Supplement to “Optimal Rates of Statistical Seriation”

Appendix A: Proof of Proposition 3.7

Recall that $V(A) = (\frac{1}{m} \sum_{j=1}^{m} V_j(A)^{2/3})^{3/2}$. Since the $\ell_2$-norm of a vector is no larger than the $\ell_2^{3/2}$-norm,

$$\sum_{j=1}^{m} V_j(A)^2 \leq \left( \sum_{j=1}^{m} V_j(A)^{2/3} \right)^3 = m^3 V(A)^2.$$

On the other hand,

$$\hat{A}_{i,j} = \frac{1}{n} \sum_{k=1}^{n} A^*_{k,j} + \frac{1}{n} \sum_{k=1}^{n} Z_{k,j},$$

so we have that

$$\|\hat{\Pi} \hat{A} - \Pi^* A^*\|_F^2 = \sum_{i \in [n], j \in [m]} \left( \frac{1}{n} \sum_{k=1}^{n} A^*_{k,j} + \frac{1}{n} \sum_{k=1}^{n} Z_{k,j} - A^*_{i,j} \right)^2 \leq 2 \sum_{i \in [n], j \in [m]} \left( \frac{1}{n} \sum_{k=1}^{n} A^*_{k,j} - A^*_{i,j} \right)^2 + \frac{2}{n^2} \sum_{i \in [n], j \in [m]} \left( \sum_{k=1}^{n} Z_{k,j} \right)^2 \leq 2n \sum_{j \in [m]} V_j(A)^2 + \frac{2}{n^2} \sum_{j \in [m]} \left( \sum_{k=1}^{n} Z_{k,j} \right)^2 \leq 2nm^3 V(A)^2 + 2 \sum_{j \in [m]} g_j^2,$$

where $g_j = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_{k,j}$ for $j \in [m]$ so that $g_1, \ldots, g_m$ are centered sub-Gaussian variables with variance proxy $\sigma^2$. It is well-known that $\mathbb{E} g_j^2 \lesssim \sigma^2$, so

$$\mathbb{E} \|\hat{\Pi} \hat{A} - \Pi^* A^*\|_F^2 \lesssim nm^3 V(A)^2 + m \sigma^2.$$

Moreover, since $(g_1, \ldots, g_m)$ is a sub-Gaussian vector with variance proxy $\sigma^2$, it follows from [4, Theorem 2.1] that $\sum_{j=1}^{m} g_j^2 \lesssim \sigma^2 m$ with probability at least $1 - \exp(-m)$. On this event,

$$\|\hat{\Pi} \hat{A} - \Pi^* A^*\|_F^2 \lesssim nm^3 V(A)^2 + m \sigma^2.$$

Dividing the previous two displays by $nm$ completes the proof.
Appendix B: Proof of Lemma 6.4

Lemma 2 of [1] and its proof extend to the unimodal case with minor modifications. We provide the proof here for completeness.

**Lemma B.1.** For $a \in \mathcal{U}$ and $k \in [n]$, there exists $\tilde{a} \in \mathcal{U}_k$ such that

$$\frac{1}{\sqrt{n}} \| \tilde{a} - a \|_2 \leq \frac{V(a)}{2k}.$$  \hspace{1cm} (B.1)

In particular, there exists $\tilde{a} \in \mathcal{U}_{k^*}$ such that

$$\frac{1}{n} \| \tilde{a} - a \|_2^2 \leq \frac{1}{4} \max \left( \left( \frac{\sigma^2 V(a) \log(en)}{n} \right)^{2/3}, \frac{\sigma^2 \log(en)}{n} \right).$$

Moreover,

$$\frac{\sigma^2 k^*}{n} \log(en) \leq 2 \max \left( \left( \frac{\sigma^2 V(a) \log(en)}{n} \right)^{2/3}, \frac{\sigma^2 \log(en)}{n} \right).$$

**Proof.** Let $a = \min(a_1, a_n)$, $\bar{a} = \max_{i \in [n]} a_i$ and $i_0 \in \arg\max_{i \in [n]} a_i$. For $j \in [k-1]$, consider the intervals

$$I_j = \left[ a + j \frac{1}{k} V(a), a + \frac{j}{k} V(a) \right].$$

and $I_k = \left[ a + \frac{k-1}{k} V(a), \bar{a} \right]$. Also for $j \in [k]$, let $J_j = \{ i \in [n] : a_i \in I_j \}$. We define the vector $\tilde{a} \in \mathbb{R}^n$ by $\tilde{a}_i = a + \frac{i-1/2}{k} V(a)$ for $i \in [n]$, where $j$ is uniquely determined by $i \in I_j$. Since $a$ is increasing on $\{1, \ldots, i_0\}$ and decreasing $\{i_0, \ldots, n\}$, so is $\tilde{a}$. Thus $\tilde{a} \in \mathcal{U}_k$. Moreover, $|\tilde{a}_i - a_i| \leq \frac{V(a)}{2k}$ for $i \in [n]$, which implies (B.1).

Next we prove the latter two assertions. Since $k^* = \lceil \left( \frac{V(a) V_n}{\sigma^2 \log(en)} \right)^{1/3} \rceil$, if $\tilde{a} \in \mathcal{U}_{k^*}$ and $k^* = 1$ then

$$\frac{1}{n} \| \tilde{a} - a \|_2^2 \leq \frac{V(a)^2}{4} \leq \frac{\sigma^2}{4n} \log(en)$$

and

$$\frac{\sigma^2 k^*}{n} \log(en) = \frac{\sigma^2}{n} \log(en).$$

On the other hand, if $k^* > 1$, then

$$\frac{1}{n} \| \tilde{a} - a \|_2^2 \leq \frac{V(a)}{4(k^*)^2} \leq \frac{1}{4} \left( \frac{\sigma^2 V(a) \log(en)}{n} \right)^{2/3}$$

and

$$\frac{\sigma^2 k^*}{n} \log(en) \leq 2 \left( \frac{\sigma^2 V(a) \log(en)}{n} \right)^{2/3}.$$
Lemma 6.4 is then an easy generalization of the previous lemma to the matrix case. Applying Lemma B.1 to columns of $A$, we see that there exists $\tilde{A} \in U_{k^*}^{m}$ such that
\[
\frac{1}{n} \| \tilde{A} \cdot j - A \cdot j \|_2^2 \leq \frac{1}{4} \max \left( \frac{\sigma^2 V(A \cdot j) \log(en)}{n}, \frac{\sigma^2}{n} \log(en) \right)
\]
and
\[
\frac{\sigma^2 k^*_j}{n} \log(en) \leq 2 \max \left( \frac{\sigma^2 V(A \cdot j) \log(en)}{n}, \frac{\sigma^2}{n} \log(en) \right).
\]
Summing over $1 \leq j \leq m$, we get that
\[
\frac{1}{nm} \| \tilde{A} - A \|_F^2 \leq \frac{1}{4m} \left( \frac{\sigma^2 \log(en)}{n} \right)^{2/3} \sum_{j=1}^{m} V(A \cdot j) + \frac{\sigma^2 \log(en)}{4n}
\]
and similarly
\[
\frac{\sigma^2 K(\tilde{A})}{nm} \log(en) \leq 2 \left( \frac{\sigma^2 V(A) \log(en)}{n} \right)^{2/3} + \frac{2\sigma^2}{n} \log(en).
\]

