

Supplementary Material

7 More examples of Assumption 1.2

Assumption 1.3 is one simple example of a support change model that ensures that, if $M_{2,t} = I_{\mathcal{T}_t}$, the assumption on $M_{2,t}$ given in Assumption 1.2 holds. If instead of one object, there are k objects, and each of their supports satisfies Assumption 1.3, then again, with some modifications, it is possible to show that both the PCA-missing and PCA-SDDC problems satisfy Assumption 1.2. Moreover, notice that Assumption 1.3 does not require the entries in \mathcal{T}_t to be contiguous at all (they need not correspond to the support of one or a few objects). Similarly, we can replace the condition that \mathcal{T}_t be constant for at most $\tilde{\beta}$ time instants in Assumption 1.3 by $|\{t : \mathcal{T}_t = \mathcal{T}^{[k]}\}| \leq \tilde{\beta}$.

Thirdly, the requirement of the object(s) always moving in one direction may seem too stringent. As explained in [4, Lemma 9.4], a Bernoulli-Gaussian “constant velocity with random acceleration” motion model will also work whp. It allows the object to move at each frame with probability p and not move with probability $1 - p$ independent of past or future frames; when the object moves, it moves with an iid Gaussian velocity that has mean $1.1s/\rho$ and variance σ^2 ; σ^2 needs to be upper bounded and p needs to be lower bounded.

Lastly, if $s < c_1\alpha$ for $c_1 \ll 1$, another model that works is that of an object of length s or less moving by at least one pixel and at most b pixels at each time [4, Lemma 9.5].

8 Proof of Theorem 2.1

This result also follows as a corollary of Theorem 3.3. We prove it separately first since its proof is short and and less notation-ally intensive. It will help understand the proof of Theorem 3.3 much more easily. Both results rely on the $\sin \theta$ theorem reviewed next.

8.1 $\sin \theta$ theorem

Davis and Kahan’s $\sin \theta$ theorem [19] studies the rotation of eigenvectors by perturbation.

Theorem 8.1 ($\sin \theta$ theorem [19]). *Consider two Hermitian matrices D and \hat{D} . Suppose that D can be decomposed as*

$$D = \begin{bmatrix} E & E_\perp \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & A_\perp \end{bmatrix} \begin{bmatrix} E' \\ E'_\perp \end{bmatrix}$$

where $\begin{bmatrix} E & E_\perp \end{bmatrix}$ is an orthonormal matrix. Suppose that \hat{D} can be decomposed as

$$\hat{D} = \begin{bmatrix} F & F_\perp \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda_\perp \end{bmatrix} \begin{bmatrix} F' \\ F'_\perp \end{bmatrix}$$

where $\begin{bmatrix} F & F_\perp \end{bmatrix}$ is another orthonormal matrix and is such that $\text{rank}(F) = \text{rank}(E)$. Let $H := \hat{D} - D$ denote the perturbation. If $\lambda_{\min}(A) > \lambda_{\max}(\Lambda_\perp)$, then

$$\|(I - FF')E\| \leq \frac{\|H\|}{\lambda_{\min}(A) - \lambda_{\max}(\Lambda_\perp)}.$$

Let $r = \text{rank}(E)$. Suppose that F is the matrix of top r eigenvectors of \hat{D} . Then Λ and Λ_\perp are diagonal and $\lambda_{\max}(\Lambda_\perp) = \lambda_{r+1}(\hat{D}) \leq \lambda_{r+1}(D) + \|H\|$. The inequality follows using Weyl’s inequality. Suppose also that $\lambda_{\min}(A) > \lambda_{\max}(A_\perp)$. Then, (i) $\lambda_r(D) = \lambda_{\min}(A)$ and $\lambda_{r+1}(D) = \lambda_{\max}(A_\perp)$ and (ii) $\text{range}(E)$ is equal to the span of the top r eigenvectors of D . Thus, $\lambda_{\max}(\Lambda_\perp) \leq \lambda_{\max}(A_\perp) + \|H\|$. With this we have the following corollary.

Corollary 8.2. *Consider a Hermitian matrix D and its perturbed version \hat{D} . Suppose that D can be decomposed as*

$$D = \begin{bmatrix} E & E_\perp \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & A_\perp \end{bmatrix} \begin{bmatrix} E' \\ E'_\perp \end{bmatrix}$$

where E is a basis matrix. Let F denote the matrix containing the top $\text{rank}(E)$ eigenvectors of \hat{D} . Let $H := \hat{D} - D$ denote the perturbation. If $\lambda_{\min}(A) - \lambda_{\max}(A_\perp) - \|H\| > 0$, then

$$\|(I - FF')E\| \leq \frac{\|H\|}{\lambda_{\min}(A) - \lambda_{\max}(A_\perp) - \|H\|}.$$

and $\text{range}(E)$ is equal to the span of the top $\text{rank}(E)$ eigenvectors of D .

8.2 Proof of Theorem 2.1

We use the $\sin \theta$ theorem [19] from Corollary 8.2. Apply it with $\hat{D} = \frac{1}{\alpha} \sum_t \mathbf{y}_t \mathbf{y}_t'$ and $D = \frac{1}{\alpha} \sum_t \ell_t \ell_t'$. Thus, $F = \hat{P}$. Recall that $\mathbf{a}_t = P' \ell_t$. Then, D can be decomposed as $P(\frac{1}{\alpha} \sum_t \mathbf{a}_t \mathbf{a}_t')P' + P_\perp \mathbf{0} P_\perp'$, and so we have $E = P$, $A = \frac{1}{\alpha} \sum_t \mathbf{a}_t \mathbf{a}_t'$ and $A_\perp = \mathbf{0}$. Moreover, it is easy to see that the perturbation $H := \frac{1}{\alpha} \sum_t \mathbf{y}_t \mathbf{y}_t' - \frac{1}{\alpha} \sum_t \ell_t \ell_t'$ satisfies

$$H = \frac{1}{\alpha} \sum_t \ell_t \mathbf{w}_t' + \frac{1}{\alpha} \sum_t \mathbf{w}_t \ell_t' + \frac{1}{\alpha} \sum_t \mathbf{w}_t \mathbf{w}_t'. \quad (7)$$

Thus,

$$\begin{aligned} \text{SE}(\hat{P}, P) &\leq \frac{2\|\frac{1}{\alpha} \sum_t \ell_t \mathbf{w}_t'\| + \|\frac{1}{\alpha} \sum_t \mathbf{w}_t \mathbf{w}_t'\|}{\lambda_r(\frac{1}{\alpha} \sum_t \ell_t \ell_t') - (2\|\frac{1}{\alpha} \sum_t \ell_t \mathbf{w}_t'\| + \|\frac{1}{\alpha} \sum_t \mathbf{w}_t \mathbf{w}_t'\|)} \end{aligned}$$

if the denominator is positive.

Remark 8.3. Because \mathbf{w}_t is correlated with ℓ_t , the $\ell_t \mathbf{w}_t'$ terms are the dominant ones in the perturbation expression given in (7). If they were uncorrelated, these two terms would be close to zero whp due to law of large numbers and the $\mathbf{w}_t \mathbf{w}_t'$ term would be the dominant one.

In the next lemma, we bound the terms in the bound on $\text{SE}(\hat{P}, P)$ using the matrix Hoeffding inequality [20].

Lemma 8.4. Let $\epsilon = 0.01r\zeta\lambda^-$.

1. With probability at least $1 - 2n \exp\left(-\alpha \frac{\epsilon^2}{32(\eta r q \lambda^+)^2}\right)$,

$$\left\| \frac{1}{\alpha} \sum_t \ell_t \mathbf{w}_t' \right\| \leq q\lambda^+ \sqrt{\frac{\beta}{\alpha}} + \epsilon = [qf \sqrt{\frac{\beta}{\alpha}} + 0.01r\zeta]\lambda^-$$

2. With probability at least $1 - 2n \exp\left(-\frac{\alpha \epsilon^2}{32(\eta r q^2 \lambda^+)^2}\right)$,

$$\left\| \frac{1}{\alpha} \sum_t \mathbf{w}_t \mathbf{w}_t' \right\| \leq \frac{\beta}{\alpha} q^2 \lambda^+ + \epsilon = [q^2 f \frac{\beta}{\alpha} + 0.01r\zeta]\lambda^-$$

3. With probability at least $1 - 2n \exp\left(-\frac{\alpha \epsilon^2}{32(\eta r \lambda^+)^2}\right)$,

$$\lambda_r\left(\frac{1}{\alpha} \sum_t \ell_t \ell_t'\right) \geq (1 - (r\zeta)^2)\lambda^- - \epsilon$$

