Convergence analysis for dictionary learning under a signal model with non-homogeneous distribution of supports & coefficient amplitudes

Morris-Luca KÜHMEIER, 1177 5648
Innsbruck, September 2023

Master Thesis

in partial fulfillment of the requirements for the degree of

Master of Science

Master programme Mathematics
Advisor: Univ.-Prof. Dr. Karin Schnass
Faculty of Mathematics, Computer Science and Physics
of the University of Innsbruck

Contents

Abstract

In this thesis we show convergence for two alternating minimisation algorithms for dictionary learning under mild conditions. To be precise we prove convergence of the Method of Optimal Directions (MOD) and the algorithm for Online Dictionary Learning (ODL) for data models with non-uniform distribution of the supports of sparse coefficients in combination with non-homogeneous distribution of the coefficient amplitudes. The innovation lies in including coefficients with non-homogeneous sizes, which is a generalization of the results in [24], in which the coefficient amplitudes are uniformly distributed. We prove that a well-behaved initial dictionary contracts to the generating dictionary with geometric convergence rate, if either their distance is not larger than $1/\log(K)$ or if it is assured that each component of the initial dictionary is associated with exactly one element of the generating dictionary.

Acknowledgements

I am extremely grateful to my supervisor Professor Karin Schnass. I want to thank her for her continuous support, invaluable feedback and especially acknowledge her kind, patient and also humorous way of explaining things. I am very much looking forward to continue my academic career under her guidance.
Additionally, I am profoundly thankful for my mother's support over the last few months.

1. Introduction

In the field of signal processing and machine learning, dictionary learning plays an important role for sparse representations and analysis of data. The goal of dictionary learning is to find a dictionary $\Phi = (\phi_1, \ldots, \phi_K) \in \mathbb{R}^{d\times K}$ for a number of given signals $y \in \mathbb{R}^d$, which are stored as columns of a matrix $Y = (y_1, \ldots, y_N)$, such that

$$Y \approx \Phi X, \quad \text{with} \quad X \in \mathbb{R}^{K\times N}\text{sparse.}$$

The coefficient matrix $X = (x_1, \ldots, x_N)$ being sparse means, that most of the entries of each x_n are zero. There is a vast variety of algorithms, which are trying to solve this problem [11, 3, 10, 27, 15, 16, 17, 29, 25] and also in a theoretical context, research is progressively increasing [12, 30, 4, 26, 13, 6, 5, 1, 2, 31, 32, 7, 28, 20, 19]. In this thesis we will discuss alternating optimisation algorithms, which use as common starting point the following programme:

$$\text{argmin}_{\Psi,X}\|Y - \Psi X\|_F^2 \quad \text{s. t.} \quad X \in \mathcal{S} \quad \text{and} \quad \|\psi_k\|_2 = 1 \quad \text{for all } k, \tag{1}$$

where $\mathcal{S}$ is a set which imposes sparsity on the coefficient matrix X, for example by assuming that every x_n has at most $S \ll d$ coefficients unequal to zero. Alternating optimisation algorithms derive their name from alternating between updating the coefficient matrix and updating the dictionary. Concretely this means for a fixed dictionary Ψ the coefficient update can be described as

$$\hat{X} = \text{argmin}_X \|Y - \Psi X\|_F^2 = \text{argmin}_X \sum_n \|y_n - \Psi x_n\|_2^2 \quad \text{s.t.} \quad \|x_n\|_0 \leq S, \tag{2}$$

whereas for a fixed coefficient matrix $\hat{X}$ the update of the dictionary underlies the following scheme:

$$\text{argmin}_\Psi \|Y - \Psi\hat{X}\|_F^2 =: f(\Psi) \quad \text{s.t.} \ \|\psi_k\|_2 = 1. \tag{3}$$

In this master thesis we study two alternating optimisation algorithms for dictionary learning, namely the Method of Optimal Directions (MOD), [9], and the algorithm for Online Dictionary Learning (ODL), [17]. The former addresses the problem in (3) by neglecting the unit norm constraint on the atoms and which has a closed form solution:

$$\text{argmin}_\Psi f(\Psi) = Y\hat{X}^\dagger.$$

After updating the dictionary as in (1) we enforce the unit norm constraint by scaling with an appropriate diagonal matrix D, which then means $Y\hat{X}^\dagger D$. On the other hand the ODL algorithm approaches the problem in (3) by projected block coordinate descent. Starting with the gradient we get

$$\nabla_\Psi f(\Psi) = -\sum (y_n - \Psi\hat{x}_n)\,\hat{x}_n^* = -Y\hat{X}^* + \Psi\hat{X}\hat{X}^*,$$

which leads to the following gradient step, where Λ defines a diagonal matrix with the length of the step sizes and D as a diagonal matrix enforcing the unit norm as above:

$$\left[\Psi + \left(Y\hat{X}^* - \Psi\hat{X}\hat{X}^*\right)\cdot\Lambda\right]\cdot D = \left[\Psi\Lambda^{-1} - Y\hat{X}^* + \Psi\hat{X}\hat{X}^*\right]\cdot\Lambda D. \tag{4}$$

Choosing Λ^{-1} as $\operatorname{diag}(\hat{X}\hat{X})$ results in the dictionary update scheme of ODL, [17].
Contribution: In the recent paper [24] convergence of the MOD and ODL algorithm was derived for a signal model which includes non-uniform distribution of the supports of sparse coefficients. The coefficient size in the signal model of [24] is independent and identically distributed. We extend these results by including non-homogeneous distributions of the coefficient amplitudes and obtain a generalisation of [24].

2. Notation and setting

Since this master thesis is a generalisation of [24], we follow their notation and setting. Let $A \in \mathbb{R}^{d\times K}$, $B \in \mathbb{R}^{K\times m}$ and A_k as well as A^k be the k-th column and row of A respectively. We denote by A^* the transpose of A. We will use the following operator norms for $1 \leq p, q, r \leq \infty$

$$\|A\|_{p,q} := \max_{\|x\|_q=1} \|Ax\|_p.$$

Note that $\|AB\|_{p,q} \leq \|A\|_{q,r}\|B\|_{r,p}$ and $\|Ax\|_q \leq \|A\|_{q,p}\|x\|_p$. Here, we mainly use

$$\|A\| := \|A\|_{2,2},$$

for the largest absolute singular values of A. Sometimes we also employ

$$\|A\|_{2,1} = \max_{k\in\{1,\dots,K\}} \|A_k\|_2 \quad \text{and} \quad \|A\|_{\infty,2} = \max_{k\in\{1,\dots,d\}} \|A^k\|_2,$$

which denote the maximal ℓ_2 norm of a column resp. row of A. Throughout this thesis we often apply the following notation

$$\underline{v} := \min_i |v_i| \quad \text{and} \quad \overline{v} := \max_i |v_i|, \tag{5}$$

where $v \in \mathbb{R}^K$. We denote by $D_v = \operatorname{diag}(v) \in \mathbb{R}^{K\times K}$ the corresponding square diagonal matrix and often abbreviate $D_{v\cdot w} := D_v \cdot D_w$ for $w \in \mathbb{R}^K$. For the so called support $I \subseteq \mathbb{K} := \{1, \dots, K\}$ we let A_I be the submatrix with columns indexed by I and $A_{I,I}$ the submatrix with rows and columns indexed by I. In this context we often use the zero-padding operator $R_I := (\mathbb{I}_I)^* \in \mathbb{R}^{|I|\times K}$, meaning

$$A_I = AR_I^*. \tag{6}$$

The zero-padding operator enables us to embed the matrix $A_I \in \mathbb{R}^{d\times|I|}$ into $\mathbb{R}^{d\times K}$ via $A_I R_I \in \mathbb{R}^{d\times K}$. We follow the convention that subscripts have higher priority than transposition, e.g. $A_I^* = (A_I)^*$. Furthermore let $\mathbf{1}_I \in \mathbb{R}^K$ be the vector with entry 1 on position i, if $i \in I$ and entry 0 otherwise. We use $\odot$ to denote the Hadamard Product or pointwise product of two matrices or vectors of the same dimension.
As already mentioned in the introduction $\Phi \in \mathbb{R}^{d\times K}$ is the generating dictionary, which we want to recover with the dictionary learning algorithms (MOD, ODL) which have as input the current guess $\Psi \in \mathbb{R}^{d\times K}$. We denote the coefficient matrix by X. For any permutation matrix P we get

$$Y = \Phi X \Rightarrow Y = (\Phi P)(PX). \tag{7}$$

So we can assume that Ψ is ordered such that $\max_j |\langle\phi_i,\psi_j\rangle| = |\langle\phi_i,\psi_i\rangle|$. In a similar way we have a sign ambiguity, since $Y = \Phi X$ implies $Y = (\Phi D)(DX)$ for $D_{ii} = \pm 1$. Hence wlog we can assume that Ψ is signed, such that $\alpha_i := \langle\phi_i,\psi_i\rangle \geq 0$ for all $i \in [K]$. We define the ℓ_2-distance between dictionary elements as

$$\varepsilon(\Psi,\Phi) := \|\Psi-\Phi\|_{2,1} = \max_i \|\psi_i - \phi_i\|_2 \quad \Leftrightarrow \quad \varepsilon(\Psi,\Phi)^2 = 2 - 2\underline{\alpha}, \tag{8}$$

which we frequently abbreviate as ε instead of $\varepsilon(\Psi,\Phi)$. We also abbreviate $Z := \Psi - \Phi$, which denotes the difference matrix between the generating dictionary Φ and the current guess Ψ. Given two vectors with positive entries $\beta,\pi \in \mathbb{R}^K$, we define the distance $\delta(\Psi,\Phi)$ between Φ and Ψ as

$$\delta(\Psi,\Phi) := \max\left\{\|(\Psi-\Phi)D_{\sqrt{\pi\cdot\beta}}\|\,\underline{\beta}^{-1/2}, \|(\Psi-\Phi)D_{\sqrt{\beta}}\|_{2,1}\underline{\beta}^{-1/2}\right\} \tag{9}$$

$$= \max\left\{\|ZD_{\sqrt{\pi\cdot\beta}}\|\,\underline{\beta}^{-1/2}, \|ZD_{\sqrt{\beta}}\|_{2,1}\underline{\beta}^{-1/2}\right\}. \tag{10}$$

The concrete choice for $\beta,\pi \in \mathbb{R}^K$ will be clarified in the signal model in (15), which we discuss in the next chapter. Further, if it is clear from context we will abbreviate $\delta(\Psi,\Phi)$ as δ. Moreover we denote the maximal absolute inner product between two non-corresponding atoms as the cross-coherence $\mu(\Psi,\Phi) := \max_{i\neq j}|\langle\psi_i,\phi_j\rangle|$ and a scaled version as $\mu_{\sqrt{\beta}}(\Psi,\Phi) := \max_{i\neq j}|\langle\psi_i,\phi_j\rangle \cdot \beta_j^{1/2}|$. We abbreviate $\mu_{\sqrt{\beta}}(\Phi,\Phi)$ as $\mu_{\sqrt{\beta}}(\Phi)$.