Appendix C: Proofs of lemmas in Section 6.2

We start with a result on the metric entropy of a ball in one norm in $\mathbb{R}^m$ with respect to another norm. This result is well-known for certain pairs of norms (e.g. [5, Lemma 7.14]), and we use the general version from [7, Lemma 5.2].

**Lemma C.1.** Let $\| \cdot \|$ and $\| \cdot \|'$ be a pair of norms on $\mathbb{R}^m$. Let $B$ and $B'$ denote the unit balls in $\| \cdot \|$ and $\| \cdot \|'$ respectively. Then for any $\varepsilon > 0$, it holds that
\[
N(B, \| \cdot \|', \varepsilon) \leq \frac{\text{vol}(\frac{3}{2}B + B')}{\text{vol}(B')},
\]
where $\text{vol}(\cdot)$ denotes the volume of the argument. In particular, for any $\varepsilon \in (0, 1)$,
\[
N\left(B^m(0, 1), \| \cdot \|_\infty, \frac{\varepsilon}{\sqrt{m}} \right) \leq (C/\varepsilon)^m,
\]
where $B^m(0, 1)$ is the unit ball in the $\ell_2$-norm and $C$ is a positive constant.

**Proof.** The proof for the general bound is a standard volume argument and can be found in [7]. To prove the second bound, note that the $\ell_\infty$ unit ball is contained in the $\ell_2$ ball of radius $\sqrt{m}/\varepsilon$. Hence the general bound with $\| \cdot \| = \| \cdot \|_2$ and $\| \cdot \|' = \| \cdot \|_\infty$ implies that
\[
N\left(B^m(0, 1), \| \cdot \|_\infty, \frac{\varepsilon}{\sqrt{m}} \right) \leq \frac{\text{vol}(\frac{3\sqrt{m}}{2}B)}{\text{vol}(B')} \leq (C/\varepsilon)^m,
\]
where the second inequality follows from the asymptotic formula for the volume of an Euclidean ball in $\mathbb{R}^m$: $\text{vol}(rB) \sim \frac{1}{\sqrt{\pi m}} \left( \frac{2\pi e}{m} \right)^{m/2} r^m$ for $r > 0$. $\square$
Define $B_m(0, \frac{x}{NM})$ as the ball in $\mathbb{R}^n$. Fortunately, the following argument corrects this dependency.

Without loss of generality, we assume that $t = 1$. We construct a $3\varepsilon$-net of $C\cap B^m(a, 1)$ as follows. First, let $N_B$ be a minimal $\frac{x}{\sqrt{m}}$-net of $B^m(0, 1)$ with respect to the $\ell_\infty$-norm. Define

$$N_D = \left\{ \mu \in N_B : \min_{i\in[m]} \mu_i \geq -\frac{1}{2\sqrt{m}} \right\}.$$ 

Note that $\mu_i + \frac{1}{\sqrt{m}} = 0$ for $\mu \in N_D$, and let $N_{\mu_i}$ be a minimal $(\mu_i + \frac{1}{\sqrt{m}})\varepsilon$-net of $C_i \cap B^m(a_i, \mu_i + \frac{1}{\sqrt{m}})$. Define $N_{\mu} = N_{\mu_1} \times \cdots \times N_{\mu_m}$, i.e.,

$$N_{\mu} = \{w \in \mathbb{R}^m : w = (w_1, \ldots, w_m), w_i \in N_{\mu_i}\}.$$ 

We claim that $\bigcup_{\mu \in N_D} N_{\mu}$ is a $3\varepsilon$-net of $C\cap B^m(a, 1)$.

Fix $v \in C \cap B^m(a, 1)$. Let $v_{I\mu}$ be the restriction of $v$ to the component space $\mathbb{R}^{n_i}$. Then $v_{I\mu} \in C_i$. Let $\lambda \in \mathbb{R}^m$ be defined by $\lambda_i = ||v_{I\mu} - a_{I\mu}||_2$, so $||\lambda||_2 = ||v - a||_2 \leq 1$. Hence we can find $\mu \in N_B$ such that $||\mu - \lambda||_\infty \leq \frac{x}{\sqrt{m}}\varepsilon$. In particular, for all $i \in [m]$, $\mu_i \geq \lambda_i - \frac{\varepsilon}{2\sqrt{m}} \geq -\frac{1}{2\sqrt{m}}$, so $\mu \in N_D$. Moreover, $||v_{I\mu} - a_{I\mu}||_2 = \lambda_i < \mu_i + \frac{1}{\sqrt{m}}$ and $v_{I\mu} \in C_i$, so by definition of $N_{\mu_i}$, there exists $w_{I\mu} \in N_{\mu_i}$ such that $||w_{I\mu} - v_{I\mu}||_2 \leq (\mu_i + \frac{1}{\sqrt{m}})\varepsilon$. Let $w = (w_1, \ldots, w_m) \in N_{\mu}$. Since

$$\sum_{i=1}^{m} \mu_i^2 \leq \sum_{i=1}^{m} (\lambda_i + |\lambda_i - \mu_i|)^2 \leq \sum_{i=1}^{m} 2\lambda_i^2 + \varepsilon^2 \leq \frac{5}{2},$$

we conclude that

$$||w - v||_2^2 \leq \sum_{i=1}^{m} \left( \mu_i + \frac{1}{\sqrt{m}} \right)^2 \varepsilon^2 \leq 7\varepsilon^2.$$ 

Therefore $\bigcup_{\mu \in N_D} N_{\mu}$ is a $3\varepsilon$-net of $C\cap B^m(a, 1)$.