Proof. This follows by using Lemma 9.6 given later with $\mathbf{G}_{\text{cur}} \equiv P$, $\mathbf{G}_{\text{det}} \equiv [\cdot]$, $\mathbf{G}_{\text{undet}} \equiv [\cdot]$, $\zeta_{\text{det}} \equiv 0$, $r\zeta \equiv 0$, $r_{\text{cur}} = r$, $g \equiv f$, $\chi \equiv 0$, $\vartheta \equiv 1$. \square

Using this lemma to bound the subspace error terms, followed by using the bounds on β/α and ζ , we conclude the following: w.p. at least $1 - 2n \exp\left(-\alpha \frac{\epsilon^2}{32(\eta r q \lambda^+)^2}\right) - 2n \exp\left(-\frac{\alpha \epsilon^2}{32(\eta r q^2 \lambda^+)^2}\right) - 2n \exp\left(-\frac{\alpha \epsilon^2}{32(\eta r \lambda^+)^2}\right)$,

$$\begin{aligned} \text{SE}(\hat{P}, P) &\leq \frac{2qf \sqrt{\frac{\beta}{\alpha}} + q^2 f \frac{\beta}{\alpha} + 0.03r\zeta}{1 - (r\zeta)^2 - 0.01r\zeta - (2qf \sqrt{\frac{\beta}{\alpha}} + q^2 f \frac{\beta}{\alpha} + 0.03r\zeta)} \\ &\leq \frac{0.75(1 - r\zeta)r\zeta + 0.03r\zeta}{1 - r\zeta} < r\zeta \end{aligned}$$

Using the bound $\alpha \geq \alpha_0$ from the theorem, the probability of the above event is at least $1 - 6n^{-10}$. We get this by bounding each of the three negative terms in the probability expression by $-2n^{-10}$.

We work this out for the first term: $\alpha \frac{\epsilon^2}{32(\eta r q \lambda^+)^2} \geq \frac{32 \cdot 11}{(0.01)^2} \frac{\eta^2 r^2 (\log n)}{(r\zeta)^2} (qf)^2 \frac{(0.01r\zeta\lambda^-)^2}{32\eta^2 r^2 q^2 \lambda^+} = 11 \log n$.

Thus, $2n \exp\left(-\alpha \frac{\epsilon^2}{32(\eta r q \lambda^+)^2}\right) \leq 2n \exp(-11 \log n) \leq 2n^{-10}$.

9 Proof of Theorem 3.3

We explain the overall idea of the proof next. In Sec. 9.2, we give a sequence of lemmas in generalized form (so that they can apply to various other problems). The proof of Theorem 3.3 is given in Sec. 9.3 and follows easily by applying these. One of the lemmas of Sec. 9.2 is proved in Sec. 10 while the others are proved there itself.

9.1 Overall idea

We need to bound $\text{SE}(\hat{P}, P)$. From Algorithm 1, $\hat{P} = [\hat{G}_1, \hat{G}_2, \dots, \hat{G}_\vartheta]$ where \hat{G}_k is the matrix of top \hat{r}_k eigenvectors of \hat{D}_k defined in Algorithm 1. Also, $P = [G_1, G_2, \dots, G_\vartheta]$ where G_k is a basis matrix with r_k columns.

Definition 9.1. Define $\zeta_k := \text{SE}([\hat{G}_1, \hat{G}_2, \dots, \hat{G}_k], G_k)$ and $\zeta_0 = 0$. Define $\zeta_k^+ := r_k \zeta$. Let $r_0 = 0$.

It is easy to see that

$$\begin{aligned} \text{SE}(\hat{P}, P) &\leq \sum_{k=1}^{\vartheta} \text{SE}(\hat{P}, G_k) \\ &\leq \sum_{k=1}^{\vartheta} \text{SE}([\hat{G}_1, \hat{G}_2, \dots, \hat{G}_k], G_k) = \sum_{k=1}^{\vartheta} \zeta_k \end{aligned} \quad (8)$$

The first inequality is triangle inequality, the second follows because $[\hat{G}_1, \hat{G}_2, \dots, \hat{G}_k]$ is orthogonal to $[\hat{G}_{k+1}, \dots, \hat{G}_\vartheta]$. Since $r = \sum_k r_k$, if we can show that $\zeta_k \leq \zeta_k^+ = r_k \zeta$ for all k we will be done.

We bound ζ_k using induction. The base case is easy and follows just from the definition, $\zeta_0 = \text{SE}([\cdot], [\cdot]) = 0 = r_0 \zeta$. For bounding ζ_k , assume that for all $i = 1, 2, \dots, k-1$, $\zeta_i \leq r_i \zeta$. This implies that

$$\begin{aligned} &\text{SE}([\hat{G}_1, \hat{G}_2, \dots, \hat{G}_{k-1}], [G_1, G_2, \dots, G_{k-1}]) \\ &\leq \sum_{i=1}^{k-1} \text{SE}([\hat{G}_1, \hat{G}_2, \dots, \hat{G}_{k-1}], G_i) \\ &\leq \sum_{i=1}^{k-1} \zeta_i \leq \sum_{i=1}^{k-1} r_i \zeta \leq r \zeta \end{aligned} \quad (9)$$

Using this, we will first show that $\hat{r}_k = r_k$, and then we will use this and the $\sin \theta$ result to bound ζ_k .

Before proceeding further, we simplify notation.

Definition 9.2.

1. Let

$$\begin{aligned} G_{\text{det}} &:= [G_1, G_2, \dots, G_{k-1}], \quad G_{\text{cur}} := G_k, \\ G_{\text{undet}} &:= [\hat{G}_{k+1}, \dots, \hat{G}_\vartheta] \end{aligned}$$

2. Similarly, let $\hat{G}_{\text{det}} := [\hat{G}_1, \hat{G}_2, \dots, \hat{G}_{k-1}]$, $\hat{G}_{\text{cur}} := \hat{G}_k$.

3. Let $\mathcal{G}_{\text{det}} := \mathcal{G}_1 \cup \mathcal{G}_2 \dots \cup \mathcal{G}_{k-1}$ and $\mathcal{G}_{\text{cur}} = \mathcal{G}_k$.

4. Let $r_{\text{cur}} := r_k = \text{rank}(G_k)$ and $\hat{r}_{\text{cur}} := \hat{r}_k$.

5. Let $\lambda_{\text{cur}}^+ := \lambda_k^+$, $\lambda_{\text{cur}}^- := \lambda_k^-$, $\lambda_{\text{undet}}^+ := \lambda_{k+1}^+$

6. Let $t_* = k\alpha$.

9.2 Main lemmas - generalized form

In this section, we give a sequence of lemmas that apply to a generic problem where $y_t = \ell_t + w_t = \ell_t + M_t \ell_t$ with ℓ_t satisfying Assumption 1.1; M_t satisfying Assumption 1.2; and with P split into three parts as $P = [G_{\text{det}}, G_{\text{cur}}, G_{\text{undet}}]$. We can correspondingly split Λ as $\Lambda = \text{diag}(\Lambda_{\text{det}}, \Lambda_{\text{cur}}, \Lambda_{\text{undet}})$.

We are given $\hat{\mathbf{G}}_{\text{det}}$ that was computed using (some or all) \mathbf{y}_t 's for $t \leq t_*$ and that satisfies $\zeta_{\text{det}} \leq r\zeta$. The goal is to estimate $\text{range}(\mathbf{G}_{\text{cur}})$ and bound the estimation error. This is done by first estimating \hat{r}_{cur} and then computing $\hat{\mathbf{G}}_{\text{cur}}$ as the top \hat{r}_{cur} eigenvectors of

$$\hat{\mathbf{D}} := \frac{1}{\alpha} \sum_{t=t_*+1}^{t_*+\alpha} \Psi \mathbf{y}_t \mathbf{y}_t' \Psi. \quad (10)$$

To bound the estimation error, we first show that, whp, $\hat{r}_{\text{cur}} = r_{\text{cur}}$ and so $\hat{\mathbf{G}}_{\text{cur}} = \mathbf{G}_{\text{cur}}$; and then we use this to show that $\zeta_{\text{cur}} \leq r_{\text{cur}}\zeta$.

Definition 9.3.

