Robust PCA in High-dimension: A Deterministic Approach

1. Proof of Corollary 1

**Lemma 1.** For any $\epsilon > 0$ and $\kappa \in [\epsilon, 1]$, we have $V(\kappa) - V(\kappa - \epsilon) \leq C\alpha \epsilon \log^2(1/\epsilon)$.

**Proof.** By monotonicity, it suffices to prove that result for $\kappa = 1$. Notice that for $K \geq 2\alpha$,

\[
V(1) - V(1 - \epsilon) \\
\leq \epsilon K^2 + \mathbb{E}_{x \sim \bar{\mu}} (x^2 \cdot 1(x > K)) \\
= \epsilon K^2 + \int_{K^2}^{\infty} \Pr(x^2 > z)dz \\
\leq \epsilon K^2 + \int_{K^2}^{\infty} \exp(1 - \sqrt{z}/\alpha)dz \\
= \epsilon K^2 + e_0 \int_{K^2/4\alpha^2}^{\infty} \exp(-2\sqrt{z})dz \\
\leq (a) \epsilon K^2 + 2e_0 \exp(-\sqrt{z})|_{K^2/4\alpha^2} \\
= \epsilon K^2 + \exp(1 + \ln 2 - K/2\alpha),
\]

where (a) holds because when $z \geq 1$, we have $\exp(-\sqrt{z}) \leq 1/\sqrt{z}$, which implies $\exp(-2\sqrt{z}) \leq \frac{d(2\exp(-\sqrt{z}))}{dz}$. Pick $K = 2\alpha \log(1/\epsilon)$, we have that

\[
V(1) - V(1 - \epsilon) \leq C\alpha \epsilon \log^2(1/\epsilon).
\]

\[\square\]

**Corollary 1.** Under the settings of the above theorem, the following holds in probability when $j \uparrow \infty$ (i.e., when $n, p \uparrow \infty$),

\[
\liminf_j \frac{\text{E.V.} \{w_1(j), \ldots, w_d(j)\}}{V(0.5)} \geq 1 - \frac{C' \sqrt{\alpha \lambda^* \log(1/\lambda^*)}}{V(0.5)}.
\]
Proof. We bound the right-hand-side of Equation (2) to establish the corollary. Notice that

\[
\frac{\mathcal{V}(1) - C_\alpha \frac{\lambda_s^*}{(1 - \lambda^*) \kappa} \log^2 \left( \frac{(1 - \lambda^*) \kappa}{\lambda^*(1 + \kappa)} \right)}{(1 + \kappa)} \times \frac{\mathcal{V}(i) - C_\alpha \frac{\lambda_s^*}{1 - \lambda^*} \log^2 \left( \frac{1 - \lambda^*}{\lambda^*} \right)}{\mathcal{V}(i)}
\]

\[\geq \left[ 1 - \frac{2C_\alpha \lambda_s^*}{\kappa} \log^2 \left( \frac{1}{\lambda^*} \right) \right] \times \left[ 1 - \frac{2C_\alpha \lambda_s^* \log^2 \left( \frac{1}{\lambda^*} \right)}{\mathcal{V}(0.5)} \right]
\]

\[\geq 1 - \frac{C'_\alpha \lambda_s^*}{\kappa} \log^2 \left( \frac{1}{\lambda^*} \right) - \frac{C'_\alpha \lambda_s^* \log^2 \left( \frac{1}{\lambda^*} \right)}{\mathcal{V}(0.5)}
\]

\[\geq 1 - \frac{2C'_\alpha \lambda_s^*}{\kappa \mathcal{V}(0.5)} \log^2 \left( \frac{1}{\lambda^*} \right).
\]

Here, (a) is due to Lemma 1; (b) is due to \( \mathcal{V}(1) = 1 \); (c) holds because \( \frac{1}{1 + \kappa} \geq 1 - \kappa, 1 - \lambda^* \geq 1/2 \), and \( \mathcal{V}(i/t) \geq \mathcal{V}(0.5) \); (d) holds because \( \kappa \) and \( \mathcal{V}(0.5) \) are both smaller than or equal to 1.