3. Probabilistic model

In order to simulate in a realistic way, how the non-zero coefficients are chosen, we want them to follow a non-uniform distribution, which is inspired for instance by [19]. Hence we define our probabilistic model for the sparse supports by the following two definitions similar to [24, Definition 1 and 2].

Definition 1 (Poisson and rejective sampling [24]) *Let δ_k denote a sequence of K independent Bernoulli $0-1$ random variables with expectations $0 \leq p_k \leq 1$ such that $\sum_{k=1}^K p_k = S$ and denote by $\mathbb{P}_B$ the probability measure of the corresponding Poisson sampling model. We say the support I follows the Poisson sampling model, if $I := \{k \mid \delta_k = 1\}$ and each support $I \subseteq \mathbb{K}$ is chosen with probability*

$$\mathbb{P}_B(I) = \prod_{i\in I} p_i \prod_{j\notin I}(1-p_j). \tag{11}$$

We say our support I follows the rejective sampling model, if each support $I \subseteq \mathbb{K}$ is chosen with probability

$$\mathbb{P}_S(I) := \mathbb{P}_B(I \mid |I| = S). \tag{12}$$

If it is clear from the context, we write $\mathbb{P}(I)$ instead of $\mathbb{P}_S(I)$.

As has already been pointed out in [24] the Poisson sampling has the major comfort that the probabilities of the atoms appearing in the support are independent from each other, but only has on average S-sparse supports. Unfortunately our requirement on the model is that the support is exactly S-sparse. Therefore we use the second model which satisfies this condition and can be related to the Poisson sampling using [23], restated in Theorem 14. This proves to be quite valuable, because we can typically reduce estimates for rejective sampling to estimates for Poisson sampling. We continue by defining our signal model based on rejective sampling.

Definition 2 (Signal model) *Let $\Phi = (\phi_1, \dots, \phi_K) \in \mathbb{R}^{d \times K}$ be the generating dictionary with K normalized atoms $\phi_i \in \mathbb{R}^d$. Additionally we choose the support $I = \{i_1, \dots i_S\} \subseteq \mathbb{K}$ according to the rejective model* (12) *with parameters $p_1, \dots, p_K$ such that $\sum_{i=1}^K p_i = S$ and $0 < p_k \leq 1/6$. Let the signals be modelled as*

$$y = \Phi_I x_I = \sum_{i \in I} \phi_i x_i, \quad x_i = c_i \sigma_i, \tag{13}$$

where the coefficients of the sequence $c = (c_i)_i \in \mathbb{R}^K$ are independent, bounded random variables c_i with $0 < c_{\min} \leq c_i \leq c_{\max} \leq 1$ and the sign sequence $\sigma \in \{-1, 1\}^K$ is a Rademacher sequence, i.e. its entries σ_i are i.i.d with $\mathbb{P}(\sigma_i = \pm 1) = 1/2$. Supports, coefficients and signs are modeled as independent and can be written as

$$x = \mathbf{1}_I \odot c \odot \sigma. \tag{14}$$

We place emphasis on the coefficient sequence c, which does not have to be i.i.d as in [24, Definition 2] and therefore is more general. To characterise our model we define the vectors $\beta, \pi \in \mathbb{R}^K$ via

$$\beta_i := \mathbb{E}[c_i^2] \quad \text{and} \quad \pi_i := \mathbb{P}(i \in I), \tag{15}$$

and the square diagonal matrices $D_{\sqrt{\beta}}, D_{\sqrt{\pi}} \in \mathbb{R}^{K \times K}$ as

$$D_{\sqrt{\beta}} := \operatorname{diag}((\sqrt{\beta_i})_i) \quad \text{and} \quad D_{\sqrt{\pi}} := \operatorname{diag}((\sqrt{\pi_i})_i),$$

which we use excessively throughout this master thesis. Sometimes we also employ the following notation for a matrix M

$$\dot{M} := M D_{\sqrt{\beta}} \quad \text{and} \quad \ddot{M} := D_{\sqrt{\beta}} M D_{\sqrt{\beta}},$$

in order to improve the readability of some calculations. With the help of our signal model we can now sketch why the output dictionary of one step of both the algorithms should be close to the generating dictionary Φ.
The update step before normalisation can be written as

$$\begin{aligned} \text{MOD:} \quad & Y\hat{X}^*(\hat{X}\hat{X}^*)^{-1} = \Phi X \hat{X}^*(\hat{X}\hat{X}^*)^{-1}, \\ \text{ODL:} \quad & \frac{1}{N}\left[Y\hat{X}^* - \Psi\hat{X}\hat{X}^* + \Psi \operatorname{diag}(\hat{X}\hat{X}^*)\right] = \frac{1}{N}\left[\Phi X\hat{X}^* - \Psi\hat{X}\hat{X}^* + \Psi \operatorname{diag}(\hat{X}\hat{X}^*)\right], \end{aligned}$$

where we use that $Y = \Phi X$. We define two averages of random matrices as

$$A := \frac{1}{N} X\hat{X}^* = \frac{1}{N}\sum_{n=1}^N x_n \hat{x}_n^* \quad \text{and} \quad B := \frac{1}{N}\hat{X}\hat{X}^* = \frac{1}{N}\sum_{n=1}^N \hat{x}_n \hat{x}_n^*, \tag{16}$$

which lead to the following expressions of the update step

$$\begin{aligned} \text{MOD:} \quad & \Phi A B^{-1}, \\ \text{ODL:} \quad & \Phi A - \Psi[B - \operatorname{diag}(B)]. \end{aligned}$$

Since we know the empirical estimators A and B are approximately $\mathbb{E}[x\hat{x}]$ and $\mathbb{E}[\hat{x}\hat{x}]$ resp., we have a closer look at these expectations. We assume that thresholding succeeds to find the correct support, meaning $\hat{I} = I$. Note that by using the zero-padding operator R_I^* we obtain

$$x\hat{x}^* = x(R_I^* \Psi_I^\dagger y)^* = xx^*\Phi^*(\Psi_I^\dagger)^* R_I$$

We assume that Ψ_I is well conditioned, which means that $\Psi_I^*\Psi_I \approx \mathbb{I}$ and therefore implies that $\Psi_I^\dagger \approx \Psi_I^*$. So we obtain

$$x\hat{x}^* \approx xx^*\Phi^*\Psi_I R_I = xx^*\Phi^*\Psi R_I^* R_I = xx^*\Phi^*\Psi \operatorname{diag}(\mathbf{1}_I).$$

According to our model we have $x = \mathbf{1}_I \odot c \odot \sigma$, where I, c, σ are independent, so

$$\mathbb{E}_{c,\sigma}[xx^*] = \mathbb{E}_c[cc^*] \odot \mathbb{E}_\sigma[\sigma\sigma^*] \odot (\mathbf{1}_I \mathbf{1}_I^*) = D_\beta \operatorname{diag}(\mathbf{1}_I) = \operatorname{diag}(\mathbf{1}_I) D_\beta,$$

and thus

$$\mathbb{E}[x\hat{x}^*] = \mathbb{E}_{I,c,\sigma}[x\hat{x}^*] \approx \mathbb{E}_I[\operatorname{diag}(\mathbf{1}_I) D_\beta \Phi^*\Psi \operatorname{diag}(\mathbf{1}_I)] = (D_\beta \Phi^*\Psi) \odot \mathbb{E}_I[\mathbf{1}_I \mathbf{1}_I^*].$$

The matrix $\mathbb{E}_I[\mathbf{1}_I \mathbf{1}_I^*]$ is diagonally dominant, which can be seen since

$$(\mathbb{E}_I[\mathbf{1}_I \mathbf{1}_I^*])_{ij} = \mathbb{P}(i, j \in I) \approx \mathbb{P}(i \in I) \cdot \mathbb{P}(j \in I) \ll \mathbb{P}(i \in I) = (\mathbb{E}_I[\mathbf{1}_I \mathbf{1}_I^*])_{ii},$$

meaning

$$\mathbb{E}[\mathbf{1}_I \mathbf{1}_I^*] \approx D_\pi + \pi\pi^* \approx D_\pi.$$

After analysing B in a similar manner and using that $D_\alpha = \operatorname{diag}(\Phi^*\Psi)$ we get

$$A \approx \mathbb{E}[x\hat{x}^*] \approx (D_\beta \Phi^*\Psi) \odot D_\pi = D_{\pi\cdot\alpha\beta} \quad \text{and} \quad B \approx \mathbb{E}[\hat{x}\hat{x}^*] \approx D_{\pi\cdot\alpha^2\beta}.$$

Substituting this into the update steps of MOD and ODL before normalisation yields

$$\Phi A B^{-1} \approx \Phi D_\alpha^{-1} \quad \text{and} \quad \Phi A - \Psi[B - \operatorname{diag}(B)] \approx \Phi D_{\pi\cdot\alpha\cdot\beta}. \tag{17}$$

This rough analysis provides insight into the MOD and ODL algorithm and suggests that both algorithms should converge to the generating dictionary. The main job is now to quantify the errors and show that they are small which proves our main result.

4. Main result and proof

Our main goal is to prove the following theorem, which is a generalization of Theorem 3 in [24].