It remains to bound the cardinality of this net. By Lemma C.1, $|N_D| \leq |N_B| \leq (C/\varepsilon)^m$. Moreover, recall that $N_{\mu}$ is a $(\mu_i + \frac{1}{\sqrt{m}})\varepsilon$-net of $C_i \cap B^m(a_i, \mu_i + \frac{1}{\sqrt{m}})$. Since $a_{I\mu} \in C_i \cap (-C_i)$, for any $t > 0$, $C_i \cap B^m(a_{I\mu}, t) = \{x + a_{I\mu} : x \in C_i \cap B^m(0, t)\}$. Hence we can choose the net so that

$$|N_{\mu_i}| = N\left(C_i \cap B^m(0, \mu_i + \frac{1}{\sqrt{m}}), \|\cdot\|_2, (\mu_i + \frac{1}{\sqrt{m}})\varepsilon\right),$$

$$= N\left(C_i \cap B^m(0, 1), \|\cdot\|_2, \varepsilon\right),$$

$$= N\left(C_i \cap B^m(a_{I\mu}, 1), \|\cdot\|_2, \varepsilon\right).$$

As $|N_{\mu}| \leq \prod_{i=1}^{m} |N_{\mu_i}|$, therefore

$$\left|\bigcup_{\mu \in N_D} N_{\mu}\right| \leq \left(\frac{C}{\varepsilon}\right)^m \prod_{i=1}^{m} N\left(C_i \cap B^m(a_{I\mu}, 1), \|\cdot\|_2, \varepsilon\right).$$
Taking the logarithm completes the proof. □

**Proof of Lemma 6.6.** Part of this proof is due to Lemma 5.1 in an old version of [3], but we improve their result by a factor $\sqrt{\log n}$ and provide the whole proof for completeness. The technique we employ here is similar to that in the proof of Lemma 6.5. Roughly speaking, we shall construct a net of the original set by carefully combining nets of sub-blocks of vectors in the original set.

The bound holds trivially if $\varepsilon > t$, since the left-hand side is zero. Hence we can assume that $\varepsilon \leq t$. We also assume that $b = 0$ since the set of interest is translation invariant. Moreover, we assume that $t = 1$ and $n$ is an even integer for simplicity, as the proof extends easily to the case where $t > 0$ or $n$ is odd. First, let $n' = n/2$ and $I = [n']$. Define $S' = \{(a_1, \ldots, a_{n'}) \in \mathbb{R}^{n'} : a \in S_n \cap B^n(0, 1)\}$. Note that by splitting the vectors into two halves and using symmetry we have

$$\log N(S_n \cap B^n(0, 1), \| \cdot \|_2, \varepsilon) \leq 2 \log N(S', \| \cdot \|_2, \varepsilon/\sqrt{2}). \quad (C.1)$$

To construct a net of $S'$, we introduce some notation. Let $k$ be the smallest integer for which $2^k > n'$, so that in particular $k \leq \log_2 n \leq C \log(en)$. We partition $I$ into $k$ blocks $I_j = I \cap [2^{j-1}, 2^j)$ for $j \in [k]$ and let $m_j = |I_j|$. Define a norm $\| \cdot \|$ on $\mathbb{R}^k$ by

$$\| \mu \| = \left( \sum_{j=1}^{k} 2^j \mu_j^2 \right)^{1/2} \quad (C.2)$$

for $\mu \in \mathbb{R}^k$. Let $B^k_{\| \cdot \|_2}(\mu, r)$ denote a ball in the norm $\| \cdot \|$ in $\mathbb{R}^k$ with radius $r > 0$ centered at $\mu$. Note that $\| \cdot \|$ is simply a weighted $\ell_2$-norm, and a ball in $\| \cdot \|$ is an ellipsoid (or more precisely, the set bounded by an ellipsoid).

Let $\mathcal{N}_\varepsilon$ be a minimal $\varepsilon$-net of $B^k_{\| \cdot \|_2}(0, \sqrt{10}) \cap \mathbb{R}^k_{\geq 0}$ with respect to the norm $\| \cdot \|$, where $\mathbb{R}^k_{\geq 0}$ is the nonnegative orthant of $\mathbb{R}^k$. For each $\mu = (\mu_1, \ldots, \mu_k) \in \mathcal{N}_\varepsilon$, let $\mathcal{N}_{\delta, \mu}$ be a minimal $\frac{\varepsilon}{\sqrt{k}}$-net of $S_{m_j} \cap [-\mu_j, \mu_j]^{m_j}$ with respect to the Euclidean distance. Then we define $\mathcal{N}_{\mu} = \mathcal{N}_{\mu_1} \times \cdots \times \mathcal{N}_{\mu_k}$, and claim that $\bigcup_{\mu \in \mathcal{N}_\varepsilon} \mathcal{N}_{\mu}$ is a $2\varepsilon$-net of $S'$ with respect to the Euclidean distance.

Fix $a \in S'$ so that $a = \tilde{a}$ for some $\tilde{a} \in S_n \cap B^n(0, 1)$. For $j \in [k]$, let $\nu_j = \max_{a_j \in I_j} a_j$. For each block $I_j$ where $j \geq 2$, the maximum $\nu_j$ is achieved either at the left boundary $i_1 = 2^{j-1}$ or the right boundary $i_2 = (2^{j-1} - 1) \wedge n'$. If we have $a_{i_1} \leq 0$ and $\nu_{i_1} = \nu_j$, then $|a_{i_1}| \geq \nu_j$ for all $i \leq i_1$ as $a$ is increasing. Otherwise, we must have $a_{i_2} \geq 0$, and so $a_{i_2} = \nu_j$. In this case, $\tilde{a}_i \geq \nu_j$ for all $i \geq i_2$, and thus $\nu_j \leq 1/\sqrt{n - i_2 + 1} \leq \sqrt{1/n'}$ since $\|a\|_2 \leq 1$ and $i_2 \leq n' = n/2$. Combining the two cases, we obtain that $a_{i_1}^2 + 1/n' \geq \nu_j^2$ for any $i \in I_{j-1}$. Summing over all $i \in [n']$ yields that

$$\sum_{j=2}^{k} 2^{j-2} \nu_j^2 \leq \sum_{j=2}^{k} \sum_{i \in I_{j-1}} (a_i^2 + 1/n') \leq 2.$$ 

Together with the trivial bound $\nu_1 \leq 1$, this implies $\sum_{j=1}^{k} 2^j \nu_j^2 \leq 10$, so we have that $\nu \in B^k_{\| \cdot \|_2}(0, \sqrt{10}) \cap \mathbb{R}^k_{\geq 0}$. By the definition of $\mathcal{N}_\varepsilon$, there exists $\mu \in \mathcal{N}_\varepsilon$ such that $\| \mu - \nu \| \leq \varepsilon$. 


