \right) \ln \frac{n}{\alpha}
$$

(50)

For the above inequality, we set $Q_{t} = Q \in \mathcal{B}_{n-k}^{n-k}$ for $t = r, \ldots, s$ and 0 elsewhere, which makes $q_{k} = q \in \mathcal{B}_{n-k}^{n-k}$ for $t = r, \ldots, s$ and 0 elsewhere. If $r \geq 2$, $m(q_{t, 1, \mathbf{T}}) = 1$, and $\|q_{1, 1, \mathbf{T}}\|_{1} = 0$. If $r = 1$, $m(q_{1, 1, \mathbf{T}}) = 0$, and $\|q_{1, 1, \mathbf{T}}\|_{1} = 1$. Thus, $m(q_{1, 1, \mathbf{T}}) + \|q_{1, 1, \mathbf{T}}\|_{1} = 1$, and the right part can be upper bounded by $\ln \frac{n}{\alpha} + T \ln \frac{1}{1-\alpha}$.

The rest of the steps follow exactly the same as in the proof of Theorem 1.

**C. Proof of Lemma 3**

**Proof.** We first deal with the term $\text{Tr}(Q_{t} \ln V_{t+1})$. According to the update in Eq.(24a), we have

$$
\text{Tr}(Q_{t} \ln V_{t+1}) = \text{Tr} \left( Q_{t} \ln \left( \frac{\text{exp}(\ln Y_{t} - \eta C_{t})}{\text{Tr}(\text{exp}(\ln Y_{t} - \eta C_{t}))} \right) \right)
$$

$$
= \text{Tr} \left( Q_{t} (\ln Y_{t} - \eta C_{t}) \right) - \text{ln} \left( \text{Tr} \left( \text{exp}(\ln Y_{t} - \eta C_{t}) \right) \right),
$$

(51)

since $Q_{t} \in \mathcal{B}_{n}^{n}$ and $\text{Tr}(Q_{t}) = 1$.

As a result, we have $\text{Tr}(Q_{t} \ln V_{t+1}) - \text{Tr}(Q_{t} \ln Y_{t}) = -\eta \text{Tr}(Q_{t} C_{t}) - \ln \left( \text{Tr} \left( \text{exp}(\ln Y_{t} - \eta C_{t}) \right) \right)$.

Thus, to prove the inequality in Lemma 3, it is enough to prove the following inequality

$$
\eta \text{Tr}(Y_{t} C_{t}) - \frac{\eta^{2}}{2} + \ln \left( \text{Tr} \left( \text{exp}(\ln Y_{t} - \eta C_{t}) \right) \right) \leq 0
$$

(52)

Before we proceed, we need the following lemma:

**Lemma 6** (Golden-Thompson inequality). For any symmetric matrices $A$ and $B$, the following inequality holds:

$$
\text{Tr} \left( \text{exp}(A + B) \right) \leq \text{Tr} \left( \text{exp}(A) \exp(B) \right)
$$

**Lemma 7** (Lemma 2.1 in [Tsuda et al., 2005]). For any symmetric matrix $A$ such that $0 \preceq A \preceq I$ and any $\rho_{1}, \rho_{2} \in \mathbb{R}$, the following holds:

$$
\text{exp} \left( \rho_{1} + (I - A) \rho_{2} \right) \leq A \text{exp}(\rho_{1}) + (I - A) \text{exp}(\rho_{2})
$$

Then we apply the Golden-Thompson inequality to the term $\text{Tr} \left( \text{exp}(\ln Y_{t} - \eta C_{t}) \right)$, which gives us the inequality below:

$$
\text{Tr} \left( \text{exp}(\ln Y_{t} - \eta C_{t}) \right) \leq \text{Tr}(Y_{t} \text{exp}(-\eta C_{t}))
$$

(53)

For the term $\text{exp}(-\eta C_{t})$, by applying the Lemma 7 with $\rho_{1} = -\eta$ and $\rho_{2} = 0$, we will have the following inequality:

$$
\text{exp}(-\eta C_{t}) \leq I - C_{t}(1 - \text{exp}(-\eta))
$$

(54)

Thus, we will have

$$
\text{Tr}(Y_{t} \text{exp}(-\eta C_{t})) \leq 1 - \text{Tr}(Y_{t} C_{t})(1 - \text{exp}(-\eta))
$$

(55)

and

$$
\text{Tr} \left( \text{exp}(\ln Y_{t} - \eta C_{t}) \right) \leq 1 - \text{Tr}(Y_{t} C_{t})(1 - \text{exp}(-\eta))
$$

(56)

since $Y_{t} \in \mathcal{B}_{n}^{n}$ and $\text{Tr}(Y_{t}) = 1$.