Theorem 3 *We assume that the given signals conform to the signal model in Definition 2. Additionally we define*

$$\underline{\alpha} := \min_k |\langle \psi_k, \phi_k \rangle| = 1 - \varepsilon^2/2, \quad \gamma := \frac{c_{\min}}{c_{\max}} \quad \text{and} \quad \rho = 2\kappa^2 S^2 \gamma^{-2} \underline{\alpha}^{-2} \underline{\pi}^{-3/2}, \tag{18}$$

where $\kappa^2 \geq 2$. *Let* δ_* *be the desired recovery accuracy, chosen such that for two universal constants* C, n *with* $C \leq 42$ *and* $n \leq 130$, δ_* *satisfies* $\delta_* C \log(nK\rho/\delta_*) \leq \delta$. *Abbreviate* $\nu = 1/\sqrt{\log(nK\rho/\delta_*)} < 1$ *meaning* $\delta_* \leq \gamma\nu^2/C$. *If the atom-wise distance* $\varepsilon = \varepsilon(\Psi, \Phi)$ *of the current guess* Ψ *to the generating dictionary* Φ *satisfies*

$$\max\{\nu \|\Phi D_{\sqrt{\pi \cdot \beta}}\|, \mu_{\sqrt{\beta}}(\Phi)\} \leq \left(1 - \frac{\varepsilon^2}{2}\right) \cdot \frac{\gamma \underline{\beta}^{1/2}}{4C \log(nK\rho/\delta_*)} \tag{19}$$

and the current guess Ψ *additionally satisfies either*

$$\max\{\nu \|\Psi D_{\sqrt{\pi \cdot \beta}}\|, \mu_{\sqrt{\beta}}(\Psi), \mu_{\sqrt{\beta}}(\Psi, \Phi)\} \leq \left(1 - \frac{\varepsilon^2}{2}\right) \cdot \frac{\gamma \underline{\beta}^{1/2}}{4C \log(nK\rho/\delta_*)} \tag{20}$$

$$\text{or} \quad \delta(\Psi, \Phi) \leq \frac{\gamma}{C \log(nK\rho/\delta_*)} =: \delta_\circ, \tag{21}$$

then the updated and normalised dictionary $\hat{\Psi}$, *which is output by ODL or MOD, satisfies*

$$\delta(\hat{\Psi}, \Phi) \leq \frac{1}{2} \cdot (\delta_*/2 + \min\{\delta_\circ, \delta(\Psi, \Phi)\}) =: \frac{\Delta}{2}, \tag{22}$$

except with probability

$$60K \exp\left(-\frac{N(\Delta/16)^2}{2\rho^2 + \rho\Delta/16}\right). \tag{23}$$

Before we proceed with the proof of the theorem, we provide an explanation of what we have achieved and compare it with [24]. In Theorem 3 we have generalised the result of [24], since for a coefficient sequence c that is i.i.d. as in [24], we obtain the same result as in [24]. This can be seen, since in the i.i.d. case, the diagonal matrix $D_{\sqrt{\beta}}$ can be simplified to a constant times the identity, which means that $\|AD_{\sqrt{\beta}}\| = \underline{\beta}^{1/2} \cdot \|A\|$, for any matrix A. This means that the factor $\underline{\beta}^{1/2}$ in (19) and (20) cancels out for all norms and also for $\mu_{\sqrt{\beta}}(\Psi, \Phi)$, $\mu_{\sqrt{\beta}}(\Phi)$ and $\mu_{\sqrt{\beta}}(\Psi)$, which we illustrate in detail for $\mu_{\sqrt{\beta}}(\Psi, \Phi)$:

$$\mu_{\sqrt{\beta}}(\Psi, \Phi) = \max_{i \neq j} |\langle \psi_i, \phi_j \rangle \cdot \beta_j^{1/2}| = \max_{i \neq j} |\langle \psi_i, \phi_j \rangle| \cdot \underline{\beta}^{1/2} \leq \underline{\alpha}\gamma\nu^2 \underline{\beta}^{1/2}/(4C).$$

The inequality implies $\mu(\Psi, \Phi) \leq \underline{\alpha}\gamma\nu^2/(4C)$, meaning the condition of [24]. In the second regime (21) it cancels out indirectly, because of our definition of δ, which is

$$\delta(\Psi, \Phi) := \max\left\{\|(\Psi - \Phi)D_{\sqrt{\pi \cdot \beta}}\| \, \underline{\beta}^{-1/2}, \|(\Psi - \Phi)D_{\sqrt{\beta}}\|_{2,1} \underline{\beta}^{-1/2}\right\}.$$

One important difference between this thesis and [24] is that D_β cannot be treated as a constant anymore. In [24] it was often used that for constant D_β we have $D_\beta^{\pm 1} M D_\beta^{\mp 1} = M$ for any square matrix, which is not true for β not being constant. So in order to deal with that we have to revise the proof strategy in [24] and rescale our matrices A, B from (16) as $D_{\sqrt{\beta}}^{-1} A D_{\sqrt{\beta}}^{-1}$ rather than AD_β^{-1} and similar for B.
Proof [Proof of Theorem 3] We define

$$\bar{B} := (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\, B\, (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\,, \tag{24}$$

$$\bar{A} := (D_{\sqrt{\pi\cdot\beta}})^{-1} A (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}, \tag{25}$$

and summarize the results of Lemma 6 to 9, which hold true except with the probability stated in (23): First by Lemma 6 and 7 we obtain

$$\|\bar{A} - \mathbb{I}\|_{2,2} \le \frac{\Delta}{8} \quad \text{and} \quad \|\bar{B} - \mathbb{I}\|_{2,2} \le \frac{\Delta}{4\gamma}, \tag{26}$$

Second by Lemma 8 and 9 we have for all $\ell \in \{1, \cdots, K\}$ and $\Lambda := \max\left\{\frac{\gamma\underline{\alpha}\nu\underline{\beta}^{1/2}}{4C}, \|\Psi D_{\sqrt{\pi\cdot\beta}}\|\right\}$ that

$$\|\Phi A\, (D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1}\, e_\ell - \phi_\ell \beta_\ell^{-1/2}\|_2 \le \frac{\Delta}{8} \quad \text{and} \quad \Lambda \cdot \|\mathbb{I}_\ell \bar{B} e_\ell \pi_\ell^{-1/2}\|_2 \le \underline{\beta}^{1/2} \cdot \frac{\Delta}{8}. \tag{27}$$

Inspired by 17 we define a scaled version of the updated dictionary $\bar{\Psi}$ as $\hat{\Psi} D_{\pi\cdot\beta\cdot\alpha}$ for ODL and as $\hat{\Psi} D_\alpha$ for MOD. We prove that $\bar{\Psi}$ contracts to the generating dictionary Φ and show that this stays true even after normalising $\bar{\Psi}$, which leads to contraction of $\hat{\Psi}$ to Φ. To put it more explicitly we start by proving the following bounds for s_ℓ either being 1 for ODL or $e_\ell^* \bar{B}^{-1} e_\ell$ for MOD:

$$\left\|(\bar{\Psi} - \Phi) D_{\sqrt{\pi\cdot\beta}}\right\| \cdot \underline{\beta}^{-1/2} \le \Delta/4 \quad \text{and} \quad \max_\ell \|(\bar{\psi}_\ell - s_\ell \phi_\ell) \beta_\ell^{-1/2}\| \cdot \underline{\beta}^{-1/2} \le \Delta/3. \tag{28}$$

ODL: Since in (17) we have seen that $\Phi A - \Psi B + \Psi\,\mathrm{diag}(B) \approx \Phi\,(D_{\pi\cdot\beta\cdot\alpha})$ we define

$$\bar{\Psi} := (\Phi A - \Psi B + \Psi\,\mathrm{diag}(B))\,(D_{\pi\cdot\beta\cdot\alpha})^{-1} \tag{29}$$

In order to use the bounds from (26) and (27) we write

$$\begin{aligned}
(\bar{\Psi} - \Phi) D_{\sqrt{\pi\cdot\beta}} &= \Phi A (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \Psi[B - \mathrm{diag}(B)]\,(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \Phi D_{\sqrt{\pi\cdot\beta}} \\
&= \Phi D_{\sqrt{\pi\cdot\beta}} \cdot \left[(D_{\sqrt{\pi\cdot\beta}})^{-1} A (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \mathbb{I}\right] \\
&\qquad - \Psi D_{\sqrt{\pi\cdot\beta}\cdot\alpha} \cdot [(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} B (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \mathbb{I}] \\
&\qquad\qquad + \Psi D_{\sqrt{\pi\cdot\beta}\cdot\alpha} \cdot [(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\,\mathrm{diag}(B) (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \mathbb{I}] \\
&= \Phi D_{\sqrt{\pi\cdot\beta}} \cdot \left[\bar{A} - \mathbb{I}\right] - \Psi D_{\sqrt{\pi\cdot\beta}} \cdot D_\alpha \cdot [\bar{B} - \mathbb{I}] + \Psi D_{\sqrt{\pi\cdot\beta}} \cdot D_\alpha \cdot [\mathrm{diag}(\bar{B}) - \mathbb{I}].
\end{aligned} \tag{30}$$

Before we continue with estimating the operator norm of the expression above we show that $\|\Psi D_{\sqrt{\pi\cdot\beta}}\| \le 2\gamma\nu\underline{\beta}^{1/2}/C$ in both regimes. For $\delta > \delta_\circ$ the assumption (20) can be used, while in the second regime, $\delta \le \delta_\circ = \gamma\nu^2/C$, we rewrite the term and use assumption (19) to get

$$\begin{aligned}
\|\Psi D_{\sqrt{\pi\cdot\beta}}\| \le \|\Phi D_{\sqrt{\pi\cdot\beta}}\| + \|(\Psi - \Phi) D_{\sqrt{\pi\cdot\beta}}\| &\le \|\Phi D_{\sqrt{\pi\cdot\beta}}\| + \delta\underline{\beta}^{1/2} \\
&\le \|\Phi D_{\sqrt{\pi\cdot\beta}}\| + \delta_\circ\underline{\beta}^{1/2} \\
&\le \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C) + \gamma\nu^2/C \cdot \underline{\beta}^{1/2} \le 2\gamma\nu\underline{\beta}^{1/2}/C.
\end{aligned}$$

h Since $\|\operatorname{diag}(\bar{B})-\mathbb{I}\| \le \|\bar{B}-\mathbb{I}\|$ and $\|D_\alpha\| \le 1$ we can use both bounds from (26) to obtain

$$\begin{aligned}\|(\bar{\Psi}-\Phi)D_{\sqrt{\pi\cdot\beta}}\| &\le \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\|\bar{A}-\mathbb{I}\| + 2\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|\cdot\|D_\alpha\|\cdot\|\bar{B}-I\| \\ &\le \frac{\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}}{4C}\cdot\frac{\underline{\alpha}\Delta}{8} + 2\cdot\frac{2\gamma\nu\underline{\beta}^{1/2}}{C}\cdot\frac{\gamma^{-1}\cdot\Delta}{4} \le \underline{\beta}^{1/2}\cdot\frac{\Delta}{4},\end{aligned} \tag{31}$$

which gets us contraction in the weighted operator norm. Next we want to show that each atom $\bar{\psi}_\ell$ of the scaled updated dictionary contracts to their partner atom ϕ_ℓ of the generating dictionary in the weighted ℓ_2-norm. Similar to before we write

$$\begin{aligned}(\bar{\psi}_\ell-\phi_\ell)\beta_\ell^{1/2} &= (\bar{\Psi}-\Phi)D_{\sqrt{\beta}}e_\ell \\ &= \Phi A(D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1}e_\ell - \Phi D_{\sqrt{\beta}}e_\ell - \Psi D_{\sqrt{\pi\cdot\beta}\cdot\alpha}[\bar{B}-\operatorname{diag}(\bar{B})]D_{\sqrt{\pi}}^{-1}e_\ell \\ &= [\Phi A(D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1}e_\ell - \Phi D_{\sqrt{\beta}}e_\ell] + \Psi D_{\sqrt{\pi\cdot\beta}\cdot\alpha}\cdot\mathbb{I}_{\ell^c}\bar{B}e_\ell\pi_\ell^{-1/2},\end{aligned} \tag{32}$$

where e_ℓ is the standard basis vector. Hence using that $\|\Psi D_{\sqrt{\beta}\cdot\alpha}\| \le \|\Psi D_{\sqrt{\beta}}\|$ and the bounds from (26) we obtain

$$\begin{aligned}&\|(\bar{\psi}_\ell-\phi_\ell)\beta_\ell^{1/2}\| \\ &\le \underbrace{\|\Phi A(D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1}e_\ell - \Phi D_{\sqrt{\beta}}e_\ell\|}_{\le\underline{\beta}^{1/2}\Delta/8} + \underbrace{\|\Psi D_{\sqrt{\pi\cdot\beta}\cdot\alpha}\|\cdot\|\mathbb{I}_{\ell^c}\bar{B}e_\ell\pi_\ell^{-1/2}\|}_{\le\underline{\beta}^{1/2}\Delta/8} \le \underline{\beta}^{1/2}\cdot\frac{\Delta}{4}.\end{aligned} \tag{33}$$