Moreover, define $a' \in S'$ by $a'_i = (a_i \land \mu_j) \lor (-\mu_j)$ for any $i \in I_j$ where $j \in [k]$. Recall that $\nu_j = \max_{i \in I_j} |a_i|$, so it holds that

$$\|a' - a\|^2 = \sum_{j=1}^{k} \sum_{i \in I_j} (a'_i - a_i)^2 \leq \sum_{j=1}^{k} \sum_{i \in I_j} (\mu_j - \nu_j)^2 \leq \sum_{j=1}^{k} 2^{j-1} (\mu_j - \nu_j)^2 \leq \|\mu - \nu\|^2 \leq \varepsilon^2.$$  

Note that $a'_j \in S_{m_j} \cap [-\mu_j, \mu_j]^{m_j}$. By the definition of $N_{\mu_j}$, there exists a vector $x_{I_j} \in N_{\mu_j}$ such that $\|x_{I_j} - a'I_j\|_2 \leq \varepsilon/\sqrt{k}$. If we define $x \in \mathbb{R}^n$ by concatenating $x_{I_j}$ for $j \in [k]$, then we have $x \in N_\mu$ and $\|x - a'\|_2 \leq \varepsilon$. Finally, the triangle inequality gives that $\|x - a\|_2 \leq 2\varepsilon$, so $\bigcup_{\mu \in \mathcal{N}\varepsilon_{\mu}} N_\mu$ is indeed a $2\varepsilon$-net of $S'$. It remains to bound the cardinality of $\bigcup_{\mu \in \mathcal{N}\varepsilon_{\mu}} N_\mu$. By the definition of $\mathcal{N}\varepsilon$, its cardinality is bounded by $N(B^k_{\|\cdot\|_2}(0, \sqrt{10}))$, so it holds that

$$\log |\mathcal{N}_{\varepsilon}| \leq \log N(B^k_{\|\cdot\|_2}(0, \sqrt{10}), \|\cdot\|_2, \varepsilon) \leq \log \frac{\text{vol}(B^k_{\|\cdot\|_2}(0, 9/\varepsilon))}{\text{vol}(B^k_{\|\cdot\|_2}(0, 1))} = -\frac{9}{\varepsilon}.$$  

because if we scale an ellipsoid in $\mathbb{R}^k$ by ratio $r > 0$ then its volume scales as $r^k$. Next, we know from [2, Lemma 4.20] that for any $d \geq c \geq 0$ and $n \geq 1$,

$$\log N(S_n \cap [c, d]^n, \|\cdot\|_2, \varepsilon) \leq C \varepsilon^{-1} \sqrt{n}(d - c).$$  

It follows that for any $\mu \in \mathcal{N}\varepsilon$ and $j \in [k]$,

$$\log |N_{\mu_j}| = \log N(S_{m_j} \cap [-\mu_j, \mu_j]^{m_j}, \|\cdot\|_2, \varepsilon/\sqrt{k}) \leq C \varepsilon^{-1} \sqrt{k} 2^{j/2} \mu_j.$$  

Summing over $j \in [k]$ and applying the Cauchy-Schwarz inequality, we get that

$$\log |N_{\mu}| \leq \sum_{j=1}^{k} \log |N_{\mu_j}| \leq \frac{C}{\varepsilon} \sqrt{k} \sum_{j=1}^{k} 2^{j/2} \mu_j \leq \frac{C}{\varepsilon} k \left( \sum_{j=1}^{k} 2^{j/2} \mu_j^2 \right)^{1/2} \leq \frac{C}{\varepsilon} k,$$  

where the last inequality holds because $\|\mu\| \leq \sqrt{10}$ by definition. Since $k \leq C \log(en)$, (C.3) and (C.4) imply that $\log |\bigcup_{\mu \in \mathcal{N}_\varepsilon} N_{\mu}| \leq C \varepsilon^{-1} \log(en)$. We complete the proof by combining this bound with (C.1). \qed

**Proof of Lemma 6.7.** Assume that $\varepsilon \leq t$ since otherwise the left-hand side is zero and the bound holds trivially. For $j \in [m]$, define $I^{1,j} = [l_j]$ and $I^{2,j} = [n] \setminus [l_j]$. Define $k_{j,1} = k(A_{j^{1,j}}, j)$ and $k_{j,2} = k(A_{j^{2,j}}, j)$. Let $\kappa = \sum_{j=1}^{m} (k_{j,1} + k_{j,2})$ and observe that $K(A) \leq \kappa \leq 2K(A)$. Moreover, let $\{I_1^{j,1}, \ldots, I_{k_{j,1}}^{j,1}\}$ be the partition of $I^{1,j}$ such that $A_{j^{1,j}}$ is a constant vector for $i \in [k_{j,1}]$. Note that elements of $I_1^{j,1}$ need not to be consecutive. Define the partition for $I^{2,j}$ analogously.
For $j \in [m]$ and $i \in [k_{j,1}]$ (resp. $[k_{j,2}]$), let $S_{I^r_j}$ (resp. $S_{I^r_j}$) denote the set of increasing (resp. decreasing) vectors in the component space $\mathbb{R}^{I^r_j}$ (resp. $\mathbb{R}^{I^r_j}$). Lemma 6.6 implies that

$$\log N(S_{I^r_j} \cap B^{I^r_j}(A_{I^r_j}, t), \| \cdot \|_F, \varepsilon) \leq C\varepsilon^{-1} t \log(e|I^r_j|).$$

As a matrix in $\mathbb{R}^{n \times m}$ can be viewed as a concatenation of $x = \sum_{j=1}^{m}(k_{j,1} + k_{j,2})$ vectors of length $|I^r_j|$, $r \in [2], j \in [m]$, we define the cone $S^*$ in $\mathbb{R}^{n \times m}$ by $S^* = \prod_{j=1}^{m} \prod_{i=1}^{k_{j,r}} S_{I^r_j}$, which is clearly a superset of $C^m_1$. It also follows that $A \in S^* \cap (-S^*)$, and thus by Lemma 6.5 and the previous display,

$$\log N(S^* \cap B^{nm}(A, t), \| \cdot \|_F, \varepsilon) \leq \varepsilon \log \frac{Ct}{\varepsilon} + \sum_{j=1}^{m} \sum_{r=1}^{2} \sum_{i=1}^{k_{j,r}} C\varepsilon^{-1} t \log(e|I^r_j|)$$

$$\leq C\varepsilon^{-1} t \varepsilon + C\varepsilon^{-1} t \varepsilon \log \frac{\varepsilon \sum_{j,r,i} |I^r_j|}{\varepsilon}$$

$$\leq C\varepsilon^{-1} t K(A) \log \frac{enm}{K(A)},$$

where we used the concavity of the logarithm and Jensen’s inequality in the second step, and that $K(A) \leq \varepsilon \leq 2K(A)$ in the last step.