1. Define $\Psi := \mathbf{I} - \hat{\mathbf{G}}_{\text{det}} \hat{\mathbf{G}}_{\text{det}}'$.
2. Define $\zeta_{\text{det}} := \text{SE}(\hat{\mathbf{G}}_{\text{det}}, \mathbf{G}_{\text{det}}) = \|\Psi \mathbf{G}_{\text{det}}\|$ and $\zeta_{\text{det}}^+ = r\zeta$.
3. Define $\zeta_{\text{cur}} := \text{SE}([\hat{\mathbf{G}}_{\text{det}}, \hat{\mathbf{G}}_{\text{cur}}], \mathbf{G}_{\text{cur}})$.
4. Let $(\Psi \mathbf{G}_{\text{cur}}) \stackrel{\text{QR}}{=} \mathbf{E}_{\text{cur}} \mathbf{R}_{\text{cur}}$ denote its reduced QR decomposition. Thus \mathbf{E}_{cur} is a basis matrix whose span equals that of $(\Psi \mathbf{G}_{\text{cur}})$ and \mathbf{R}_{cur} is a square upper triangular matrix with $\|\mathbf{R}_{\text{cur}}\| = \|\Psi \mathbf{G}_{\text{cur}}\| \leq 1$.
5. Let $\lambda_{\text{cur}}^+ = \lambda_{\max}(\mathbf{\Lambda}_{\text{cur}})$, $\lambda_{\text{cur}}^- = \lambda_{\min}(\mathbf{\Lambda}_{\text{cur}})$, $\lambda_{\text{undet}}^+ = \lambda_{\max}(\mathbf{\Lambda}_{\text{undet}})$.
6. Let $r_{\text{cur}} = \text{rank}(\mathbf{G}_{\text{cur}})$. Clearly, $r_{\text{cur}} \leq r$.

Remark 9.4. In special cases, \mathbf{G}_{det} (and hence $\hat{\mathbf{G}}_{\text{det}}$) could be empty; and/or $\mathbf{G}_{\text{undet}}$ could be empty.

- Since $\mathbf{\Lambda}$ contains eigenvalues in decreasing order, when $\mathbf{G}_{\text{undet}}$ is not empty, $\lambda^- \leq \lambda_{\text{undet}}^+ \leq \lambda_{\text{cur}}^- \leq \lambda_{\text{cur}}^+ \leq \lambda^+$.
- When $\mathbf{G}_{\text{undet}}$ is empty, $\lambda_{\text{undet}}^+ = 0$ and $\lambda^- \leq \lambda_{\text{cur}}^- \leq \lambda_{\text{cur}}^+ \leq \lambda^+$.

Using $\|\mathbf{R}_{\text{cur}}\| = \|\Psi \mathbf{G}_{\text{cur}}\| \leq 1$,

$$\begin{aligned} \zeta_{\text{cur}} &= \|(I - \hat{\mathbf{G}}_{\text{cur}} \hat{\mathbf{G}}_{\text{cur}}') \Psi \mathbf{G}_{\text{cur}}\| \\ &= \|(I - \hat{\mathbf{G}}_{\text{cur}} \hat{\mathbf{G}}_{\text{cur}}') \mathbf{E}_{\text{cur}} \mathbf{R}_{\text{cur}}\| \\ &\leq \|(I - \hat{\mathbf{G}}_{\text{cur}} \hat{\mathbf{G}}_{\text{cur}}') \mathbf{E}_{\text{cur}}\| = \text{SE}(\hat{\mathbf{G}}_{\text{cur}}, \mathbf{E}_{\text{cur}}). \end{aligned}$$

Thus, to bound ζ_{cur} we need to bound $\text{SE}(\hat{\mathbf{G}}_{\text{cur}}, \mathbf{E}_{\text{cur}})$. $\hat{\mathbf{G}}_{\text{cur}}$ is the matrix of top \hat{r}_{cur} eigenvectors of $\hat{\mathbf{D}}$. From its definition, \mathbf{E}_{cur} is a basis matrix with r_{cur} columns. Suppose for a moment that $\hat{r}_{\text{cur}} = r_{\text{cur}}$. Then, in order to bound $\text{SE}(\hat{\mathbf{G}}_{\text{cur}}, \mathbf{E}_{\text{cur}})$, we can use the $\sin \theta$ result, Corollary 8.2. To do this, we need to define a matrix \mathbf{D} so that, under appropriate assumptions, the span of its top r_{cur} eigenvectors equals $\text{range}(\mathbf{E}_{\text{cur}})$. For the simple EVD proof, we used $\frac{1}{\alpha} \sum_{t=t_*+1}^{t_*+\alpha} \Psi \ell_t \ell_t' \Psi$ as the matrix \mathbf{D} . However, this will not work now since \mathbf{E}_{cur} is not orthonormal to $\Psi \mathbf{G}_{\text{det}}$ or to $\Psi \mathbf{G}_{\text{undet}}$. But, instead we can use

$$\begin{aligned} \mathbf{D} &= \mathbf{E}_{\text{cur}} \mathbf{A} \mathbf{E}_{\text{cur}}' + \mathbf{E}_{\text{cur},\perp} \mathbf{A}_{\perp} \mathbf{E}_{\text{cur},\perp}', \text{ where} \\ \mathbf{A} &:= \mathbf{E}_{\text{cur}}' \left(\frac{1}{\alpha} \sum_{t=t_*+1}^{t_*+\alpha} \Psi \ell_t \ell_t' \Psi \right) \mathbf{E}_{\text{cur}} \text{ and} \\ \mathbf{A}_{\perp} &:= \mathbf{E}_{\text{cur},\perp}' \left(\frac{1}{\alpha} \sum_{t=t_*+1}^{t_*+\alpha} \Psi \ell_t \ell_t' \Psi \right) \mathbf{E}_{\text{cur},\perp} \end{aligned} \quad (11)$$

Now, by construction, \mathbf{D} is in the desired form.

With the above choice of \mathbf{D} , $\mathbf{H} := \hat{\mathbf{D}} - \mathbf{D}$ satisfies¹ $\mathbf{H} = \text{term1} + \text{term1}' + \text{term2} + \text{term3} + \text{term3}'$ where $\text{term1} := \frac{1}{\alpha} \sum_t \Psi \ell_t \ell_t' \Psi$, $\text{term2} := \frac{1}{\alpha} \sum_t \Psi \mathbf{w}_t \mathbf{w}_t' \Psi$ and $\text{term3} = \mathbf{E}_{\text{cur}} \mathbf{E}_{\text{cur}}' \left(\frac{1}{\alpha} \sum_t \Psi \ell_t \ell_t' \Psi \right) \mathbf{E}_{\text{cur},\perp} \mathbf{E}_{\text{cur},\perp}'$.

¹This follows easily by writing $\mathbf{H} = (\hat{\mathbf{D}} - \frac{1}{\alpha} \sum_t \Psi \ell_t \ell_t' \Psi) + (\frac{1}{\alpha} \sum_t \Psi \ell_t \ell_t' \Psi - \mathbf{D})$ and using the fact that $\mathbf{M} = (\mathbf{E} \mathbf{E}' + \mathbf{E}_{\perp} \mathbf{E}_{\perp}') \mathbf{M} (\mathbf{E} \mathbf{E}' + \mathbf{E}_{\perp} \mathbf{E}_{\perp}')$ for $\frac{1}{\alpha} \sum_t \Psi \ell_t \ell_t' \Psi$.

Thus, using the above along with Corollary 8.2, we can conclude the following.

Fact 9.5.

1. If $\hat{r}_{\text{cur}} = r_{\text{cur}}$, and $\lambda_{\min}(\mathbf{A}) - \lambda_{\max}(\mathbf{A}_{\perp}) - \|\mathbf{H}\| > 0$,

$$\zeta_{\text{cur}} \leq \text{SE}(\hat{\mathbf{G}}_{\text{cur}}, \mathbf{E}_{\text{cur}}) \leq \frac{\|\mathbf{H}\|}{\lambda_{\min}(\mathbf{A}) - \lambda_{\max}(\mathbf{A}_{\perp}) - \|\mathbf{H}\|}.$$

2. Let $\mathbf{Q} := \mathbf{E}_{\text{cur}} \mathbf{E}_{\text{cur}}' (\frac{1}{\alpha} \sum_t \mathbf{\Psi} \ell_t \ell_t' \mathbf{\Psi}) \mathbf{E}_{\text{cur}, \perp} \mathbf{E}_{\text{cur}, \perp}'$. We have

$$\|\mathbf{H}\| \leq 2 \left\| \frac{1}{\alpha} \sum_t \mathbf{\Psi} \ell_t \mathbf{w}_t' \right\| + \left\| \frac{1}{\alpha} \sum_t \mathbf{w}_t \mathbf{w}_t' \right\| + 2 \|\mathbf{Q}\|.$$

The next lemma bounds the RHS terms in the above lemma and a few other quantities needed for showing $\hat{r}_{\text{cur}} = r_{\text{cur}}$.

Lemma 9.6. (1) Assume that $\mathbf{y}_t = \ell_t + \mathbf{w}_t = \ell_t + \mathbf{M}_t \ell_t$ with ℓ_t satisfying Assumption 1.1 and \mathbf{M}_t satisfying Assumption 1.2.

(2) Assume that we are given $\hat{\mathbf{G}}_{\text{det}}$ that was computed using (some or all) \mathbf{y}_t 's for $t \leq t_*$ and that satisfies $\zeta_{\text{det}} \leq r\zeta$.