2. Proof of Theorem 5

Theorem 5 The event \( \mathcal{E}(s) \) is true for some \( 1 \leq s \leq s_0 \), where \( s_0 \leq \frac{\lambda n(1 + \kappa)}{\kappa} \).

Proof. If \( \mathcal{E}'(s) \) is true, then

\[
\sum_{j=1}^{d} \sum_{i \in \mathcal{Z}} \alpha_i^{(s)} (w_j(s)^T w_i)^2 < \frac{1}{\kappa} \sum_{j=1}^{d} \sum_{i \in \mathcal{O}} \alpha_i^{(s)} (w_j(s)^T w_i)^2.
\]

Since \( \Delta \alpha_i^{(s)} = \eta \alpha_i^{(s)} \sum_{j=1}^{d} (w_j(s)^T \hat{y}_i)^2 \), we have

\[
\sum_{i \in \mathcal{Z}} \Delta \alpha_i^{(s)} < \frac{1}{\kappa} \sum_{i \in \mathcal{O}} \Delta \alpha_i^{(s)}.
\]

If \( \bigcap_{s=1}^{s_0} \mathcal{E}'(s) \) is true,

\[
\sum_{s=1}^{s_0} \sum_{i \in \mathcal{Z}} \Delta \alpha_i^{(s)} < \frac{1}{\kappa} \sum_{s=1}^{s_0} \sum_{i \in \mathcal{O}} \Delta \alpha_i^{(s)}.
\]

In the Algorithm 1, we eliminate at least one weight coefficient in each iteration. Therefore, to step \( s_0 \), we have \( \sum_{s=1}^{s_0} \sum_{i \in \mathcal{Z}} \Delta \alpha_i^{(s)} \geq s_0 \). Namely,

\[
\sum_{s=1}^{s_0} \sum_{i \in \mathcal{Z}} \Delta \alpha_i^{(s)} + \sum_{s=1}^{s_0} \sum_{i \in \mathcal{O}} \Delta \alpha_i^{(s)} \geq s_0.
\]

Thus,

\[
\frac{1}{\kappa} \sum_{s=1}^{s_0} \sum_{i \in \mathcal{O}} \Delta \alpha_i^{(s)} + \frac{1}{\kappa} \sum_{s=1}^{s_0} \sum_{i \in \mathcal{O}} \Delta \alpha_i^{(s)} \geq s_0.
\]
From the above inequality, we can obtain
\[ \lambda n \geq \sum_{s=1}^{s_0} \sum_{i \in \mathcal{O}} \Delta \alpha_i^{(s)} \geq \frac{s_0 \kappa}{1 + \kappa}. \]

Therefore, we can conclude bound $s_0 \leq \frac{\lambda n (1+\kappa)}{\kappa}$.

\[ \square \]

3. Proof of Theorem 6

As stated in the main body, our proof comprises following two steps.

**Lemma 2.** If $\mathcal{E}(s)$ is true for some $s \leq s_0$, and there exist $\epsilon_1$ such that $\sup_{w \in S_d} \left| \frac{1}{t} \sum_{i=1}^{t-s_0} |w^T x_i^2 - \mathcal{V}(\frac{t-s_0}{t})| \right| \leq \epsilon_1$ and $\epsilon_2, \bar{c}$ satisfying conditions (II) and (III) in Theorem 4, then
\[ \frac{1}{1 + \kappa} \left[ (1 - \epsilon_1) \mathcal{V}(\frac{t-s_0}{t}) \bar{H} - 2 \sqrt{(1 + \epsilon_2) \bar{c} \bar{H}} \right] \leq (1 + \epsilon_2) H_s + 2 \sqrt{(1 + \epsilon_2) \bar{c} \bar{H}} s + \bar{c}. \]

**Proof.** If $\mathcal{E}(s)$ is true, then we have
\[ \sum_{j=1}^{d} \sum_{i=1}^{t} \alpha_i^{(s)} (w_j(s)^T z_i)^2 \geq \frac{1}{\kappa} \sum_{j=1}^{d} \sum_{i=1}^{t} \alpha_i^{(s)} (w_j(s)^T y_i)^2. \]