So we proved the desired bounds in (28) for the ODL algorithm with $s_\ell = 1$. Before we take care of the normalisation, we want to obtain the same bounds for MOD.
MOD: We recall that the dictionary update of MOD corresponds to
$\Phi X\hat{X}^*(\hat{X}\hat{X}^*)^{-1} = \Phi AB^{-1}$ as long as $\hat{X}\hat{X}^*$ is invertible. This is given due to the right inequality in (26) together with the Neumann-series, which we will verify below. Inspired by (17) we define our scaled version of the updated dictionary as

$$\bar{\Psi} := \Phi AB^{-1}D_\alpha$$

and decompose $\bar{\Psi}D_{\sqrt{\pi\cdot\beta}}$ as

$$\begin{aligned}\hat{\Psi}D_{\sqrt{\pi\cdot\beta}} &= \Phi AB^{-1}D_{\sqrt{\pi\cdot\beta}\cdot\alpha} = \Phi D_{\sqrt{\pi\cdot\beta}}\cdot((D_{\sqrt{\pi\cdot\beta}})^{-1}AB^{-1}D_{\sqrt{\pi\cdot\beta}\cdot\alpha}) \\ &= \Phi D_{\sqrt{\pi\cdot\beta}}\Bigg[(D_{\sqrt{\pi\cdot\beta}})^{-1}A\,(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} \\ &\qquad + (D_{\sqrt{\pi\cdot\beta}})^{-1}A\,(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\left(\left[(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}B\,(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\right]^{-1} - \mathbb{I}\right)\Bigg] \\ &= \Phi D_{\sqrt{\pi\cdot\beta}}\cdot\left(\bar{A}+\bar{A}\left(\bar{B}^{-1}-\mathbb{I}\right)\right).\end{aligned} \tag{34}$$

Indeed since by Lemma 7 we have $\|\bar{B}-I\| \le \gamma^{-1}\cdot\Delta/4$, the matrix $\bar{B}$ can be inverted by applying the Neumann-series $\bar{B}^{-1} = [\mathbb{I}-(\mathbb{I}-\bar{B})]^{-1} = \sum_{k\ge0}(\mathbb{I}-\bar{B})^k$ and bounded as

$$\|\bar{B}^{-1}\| = \|\sum_{k\ge0}(\mathbb{I}-\bar{B})^k\| \le \sum_{k\ge0}\|\mathbb{I}-\bar{B}\|^k \le \frac{1}{1-\gamma^{-1}\Delta/4}. \tag{35}$$

Further, we get

$$\|\bar{B}^{-1}-\mathbb{I}\| = \|\sum_{k\geq 1}(\mathbb{I}-\bar{B})^k\| \leq \sum_{k\geq 1}\|\bar{B}-\mathbb{I}\|^k \leq \frac{\gamma^{-1}\Delta}{4-\gamma^{-1}\Delta} = \frac{\Delta}{4\gamma-\Delta} \leq \frac{\Delta}{3\gamma}, \tag{36}$$

where the last inequality follows from the fact that $\delta_\circ = \gamma\nu^2/C$ and $\nu \leq 1/3$ and hence $\Delta \leq \frac{3}{2}\delta_\circ \leq \gamma$. Before we estimate $\|(\bar{\Psi}-\Phi)D_{\sqrt{\pi\cdot\beta}}\|$ we also bound the scaled version $\bar{A}$ using Lemma 6 or the first inequality in (26), which yields

$$\|\bar{A}\| \leq \|\bar{A}-\mathbb{I}+\mathbb{I}\| \leq \|\bar{A}-\mathbb{I}\| + \|I\| \leq \frac{\Delta}{8}+1. \tag{37}$$

Next we use these observations to show contraction of the scaled updated dictionary $\bar{\Psi}$ to the generating dictionary:

$$\begin{aligned}\|(\bar{\Psi}-\Phi)D_{\sqrt{\pi\cdot\beta}}\| &\leq \|\Phi D_{\sqrt{\pi\cdot\beta}}\| \cdot \left(\|\bar{A}-\mathbb{I}\| + \|\bar{A}\|\cdot\|\bar{B}^{-1}-\mathbb{I}\|\right)\\ &\leq \frac{\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}}{4C}\left(\frac{\Delta}{8}+\left(\frac{\Delta}{8}+1\right)\cdot\frac{\Delta}{3\gamma}\right) \leq \underline{\beta}^{1/2}\cdot\frac{\Delta}{4}.\end{aligned} \tag{38}$$

This shows that under the assumptions of the theorem, the weighted operator norm of the distance between the generating dictionary and the scaled update decreases in each iteration. We proceed with the atomwise ℓ_2-norm: We define $\mathbb{I} = e_\ell e_\ell^* + \mathbb{I}_{\ell^c}$ as well as $s_\ell := e_\ell^*\bar{B}^{-1}e_\ell$ and decompose $\bar{\psi}_\ell$ as

$$\begin{aligned}\bar{\psi}_\ell\beta_\ell^{1/2} &= \bar{\Psi}D_{\sqrt{\beta}}e_\ell = \Phi AB^{-1}D_{\sqrt{\pi\cdot\beta}\cdot\alpha}e_\ell\pi_\ell^{-1/2} = \Phi A(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\bar{B}^{-1}e_\ell\pi_\ell^{-1/2}\\ &= \Phi A(D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1}e_\ell\cdot s_\ell + \Phi D_{\sqrt{\pi\cdot\beta}}\cdot(D_{\sqrt{\pi\cdot\beta}})^{-1}A(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\cdot\mathbb{I}_{\ell^c}\bar{B}^{-1}e_\ell\pi_\ell^{-1/2},\end{aligned} \tag{39}$$

which yields

$$\begin{aligned}&\|(\bar{\psi}_\ell - s_\ell\phi_\ell)\beta_\ell^{1/2}\|\\ &\quad\leq |s_\ell|\cdot\|\Phi A(D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1}e_\ell - \Phi D_{\sqrt{\beta}}e_\ell\| + \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\|\bar{A}\|\cdot\|\mathbb{I}_{\ell^c}\bar{B}^{-1}e_\ell\pi_\ell^{-1/2}\|.\end{aligned} \tag{40}$$

In order to establish a suitable bound for the atomwise difference, we still need to bound the first and last term on the right-hand side above. It is easy to see that $|s_\ell| \leq \|\bar{B}^{-1}\|$. Combining the bound for $\|\bar{B}^{-1}\|$ from (35) and Lemma 8 we obtain

$$|s_\ell|\cdot\|\Phi A\left(D_{\sqrt{\beta}\cdot\pi\cdot\alpha}\right)^{-1}e_\ell - \Phi D_{\sqrt{\beta}}e_\ell\| \leq (1-\gamma^{-1}\Delta/4)^{-1}\cdot\underline{\beta}^{1/2}\cdot\Delta/8. \tag{41}$$

For the second term we need to have a closer look at $\|\mathbb{I}_{\ell^c}\bar{B}^{-1}e_\ell\pi_\ell^{-1/2}\|$. Since $\bar{B}^{-1} = \mathbb{I}+\bar{B}^{-1}(\mathbb{I}-\bar{B})$ and $\mathbb{I}_{\ell^c}e_\ell = 0$ we get

$$\begin{aligned}\|\mathbb{I}_{\ell^c}\bar{B}^{-1}e_\ell\pi_\ell^{-1/2}\| &= \|\mathbb{I}_{\ell^c}\bar{B}^{-1}(\mathbb{I}-\bar{B})e_\ell\pi_\ell^{-1/2}\| = \|\mathbb{I}_{\ell^c}\bar{B}^{-1}\left(e_\ell e_\ell^*+\mathbb{I}_{\ell^c}\right)(\mathbb{I}-\bar{B})e_\ell\pi_\ell^{-1/2}\|\\ &\leq \|\mathbb{I}_{\ell^c}\bar{B}^{-1}e_\ell\pi_\ell^{-1/2}\|\cdot\|\mathbb{I}-\bar{B}\| + \|\bar{B}^{-1}\|\cdot\|\mathbb{I}_{\ell^c}\bar{B}e_\ell\pi_\ell^{-1/2}\|.\end{aligned}$$

Restructuring this inequality and applying the bound from Lemma 7 leads to

$$\|\mathbb{I}_{\ell^c}\bar{B}^{-1}e_\ell\pi_\ell^{-1/2}\| \leq \frac{\|\bar{B}^{-1}\|}{1-\|\mathbb{I}-\bar{B}\|}\cdot\|\mathbb{I}_{\ell^c}\bar{B}e_\ell\pi_\ell^{-1/2}\| \leq \frac{1}{(1-\gamma^{-1}\Delta/4)^2}\cdot\|\mathbb{I}_{\ell^c}\bar{B}e_\ell\pi_\ell^{-1/2}\|. \tag{42}$$