Define $g := \lambda_{\text{cur}}^+ / \lambda_{\text{cur}}^-$, $\chi := \lambda_{\text{undet}}^+ / \lambda_{\text{cur}}^-$. Set $\epsilon := 0.01 r_{\text{cur}} \zeta \lambda_{\text{cur}}^-$.

Then, the following hold:

1. Let $p_1 := 2n \exp(-\frac{\alpha \epsilon^2}{32 b_{\text{prob}}^2})$ where $b_{\text{prob}} := \eta r q ((r\zeta) \lambda^+ + \lambda_{\text{cur}}^+ + (r\zeta) \sqrt{\lambda^+ \lambda_{\text{cur}}^+} + \sqrt{\lambda^+ \lambda_{\text{cur}}^+})$. Conditioned on $\{\zeta_{\text{det}} \leq r\zeta\}$, with probability at least $1 - p_1$

$$\begin{aligned} \left\| \frac{1}{\alpha} \sum_t \mathbf{\Psi} \ell_t \mathbf{w}_t' \right\| &\leq q((r\zeta) \lambda^+ + \lambda_{\text{cur}}^+) \sqrt{\frac{\beta}{\alpha}} + \epsilon \\ &\leq [q(r\zeta) f \sqrt{\frac{\beta}{\alpha}} + q g \sqrt{\frac{\beta}{\alpha}} + 0.01 r_{\text{cur}} \zeta] \lambda_{\text{cur}}^-. \end{aligned}$$

2. Let $p_2 := 2n \exp(-\frac{\alpha \epsilon^2}{32 (q^2 \eta r \lambda^+)^2})$. Conditioned on $\{\zeta_{\text{det}} \leq r\zeta\}$, with probability (w.p.) at least $1 - p_2$,

$$\left\| \frac{1}{\alpha} \sum_t \mathbf{w}_t \mathbf{w}_t' \right\| \leq \frac{\beta}{\alpha} q^2 \lambda^+ + \epsilon \leq \left[\frac{\beta}{\alpha} q^2 f + 0.01 r_{\text{cur}} \zeta \right] \lambda_{\text{cur}}^-.$$

3. Let $p_3 := 2n \exp(-\frac{\alpha \epsilon^2}{32 b_{\text{prob}}^2})$ with $b_{\text{prob}} := \eta r ((r\zeta)^2 \lambda^+ + \lambda_{\text{cur}}^+ + 2(r\zeta) \sqrt{\lambda^+ \lambda_{\text{cur}}^+})$. Conditioned on $\{\zeta_{\text{det}} \leq r\zeta\}$, with probability at least $1 - p_3$,

$$\begin{aligned} &\left\| \mathbf{E}_{\text{cur}} \mathbf{E}_{\text{cur}}' \left(\frac{1}{\alpha} \mathbf{\Psi} \ell_t \ell_t' \mathbf{\Psi} \right) \mathbf{E}_{\text{cur}, \perp} \mathbf{E}_{\text{cur}, \perp}' \right\| \\ &\leq (r\zeta)^2 \lambda^+ + \frac{(r\zeta)^2}{\sqrt{1 - (r\zeta)^2}} \lambda_{\text{undet}}^+ + \epsilon \\ &\leq [(r\zeta)^2 f + \frac{(r\zeta)^2}{\sqrt{1 - (r\zeta)^2}} \chi + 0.01 r_{\text{cur}} \zeta] \lambda_{\text{cur}}^-. \end{aligned}$$

4. Conditioned on $\{\zeta_{\text{det}} \leq r\zeta\}$, w.p. at least $1 - p_3$,

$$\begin{aligned} \lambda_{\min}(\mathbf{A}) &\geq (1 - (r\zeta)^2) \lambda_{\text{cur}}^- - \epsilon \\ &= [1 - (r\zeta)^2 - 0.01 r_{\text{cur}} \zeta] \lambda_{\text{cur}}^- \end{aligned}$$

5. Conditioned on $\{\zeta_{\text{det}} \leq r\zeta\}$, w.p. at least $1 - p_3$,

$$\begin{aligned} \lambda_{\max}(\mathbf{A}_{\perp}) &\leq ((r\zeta)^2 \lambda^+ + \lambda_{\text{undet}}^+) + \epsilon \\ &\leq [(r\zeta)^2 f + \chi + 0.01 r_{\text{cur}} \zeta] \lambda_{\text{cur}}^-. \end{aligned}$$

6. Conditioned on $\{\zeta_{\text{det}} \leq r\zeta\}$, with probability at least $1 - p_3$,

$$\lambda_{\max}(\mathbf{A}_{\perp}) \geq (1 - (r\zeta)^2 - \frac{(r\zeta)^2}{\sqrt{1 - (r\zeta)^2}})\lambda_{\text{undet}}^+ - \epsilon.$$

7. Conditioned on $\{\zeta_{\text{det}} \leq r\zeta\}$, w.p. at least $1 - p_3$,

$$\begin{aligned}\lambda_{\max}(\mathbf{A}) &\geq (1 - (r\zeta)^2)\lambda_{\text{cur}}^+ - \epsilon \\ &= [(1 - (r\zeta)^2)g - 0.01r_{\text{cur}}\zeta]\lambda_{\text{cur}}^-.\end{aligned}$$

8. Conditioned on $\{\zeta_{\text{det}} \leq r\zeta\}$, w.p. at least $1 - p_3$,

$$\begin{aligned}\lambda_{\max}(\mathbf{A}) &\leq \lambda_{\text{cur}}^+ + (r\zeta)^2\lambda^+ + \frac{1}{1 - r^2\zeta^2}(r\zeta)^2\lambda_{\text{undet}}^+ + \epsilon \\ &\leq [g + (r\zeta)^2f + \frac{(r\zeta)^2}{1 - (r\zeta)^2}\chi + 0.01r_{\text{cur}}\zeta]\lambda_{\text{cur}}^-.\end{aligned}$$

Proof. The proof is in Section 10. □

Corollary 9.7. Consider the setting of Lemma 9.6. Assume

1. $r(r\zeta) \leq 0.0001$, and $r(r\zeta)f \leq 0.01$. Since $r_{\text{cur}} \leq r$, this implies that $r_{\text{cur}}\zeta \leq 0.0001$, and
2. $\beta \leq \left(\frac{(1 - r_{\text{cur}}\zeta - \chi)}{2}\right)^2 \min\left(\frac{(r_{\text{cur}}\zeta)^2}{4.1q^2g^2}, \frac{(r_{\text{cur}}\zeta)}{q^2f}\right)\alpha$.

Using these and using $g \geq 1$, $g \leq f$, $\chi \leq 1$ (these hold by definition), with probability at least $1 - p_1 - p_2 - 4p_3$,

$$\begin{aligned}\|\mathbf{H}\| &\leq [2.02qg\sqrt{\frac{\beta}{\alpha}} + \frac{\beta}{\alpha}q^2f + 0.08r_{\text{cur}}\zeta]\lambda_{\text{cur}}^- \\ &\leq [0.75(1 - r\zeta - \chi)r_{\text{cur}}\zeta + 0.08r_{\text{cur}}\zeta]\lambda_{\text{cur}}^- \\ &\leq 0.83r_{\text{cur}}\zeta\lambda_{\text{cur}}^-, \\ \lambda_{\max}(\mathbf{A}_{\perp}) &\leq [\chi + 0.02r_{\text{cur}}\zeta]\lambda_{\text{cur}}^-, \\ \lambda_{\max}(\mathbf{A}_{\perp}) &\geq [\chi - 0.02r_{\text{cur}}\zeta]\lambda_{\text{cur}}^-, \\ \lambda_{\min}(\mathbf{A}) &\geq [1 - 0.0101r_{\text{cur}}\zeta]\lambda_{\text{cur}}^-, \\ \lambda_{\max}(\mathbf{A}) &\leq [g + 0.0202r_{\text{cur}}\zeta]\lambda_{\text{cur}}^-, \\ \lambda_{\max}(\mathbf{A}) &\geq [g - 0.02r_{\text{cur}}\zeta]\lambda_{\text{cur}}^-.\end{aligned}$$

Lemma 9.8. Consider the setting of Corollary 9.7. In addition, also assume that

1. $\hat{g} = 1.01g + 0.0001$ and
2. $\chi \leq \min\left(\frac{g - 0.0001}{1.01g + 0.0001} - 0.0001, 1 - r_{\text{cur}}\zeta - \frac{0.08}{0.25}\right)$.

Let $\hat{\lambda}_i := \lambda_i(\hat{\mathbf{D}})$. Then, with probability at least $1 - p_1 - p_2 - 4p_3$, the following hold.