Thus we have
\[ \frac{1}{1 + \kappa} \sum_{j=1}^{d} \sum_{i=1}^{n} \alpha_i (w_j(s)^T y_i)^2 \leq \sum_{j=1}^{d} \sum_{i=1}^{n} \alpha_i (w_j(s)^T z_i)^2. \]

Since $w_1(s), \ldots, w_d(s)$ is the solution of the $s^{th}$ stage, the following holds by definition of the algorithm
\[ \sum_{j=1}^{d} \sum_{i=1}^{n} \alpha_i (w_j(s)^T y_i)^2 \leq \sum_{j=1}^{d} \sum_{i=1}^{n} (w_j(s)^T y_i)^2. \]

Since $0 \leq \alpha_i \leq 1, \forall i = 1, \ldots, n$, we have
\[ \sum_{j=1}^{d} \sum_{i=1}^{n} \alpha_i (w_j(s)^T y_i)^2 \leq \sum_{j=1}^{d} \sum_{i=1}^{n} (w_j(s)^T y_i)^2. \]

Since $1 \leq s \leq s_0$, from the definition of the algorithm, we have $\sum_{i \in \mathcal{Z}} \alpha_i \geq t - s_0$. Thus
\[ \sum_{i=1}^{t} \alpha_i (w_j^T z_i)^2 - \sum_{i=1}^{t-s_0} |w_j^T z_i|_{(i)}^2 = \sum_{i=1}^{t-s_0} (\alpha(i) - 1) |w_j^T z_i|_{(i)}^2 + \sum_{i=t-s_0+1}^{t} \alpha(i) |w_j^T z_i|_{(i)}^2. \]

\[ \geq \sum_{i=1}^{t-s_0} (\alpha(i) - 1) |w_j^T z_i|_{(t-s_0)}^2 + \sum_{i=t-s_0+1}^{t} \alpha(i) |w_j^T z_i|_{(t-s_0)}^2 = \left[ \sum_{i=1}^{t} \alpha(i) (t - s_0) \right] |w_j^T z_i|_{(t-s_0)}^2 \geq 0. \]
Thus we have
\[ \sum_{j=1}^{d} \sum_{i=1}^{t-s_0} |w_j^T z_i|^2 \leq \sum_{j=1}^{d} \sum_{i=1}^{t} \alpha_i (w_j^T z_i)^2 \leq \sum_{j=1}^{d} \sum_{i=1}^{n} \alpha_i (w_j^T y_i)^2. \]

Combining the above inequalities, we get
\[ \frac{1}{1 + \kappa} \sum_{j=1}^{d} \sum_{i=1}^{t-s_0} |\hat{w}_j^T z_i|^2 \leq \sum_{j=1}^{d} \sum_{i=1}^{t} (w_j(s)^T z_i)^2. \]

By Corollary 1 we complete the proof. \(\square\)

The following lemma guarantees that the value \(H^\ast\) of the algorithm’s output is lower bounded in terms of the value \(H\) of any output that has a smaller value of the robust variance estimator.

**Lemma 3.** Fix a \(\hat{t} \leq t\). If \(\sum_{j=1}^{d} \hat{V}_i(w_j) \geq \sum_{j=1}^{d} \hat{V}_i(w_j)\), and there exists \(\epsilon_1, \epsilon_2\) and \(\bar{c}\) such that \(\sup_{w \in S_d} \left| \frac{1}{1 + \kappa} \sum_{i=1}^{t-s_0} |w^T x_i|^2 - \mathcal{V} \left( \frac{\hat{t}}{t} - \frac{\lambda}{1 - \lambda} \right) \right| \leq \epsilon_1\) and conditions in Theorem 4 are satisfied, then
\[
(1 - \epsilon_1) \mathcal{V} \left( \frac{\hat{t}}{t} - \frac{\lambda}{1 - \lambda} \right) H(w') - 2\sqrt{(1 + \epsilon_2)\bar{c}dH(w')} \leq (1 + \epsilon_1) H(w) \mathcal{V} \left( \frac{\hat{t}}{t} \right) + 2\sqrt{(1 + \epsilon_2)\bar{c}dH(w) + \bar{c}}.
\]