Note that $\gamma^{-1}\Delta \le \gamma^{-1} \cdot \frac{3}{2}\delta_\circ \le \gamma^{-1} \cdot \frac{3}{2}\frac{\gamma\nu^2}{C} \le \frac{1}{4}$ which implies $\frac{1}{1-\gamma^{-1}\Delta/4} \le \frac{16}{15}$. Putting the results of (41) and (42) together and plugging it into (40) yields

$$\begin{aligned}\|(\bar\psi_\ell - s_\ell\phi_\ell)\beta_\ell^{1/2}\| &\le \frac{1}{1-\gamma^{-1}\Delta/4} \cdot \frac{\underline{\beta}^{1/2}\Delta}{8} \\ &\qquad + \frac{1}{(1-\gamma^{-1}\Delta/4)^2} \cdot \left(\frac{\Delta}{8}+1\right) \cdot \underbrace{\frac{\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}}{4C} \cdot \|\mathbb{I}_{\ell^c}\bar B e_\ell \pi_\ell^{-\frac{1}{2}}\|}_{\le \underline{\beta}^{1/2}\cdot\Delta/8 \quad (27)} \\ &\le \left[\frac{16}{15} + \frac{16^2}{15^2}\cdot\frac{5}{4}\right] \cdot \frac{\underline{\beta}^{1/2}\Delta}{8} \le \underline{\beta}^{1/2}\cdot\frac{\Delta}{3}. \end{aligned} \tag{43}$$

Note that $|s_\ell - 1| \le \|\bar B^{-1} - \mathbb{I}\| \le \gamma^{-1}\cdot\Delta/3$ in MOD. Hence, in summary our results so far are the following, where $s_\ell = 1$ for the ODL algorithm and close to 1 for MOD:

$$\big\|(\bar\Psi - \Phi)D_{\sqrt{\pi\cdot\beta}}\big\| \le \underline{\beta}^{1/2}\cdot\Delta/4 \quad \text{and} \quad \max_\ell \big\|(\bar\psi_\ell - s_\ell\phi_\ell)\beta_\ell\big\| \le \underline{\beta}^{1/2}\cdot\Delta/3.$$

Normalisation: In the last step of the proof, we show that normalizing the scaled versions of the updated dictionary does not affect the convergence. We define $F := \operatorname{diag}\left(\|\bar\psi_i\|_2\right)^{-1}$ to be the normalization matrix, such that $\hat\Psi := \bar\Psi F$. Since $\|\phi_\ell\| = 1$ and $\gamma \le 1$ we obtain

$$\begin{aligned}\big|\|\bar\psi_\ell\| - 1\big| \le \|\bar\psi_\ell - \phi_\ell\| &\le \|(\bar\psi_\ell - s_\ell\phi_\ell)\beta_\ell^{1/2}\| \cdot \beta_\ell^{-1/2} + \|(s_\ell - 1)\phi_\ell\| \\ &\le (\underline{\beta}^{1/2}\cdot\Delta/3)\cdot\beta_\ell^{-1/2} + \gamma^{-1}\cdot\Delta/3 \le 2/3\cdot\gamma^{-1}\Delta \le \gamma^{-1}\Delta.\end{aligned}$$

Therefore we get $\|F\| \le (1-\gamma^{-1}\Delta)^{-1}$ and $\|\mathbb{I} - F\| \le \gamma^{-1}\Delta\cdot(1-\gamma^{-1}\Delta)^{-1}$. With these observations we can show that the updated dictionary $\hat\Psi$ contracts towards the generating dictionary Φ in the weighted operator norm: Note that by assumption $\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \le \gamma\nu\underline{\beta}^{1/2}/(4C) \le \gamma\underline{\beta}^{1/2}/(12C)$ and $1 - \gamma^{-1}\Delta \ge 1 - \frac{3\delta_\circ}{2\gamma} \ge 1 - \frac{1}{6C}$, which leads to

$$\begin{aligned}\|(\hat\Psi - \Phi)D_{\sqrt{\pi\cdot\beta}}\| &\le \|(\bar\Psi - \Phi)D_{\sqrt{\pi\cdot\beta}}\|\cdot\|F\| + \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\|(\mathbb{I} - F)\| \\ &\le \underline{\beta}^{1/2}\cdot\frac{\Delta}{4}\cdot\frac{1}{1-\gamma^{-1}\Delta} + \frac{\gamma\underline{\beta}^{1/2}}{12C}\cdot\frac{\gamma^{-1}\Delta}{1-\gamma^{-1}\Delta} \le \underline{\beta}^{1/2}\cdot\frac{\Delta}{2}.\end{aligned} \tag{44}$$

For the ℓ_2-norm we use the help of Lemma B.10 from [27]. If $\|\phi_\ell\| = 1$ and $\|\psi_\ell - s_\ell\phi_\ell\| \le t$, then $\hat\psi_\ell = \psi_\ell/\|\psi_\ell\|$ satisfies

$$\|\hat\psi_\ell - \phi_\ell\|^2 \le 2 - 2\sqrt{1 - \frac{t^2}{s_\ell^2}} \le 2 - 2\left(1 - \frac{t^2/s_\ell^2}{2 - t^2/s_\ell^2}\right) = t^2\cdot\left(s_\ell^2 - \frac{t^2}{2}\right), \tag{45}$$

where the second inequality holds, if $t^2/s_\ell^2 \le 1$. In order to apply the lemma we note that

$$\|\bar\psi_\ell - s_\ell\phi_\ell\|\cdot\beta_\ell^{1/2} = \|(\bar\psi_\ell - s_\ell\phi_\ell)\beta_\ell^{1/2}\| \le \underline{\beta}^{1/2}\cdot\Delta/3 \tag{46}$$

Hence we set $t := \beta_\ell^{-1/2}\underline{\beta}^{1/2}\Delta/3 \le \Delta/3$ while $|1 - s_\ell| \le \Delta/3$ and therefore $s_\ell \ge 1 - \Delta/3 \ge t$, so we get

$$\|\hat\psi_\ell - \phi_\ell\| \le t\cdot\left(s_\ell^2 - \frac{t^2}{2}\right)^{-1/2} \le \frac{\underline{\beta}^{1/2}\Delta}{3\beta_\ell^{1/2}}\cdot\left(\left(1 - \frac{\Delta}{3}\right)^2 - \frac{\Delta^2}{18}\right)^{-1/2} \le \frac{\underline{\beta}^{1/2}}{\beta_\ell^{1/2}}\cdot\frac{\Delta}{2}. \tag{47}$$

This implies that $\|(\hat{\psi}_\ell - \phi_\ell)\beta_\ell^{1/2}\| = \|\hat{\psi}_\ell - \phi_\ell\| \cdot \beta_\ell^{1/2} \leq \underline{\beta}^{1/2} \cdot \Delta/2$. So we can conclude that

$$\|(\hat{\Psi} - \Phi)D_{\sqrt{\pi \cdot \beta}}\| \cdot \underline{\beta}^{-1/2} \leq \frac{\Delta}{2} \quad \text{and} \quad \|(\hat{\Psi} - \Phi)D_{\sqrt{\beta}}\|_{2,1} \cdot \underline{\beta}^{-1/2} \leq \frac{\Delta}{2}, \tag{48}$$

hence $\delta(\hat{\Psi}, \Phi) \leq \Delta/2$, which finishes the proof of Theorem 3. ■

In Theorem 3 we have sufficient conditions for the initial dictionary Ψ and the generating dictionary Φ, ensuring that one iteration of MOD and ODL algorithm in combination with thresholding as sparse approximation algorithm decreases the distance of the initial dictionary and the ground truth Φ. This is shown in a scaled operator norm as well as in the ℓ_2-norm. Applying Theorem 3 iteratively yields convergence of MOD and ODL up to the desired accuracy δ_*. The proof is mainly based on the four inequalities in (26) and (27), which we discuss in the following chapter.

5. Four bounds needed in the proof

Like in the chapter before, we closely follow the approach of [24] to prove the following lemmas. Since for each update step of the coefficients there is a need to project them onto specific submatrices derived from the current guess Ψ, we define the index sets $\mathcal{F}_\Phi, \mathcal{F}_\Psi$ where the random variables Φ_I, Ψ_I are well conditioned:

$$\mathcal{F}_\Phi := \{I \ : \ \|\Phi_I^*\Phi_I - \mathbb{I}\| \leq \vartheta\} \quad \text{and} \quad \mathcal{F}_\Psi := \{I \ : \ \|\Psi_I^*\Psi_I - \mathbb{I}\| \leq \vartheta\}, \tag{49}$$

where we set $\vartheta := 1/4$ for the rest of the chapter. Moreover we define the index set $\mathcal{F}_{\dot{Z}}$ where the norm of the random variable $\dot{Z}_I = \dot{\Psi}_I - \dot{\Phi}_I = (\Psi_I - \Phi_I)(D_{\sqrt{\beta}})_{I,I}$ is of a similar scale to δ:

$$\mathcal{F}_{\dot{Z}} := \left\{I \ : \ \|\dot{Z}_I\| \leq \delta \cdot \sqrt{2\underline{\beta}\log(nK\rho/\delta_*)}\right\}. \tag{50}$$

Imposing these conditions on the index set I together we set

$$\mathcal{G} := \mathcal{F}_\Phi \cap \mathcal{F}_\Psi \cap \mathcal{F}_{\dot{Z}}. \tag{51}$$

Next we want to define a set, where thresholding recovers the correct support in the vector $\hat{x}$. We recall that, thresholding finds the largest S entries and gathers them in the index set $\hat{I}$. The optimal coefficients then are

$$\hat{x}_{\hat{I}} = \Psi_{\hat{I}}^\dagger y.$$

We define the set which contains all index, sign and coefficient triplet, where thresholding finds the correct support, as

$$\mathcal{H} := \left\{(I, \sigma, c) \ : \ \hat{I} = I\right\}. \tag{52}$$

In the proofs of Lemmas 6 to 9 later on we will often restrict the dictionaries Ψ and Φ to the sets above and bound the probability that thresholding fails as well as the probability that the submatrices Φ_I, Ψ_I are ill-conditioned by Lemma 4.