Since $A \in S^* \cap (-S^*)$ (the cone $S^*$ is pointed at $A$) we have that $S^* \cap B^{nm}(\lambda A, t) - \lambda A = S^* \cap B^{nm}(0, t)$ for any $\lambda \geq 0$. In view of the definition of $\Theta$, it holds

$$\Theta_{S^*}(A, t) = \bigcup_{\lambda \geq 0} S^* \cap B^{nm}(\lambda A, t) - \lambda A = S^* \cap B^{nm}(\lambda A, t) - \lambda A, \quad \forall \lambda \geq 0.$$

In particular, taking $\lambda = 1$, we get $\Theta_{S^*}(A, t) = S^* \cap B^{nm}(A, t) - A$. Moreover, $C^m_1 \subset S^*$, so that $\Theta_{C^m_1}(A, t) \subset \Theta_{S^*}(A, t) = S^* \cap B^{nm}(A, t) - A$. Thus the metric entropy of $\Theta_{C^m_1}(A, t)$ is subject to the above bound as well.

**Proof of Lemma 6.8.** Assume that $\varepsilon \leq t$ since otherwise the left-hand side is zero and the bound holds trivially. Note that $U^m = \bigcup_{i \in [m]} C^m_i$, and that $M = \bigcup_{i \in \mathbb{R}} \mathbb{R} U^m$. Thus $M$ is the union of $nm^n$ cones of the form $\Pi C^m_i$. By definition, $\Theta_M(A, t)$ is also the union of $nm^n$ sets $\Theta_{\Pi C^m_i}(A, t)$, each having metric entropy subject to the bound in Lemma 6.7. Therefore, a union bound implies that

$$\log N(\Theta_M(A, t), \| \cdot \|_F, \varepsilon) \leq \log N(\Theta_{C^m_1}(A, t), \| \cdot \|_F, \varepsilon) + \log(nm^n)$$

$$\leq C\varepsilon^{-1} t K(A) \log \frac{enm}{K(A)} + m \log n + n \log n$$

$$\leq C\varepsilon^{-1} t K(A) \log \frac{enm}{K(A)} + n \log n,$$

where the last step follows from that $K \log(enm/K) \geq m \log n$ for $m \leq K \leq nm$ and that $\varepsilon \leq t$. 

\[ \square \]
Appendix D: Proofs of lemmas in Section 6.3

The Varshamov-Gilbert lemma [5, Lemma 4.7] is a standard tool for proving lower bounds.

**Lemma D.1 (Varshamov-Gilbert).** Let $\delta$ denote the Hamming distance on $\{0,1\}^d$ where $d \geq 2$. Then there exists a subset $\Omega \subset \{0,1\}^d$ such that $\log |\Omega| \geq d/8$ and $\delta(\omega,\omega') \geq d/4$ for distinct $\omega,\omega' \in \Omega$.

We also need the following useful lemma.

**Lemma D.2.** Consider the model $y = \theta + z$ where $\theta \in \Theta \subset \mathbb{R}^d$ and $z \sim N(0,\sigma^2 I_d)$. Suppose that $|\Theta| \geq 3$ and for distinct $\theta,\theta' \in \Theta$, $4\phi \leq \|\theta - \theta'\|_2^2 \leq 2\sigma^2 \log |\Theta|$ where $\phi > 0$. Then there exists $c > 0$ such that

$$\inf_{\theta} \sup_{\theta' \in \Theta} \mathbb{P}_{\theta}[\|\theta - \theta'\|_2^2 \geq \phi] \geq c.$$

**Proof.** Let $\mathbb{P}_{\theta}$ denote the probability with respect to $\theta + z$. Then the Kullback-Leibler divergence between $\mathbb{P}_{\theta}$ and $\mathbb{P}_{\theta'}$ satisfies

$$\text{KL}(\mathbb{P}_{\theta}, \mathbb{P}_{\theta'}) = \frac{\|\theta - \theta'\|_2^2}{2\sigma^2} \leq \frac{\log |\Theta|}{16} \leq \frac{\log(|\Theta| - 1)}{10},$$

since $|\Theta| \geq 3$. Applying [6, Theorem 2.5] with $\alpha = \frac{1}{10}$ gives the conclusion. \qed

**Proof of Lemma 6.10.** We adapt the proof of [1, Theorem 4] to the case of matrices. Let $V_j = V_0$ for all $j \in [m]$. Since

$$K_0 \leq m\left(\frac{16n}{\sigma^2}\right)^{1/3}V_0^{2/3} - m = \sum_{j=1}^{m} \left[\left(\frac{16n}{\sigma^2}\right)^{1/3}V_j^{2/3} - 1\right],$$

we can choose $k_j \in [n]$ so that $k_j \leq \left(\frac{16n}{\sigma^2}\right)^{1/3}V_j^{2/3}$ and $K_0 = \sum_{j=1}^{m} k_j$. According to Lemma D.1, there exists $\Omega \subset \{0,1\}^{K_0}$ such that $\log |\Omega| \geq K_0/8$ and $\delta(\omega,\omega') \geq K_0/4$ for distinct $\omega,\omega' \in \Omega$. Consider the partition $[K_0] = \bigcup_{j=1}^{m} I_j$ with $|I_j| = k_j$. For each $\omega \in \Omega$, let $\omega^j \in \{0,1\}^{k_j}$ be the restriction of $\omega$ to coordinates in $I_j$. Define $M^\omega \in \mathbb{R}^{n \times m}$ by

$$M^\omega_{i,j} = \frac{(i-1)k_j/n\cdot V_j}{2k_j} + \gamma_j\omega_{(i-1)k_j/n}^{(i-1)k_j},$$

where $\gamma_j = \frac{\sigma}{8}\sqrt{k_j}/2n$. It is straightforward to check that $k(M_{i,j}) \leq k_j$, $V(M_{i,j}) \leq V_j$ and $M_{i,j}$ is increasing, so $M$ is in the parameter space. Moreover, for distinct $\omega,\omega' \in \Omega$,

$$\|M^\omega - M'^\omega\|_2^2 \geq c \sum_{j=1}^{m} \frac{n}{k_j} \gamma_j^2 \delta(\omega^j,(\omega')^j) \geq c\sigma^2 \sum_{j=1}^{m} \delta(\omega^j,(\omega')^j) = c\sigma^2 K_0.$$
On the other hand,

\[ \|M^\omega - M^{\omega'}\|_F^2 \leq \frac{2}{m} \sum_{j=1}^{\lfloor n/2 \rfloor} \gamma_j^2 \delta(\omega^j, (\omega')^j) \leq \frac{\sigma^2}{64} \delta(\omega, \omega') \leq \frac{\sigma^2 K_0}{64} \leq \frac{\sigma^2}{8} \log |\Omega|. \]

Applying Lemma D.2 completes the proof.