**Theorem 6** If \(\bigcup_{s=1}^{n} \mathcal{E}(s)\) is true, and there exist \(\epsilon_1 < 1, \epsilon_2, \bar{c}\) such that \(\sup_{w \in S_d} \left| \frac{1}{1 + \kappa} \sum_{i=1}^{t-s_0} |w^T x_i|^2 - \mathcal{V} \left( \frac{t-s_0}{t} \right) \right| \leq \epsilon_1\) and Condition 1 holds, then
\[
\frac{H^\ast}{H} \geq \frac{(1 - \epsilon_1)^2 \mathcal{V} \left( \frac{\hat{t}}{t} - \frac{\lambda}{1 - \lambda} \right) \mathcal{V} \left( \frac{t-s_0}{t} \right)}{(1 + \epsilon_1)(1 + \epsilon_2)(1 + \kappa) \mathcal{V} \left( \frac{\hat{t}}{t} \right)} - \left[ \frac{(2\kappa + 4)(1 - \epsilon_1) \mathcal{V} \left( \frac{\hat{t}}{t} - \frac{\lambda}{1 - \lambda} \right) + 4(1 + \epsilon_2)(1 + \kappa) \sqrt{(1 + \epsilon_2)\bar{c}d}}{(1 + \epsilon_1)(1 + \epsilon_2)(1 + \kappa)} \right] (H)^{-1/2}
\]
\[
- \left[ \frac{(1 - \epsilon_1) \mathcal{V} \left( \frac{\hat{t}}{t} - \frac{\lambda}{1 - \lambda} \right) \bar{c} + (1 + \epsilon_2) \bar{c}}{(1 + \epsilon_1)(1 + \epsilon_2) \mathcal{V} \left( \frac{\hat{t}}{t} \right)} \right] (H)^{-1},
\]

**Proof.** Since \(\bigcup_{s=1}^{n} \mathcal{E}(s)\) is true, there exists a \(s' \leq s_0\) such that \(\mathcal{E}(s')\) is true. By Lemma 2 we have
\[
\frac{1}{1 + \kappa} \left[ (1 - \epsilon_1) \mathcal{V} \left( \frac{t-s_0}{t} \right) \bar{H} - 2\sqrt{(1 + \epsilon_2)\bar{c}d\bar{H}} \right] \leq (1 + \epsilon_2) H_{s'} + 2\sqrt{(1 + \epsilon_2)\bar{c}dH_{s'} + \bar{c}}.
\]

By the definition of the algorithm, we have \(\sum_{j=1}^{d} \hat{V}_i(w_j^\ast) \geq \sum_{j=1}^{d} \hat{V}_i(w_j(s'))\), which by Lemma 3 implies
\[
(1 - \epsilon_1) \mathcal{V} \left( \frac{\hat{t}}{t} - \frac{\lambda}{1 - \lambda} \right) H_{s'} - 2\sqrt{(1 + \epsilon_2)\bar{c}dH_{s'}} \leq (1 + \epsilon_1) H^\ast \mathcal{V} \left( \frac{\hat{t}}{t} \right) + 2\sqrt{(1 + \epsilon_2)\bar{c}dH^\ast + \bar{c}}.
\]