Lemma 4 *Given the conditions stated in Theorem 3, we have*

$$[2\mathbb{P}(\mathcal{H}^c) + \mathbb{P}(\mathcal{G}^c)] \cdot \rho \leq \delta_*/32.$$

Proof For this proof we set $N := \Psi^*\Phi - \operatorname{diag}(\Psi^*\Phi)$ and recall that $\underline{\alpha} = \min_i |\langle \psi_i, \phi_i \rangle|$. Thresholding retrieves the complete support of a signal $y = \Phi_I x_I$, if

$$\|\Psi_{I^c}^* y\|_\infty < \|\Psi_I^* y\|_{\min}.$$

Therefore in order to find an upper bound on the probability that thresholding fails to recover the correct support that is on $\mathbb{P}(\mathcal{H}^c)$, we bound the following

$$\begin{aligned}
\mathbb{P}_y\left(\|\Psi_I^* y\|_{\min} < \|\Psi_{I^c}^* y\|_\infty\right) &= \mathbb{P}_y\left(\|\Psi_I^*\Phi_I x_I\|_{\min} < \|\Psi_{I^c}^*\Phi_I x_I\|_\infty\right) \\
&\leq \mathbb{P}_y\left(\|\operatorname{diag}(\Psi_I^*\Phi_I)x_I\|_{\min} - \|N_{I,I}x_I\|_\infty < \|\Psi_{I^c}^*\Phi_I\|_\infty\right) \\
&\leq \mathbb{P}_y\left(c_{\min}\|\operatorname{diag}(\Psi_I^*\Phi_I)\|_{\min} - \|N_{I,I}x_I\|_\infty < \|\Psi_{I^c}^*\Phi_I x_I\|_\infty\right) \\
&\leq \mathbb{P}_y\left(c_{\min} \cdot \underline{\alpha} < 2\|N_I x_I\|_\infty\right) \\
&\leq \mathbb{P}_y(2\|N_I x_I\|_\infty \geq c_{\min} \cdot \underline{\alpha} \mid \|N_I\|_{\infty,2} < \eta) \\
&\qquad + \mathbb{P}_S(\|N_I\|_{\infty,2} \geq \eta). \qquad (53)
\end{aligned}$$

For the first term we can use Hoeffding's inequality stated in [22, Lemma 23]. Since $\operatorname{sign}(x_k) = \sigma_k$ is a Rademacher sequence independent of I, this means that

$$\mathbb{P}_y\left(\|N_I x_I\|_\infty \geq c_{\min} \cdot \underline{\alpha}^2/2 \mid \|N_I\|_{\infty,2} < \eta\right) \leq 2K \exp\left(-\frac{1/4 \cdot c_{\min}^2 \underline{\alpha}^2}{2 \cdot \|N_I\|_{\infty,2}^2 \cdot \|x_I\|_\infty^2}\right). \qquad (54)$$

To our second term we first apply the Poissonisation trick, which can be found in [22, Lemma 7]. If $\mathbb{P}_B$ is the Poisson sampling model corresponding to $p_1, \cdots, p_K$, we obtain

$$\mathbb{P}\left(\|N_I\|_{\infty,2} \geq \eta\right) \leq 2\mathbb{P}_B\left(\|N_I\|_{\infty,2} \geq \eta\right).$$

Applying Lemma 21 from [22], which is in itself a consequence of the Chernoff inequality, we get

$$\mathbb{P}_B\left(\|N_I\|_{\infty,2} \geq \eta\right) \leq K\left(e\frac{\|ND_{\sqrt{p}}\|_{\infty,2}^2}{\eta^2}\right)^{\frac{\eta^2}{\mu(\Psi,\Phi)^2}}. \qquad (55)$$

Next we use $(1 - \|p\|_\infty) \cdot p_i \leq \pi_i$, from [23] resp. the condition from Theorem 14 in the Appendix A.3, to get $p_i \leq \frac{6}{5}\pi_i < 2\pi_i$ as well as

$$\|ND_{\sqrt{p}}\|_{\infty,2}^2 \leq 2\|ND_{\sqrt{\pi}}\|_{\infty,2}^2 = 2\|[\Psi^*\Phi - \operatorname{diag}(\Psi^*\Phi)]D_{\sqrt{\pi}}\|_{\infty,2}^2 \leq 2\|\Phi D_{\sqrt{\pi\cdot\beta}}\|^2 \cdot \|D_{\sqrt{\beta}}^{-1}\|^2.$$

Substituting (54) and (55) into (53) yields

$$\mathbb{P}_y\left(\|\Psi_I^* y\|_{\min} < \|\Psi_{I^c}^* y\|_\infty\right) \qquad (56)$$

$$\leq 2K\exp\left(-\frac{\gamma^2}{8\eta^2}\cdot\underline{\alpha}^2\right) + 2K\left(2e\frac{\|\Phi D_{\sqrt{\pi\cdot\beta}}\|_{\infty,2}^2 \cdot \underline{\beta}^{-1}}{\eta^2}\right)^{\frac{\eta^2}{\mu(\Psi,\Phi)^2}}. \qquad (57)$$

In order to determine η we recall the following abbreviations

$$\nu = \frac{1}{\sqrt{\log(nK\rho/\delta_*)}} \leq \frac{1}{3} \quad \text{and} \quad \delta_\circ = \frac{\gamma}{C\log(nK\rho/\delta_*)} = \frac{\gamma\nu^2}{C}, \tag{58}$$

where the upper bound for ν is valid for $n\rho\delta_*^{-1} \geq 130 \cdot 2 \cdot 42$. We choose $\eta := \underline{\alpha}\gamma\nu/4$ and substitute it into our first term in (57) which becomes $2K\delta_*^2/(n\rho K)^2$. For the second term we use the condition on the generating dictionary in (19), which is $\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \leq \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)$ and obtain $2e\|\Phi D_{\sqrt{\pi\cdot\beta}}\|_{\infty,2}^2\underline{\beta}^{-1}\eta^{-2} \leq e^{-2}$. For the exponent of the second term we separate the cases $\delta > \delta_\circ$ and $\delta \leq \delta_\circ$, since we use different conditions described in Theorem 3. While in the first regime, $\delta > \delta_\circ$, by (20), we get that

$$\mu(\Psi,\Phi) = \max_{i\neq j} |\langle\psi_i,\phi_j\rangle \cdot \beta_j^{1/2}\beta_j^{-1/2}| \leq \mu_{\sqrt{\beta}}(\Psi,\Phi)\cdot\underline{\beta}^{-1/2} \leq \underline{\alpha}\gamma\nu^2/(4C),$$

the second regime, $\delta \leq \delta_\circ$, requires further attention:

$$\begin{aligned}
\mu(\Psi,\Phi) = \mu(\Phi,\Psi) &\leq \mu_{\sqrt{\beta}}(\Phi,\Psi)\cdot\underline{\beta}^{-1/2} = \max_{i\neq j}|\langle\phi_i,\psi_j\rangle\beta_j^{1/2}|\cdot\underline{\beta}^{-1/2} \\
&\leq \max_{i\neq j}|\langle\phi_i,\phi_j\rangle\beta_j^{1/2}|\cdot\underline{\beta}^{-1/2} + \max_{i\neq j}|\langle\ \psi_i,\psi_j-\phi_j\rangle\beta_j^{1/2}|\cdot\underline{\beta}^{-1/2} \\
&\leq \mu_{\sqrt{\beta}}(\Phi)\cdot\underline{\beta}^{-1/2} + \max_j\|(\psi_j-\phi_j)\beta_j^{1/2}\|\cdot\underline{\beta}^{-1/2} \leq \mu_{\sqrt{\beta}}(\Phi)\cdot\underline{\beta}^{-1/2}+\delta \\
&\leq \underline{\alpha}\gamma\nu^2/(4C)+\delta,
\end{aligned}$$

where in the last inequality we used the condition on $\mu_{\sqrt{\beta}}(\Phi)$ in (21). Since $\delta \leq \delta_\circ \leq 1/C$ and $\underline{\alpha} = 1-\varepsilon^2/2 \geq 1-\delta^2/2 \geq (C-1)/C$, we get

$$\delta \leq \frac{\underline{\alpha}\gamma}{\underline{\alpha}C\log(nK\rho/\delta_*)} \leq \frac{\underline{\alpha}\gamma}{(C-1)\log(nK\rho/\delta_*)} = \frac{\underline{\alpha}\gamma\nu^2}{(C-1)},$$

and therefore $\mu(\Psi,\Phi) \leq 2\underline{\alpha}\gamma\nu^2/(C-1)$. Hence we found a lower bound for the exponent in both regimes

$$\frac{\eta^2}{\mu(\Psi,\Phi)^2} \geq \frac{\underline{\alpha}^2\gamma^2\nu^2(C-1)^2}{16\cdot 4\underline{\alpha}^2\gamma^2\nu^4} \geq \frac{(C-1)^2}{64\nu^2}, \tag{59}$$

which yields

$$\mathbb{P}(\mathcal{H}^c) \leq 2K\left(\frac{\delta_*}{n\rho K}\right)^2 + 2K\left(\frac{\delta_*}{n\rho K}\right)^{(C-1)^2/32} \leq 4K\left(\frac{\delta_*}{n\rho K}\right)^2. \tag{60}$$

In the second part of the proof we analyze the probability $\mathbb{P}(\mathcal{G}^c)$ which can be bounded by $\mathbb{P}(\mathcal{F}_{\dot{Z}}^c) + \mathbb{P}(\mathcal{F}_\Phi^c \cup \mathcal{F}_\Psi^c)$. We begin with $\mathbb{P}(\mathcal{F}_{\dot{Z}}^c)$ where we employ the Poissonisation trick as before, which yields

$$\mathbb{P}(\mathcal{F}_{\dot{Z}}^c) \leq 2\mathbb{P}_B(\mathcal{F}_{\dot{Z}}^c).$$

We continue by applying the matrix Chernoff inequality from 10. Therefore we note for $Z = (z_1,\ldots,z_K)$, where $z_i \in \mathbb{R}^d$, that

$$\dot{Z}_I\dot{Z}_I^* = (Z_I(D_\beta)_{I,I}Z_I^*) = \sum_{i\in I} z_i\beta_i z_i^*. \tag{61}$$

Recalling that $\delta = \max\left\{\|ZD_{\sqrt{\pi\cdot\beta}}\|\cdot\underline{\beta}^{-1/2}, \|Z\|_{2,1}\cdot\underline{\beta}^{-1/2}\right\}$ by Definition (10), we can estimate

$$\|z_i\beta_i z_i^*\| = \|\beta_i^{1/2}z_i^* z_i\beta_i^{1/2}\| = |\langle z_i\beta_i^{1/2}, z_i\beta_i^{1/2}\rangle| \leq \max_i \|(\psi_i - \phi_i)\beta_i^{1/2}\|^2 \leq \delta^2\underline{\beta} \quad \text{and}$$
$$\|\sum_{i\in I}\mathbb{E}_B[z_i\beta_i z_i^*]\| = \|\mathbb{E}_B[Z_I(D_\beta)_{I,I}Z_I^*]\| = \|ZD_{\sqrt{\beta}}\mathbb{E}_B[R_I^*R_I]D_{\sqrt{\beta}}Z^*\| = \|ZD_{p\cdot\beta}Z\|.$$