1. When $\mathbf{G}_{\text{undet}}$ is not empty: $\frac{\hat{\lambda}_1}{\hat{\lambda}_{r_{\text{cur}}}} \leq \hat{g}$, $\frac{\hat{\lambda}_1}{\hat{\lambda}_{r_{\text{cur}}+1}} > \hat{g}$, and $\hat{\lambda}_{r_{\text{cur}}+1} \geq \lambda_{\text{thresh}}$.
2. When $\mathbf{G}_{\text{undet}}$ is empty: $\frac{\hat{\lambda}_1}{\hat{\lambda}_{r_{\text{cur}}}} \leq \hat{g}$ and $\hat{\lambda}_{r_{\text{cur}}+1} < \lambda_{\text{thresh}} < \hat{\lambda}_{r_{\text{cur}}}$.
3. If $\hat{r}_{\text{cur}} = r_{\text{cur}}$, then $\zeta_{\text{cur}} \leq \frac{\|\mathbf{H}\|}{\lambda_{\min}(\mathbf{A}) - \lambda_{\max}(\mathbf{A}_{\perp}) - \|\mathbf{H}\|} \leq 0.75r_{\text{cur}}\zeta + \frac{0.08r_{\text{cur}}\zeta}{(1 - r_{\text{cur}}\zeta - \chi)} \leq r_{\text{cur}}\zeta$.

Proof.

Fact 9.9. From the bound on χ , $\chi \leq 1 - 0.0001 \leq 1 - r_{\text{cur}}\zeta$. Thus, using Corollary 9.7, $\lambda_{\min}(\mathbf{A}) > \lambda_{\max}(\mathbf{A}_{\perp})$ and so $\lambda_{r_{\text{cur}}}(\mathbf{D}) = \lambda_{\min}(\mathbf{A})$, $\lambda_{r_{\text{cur}}+1}(\mathbf{D}) = \lambda_{\max}(\mathbf{A}_{\perp})$, and $\lambda_1(\mathbf{D}) = \lambda_{\max}(\mathbf{A})$. Recall: $\lambda_1(\cdot)$ is the same as $\lambda_{\max}(\cdot)$.

Proof of item 1. Recall that $\hat{\mathbf{D}}$ and \mathbf{D} are defined in (10) and (11). Using Weyl's inequality, Fact 9.9, and Corollary 9.7, with the probability given there,

$$\frac{\hat{\lambda}_1}{\hat{\lambda}_{r_{\text{cur}}}} \leq \frac{\lambda_{\max}(\mathbf{A}) + \|\mathbf{H}\|}{\lambda_{\min}(\mathbf{A}) - \|\mathbf{H}\|} \leq \frac{g + 0.86r_{\text{cur}}\zeta}{1 - 0.85r_{\text{cur}}\zeta}$$

and

$$\frac{\hat{\lambda}_1}{\hat{\lambda}_{r_{\text{cur}}+1}} > \frac{\lambda_{\max}(\mathbf{A}) - \|\mathbf{H}\|}{\lambda_{\max}(\mathbf{A}_{\perp}) + \|\mathbf{H}\|} > \frac{g - 0.85r_{\text{cur}}\zeta}{\chi + 0.85r_{\text{cur}}\zeta}$$

Thus, if

$$\frac{g + 0.85r_{\text{cur}}\zeta}{1 - 0.85r_{\text{cur}}\zeta} \leq \hat{g} \leq \frac{g - 0.85r_{\text{cur}}\zeta}{\chi + 0.85r_{\text{cur}}\zeta} \quad (12)$$

holds, we will be done. The above requires χ to be small enough so that the lower bound is not larger than the upper bound and it requires \hat{g} to be appropriately set. Both are ensured by the assumptions in the lemma.

Since $\mathbf{G}_{\text{undet}}$ is not empty, $\lambda_{\text{undet}}^+ = \chi\lambda_{\text{cur}}^- > \lambda^-$. Thus, using Weyl's inequality followed by Corollary 9.7, with the probability given there,

$$\begin{aligned} \hat{\lambda}_{r_{\text{cur}}+1} &\geq \lambda_{r_{\text{cur}}+1}(\mathbf{D}) - \|\mathbf{H}\| = \lambda_{\max}(\mathbf{A}_{\perp}) - \|\mathbf{H}\| \\ &\geq [\chi - 0.02r_{\text{cur}}\zeta]\lambda_{\text{cur}}^- - 0.83r_{\text{cur}}\zeta\lambda_{\text{cur}}^- \\ &\geq (1 - 0.85r_{\text{cur}}\zeta)\lambda^- > \lambda_{\text{thresh}} \end{aligned}$$

Proof of item 2. Since $\mathbf{G}_{\text{undet}}$ is empty, $\lambda_{\text{undet}}^+ = 0$ and so $\chi = 0$. Thus, using Corollary 9.7, with probability given there,

$$\begin{aligned} \hat{\lambda}_{r_{\text{cur}}+1} &\leq \lambda_{r_{\text{cur}}+1}(\mathbf{D}) + \|\mathbf{H}\| = \lambda_{\max}(\mathbf{A}_{\perp}) + \|\mathbf{H}\| \\ &\leq 0 + 0.02r_{\text{cur}}\zeta\lambda^- + \|\mathbf{H}\| \leq 0.85r_{\text{cur}}\zeta\lambda^- \\ &< \lambda_{\text{thresh}}, \end{aligned}$$

$$\begin{aligned} \hat{\lambda}_{r_{\text{cur}}} &\geq \lambda_{r_{\text{cur}}}(\mathbf{D}) - \|\mathbf{H}\| = \lambda_{\min}(\mathbf{A}) - \|\mathbf{H}\| \\ &\geq \lambda_{\text{cur}}^- - 0.085r_{\text{cur}}\zeta\lambda_{\text{cur}}^- \geq (1 - 0.85r_{\text{cur}}\zeta)\lambda^- \\ &> \lambda_{\text{thresh}}, \end{aligned}$$

and

$$\frac{\hat{\lambda}_1}{\hat{\lambda}_{r_{\text{cur}}}} \leq \frac{\lambda_{\max}(\mathbf{A}) + \|\mathbf{H}\|}{\lambda_{\min}(\mathbf{A}) - \|\mathbf{H}\|} \leq \frac{g + 0.85r_{\text{cur}}\zeta}{1 - 0.85r_{\text{cur}}\zeta} \leq \hat{g}$$

Proof of item 3. Using Fact 9.5 and Corollary 9.7, since $\hat{r}_{\text{cur}} = r_{\text{cur}}$ is assumed, we get

$$\begin{aligned} \zeta_{\text{cur}} &\leq \frac{[0.75(1 - r_{\text{cur}}\zeta - \chi)r_{\text{cur}}\zeta + 0.08r_{\text{cur}}\zeta]\lambda_{\text{cur}}^-}{\lambda_{\text{cur}}^-[1 - 0.0101r_{\text{cur}}\zeta - \chi - 0.02r_{\text{cur}}\zeta - 0.83r\zeta]} \\ &\leq \frac{0.75(1 - r\zeta - \chi)r_{\text{cur}}\zeta + 0.08r_{\text{cur}}\zeta}{(1 - r_{\text{cur}}\zeta - \chi)} \leq r_{\text{cur}}\zeta \end{aligned} \quad (13)$$

The last inequality used the bound on χ . \boxtimes

9.3 Proof of Theorem 3.3

The theorem is a direct consequence of using (9) and applying Lemma 9.8 for each of the k steps with the substitutions given in Definition 9.2; along with picking α appropriately. A detailed proof is in Sec. 11.

10 Proof of Hoeffding lemma, Lemma 9.6

The following lemma, which is a modification of [3, Lemma 8.15], will be used in our proof. It is proved in Sec. 11. The proof uses [3, Lemma 2.10].

Lemma 10.1. *Given $\zeta_{\text{det}} \leq r\zeta$.*

1. $\|\Psi \mathbf{G}_{\text{det}}\| \leq r\zeta$ and $\|\Psi \mathbf{G}_{\text{cur}}\| \leq 1$.
2. $\sqrt{1 - (r\zeta)^2} \leq \sigma_i(\mathbf{R}_{\text{cur}}) = \sigma_i(\Psi \mathbf{G}_{\text{cur}}) \leq 1$ and $\sqrt{1 - (r\zeta)^2} \leq \sigma_i(\Psi \mathbf{G}_{\text{undet}}) \leq 1$
3. $\|\mathbf{E}_{\text{cur}}' \Psi \mathbf{G}_{\text{undet}}\| \leq \frac{(r\zeta)^2}{\sqrt{1 - (r\zeta)^2}}$
- 4.

$$\Psi \Sigma \Psi = [\Psi \mathbf{G}_{\text{det}} \quad \Psi \mathbf{G}_{\text{cur}} \quad \Psi \mathbf{G}_{\text{undet}}] \begin{bmatrix} \Lambda_{\text{det}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Lambda_{\text{cur}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Lambda_{\text{undet}} \end{bmatrix} \begin{bmatrix} \Psi \mathbf{G}_{\text{det}} \\ \Psi \mathbf{G}_{\text{cur}} \\ \Psi \mathbf{G}_{\text{undet}} \end{bmatrix}'$$

with $\lambda_{\max}(\Lambda_{\text{det}}) \leq \lambda^+$, $\lambda_{\text{cur}}^- \leq \lambda_{\min}(\Lambda_{\text{cur}}) \leq \lambda_{\max}(\Lambda_{\text{cur}}) \leq \lambda_{\text{cur}}^+$, $\lambda_{\max}(\Lambda_{\text{undet}}) \leq \lambda_{\text{undet}}^+$.