By definition, \(H_{s'}, H^\ast \leq \bar{H}\). Thus we have
\[
(I) \quad \frac{1}{1 + \kappa} \left[ (1 - \epsilon_1) \mathcal{V} \left( \frac{t-s_0}{t} \right) \bar{H} - 2\sqrt{(1 + \epsilon_2)\bar{c}d\bar{H}} \right] \leq (1 + \epsilon_2) H_{s'} + 2\sqrt{(1 + \epsilon_2)\bar{c}d\bar{H} + \bar{c}};
\]
\[
(II) \quad (1 - \epsilon_1) \mathcal{V} \left( \frac{\hat{t}}{t} - \frac{\lambda}{1 - \lambda} \right) H_{s'} - 2\sqrt{(1 + \epsilon_2)\bar{c}dH_{s'}} \leq (1 + \epsilon_1) H^\ast \mathcal{V} \left( \frac{\hat{t}}{t} \right) + 2\sqrt{(1 + \epsilon_2)\bar{c}dH^\ast + \bar{c}}.
\]
Deterministic Robust PCA

Rearrange the inequalities, we have

\[(I) \quad (1 - \epsilon_1) \mathcal{V} \left( \frac{t - s_0}{t} \right) \bar{H} - (2\kappa + 4) \sqrt{(1 + \epsilon_2)\bar{c}d\bar{H}} - (1 + \kappa)\bar{c} \leq (1 + \kappa)(1 + \epsilon_2)H_s';\]

\[(II) \quad (1 - \epsilon_1) \mathcal{V} \left( \frac{i}{t} - \frac{\lambda}{1 - \lambda} \right) H_{s'} \leq (1 + \epsilon_1) \mathcal{V} \left( \frac{i}{t} \right) H^* + 4 \sqrt{(1 + \epsilon_2)\bar{c}d\bar{H}} + \bar{c}.

Simplify the inequality, we get

\[
\frac{H^*}{\bar{H}} \geq \frac{(1 - \epsilon_1)^2 \mathcal{V} \left( \frac{i}{t} - \frac{\lambda}{1 - \lambda} \right) \mathcal{V} \left( \frac{t - s_0}{t} \right)}{(1 + \epsilon_1)(1 + \epsilon_2)(1 + \kappa) \mathcal{V} \left( \frac{i}{t} \right)}
- \left[ \frac{(2\kappa + 4)(1 - \epsilon_1) \mathcal{V} \left( \frac{i}{t} - \frac{\lambda}{1 - \lambda} \right) + 4(1 + \epsilon_2)(1 + \kappa)}{(1 + \epsilon_1)(1 + \epsilon_2)(1 + \kappa)} \sqrt{(1 + \epsilon_2)\bar{c}d} \right] (\bar{H})^{-1/2}
- \left[ \frac{(1 - \epsilon_1) \mathcal{V} \left( \frac{i}{t} - \frac{\lambda}{1 - \lambda} \right) \bar{c} + (1 + \epsilon_2)\bar{c}}{(1 + \epsilon_1)(1 + \epsilon_2) \mathcal{V} \left( \frac{i}{t} \right)} \right] (\bar{H})^{-1},
\]

\(4. \) Simulations

In the following figures, we provide more simulation results for comparison between DHR-PCA and HR-PCA.

\(\lambda = 0.01\)
\(\lambda = 0.03\)
\(\lambda = 0.05\)
\(\lambda = 0.08\)
\(\lambda = 0.10\)
\(\lambda = 0.15\)
\(\lambda = 0.20\)
\(\lambda = 0.30\)
\(\lambda = 0.40\)

Figure 1. DHR-PCA (red line) vs. HR-PCA (black line). \(m = n = 100, \sigma = 2\). The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.
Figure 2. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 100, \sigma = 3$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.
Figure 3. DHR-PCA (red line) vs. HR-PCA (black line). \( m = n = 100, \sigma = 10 \). The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.
Figure 4. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 100, \sigma = 20$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.
Figure 5. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 1000, \sigma = 2$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.
Figure 6. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 1000, \sigma = 3$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.
Figure 7. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 1000, \sigma = 10$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.
Figure 8. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 1000, \sigma = 20$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.
Figure 9. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 10000, \sigma = 2$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.

Figure 10. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 10000, \sigma = 3$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.
Figure 11. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 10000, \sigma = 10$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.

Figure 12. DHR-PCA (red line) vs. HR-PCA (black line). $m = n = 10000, \sigma = 20$. The horizontal axis is the iteration and the vertical axis is the expressive variance value. Please refer to the color version.