Using the bound $p_i \leq 2\pi_i$ we obtain

$$\begin{aligned}\mathbb{P}\left(\|Z_I(D_\beta)_{I,I}Z_I^*\| > t\right) &\leq 2\mathbb{P}_B\left(\|Z_I(D_\beta)_{I,I}Z_I^*\| > t\right)\\ &\leq 2K\left(\frac{e\|ZD_{p\cdot\beta}Z^*\|}{t}\right)^{t/(\delta^2\underline{\beta})} \leq 2K\left(\frac{2e\|ZD_{\pi\cdot\beta}Z^*\|}{t}\right)^{t/(\delta^2\underline{\beta})}\\ &\leq 2K\left(\frac{2e\|ZD_{\sqrt{\pi\cdot\beta}}\|^2}{t}\right)^{t/(\delta^2\underline{\beta})} \leq 2K\left(\frac{2e\delta^2\underline{\beta}}{t}\right)^{t/(\delta^2\underline{\beta})}. \end{aligned} \tag{62}$$

We choose $t = 2\delta^2\underline{\beta}\max\left\{e^2, \log(nK\rho/\delta_*)\right\} = 2\delta^2\underline{\beta}\log(nK\rho/\delta_*)$ which gives us the bound

$$\mathbb{P}\left(\mathcal{F}_{\dot{Z}^c}\right) = \mathbb{P}\left(\|\dot{Z}_I\| \geq \delta\cdot\sqrt{2\underline{\beta}\log(nK\rho/\delta_*)}\right) \leq 2K\left(\frac{\delta_*}{nK\rho}\right)^2$$

The last inequality holds if $\delta_*/(nK\rho) \leq 1/1619$, which is always satisfied since $n\rho\delta_*^{-1} \geq 130\cdot 2\cdot 42$ as before. For $\mathbb{P}(\mathcal{F}_\Phi^c \cup \mathcal{F}_\Psi^c)$ we can apply Theorem 12 taken from [22], and obtain

$$\mathbb{P}\left(\mathcal{F}_\Phi \cup \mathcal{F}_\Psi\right) \leq 512K\exp\left(-\min\left\{\frac{5\vartheta^2}{24e^2\|\Phi D_{\sqrt{\pi}}\|^2}, \frac{\vartheta}{2\mu(\Phi)}, \frac{5\vartheta^2}{24e^2\|\Psi D_{\sqrt{\pi}}\|^2}, \frac{\vartheta}{2\mu(\Psi)}\right\}\right).$$

Using the condition on the generating dictionary we obtain $\|\Phi D_{\sqrt{\pi}}\| \leq \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\|D_{\sqrt{\beta}}^{-1}\| \leq \gamma\underline{\alpha}\nu/(4C)$ and $\mu(\Phi) \leq \mu_{\sqrt{\beta}}(\Phi)\cdot\underline{\beta}^{-1/2} \leq \underline{\alpha}\gamma\nu^2/(4C)$. In the first regime, where $\delta > \delta_\circ$, we can employ a similar approach as for $\|\Psi D_{\sqrt{\pi}}\|$ and $\mu(\Psi)$. In the second regime, $\delta \leq \delta_\circ \leq \underline{\alpha}\gamma\nu^2/(C-1)$, we have to bound $\|\Psi D_{\sqrt{\pi}}\|$ and $\mu(\Psi)$ in a different manner

$$\begin{aligned}\|\Psi D_{\sqrt{\pi}}\| &\leq \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\underline{\beta}^{-1/2} + \|(\Psi-\Phi)D_{\sqrt{\pi\cdot\beta}}\|\cdot\underline{\beta}^{-1/2}\\ &\leq \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\underline{\beta}^{-1/2} + \delta \leq \underline{\alpha}\gamma\nu\cdot\left(\frac{1}{4C} + \frac{\nu}{C-1}\right) \leq \underline{\alpha}\gamma\nu/C,\end{aligned} \tag{63}$$

$$\begin{aligned}\mu(\Psi) &= \max_{i\neq j}|\langle\psi_i,\psi_j\rangle|\\ &\leq \max_{i\neq j}\left(|\langle\phi_i,\phi_j\rangle|\beta_j^{1/2} + |\langle\phi_i,\psi_j-\phi_j\rangle|\beta_j^{1/2} + \beta_i^{1/2}|\langle\psi_i-\phi_i,\psi_j\rangle|\right)\cdot\underline{\beta}^{-1/2}\\ &\leq \mu_{\sqrt{\beta}}(\Phi)\cdot\underline{\beta}^{-1/2} + 2\delta \leq \frac{9}{4}\cdot\nu^2/C.\end{aligned} \tag{64}$$

With these bounds and $\vartheta = 1/4$ we obtain $2\log(nK\rho/\delta_*)$ as a lower bound for the exponent, which results in $\mathbb{P}(\mathcal{F}_\Phi^c \cup \mathcal{F}_\Psi^c) \leq 512K\left(\delta_*/(n\rho K)\right)^2$. We recall that $n \geq 130$ and put all bounds together, to obtain our desired bound

$$\mathbb{P}\left(\mathcal{H}^c\right)\cdot 2\rho + \mathbb{P}\left(\mathcal{G}^c\right)\cdot\rho \leq 522K\rho\left(\frac{\delta_*}{n\rho K}\right)^2 \leq \frac{\delta_*}{32}.$$

■

We note that in Lemma 4 the bound on the probability that thresholding recovers the correct support in (57) is quite rough similar as in [24]. This can be optimized, for instance by defining

$$\hat{c}_{\min} := \min_i \frac{c_i}{\beta_i^{1/2}} \quad \text{and} \quad \hat{c}_{\max} := \max_i \frac{c_i}{\beta_i^{1/2}},$$

and therefore $\hat{\gamma} := \hat{c}_{\min}/\hat{c}_{\max}$. Unfortunately, we then lose the property that $(\bar{\beta}/\underline{\beta})^{1/2}$ can be bounded by $\hat{\gamma}^{-1}$ on which we rely in the Lemmas 6 to 9. For Lemma 6 to Lemma 9 we often use the following corollary from [23, 24].

Corollary 5 ([23],[24]) *Denote by $\mathbb{E}$ the expectation according to the rejective sampling probability with level S and by $\pi \in \mathbb{R}^K$ the first order inclusion probabilities of level S. Let $\mathcal{H}$ be a $K \times K$ matrix with zero diagonal, $W = (w_1 \dots, w_K)$ and $V = (v_1, \dots, v_K)$ a pair of $d \times K$ matrices and $\mathcal{G}$ a subset of all supports of size S, meaning $\mathcal{G} \subseteq \{I : |I| = S\}$. If $\|\pi\|_\infty \leq 1/3$, we have*

$$\|\mathbb{E}[D_{\sqrt{\pi}}^{-1} R_I^* \mathcal{H}_{I,I} R_I D_{\sqrt{\pi}}^{-1}]\| \leq 3 \cdot \|D_{\sqrt{\pi}} \mathcal{H} D_{\sqrt{\pi}}\|, \tag{a}$$

$$\|\mathbb{E}[D_{\sqrt{\pi}}^{-1} R_I^* \mathcal{H}_{I,I} \mathcal{H}_{I,I}^* R_I D_{\sqrt{\pi}}^{-1}]\| \leq \frac{9}{2} \cdot \|D_{\sqrt{\pi}} \mathcal{H} D_{\sqrt{\pi}}\|^2 + \frac{3}{2} \cdot \max_k \|e_k^* \mathcal{H} D_{\sqrt{\pi}}\|^2, \tag{b}$$

$$\|\mathbb{E}[W R_I^* R_I V^* \cdot \mathbb{1}_I(\ell) \mathbb{1}_{\mathcal{G}}(I)]\| \leq \pi_\ell \cdot (\|W D_{\sqrt{\pi}}\| \cdot \|V D_{\sqrt{\pi}}\| + \|w_\ell\| \cdot \|v_\ell\|), \tag{c}$$

as well as

$$\begin{aligned} &\|\mathbb{E}[D_{\sqrt{\pi}}^{-1} \mathbb{I}_{\ell^c} R_I^* \mathcal{H}_{I,I} \mathcal{H}_{I,I}^* R_I \mathbb{I}_{\ell^c} D_{\sqrt{\pi}}^{-1} \cdot \mathbb{1}_I(\ell)]\| \\ &\leq \frac{3}{2} \cdot \pi_\ell \cdot \left(3 \cdot \|D_{\sqrt{\pi}} \mathcal{H} e_\ell\|^2 + \max_k \mathcal{H}_{k\ell}^2 + \frac{9}{2} \cdot \|D_{\sqrt{\pi}} \mathcal{H} D_{\sqrt{\pi}}\|^2 + \frac{3}{2} \cdot \max_k \|e_k^* \mathcal{H} D_{\sqrt{\pi}}\|^2\right). \end{aligned} \tag{d}$$

Now we are ready to prove the first of the four inequalities used in the proof of the main theorem.

Lemma 6 *Given the conditions stated in Theorem 3, we have*

$$\mathbb{P}(\|(D_{\sqrt{\pi \cdot \beta}})^{-1} A (D_{\sqrt{\pi \cdot \beta} \cdot \alpha})^{-1} - \mathbb{I}\| > \Delta/8) \leq (d + K) \exp\left(-\frac{N(\Delta/16)^2}{2\rho^2 + \rho\Delta/16}\right).$$

Proof The fundamental idea behind the proof is to rewrite $(D_{\sqrt{\pi \cdot \beta}})^{-1} A (D_{\sqrt{\pi \cdot \beta} \cdot \alpha})^{-1} - \mathbb{I}$ as $N^{-1} \sum_n \hat{Y}_n$ and apply the matrix Bernstein inequality, restated in Theorem11. Therefore the matrices $\hat{Y}_n$ need to be independent and $\|\hat{Y}_n\|$ as well as $\|\mathbb{E}[\hat{Y}_n]\|$ must be bounded. In order to define such $\hat{Y}_n$ we first recall that $A = N^{-1} \sum_{n=1}^N x_n \hat{x}_n^*$ for well behaved $\hat{x}_n$. By the algorithm the norm of the estimated coefficients have to be smaller than the signals times κ or are set to zero. Hence we define the set of all stable supports as $\mathcal{B}(v) := \{I : \|\Psi_I^\dagger v\| \leq \kappa \|v\|\}$ for $v \in \mathbb{R}^d$. We denote by $\hat{I}_n$ the set determined by the thresholding algorithm and define

$$\hat{Y}_n := (D_{\sqrt{\pi \cdot \beta}})^{-1} R_{\hat{I}_n}^* x_{\hat{I}_n} y_n^* (\Psi_{\hat{I}_n}^\dagger)^* R_{\hat{I}_n} (D_{\sqrt{\pi \cdot \beta} \cdot \alpha})^{-1} \cdot \mathbb{1}_{\mathcal{B}(y_n)}(\hat{I}_n) - \mathbb{I}. \tag{65}$$