5. *Using the first four claims, it is easy to see that*

- (a) $\|\mathbf{E}_{\text{cur},\perp}' \Psi \Sigma \Psi \mathbf{E}_{\text{cur},\perp}\| \leq (r\zeta)^2 \lambda^+ + \lambda_{\text{undet}}^+$
- (b) $\|\mathbf{E}_{\text{cur},\perp}' \Psi \Sigma \Psi \mathbf{E}_{\text{cur}}\| \leq (r\zeta)^2 \lambda^+ + \frac{(r\zeta)^2}{\sqrt{1 - (r\zeta)^2}} \lambda_{\text{undet}}^+$
- (c) $\|\Psi \Sigma\| \leq (r\zeta) \lambda^+ + \lambda_{\text{cur}}^+$ and $\|\Psi \Sigma \mathbf{M}_{1,t}'\| \leq q((r\zeta) \lambda^+ + \lambda_{\text{cur}}^+)$
- (d) $\|\mathbf{M}_{1,t} \Sigma\| \leq q \lambda^+$ and $\|\mathbf{M}_{1,t} \Sigma \mathbf{M}_{1,t}'\| \leq q^2 \lambda^+$

If $\hat{\mathbf{G}}_{\text{det}} = \mathbf{G}_{\text{det}} = [\cdot]$, then all the terms containing $(r\zeta)$ disappear.

6. $\lambda_{\min}(A + B) \geq \lambda_{\min}(A) + \lambda_{\min}(B)$
7. Let $\mathbf{a}_t := \mathbf{P}' \ell_t$, $\mathbf{a}_{t,\text{det}} := \mathbf{G}_{\text{det}}' \ell_t$, $\mathbf{a}_{t,\text{cur}} := \mathbf{G}_{\text{cur}}' \ell_t$ and $\mathbf{a}_{t,\text{undet}} := \mathbf{G}_{\text{undet}}' \ell_t$. Also let $\mathbf{a}_{t,\text{rest}} := [\mathbf{a}_{t,\text{cur}}', \mathbf{a}_{t,\text{undet}}']'$. Then $\|\mathbf{a}_{t,\text{rest}}\|^2 \leq r\eta \lambda_{\text{cur}}^+$ and $\|\mathbf{a}_{t,\text{det}}\|^2 \leq \|\mathbf{a}_t\|^2 \leq r\eta \lambda^+$.
8. $\sigma_{\min}(\mathbf{E}_{\text{cur},\perp}' \Psi \mathbf{G}_{\text{undet}})^2 \geq 1 - (r\zeta)^2 - \frac{(r\zeta)^2}{\sqrt{1 - (r\zeta)^2}}$.

The following corollaries of the matrix Hoeffding inequality [20], proved in [3], will be used in the proof.

Corollary 10.2. *Given an α -length sequence $\{\mathbf{Z}_t\}$ of random Hermitian matrices of size $n \times n$, a r.v. X , and a set \mathcal{C} of values that X can take. For all $X \in \mathcal{C}$, (i) \mathbf{Z}_t 's are conditionally independent given X ; (ii) $\mathbb{P}(b_1 \mathbf{I} \preceq \mathbf{Z}_t \preceq b_2 \mathbf{I} | X) = 1$ and (iii) $b_3 \mathbf{I} \preceq \mathbb{E}[\frac{1}{\alpha} \sum_t \mathbf{Z}_t | X] \preceq b_4 \mathbf{I}$. For any $\epsilon > 0$, for all $X \in \mathcal{C}$,*

$$\mathbb{P}\left(\lambda_{\max}\left(\frac{1}{\alpha} \sum_t \mathbf{Z}_t\right) \leq b_4 + \epsilon \mid X\right) \geq 1 - n \exp\left(\frac{-\alpha \epsilon^2}{8(b_2 - b_1)^2}\right),$$

$$\mathbb{P}\left(\lambda_{\min}\left(\frac{1}{\alpha} \sum_t \mathbf{Z}_t\right) \geq b_3 - \epsilon \mid X\right) \geq 1 - n \exp\left(\frac{-\alpha \epsilon^2}{8(b_2 - b_1)^2}\right).$$

Corollary 10.3. *Given an α -length sequence $\{\mathbf{Z}_t\}$ of random matrices of size $n_1 \times n_2$. For all $X \in \mathcal{C}$, (i) \mathbf{Z}_t 's are conditionally independent given X ; (ii) $\mathbb{P}(\|\mathbf{Z}_t\| \leq b_1 | X) = 1$ and (iii) $\|\mathbb{E}[\frac{1}{\alpha} \sum_t \mathbf{Z}_t | X]\| \leq b_2$. For any $\epsilon > 0$, for all $X \in \mathcal{C}$,*

$$\mathbb{P}\left(\left\|\frac{1}{\alpha} \sum_t \mathbf{Z}_t\right\| \leq b_2 + \epsilon \mid X\right) \geq 1 - (n_1 + n_2) \exp\left(\frac{-\alpha \epsilon^2}{32b_1^2}\right).$$

Proof of Lemma 9.6. Recall that we are given \hat{G}_{det} that was computed using (some or all) \mathbf{y}_t 's for $t \leq t_*$ and that satisfies $\zeta_{\text{det}} \leq r\zeta$. From (2), \mathbf{y}_t is a linear function of ℓ_t . Thus, we can let $X := \{\ell_1, \ell_2, \dots, \ell_{t_*}\}$ denote all the random variables on which the event $\{\zeta_{\text{det}} \leq r\zeta\}$ depends. In each item of this proof, we need to lower bound the probability of the desired event conditioned on $\zeta_{\text{det}} \leq r\zeta$. To do this, we first lower bound the probability of the event conditioned on X that is such that $X \in \{\zeta_{\text{det}} \leq r\zeta\}$. We get a lower bound that does not depend on X as long as $X \in \{\zeta_{\text{det}} \leq r\zeta\}$. Thus, the same probability lower bound holds conditioned on $\{\zeta_{\text{det}} \leq r\zeta\}$.

Fact 10.4. *For an event \mathcal{E} and random variable X , $\mathbb{P}(\mathcal{E}|X) \geq p$ for all $X \in \mathcal{C}$ implies that $\mathbb{P}(\mathcal{E}|X \in \mathcal{C}) \geq p$.*

Proof of Lemma 9.6, item 1. Let

$$\text{term} := \frac{1}{\alpha} \sum_t \Psi \ell_t \mathbf{w}_t' = \frac{1}{\alpha} \sum_t \Psi \ell_t \ell_t' \mathbf{M}_{1,t}' \mathbf{M}_{2,t}'$$

Since Ψ is a function of X , since ℓ_t 's used in the summation above are independent of X and $\mathbb{E}[\ell_t \ell_t'] = \Sigma$,

$$\mathbb{E}[\text{term}|X] = \frac{1}{\alpha} \sum_t \Psi \Sigma \mathbf{M}_{1,t}' \mathbf{M}_{2,t}'$$

Next, we use Cauchy-Schwartz for matrices:

$$\left\| \sum_{t=1}^{\alpha} \mathbf{X}_t \mathbf{Y}_t' \right\|^2 \leq \lambda_{\max} \left(\sum_{t=1}^{\alpha} \mathbf{X}_t \mathbf{X}_t' \right) \lambda_{\max} \left(\sum_{t=1}^{\alpha} \mathbf{Y}_t \mathbf{Y}_t' \right) \quad (14)$$

Using (14), with $\mathbf{X}_t = \Psi \Sigma \mathbf{M}_{1,t}'$ and $\mathbf{Y}_t = \mathbf{M}_{2,t}$, followed by using $\sqrt{\|\frac{1}{\alpha} \sum_t \mathbf{X}_t \mathbf{X}_t'\|} \leq \max_t \|\mathbf{X}_t\|$, Assumption 1.2 with $\mathbf{A}_t \equiv \mathbf{I}$, and Lemma 10.1,

$$\begin{aligned} \|\mathbb{E}[\text{term}|X]\| &\leq \max_t \|\Psi \Sigma \mathbf{M}_{1,t}'\| \sqrt{\frac{\beta}{\alpha}} \\ &\leq q((r\zeta)\lambda^+ + \lambda_{\text{cur}}^+) \sqrt{\frac{\beta}{\alpha}} \end{aligned}$$

for all $X \in \{\zeta_{\text{det}} \leq r\zeta\}$. To bound $\|\Psi \ell_t \mathbf{w}_t'\|$, rewrite it as $\Psi \ell_t \mathbf{w}_t' = [\Psi \mathbf{G}_{\text{det}} a_{t,\text{det}} + \Psi \mathbf{G}_{\text{rest}} a_{t,\text{rest}}][a_{t,\text{det}}' \mathbf{G}_{\text{det}}' + a_{t,\text{rest}}' \mathbf{G}_{\text{rest}}'] \mathbf{M}_{1,t}' \mathbf{M}_{2,t}'$. Thus, using $\|\mathbf{M}_{2,t}\| \leq 1$, $\|\mathbf{M}_{1,t} \mathbf{P}\| \leq q < 1$, and Lemma 10.1,

$$\|\Psi \ell_t \mathbf{w}_t'\| \leq q r \eta ((r\zeta)\lambda^+ + \lambda_{\text{cur}}^+ + (r\zeta)\sqrt{\lambda^+ \lambda_{\text{cur}}^+} + \sqrt{\lambda^+ \lambda_{\text{cur}}^+})$$

holds w.p. one when $\{\zeta_{\text{det}} \leq r\zeta\}$.