**Proof of Lemma 6.11.** By Lemma D.1, there exists \( \Omega \subset \{0,1\}^n \) such that \( \log |\Omega| \geq n/8 \) and \( \delta(\omega, \omega') \geq n/4 \) for distinct \( \omega, \omega' \in \Omega \). For each \( \omega \in \Omega \), define \( M^\omega \in \mathbb{R}^{n \times m} \) by setting the first column of \( M^\omega \) to be \( \alpha \omega \) and all other entries to be zero, where \( \alpha = \min \left( \frac{\sigma}{8}, m^{1/2} \right) \). Then

1. \( M^\omega \in M_{K_0}(\Omega) \) since \( K(M) = m + 1 \leq K_0 \), \( V(M) \leq V_0 \) and we can permute the rows of \( M^\omega \) so that its first column is increasing;
2. \( \|M^\omega - M^{\omega'}\|_F^2 \geq \min \left( \frac{\sigma^2}{64}, m^3 V_0^2 \right) \delta(\omega, \omega') \geq \min \left( \frac{\sigma^2}{256}, \frac{m^3}{4} V_0^2 \right) \) for distinct \( \omega, \omega' \in \\Omega \);
3. \( \|M^\omega - M^{\omega'}\|_F^2 \leq \frac{\sigma^2}{64} \delta(\omega, \omega') \leq \frac{\sigma^2}{64} n \leq \frac{\sigma^2}{8} \log |\Omega| \) for \( \omega, \omega' \in \Omega \).

Applying Lemma D.2 completes the proof.

The following packing lemma is the key to the proof of Lemma 6.12.

**Lemma D.3.** For \( l \in [m] \), consider the set \( \mathfrak{M} \) of \( n \times m \) matrices of the form

\[ M = \begin{cases} 1 & \text{for exactly one } j_i \in [l] \text{ for each } i \in [n], \\ 0 & \text{otherwise}. \end{cases} \]

For \( \varepsilon > 0 \), define \( k = \left\lfloor \frac{\varepsilon n^2}{4} \right\rfloor \). Then there exists an \( \varepsilon \sqrt{n} \)-packing \( \mathcal{P} \) of \( \mathfrak{M} \) such that \( |\mathcal{P}| \geq l^{n-k} \left( \frac{k}{\varepsilon n} \right)^k \) if \( k \geq 1 \) and \( |\mathcal{P}| \leq l^n \) if \( k = 0 \).

**Proof.** There are \( l \) choices of entries to put the one in each row of \( M \), so \( |\mathfrak{M}| = l^n \). Fix \( M_0 \in \mathfrak{M} \). If \( \|M - M_0\|_F \leq \varepsilon \sqrt{n} \) where \( M \in \mathfrak{M} \), then \( M \) differs from \( M_0 \) in at most \( k \) rows. If \( k = 0 \), taking \( \mathcal{P} = \mathfrak{M} \) gives the result. If \( k \geq 1 \) then

\[ |\mathfrak{M} \cap B^{nm}(M_0, \varepsilon \sqrt{n})| \leq \left( \frac{n}{k} \right)^k \leq \left( \frac{\varepsilon n}{k} \right)^k. \]

Moreover, let \( \mathcal{P} \) be a maximal \( \varepsilon \sqrt{n} \)-packing of \( \mathfrak{M} \). Then \( \mathcal{P} \) is also an \( \varepsilon \sqrt{n} \)-net, so \( \mathfrak{M} \subset \bigcup_{M_0 \in \mathcal{P}} B^{nm}(M_0, \varepsilon \sqrt{n}) \). It follows that

\[ l^n = |\mathfrak{M}| \leq \sum_{M_0 \in \mathcal{P}} |\mathfrak{M} \cap B^{nm}(M_0, \varepsilon \sqrt{n})| \leq |\mathcal{P}| \cdot \left( \frac{\varepsilon n}{k} \right)^k. \]

We conclude that \( |\mathcal{P}| \geq l^{n-k} \left( \frac{k}{\varepsilon n} \right)^k \).
Proof of Lemma 6.12. For notational simplicity, we consider $2 \leq l \leq \min(K_0 - m, m)$ instead of $3 \leq l \leq \min(K_0 - m, m) + 1$.

Set $\varepsilon = 1/2$ and let $\mathcal{P}$ be the $\sqrt{n}/2$-packing given by Lemma D.3. If $k = \lfloor \frac{n}{x} \rfloor = 0$, then $\log |\mathcal{P}| = n \log l$. Now assume that $k \geq 1$. Since $(\frac{x}{\varepsilon n})^x$ is decreasing on $[1, n]$, we have that $|\mathcal{P}| \geq \frac{l^n n^{8/3}}{\varepsilon^3}$. Hence for $l \geq 2$,
\[\log |\mathcal{P}| \geq \frac{7n}{8} \log l - \frac{n}{8} \log (8e) \geq \frac{n}{4} \log l. \tag{D.1}\]
Moreover, for each $M_0 \in \mathcal{P}$, consider the rescaled matrix
\[M = \min \left( \frac{\sigma}{8} \sqrt{\frac{\log l}{2}}, (\frac{m}{l})^{3/2} V_0 \right) M_0.\]

1. We can permute the rows of $M_0$ so that each column has consecutive ones (or all zeros), so $M \in \mathcal{M}$. Moreover,
\[K(M) = 2l + m - l \leq \min(m, K_0 - m) + m \leq K_0\]
and
\[V(M) \leq \left( \frac{1}{m} \sum_{j=1}^{l} ((m/l)^{3/2} V_0)^{2/3} \right)^{3/2} = V_0,\]
so $M \in \mathcal{M}_{V_0}(V_0)$ for $M_0 \in \mathcal{P}$.

2. For $M_0, M'_0 \in \mathcal{P}$, $\|M_0 - M'_0\|_F^2 \geq n/4$, so
\[\|M - M'\|_F^2 = \min \left( \frac{\sigma^2 \log l}{128}, (\frac{m}{l})^{3/2} V_0^2 \right) \|M_0 - M'_0\|_F^2 \geq \min \left( \frac{\sigma^2}{512} n \log l, \frac{n}{4} (\frac{m}{l})^{3/2} V_0^2 \right).\]

3. For $M_0, M'_0 \in \mathcal{P}$, $\|M_0 - M'_0\|_F^2 \leq 2\|M_0\|_F^2 + 2\|M'_0\|_F^2 \leq 4n$, so by (D.1),
\[\|M - M'\|_F^2 \leq \frac{\sigma^2 \log l}{128} \|M_0 - M'_0\|_F^2 \leq \frac{\sigma^2}{32} n \log l \leq \frac{\sigma^2}{8} \log |\mathcal{P}|.\]
Since $\log l \geq \frac{1}{2} \log (l + 1)$ for $l \geq 2$, applying Lemma D.2 completes the proof.