Thus, it is enough to prove the following inequality

$$
\eta \text{Tr}(Y_{t} C_{t}) - \frac{\eta^{2}}{2} + \ln \left( 1 - \text{Tr}(Y_{t} C_{t}) \right)(1 - \text{exp}(-\eta)) \leq 0
$$

(57)

Since $\ln(1 - x) \leq -x$, we have

$$
\ln \left( 1 - \text{Tr}(Y_{t} C_{t}) \right)(1 - \text{exp}(-\eta)) \leq - \text{Tr}(Y_{t} C_{t})(1 - \text{exp}(-\eta))
$$

(58)

Thus, it suffices to prove the following inequality:

$$
(\eta - 1 + \text{exp}(-\eta)) \text{Tr}(Y_{t} C_{t}) - \frac{\eta^{2}}{2} \leq 0
$$

(59)

Note that by using convexity of $\text{exp}(-\eta)$, $\eta - 1 + \text{exp}(-\eta) \geq 0$.

By applying Lemma 5 with $A = Y_{t}$, $B = C_{t}$, and $C = I$, we have $\text{Tr}(Y_{t} C_{t}) \leq \text{Tr}(Y_{t}) = 1$. Thus, when $\eta \geq 0$, it is enough to prove the following inequality

$$
\eta - 1 + \text{exp}(-\eta) - \frac{\eta^{2}}{2} \leq 0
$$

(60)

This inequality follows from convexity of $\frac{\eta^{2}}{2} - \text{exp}(-\eta)$ over $\eta \geq 0$. 

\qed
D. Proof of Theorem 6

Proof. First, since $0 \leq C_t \leq I$, we have $\max_{i,j} |C_t(i,j)| \leq 1$.

Before we proceed, we need the following lemma from (Warmuth & Kuzmin, 2006)

**Lemma 8** (Lemma 1 in (Warmuth & Kuzmin, 2006)). Let $\max_{i,j} |C_t(i,j)| \leq \frac{\gamma}{2}$, then for any $u_t \in \mathbb{B}^r_1$, any constants $a$ and $b$ such that $0 \leq a \leq \frac{b}{1+b^r}$, and $\eta = \frac{2b}{1+b^r}$, we have

$$a y_t^T C_t y_t - b u_t^T C_t u_t \leq d(u_t, y_t) - d(u_t, v_{t+1})$$

Now we apply Lemma 8 under the conditions $r = 2$, $a = \frac{b}{2b+1}$, $\eta = 2a$, and $b = \frac{\gamma}{2}$.

Recall that $d(u_t, y_t) - d(u_t, v_{t+1}) = \sum_i u_{t,i} \ln \left( \frac{y_{t+1,i}}{y_{t,i}} \right)$. Combining this with the inequality in Lemma 8 and the fact that $\|u_t\|_1 = 1$, we have

$$a \|u_t\|_1 y_t^T C_t y_t - b u_t^T C_t u_t \leq \sum_i u_{t,i} \ln \left( \frac{y_{t+1,i}}{y_{t,i}} \right)$$

(61)

Note that the above inequality is also true when $u_t = 0$.

Note that the right side of the above inequality is the same as the right part of the Eq.(32) in the proof of Theorem 1.

As a result, we will use the same steps as in the proof of Theorem 1. Then we will set $u_t = u = \arg\min_{q \in B^r_1} \sum_{t=r}^s q^T C_t q$ for $t = r, \ldots, s$, and 0 elsewhere.

Summing from $t = 1$ up to $T$, gives the following inequality:

$$a \left[ \sum_{t=r}^s y_t^T C_t y_t \right] - b \left[ \min_{u \in B^r_1} \sum_{t=r}^s u^T C_t u \right] \leq \ln \frac{n}{\alpha} + T \ln \frac{1}{1-\alpha}$$

(62)

Since $\alpha = 1/(T+1)$, $T \ln \frac{1}{1-\alpha} \leq 1$. Then the above inequality becomes

$$a \left[ \sum_{t=r}^s y_t^T C_t y_t \right] - b \left[ \min_{u \in B^r_1} \sum_{t=r}^s u^T C_t u \right] \leq \ln ((1+T)n) + 1$$

(63)

Plugging in the expressions of $a = c/(2c+2)$, $b = c/2$, and $c = \sqrt{2 \ln (2c+2) + 2}$, we will have

$$\sum_{t=r}^s y_t^T C_t y_t - \min_{u \in B^r_1} \sum_{t=r}^s u^T C_t u$$

$$\leq c \left[ \min_{u \in B^r_1} u^T C_t u \right] + 2 + 2 \ln ((1+T)n) + 1$$

$$\leq c L + 2 \ln ((1+T)n) + 1$$

$$= 2 \sqrt{2 L \ln ((1+T)n) + 1} + 2 \ln ((1+T)n)$$

(64)