Finally, conditioned on X , the individual summands in term are conditionally independent. Using matrix Hoeffding, Corollary 10.3, followed by Fact 10.4, the result follows.

Proof of Lemma 9.6, item 2.

$$\mathbb{E}[\frac{1}{\alpha} \sum_t \mathbf{w}_t \mathbf{w}_t' | X] = \frac{1}{\alpha} \sum_t \mathbf{M}_{2,t} \mathbf{M}_{1,t} \Sigma \mathbf{M}_{1,t}' \mathbf{M}_{2,t}'$$

By Lemma 10.1, $\|\mathbf{M}_{1,t} \Sigma \mathbf{M}_{1,t}'\| \leq q^2 \lambda^+$. Thus, using Assumption 1.2 with $\mathbf{A}_t \equiv \mathbf{M}_{1,t} \Sigma \mathbf{M}_{1,t}'$,

$$\|\mathbb{E}[\frac{1}{\alpha} \sum_t \mathbf{w}_t \mathbf{w}_t' | X]\| \leq \frac{\beta}{\alpha} q^2 \lambda^+.$$

Using Assumption 1.2 and Lemma 10.1,

$$\|\mathbf{w}_t \mathbf{w}_t'\| = \|\mathbf{M}_{2,t} \mathbf{M}_{1,t} \mathbf{P} \mathbf{a}_t\|^2 \leq q^2 \eta r \lambda^+.$$

Conditional independence of the summands holds as before. Thus, using Corollary 10.3 and Fact 10.4, the result follows.

Proof of Lemma 9.6, item 3.

$$\begin{aligned} \mathbb{E}\left[\frac{1}{\alpha} \sum_t \mathbf{E}_{\text{cur}} \mathbf{E}_{\text{cur}}' \Psi \ell_t \ell_t' \Psi \mathbf{E}_{\text{cur},\perp} \mathbf{E}_{\text{cur},\perp}' \middle| X\right] \\ = \mathbf{E}_{\text{cur}} \mathbf{E}_{\text{cur}}' \Psi \Sigma \Psi \mathbf{E}_{\text{cur},\perp} \mathbf{E}_{\text{cur},\perp}' \end{aligned}$$

Using Lemma 10.1, $\|\mathbf{E}_{\text{cur}} \mathbf{E}_{\text{cur}}' \Psi \Sigma \Psi \mathbf{E}_{\text{cur},\perp} \mathbf{E}_{\text{cur},\perp}'\| \leq (r\zeta)^2 \lambda^+ + \frac{(r\zeta)^2}{\sqrt{1-(r\zeta)^2}} \lambda_{\text{undet}}^+$ when $\{\zeta_{\text{det}} \leq r\zeta\}$. Also, $\|\mathbf{E}_{\text{cur}}' \Psi \ell_t \ell_t' \Psi \mathbf{E}_{\text{cur},\perp}\| \leq \|\Psi \ell_t \ell_t' \Psi\| \leq \eta r((r\zeta)^2 \lambda^+ + \lambda_{\text{cur}}^+ + 2(r\zeta)\sqrt{\lambda^+ \lambda_{\text{cur}}^+}) := b_{\text{prob}}$ holds w.p. one when $\{\zeta_{\text{det}} \leq r\zeta\}$. In the above bound, the first inequality is used to get a loose bound, but one that will also apply for the proofs of the later items given below. The rest is the same as in the proofs of the earlier parts.

Proof of Lemma 9.6, item 4. Using Ostrowski's theorem,

$$\begin{aligned} \lambda_{\min}(\mathbb{E}[\mathbf{A}|X]) &= \lambda_{\min}(\mathbf{E}_{\text{cur}}' \Psi(\Sigma) \Psi \mathbf{E}_{\text{cur}}) \\ &\geq \lambda_{\min}(\mathbf{E}_{\text{cur}}' \Psi \mathbf{G}_{\text{cur}} \Lambda_{\text{cur}} \mathbf{G}_{\text{cur}}' \Psi \mathbf{E}_{\text{cur}}) \\ &= \lambda_{\min}(\mathbf{R}_{\text{cur}} \Lambda_{\text{cur}} \mathbf{R}_{\text{cur}}') \\ &\geq \lambda_{\min}(\mathbf{R}_{\text{cur}} \mathbf{R}_{\text{cur}}') \lambda_{\min}(\Lambda_{\text{cur}}) \geq (1 - (r\zeta)^2) \lambda_{\text{cur}}^- \end{aligned}$$

for all $X \in \{\zeta_{\text{det}} \leq r\zeta\}$. Ostrowski's theorem is used to get the second-last inequality, while Lemma 10.1 helps get the last one.

As in the proof of item 3, $\|\mathbf{E}_{\text{cur}}' \Psi \ell_t \ell_t' \Psi \mathbf{E}_{\text{cur}}\| \leq \|\Psi \ell_t \ell_t' \Psi\| \leq b_{\text{prob}}$ holds w.p. one when $\{\zeta_{\text{det}} \leq r\zeta\}$. Conditional independence of the summands holds as before. Thus, by matrix Hoeffding, Corollary 10.2, the result follows.

Proof of Lemma 9.6, item 5. By Lemma 10.1,

$$\begin{aligned} \lambda_{\max}(\mathbb{E}[\mathbf{A}_{\perp}|X]) &= \lambda_{\max}(\mathbf{E}_{\text{cur},\perp}' \Psi \Sigma \Psi \mathbf{E}_{\text{cur},\perp}) \\ &\leq ((r\zeta)^2 \lambda^+ + \lambda_{\text{undet}}^+) \end{aligned}$$

when $\{\zeta_{\text{det}} \leq r\zeta\}$. The rest of the proof is the same as that of the previous part.

Proof of Lemma 9.6, item 6. Using Ostrowski's theorem, $\lambda_{\max}(\mathbb{E}[\mathbf{A}_{\perp}|X]) \geq \lambda_{\max}(\mathbf{E}_{\text{cur},\perp}' \Psi \mathbf{G}_{\text{undet}} \Lambda_{\text{undet}} \mathbf{G}_{\text{undet}}' \Psi \mathbf{E}_{\text{cur},\perp}) \geq \lambda_{\min}(\mathbf{E}_{\text{cur},\perp}' \Psi \mathbf{G}_{\text{undet}} \mathbf{G}_{\text{undet}}' \Psi \mathbf{E}_{\text{cur},\perp}) \lambda_{\max}(\Lambda_{\text{undet}})$. By definition, $\lambda_{\max}(\Lambda_{\text{undet}}) = \lambda_{\text{undet}}^+$. By Lemma 10.1, $\lambda_{\min}(\mathbf{E}_{\text{cur},\perp}' \Psi \mathbf{G}_{\text{undet}} \mathbf{G}_{\text{undet}}' \Psi \mathbf{E}_{\text{cur},\perp}) = \sigma_{\min}(\mathbf{E}_{\text{cur},\perp}' \Psi \mathbf{G}_{\text{undet}})^2 \geq (1 - (r\zeta)^2 - \frac{(r\zeta)^2}{\sqrt{1-(r\zeta)^2}})$ when $\{\zeta_{\text{det}} \leq r\zeta\}$. The rest of the proof is the same as above.