Proof of Theorem 3.6. The last term $\min(\frac{\sigma^2}{m}, m^2 V_0^2)$ is achieved by Lemma 6.11, so we focus on the trade-off between the first two terms. Suppose that $(\frac{16n}{\sigma^2})^{1/3} V_0^{2/3} \geq 3$, in which case the first term $(\frac{\sigma^2 V_0}{m})^{2/3}$ dominates the second term. Then $m(\frac{16n}{\sigma^2})^{1/3} V_0^{2/3} - m \geq 2m$. Setting
\[K_0 = \left \lfloor m \left( \frac{16n}{\sigma^2} \right)^{1/3} V_0^{2/3} - m \right \rfloor,\]
we see that $K_0 \geq \left \lfloor \frac{m}{2} \left( \frac{16n}{\sigma^2} \right)^{1/3} V_0^{2/3} \right \rfloor$. Lemma 6.10 can be applied with this choice of $K_0$. Then the term $\sigma^2 K_0^2/m$ is lower bounded by $c(\frac{\sigma^2 V_0}{m})^{2/3}$.
On the other hand, if \((\frac{1}{\sqrt{n}})^{1/3}V_0^{2/3} \leq 3\), then the second term \(\frac{\sigma^2}{n}\) dominates the first up to a constant. To deduce a lower bound of this rate, we apply Lemma D.1 to get \(\Omega \subset \{0,1\}^m\) such that \(\log |\Omega| \geq m/8\) and \(\delta(\omega,\omega') \geq m/4\) for distinct \(\omega,\omega' \in \Omega\). For each \(\omega \in \Omega\), define \(M^\omega \in \mathbb{R}^{n \times m}\) by setting every row of \(M^\omega\) equal to \(\frac{\sigma}{8\sqrt{n}}\omega^\top\). Then

1. \(M^\omega \in \mathcal{U}^m(V_0)\) since \(V(M^\omega) = 0\);
2. \(\|M^\omega - M^{\omega'}\|^2_F = \frac{\sigma^2}{\sqrt{m}}\delta(\omega,\omega') \geq c\sigma^2 m\);
3. \(\|M^\omega - M^{\omega'}\|^2_F = \frac{\sigma^2}{\sqrt{m}}\delta(\omega,\omega') \leq \frac{\sigma^2}{\sqrt{m}} m \leq \frac{\sigma^2}{n} \log |\Omega|\).

Hence Lemma D.2 implies a lower bound on \(\frac{\sigma^2 m}{n}\) of rate \(\frac{\sigma^2 m}{n}\). \(\square\)

**Appendix E: Proofs of lemmas in Section 6.4**

**Proof of Lemma 6.13.** The proof follows that of Theorem 3.1 with appropriate adaptation, so for simplicity we will not detail every step. Assume without loss of generality that \(\Pi^* = I_n\). Fix \(A \in \mathcal{S}^m\) and \(\tilde{\Pi} \in \mathcal{G}_n\). Define

\[
\tilde{f}_{\Pi A}(t) = \sup_{M \in \Pi S^m \cap B_{\mathbb{R}^{n \times m}}(\tilde{\Pi}A,t)} (M - \tilde{\Pi}A, Y - \tilde{\Pi}A) - \frac{t^2}{2}.
\]

Since \(\mathcal{S}^m = C_1^m\) with \(I = (n, \ldots, n)\), by Lemma 6.7,

\[
\log N(\Theta_{\tilde{\Pi}S^m}(A,t), \|\cdot\|_F, \varepsilon) \leq C\varepsilon^{-1} t K(A) \log \frac{enm}{K(A)}.
\]

Following the proof of Lemma 6.3, we see that

\[
\tilde{f}_{\Pi A}(t) \leq C\sigma t \sqrt{K(A) \log \frac{enm}{K(A)}} + t\|\tilde{\Pi}A - A^*\|_F - \frac{t^2}{2} + st
\]

with probability at least \(1 - C\exp(-\frac{\sigma^2}{\varepsilon^2})\). Lemma 6.1 then implies that on this event

\[
\|\tilde{\Pi}A - A^*\|_F \leq 2C\sigma \sqrt{K(A) \log \frac{enm}{K(A)}} + 3\|\tilde{\Pi}A - A^*\|_F + 2s. \tag{E.1}
\]

Taking \(s = \sigma \left[\sqrt{K(A) \log \frac{enm}{K(A)}} + C_2 \sqrt{n \log n}\right]\) for a sufficiently large constant \(C_2 > 0\), we see that with probability at least \(1 - \exp(-c(m + n) - n \log n)\),

\[
\|\tilde{\Pi}A - A^*\|_F^2 \leq \sigma^2 K(A) \log \frac{enm}{K(A)} + \sigma^2 n \log n + \|\tilde{\Pi}A - A^*\|_F^2
\]

\[
\leq \sigma^2 K(A) \log \frac{enm}{K(A)} + \sigma^2 n \log n + \|A - A^*\|_F^2 + \|\tilde{\Pi}A - A^*\|_F^2.
\]

Minimizing over \(A \in \mathcal{S}^m\) yields the desired bound for a fixed \(\tilde{\Pi}\). Finally, the bound holds simultaneously for all \(\tilde{\Pi} \in \mathcal{G}_n\) with probability at least \(1 - e^{-c(m+n)}\) by a union bound since \(n! < n^n = \exp(n \log n)\). \(\square\)
Proof of Lemma 6.14. Since $Z \sim \text{subG}(\sigma^2)$, $Z_{i,j}$ and $\frac{1}{\sqrt{m}} \sum_{j=1}^{m} Z_{i,j}$ are sub-Gaussian random variables with variance proxy $\sigma$. A standard union bound yields that

$$
\max \left( \max_{i \in [n], j \in [m]} |Z_{i,j}|, \max_{i \in [n]} \frac{1}{\sqrt{m}} \left| \sum_{j=1}^{m} Z_{i,j} \right| \right) \leq \tau = 3\sigma \sqrt{\log(nm\delta^{-1})}
$$

on an event $\mathcal{E}$ of probability at least $1 - 2(nm + n) \exp(-\frac{\tau^2}{2\sigma^2}) \geq 1 - \delta$.