Proof of Lemma 9.6, item 7. Using Ostrowski's theorem and Lemma 10.1, $\lambda_{\max}(\mathbb{E}[\mathbf{A}|X]) \geq \lambda_{\max}(\mathbf{E}_{\text{cur}}' \Psi \mathbf{G}_{\text{cur}} \Lambda_{\text{cur}} \mathbf{G}_{\text{cur}}' \Psi \mathbf{E}_{\text{cur}}) \geq \lambda_{\min}(\mathbf{R}_{\text{cur}} \mathbf{R}_{\text{cur}}') \lambda_{\max}(\Lambda_{\text{cur}}) \geq (1 - (r\zeta)^2) \lambda_{\text{cur}}^+$ when $\{\zeta_{\text{det}} \leq r\zeta\}$. The rest of the proof is the same as above. \square

11 Detailed Proof of Theorem 3.3 and Proof of Lemma 10.1

Proof of Theorem 3.3. Recall that we need to show that $\zeta_k \leq r_k \zeta$. Assume the substitutions given in Definition 9.2. We will use induction.

Consider a $k < \vartheta$. For the k -th step, assume that $\zeta_i \leq r_i \zeta$ for $i = 1, 2, \dots, k-1$. Thus, using (9), $\zeta_{\text{det}} \leq r\zeta$ and so Lemma 9.8 is applicable. We first show that $\hat{r}_k = r_k$ and that Algorithm 1 does not stop (proceeds to $(k+1)$ -th step). From Algorithm 1, $\hat{r}_k = r_k$ if $\frac{\hat{\lambda}_1}{\hat{\lambda}_{r_k}} \leq \hat{g}$, and $\frac{\hat{\lambda}_1}{\hat{\lambda}_{r_k+1}} > \hat{g}$. Also it will not stop if $\hat{\lambda}_{r_k+1} \geq \lambda_{\text{thresh}}$. Since $k < \vartheta$, $\mathbf{G}_{\text{undet}}$ is not empty. Thus, item 1 of Lemma 9.8 shows that all these hold. Hence $\hat{r}_k = r_k$ and algorithm does not stop w.p. at least $1 - p_1 - p_2 - 4p_3$. Thus, by item 3 of the same lemma, with the same probability, $\zeta_k \leq r_k \zeta$.

Now consider $k = \vartheta$. We first show $\hat{r}_k = r_k$ and that Algorithm 1 does stop, i.e., $\hat{\vartheta} = \vartheta$. This will be true if $\frac{\hat{\lambda}_1}{\hat{\lambda}_{r_k}} \leq \hat{g}$ and $\hat{\lambda}_{r_k+1} < \lambda_{\text{thresh}}$. For $k = \vartheta$, $\mathbf{G}_{\text{undet}}$ is empty. Thus, item 2 of Lemma 9.8 shows that this holds w.p. at least $1 - p_1 - p_2 - 4p_3$. Thus, by item 3 of the same lemma, with the same probability, $\zeta_k \leq r_k \zeta$.

Thus, using the union bound, w.p. at least $1 - \vartheta(p_1 + p_2 + 4p_3)$, $\hat{r}_k = r_k$ and $\zeta_k \leq r_k \zeta$ for all k . Using (8), this implies that $\text{SE} \leq r\zeta$ with the same probability.

Finally, the choice $\alpha \geq \alpha_0$, implies that $p_1 \leq \frac{1}{\vartheta} 2n^{-10}$, $p_2 \leq \frac{1}{\vartheta} 2n^{-10}$, $p_3 \leq \frac{1}{\vartheta} 2n^{-10}$. Hence $\text{SE} \leq r\zeta$ w.p. at least $1 - 12n^{-10}$. We work this out for p_1 below. The others follow similarly.

Recall that $p_1 = 2n \exp(-\alpha \frac{\epsilon^2}{32b_{prob}^2})$, $\epsilon = 0.01(r\zeta)\lambda^-$ and $b_{prob} = \eta r q((r\zeta)\lambda^+ + \lambda_{\text{cur}}^+ + (r\zeta)\sqrt{\lambda^+ \lambda_{\text{cur}}^+} + \sqrt{\lambda^+ \lambda_{\text{cur}}^+})$. Thus, $\frac{b_{prob}^2}{(\lambda^-)^2} \leq (4\eta r \max(q(r\zeta)f, qg, q\sqrt{fg}, q(r\zeta)\sqrt{fg}))^2 \leq 16\eta^2 r^2 \max(q(r\zeta)f, qg, q\sqrt{fg})^2$

Thus, $\alpha \frac{\epsilon^2}{32b_{prob}^2} \geq \frac{32 \cdot 16}{(0.01)^2} \frac{\eta^2 r^2 (11 \log n + \log \vartheta)}{(r\zeta)^2} \max(q(r\zeta)f, qg, q\sqrt{fg}) \frac{(0.01(r\zeta))^2}{32 \cdot 16 \eta^2 r^2 \max(q(r\zeta)f, qg, q\sqrt{fg})^2} \geq 11 \log n + \log \vartheta$. Thus, $p_1 \leq \frac{1}{\vartheta} 2n^{-10}$. \square

Proof of Lemma 10.1. The first claim is obvious. The next two claims follow using the following lemma:

Lemma 11.1 ([3], Lemma 2.10). *Suppose that P , \hat{P} and Q are three basis matrices. Also, P and \hat{P} are of the same size, $Q'P = 0$ and $\|(I - \hat{P}\hat{P}')P\| = \zeta_*$. Then,*

1. $\|(I - \hat{P}\hat{P}')PP'\| = \|(I - PP')\hat{P}\hat{P}'\| = \|(I - PP')\hat{P}\| = \|(I - \hat{P}\hat{P}')P\| = \zeta_*$
2. $\|PP' - \hat{P}\hat{P}'\| \leq 2\|(I - \hat{P}\hat{P}')P\| = 2\zeta_*$
3. $\|\hat{P}'Q\| \leq \zeta_*$
4. $\sqrt{1 - \zeta_*^2} \leq \sigma_i((I - \hat{P}\hat{P}')Q) \leq 1$

Use item 4 of Lemma 11.1 and the fact that $G_{\text{det}}'G_{\text{cur}} = 0$ and $G_{\text{det}}'G_{\text{undet}} = 0$ to get the second claim.

For the third claim, notice that $E_{\text{cur}}'\Psi G_{\text{undet}} = R_{\text{cur}}^{-1}G_{\text{cur}}'\Psi G_{\text{undet}} = R_{\text{cur}}^{-1}G_{\text{cur}}'\hat{G}_{\text{det}}\hat{G}_{\text{det}}'G_{\text{undet}}$. since $\Psi^2 = \Psi$ and $G_{\text{cur}}'G_{\text{undet}} = 0$. Using the second claim, $\|R_{\text{cur}}^{-1}\| \leq \frac{1}{\sigma_{\min}(R_{\text{cur}})} \leq \frac{1}{1 - (r\zeta)^2}$. Use item 3 of Lemma 11.1 and the facts that $G_{\text{cur}}'G_{\text{det}} = 0$ and $G_{\text{undet}}'G_{\text{det}} = 0$ to bound $\|G_{\text{cur}}'\hat{G}_{\text{det}}\|$ and $\|\hat{G}_{\text{det}}'G_{\text{undet}}\|$ respectively.

The fourth claim just uses the definitions. The fifth claim uses the previous claims and the assumptions on M_t from Assumption 1.2. The sixth claim follows using Weyl's inequality.

The second last claim: We show how to bound $\mathbf{a}_{t,\text{rest}}$: $\|\mathbf{a}_{t,\text{rest}}\|^2 = \|\mathbf{a}_{t,\text{cur}}\|^2 + \|\mathbf{a}_{t,\text{undet}}\|^2 \leq \sum_{j \in \mathcal{G}_{\text{cur}}} \eta \lambda_j + \sum_{j \in \mathcal{G}_{\text{undet}}} \eta \lambda_j \leq r\eta \lambda_{\text{cur}}^+$ (since $\lambda_j \leq \lambda_{\text{cur}}^+$ for all the j 's being summed over). The other bounds follow similarly.

Last claim:

$$\begin{aligned}
& \sigma_{\min}(E_{\text{cur},\perp}'\Psi G_{\text{undet}})^2 \\
&= \lambda_{\min}(G_{\text{undet}}'\Psi E_{\text{cur},\perp} E_{\text{cur},\perp}'\Psi G_{\text{undet}}) \\
&= \lambda_{\min}(G_{\text{undet}}'\Psi(I - E_{\text{cur}}E_{\text{cur}}')\Psi G_{\text{undet}}) \\
&\geq \lambda_{\min}(G_{\text{undet}}'\Psi\Psi G_{\text{undet}}) - \\
&\quad \lambda_{\max}(G_{\text{undet}}'\Psi E_{\text{cur}}E_{\text{cur}}'\Psi G_{\text{undet}}) \\
&= \sigma_{\min}(\Psi G_{\text{undet}})^2 - \|E_{\text{cur}}'\Psi G_{\text{undet}}\|^2 \\
&\geq 1 - (r\zeta)^2 - \frac{(r\zeta)^2}{\sqrt{1 - (r\zeta)^2}}.
\end{aligned}$$

The last inequality follows using the second and the third claim. \square