In the sequel, we make statements that are valid on the event $\mathcal{E}$. Since $Y_{\tau^*(i),j} = A_{*i,j} + Z_{i,j}$, by the triangle inequality,

$$
|\Delta_Y(\tau^*(i), \tau^*(i')) - \Delta_{A^*}(i, i')| \leq 2\tau. \quad (E.2)
$$

Suppose that $\Delta_{A^*}(i, i') \geq 4\tau$. We claim that $s_{\tau^*(i)} < s_{\tau^*(i')}$.

If for $l \in [n]$ we have $\Delta_Y(\tau^*(l), \tau^*(i)) \geq 2\tau$, then $\Delta_{A^*}(l, i) \geq 0$ by (E.2). Since $A^*$ has increasing columns, $\Delta_{A^*}(l, i') \geq 4\tau$.

Again by (E.2), $\Delta_Y(\tau^*(l), \tau^*(i')) \geq 2\tau$. By the definition of the score, we see that $s_{\tau^*(i)} \leq s_{\tau^*(i')}$. Moreover, $\Delta_{A^*}(i, i') \geq 4\tau$ so $\Delta_Y(\tau^*(i), \tau^*(i')) \geq 2\tau$. Therefore $s_{\tau^*(i)} < s_{\tau^*(i')}$. According to the construction of $\pi, \pi^* = \pi^{-1} \circ \tau^*(i) < \pi^{-1} \circ \tau^*(i')$.

Proof of Lemma 6.15. Throughout the proof, we restrict ourselves to the event $\mathcal{E}$ defined in Lemma 6.14. To simplify the notation, we define $\alpha_i = A^*_{\pi(i),i} - A^*_{\pi^*(i),i}$. Then

$$
||\Pi A^* - \Pi^* A^*||^2_F = \sum_{i=1}^{n} ||A^*_{\pi(i),i} - A^*_{\pi^*(i),i}||^2_2 = \sum_{i \in I} ||\alpha_i||^2_2, \quad (E.3)
$$

where $I$ is the set of indices $i$ for which $\alpha_i$ is nonzero. For each $i \in I$,

$$
||\alpha_i||^2_2 = \min \left( \frac{||\alpha_i||^2_2}{||\alpha_i||^2_{\infty}}, \frac{m||\alpha_i||^2_2}{||\alpha_i||^2_1} \right) \cdot \max \left( \frac{||\alpha_i||^2_2}{||\alpha_i||^2_{\infty}}, \frac{m||\alpha_i||^2_2}{||\alpha_i||^2_1} \right)
$$

by (6.11).

Next, we proceed to showing that $|\Delta_{A^*}(i, \nu(i))| \leq 4\tau$ for any $i \in [n]$, where $\nu = \pi^{-1} \circ \tau^*$. To that end, note that if $\Delta_{A^*}(i, \nu(i)) > 4\tau$, in which case $\Delta_{A^*}(i, i') > 4\tau$ for all $i' \in I' := \{i' \in [n] : i' \geq \nu(i)\}$, then it follows from Lemma 6.14 that on $\mathcal{E}$, $\nu(i) < \nu(i'), \forall i \in I'$. Note that $|\nu(I')| = |I'| = n - \nu(i) + 1$. Hence $\nu(i) \leq \nu(i'), \forall i \in I'$ implies that $\nu(i) \leq n - \nu(i') = \nu(i) - 1$, which is a contradiction. Therefore, there does not exist such $i \in [n]$ on $\mathcal{E}$. The case where $\Delta_{A^*}(i, \nu(i)) < -4\tau$ is treated in a symmetric manner.

Combining this bound with (E.3) and (E.4), we conclude that

$$
||\Pi A^* - \Pi^* A^*||^2_F \leq \sum_{i \in I} \min \left( \frac{||\alpha_i||^2_2}{||\alpha_i||^2_{\infty}}, \frac{m||\alpha_i||^2_2}{||\alpha_i||^2_1} \right) \cdot \tau^2 \leq \sigma^2 R(A^*) n \log(nm\delta^{-1}),
$$

by the definitions of $R(A^*)$ and $\tau$. \qed
Appendix F: Proof of Corollary 5.1

The proof closely follows that of Theorem 3.1 and Theorem 3.3.

First note that the term \( n \log n \) in the bound of Lemma 6.8 comes from a union bound applied to the set of permutations, so it is not present if we consider only the set of unimodal matrices \( \mathcal{U}^m \) instead of \( \mathcal{M} \). Hence taking \( m = 1 \) in the lemma yields that

\[
\log N(\Theta_{\mathcal{U}}(\hat{\theta}, t), \| \cdot \|_2, \varepsilon) \leq C\varepsilon^{-1} t k(\hat{\theta}) \log \frac{en}{k(\theta)} .
\]

For \( \tilde{\theta} \in \mathcal{U} \), define

\[
f_{\tilde{\theta}}(t) = \sup_{\theta \in \mathcal{U} \cap B^n(\tilde{\theta}, t)} \langle \theta - \tilde{\theta}, y - \tilde{\theta} \rangle - \frac{t^2}{2}.
\]

Following the proof of Lemma 6.3 and using the above metric entropy bound, we see that

\[
f_{\tilde{\theta}}(t) \leq C\sigma t \sqrt{k(\tilde{\theta})} \log \frac{en}{k(\theta)} + t\|\tilde{\theta} - \theta^*\|_2 - \frac{t^2}{2} + st
\]

with probability at least \( 1 - C \exp(-\frac{ct^2}{\sigma^2}) \). Then the proof of Theorem 3.1 gives that with probability at least \( 1 - C \exp(-\frac{c\alpha n^2}{\sigma^2}) \),

\[
\|\hat{\theta} - \theta^*\|_2 \leq C\left( \sigma \sqrt{k(\hat{\theta})} \log \frac{en}{k(\hat{\theta})} + \|\hat{\theta} - \theta^*\|_2 \right) + 2s.
\]

Taking \( s = C\sigma \sqrt{\alpha \log n} \) for \( \alpha \geq 1 \) and \( C \) sufficiently large, we get that with probability at least \( 1 - n^{-\alpha} \),

\[
\|\hat{\theta} - \theta^*\|_2^2 \lesssim \sigma^2 k(\hat{\theta}) \log \frac{en}{k(\hat{\theta})} + \|\hat{\theta} - \theta^*\|_2^2 + \alpha \sigma^2 \log n.
\]

Minimizing over \( \tilde{\theta} \in \mathcal{U} \) yields the first bound of the corollary. The corresponding bound in expectation follows from integrating the tail probability as in the proof of Theorem 3.1.

Finally, we can apply the proof of Theorem 3.3 with \( m = 1 \) to achieve the global bound.

References