The matrices $\hat{Y}_n$ are indeed independent, since each $\hat{Y}_n$ only depends on the signal y_n. We continue with deriving a suitable bound for $\|\hat{Y}_n\|$ and $\|\mathbb{E}[\hat{Y}_n]\|$. Since in the algorithm large estimated coefficients are eliminated, we can use that $\|\hat{x}_n\| \leq \kappa\|y_n\|$ and $\|y_n\| \leq |I_n| \cdot \|x_n\|_\infty \leq Sc_{\max}$. This yields

$$\|\hat{Y}_n\| \leq \kappa S^2 c_{\max}^2 \|D_\beta^{-1}\|\|D_\alpha^{-1}\|\|D_{\sqrt{\pi}}^{-1}\| + 1 \leq \rho\underline{\alpha}\underline{\pi} + 1 \leq \rho/2 =: r, \tag{66}$$

where we employed from (18) that $\rho = 2\kappa^2 S^2 \gamma^{-2}\underline{\alpha}^{-2}\underline{\pi}^{-3/2}$ and $\underline{\pi} < 1/3$. The bound on $\|\mathbb{E}[\hat{Y}_n]\|$ entails a more complex approach. First we want to replace the estimated support $\hat{I}_n$ with the correct support I_n and define

$$Y_n := (D_{\sqrt{\pi\cdot\beta}})^{-1} R_{I_n}^* x_{I_n} y_n^* (\Psi_{I_n}^\dagger)^* R_{I_n} (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} \cdot \mathbb{1}_{\mathcal{B}(y_n)}(I_n) - \operatorname{diag}(\mathbf{1}_{I_n}) D_\pi^{-1} \tag{67}$$

Note that we have

$$\|Y_n\| \leq \kappa S^2 c_{\max}^2 \underline{\pi}^{-1}\underline{\alpha}^{-1}\underline{\beta}^{-1} \ + \underline{\pi}^{-1} \leq \rho/2 = r. \tag{68}$$

Recalling that $\mathcal{H}$ is the set, where thresholding recovers the correct support, and since $\mathbb{I} = \mathbb{E}[\operatorname{diag}(\mathbf{1}_{I_n})D_\pi^{-1}]$, we see that the left parts of Y_n and $\hat{Y}_n$ coincide on $\mathcal{H}$ and the right part coincide in expectation. For the rest of the proof we simplify matters by omitting the index n, since each signal shares the same distribution. We recall that $\mathcal{G}$ is the set, where Φ_I and Ψ_I are well conditioned, to get

$$\begin{aligned}\|\mathbb{E}[\hat{Y}]\| \leq \|\mathbb{E}[\hat{Y} - Y]\| + \|\mathbb{E}[Y]\| &\leq \mathbb{P}(\mathcal{H}^c) \cdot 2r + \|\mathbb{E}[\mathbb{1}_{\mathcal{G}^c}(I)Y]\| + \|\mathbb{E}[\mathbb{1}_{\mathcal{G}}(I)Y]\| \\ &\leq \mathbb{P}(\mathcal{H}^c) \cdot 2r + \mathbb{P}(\mathcal{G}^c) \cdot r + \|\mathbb{E}[\mathbb{1}_{\mathcal{G}}(I)Y]\|.\end{aligned} \tag{69}$$

Using the singular value decomposition and the fact that $\|\Psi_I^*\Psi_I - \mathbb{I}_S\| \leq \vartheta$ on $\mathcal{G}$ we get for $I \in \mathcal{G}$ that $\|\Psi_I^\dagger y\| \leq (1-\vartheta)^{-1/2} \cdot \|y\| \leq \kappa\|y\|$. This implies that $\mathcal{G} \subseteq \mathcal{B}(y)$ and therefore $\mathbb{1}_{\mathcal{B}(y)}\mathbb{1}_{\mathcal{G}} = \mathbb{1}_{\mathcal{G}}$. We recall the notation $\dot{M} := MD_{\sqrt{\beta}}$ to show that

$$\begin{aligned}\mathbb{E}[x_I y^*] &= \mathbb{E}_{I,\sigma,c}[x_I x_I \Phi_I^*] \\ &= \mathbb{E}_I[R_I \mathbb{E}_{\sigma,c}[xx^*] \cdot R_I^* \Phi_I^*] = \mathbb{E}_I[(D_\beta)_{I,I}\Phi_I^*] = \mathbb{E}_I[(D_{\sqrt{\beta}})_{I,I}\dot{\Phi}_I^*],.\end{aligned} \tag{70}$$

Further, we have

$$\begin{aligned}(\Psi_I^\dagger)^*(D_{\sqrt{\beta}})_{I,I}^{-1} &= ((\Psi_I^*\Psi_I)^{-1}\Psi_I)^*(D_{\sqrt{\beta}})_{I,I}^{-1} \\ &= \Psi_I (D_{\sqrt{\beta}})_{I,I}(D_{\sqrt{\beta}})_{I,I}^{-1}(\Psi_I^*\Psi_I)^{-1}(D_{\sqrt{\beta}})_{I,I}^{-1} = \dot{\Psi}_I(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}.\end{aligned}$$

We define $\mathbb{E}_{\mathcal{G}}[f(I)] := \mathbb{E}_I[\mathbb{1}_{\mathcal{G}}(I)f(I)]$ and use the calculations above to get

$$\begin{aligned}\|\mathbb{E}[\mathbb{1}_{\mathcal{G}}(I)Y]\| &= \|\mathbb{E}_I[\mathbb{1}_{\mathcal{G}}(I) \cdot \mathbb{E}_{\sigma,c}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* x_I y^* (\Psi_I^\dagger)^* R_I (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \operatorname{diag}(\mathbf{1}_I)D_\pi^{-1}]\| \\ &= \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* [\dot{\Phi}_I^*\dot{\Psi}_I(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1} - (D_\alpha)_{I,I}] R_I (D_{\sqrt{\pi}\cdot\alpha})^{-1}]\|.\end{aligned} \tag{71}$$

Next we rewrite $\dot{\Phi}_I^*\dot{\Psi}_I(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}$ with the help of zero diagonal matrices in order to employ Corollary 5. We recall that $Z = \Psi - \Phi$ as well as $\ddot{M} := D_{\sqrt{\beta}} M D_{\sqrt{\beta}}$ and define the following matrices which have zero diagonal

$$\begin{aligned}\ddot{H} &:= D_{\sqrt{\beta}}(\mathbb{I} - \Psi^*\Psi)D_{\sqrt{\beta}} = D_\beta - \dot{\Psi}^*\dot{\Psi} \\ \mathcal{E} &:= \operatorname{diag}(Z^*\Psi) = \mathbb{I} - \operatorname{diag}(\Phi^*\Psi) = \mathbb{I} - D_\alpha \\ \text{И} &:= (\Psi\mathcal{E} - Z)^*\Psi = (\Phi\mathcal{E} - ZD_\alpha)^*\Psi = (\Phi - \Psi D_\alpha)^*\Psi \\ \ddot{\text{И}} &:= D_{\sqrt{\beta}} \text{И} D_{\sqrt{\beta}} = (\dot{\Psi}\mathcal{E} - \dot{Z})^*\dot{\Psi}.\end{aligned}$$

Using that

$$\mathbb{I}_S = \Big((D_\beta)_{I,I} - (\dot\Psi_I^*\dot\Psi_I) + (\dot\Psi_I^*\dot\Psi_I)\Big)(D_\beta)^{-1}_{I,I} = \Big(\ddot{H}_{I,I} + \dot\Psi_I^*\dot\Psi_I\Big)(D_\beta)^{-1}_{I,I}, \tag{72}$$

we can write

$$\begin{aligned}
\dot\Phi_I^*\dot\Psi_I(\dot\Psi_I^*\dot\Psi_I)^{-1} &= (\dot\Psi_I^* - \dot Z_I^*)\dot\Psi_I(\dot\Psi_I^*\dot\Psi_I)^{-1} = \mathbb{I}_S - \dot Z_I^*\dot\Psi_I(\dot\Psi_I^*\dot\Psi_I)^{-1}\\
&= (D_\alpha)_{I,I} + \mathcal{E}_{I,I}\dot\Psi_I^*\dot\Psi_I(\dot\Psi_I^*\dot\Psi_I)^{-1} - \dot Z_I^*\dot\Psi_I(\dot\Psi_I^*\dot\Psi_I)^{-1}\\
&= (D_\alpha)_{I,I} + \ddot{\not{H}}_{I,I}(\dot\Psi_I^*\dot\Psi_I)^{-1}\cdot\mathbb{I}_S \qquad (74\text{a})\\
&= (D_\alpha)_{I,I} + \ddot{\not{H}}_{I,I}(D_\beta^{-1})_{I,I} + \ddot{\not{H}}_{I,I}(\dot\Psi_I^*\dot\Psi_I)^{-1}\ddot{H}_{I,I}(D_\beta^{-1})_{I,I}. \qquad (74\text{b})
\end{aligned}$$

Observing that for any I, we have $\|\ddot{\not{H}}_{I,I}\| \le \|(\dot\Psi\mathcal{E} - \dot Z)_I\|\cdot\|\dot\Psi_I\| \le \bar\beta\cdot\varepsilon\sqrt{S}\cdot\sqrt{S} < 2S$ and therefore $\|\ddot{\not{H}}_{I,I}R_I(D_{\sqrt{\pi}\alpha})^{-1}\| \le \rho/2 = r$ we get that $\|\mathbb{E}[\mathbb{1}_{\mathcal{G}^c}(I)\cdot\ddot{\not{H}}_{I,I}R_I(D_{\sqrt{\pi}\alpha})^{-1}]\| \le \mathbb{P}(\mathcal{G}^c)]\cdot r$. We substitute (74b) into (71) and use the bound established above to obtain

$$\begin{aligned}
\|\mathbb{E}[\hat Y]\| &\le [2\mathbb{P}(\mathcal{H}^c) + \mathbb{P}(\mathcal{G}^c)]\cdot r + \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1}R_I^*(\ddot{\not{H}}_{I,I} + \ddot{\not{H}}_{I,I}(\dot\Psi_I^*\dot\Psi_I)^{-1}\ddot{H}_{I,I})R_I(D_{\sqrt{\pi}\cdot\beta\cdot\alpha})^{-1}]\|\\
&\le [2\mathbb{P}(\mathcal{H}^c) + \mathbb{P}(\mathcal{G}^c)]\cdot r + \mathbb{P}(\mathcal{G}^c)\cdot r + \|\mathbb{E}[D_{\sqrt{\pi}}^{-1}R_I^*\ddot{\not{H}}_{I,I}R_I(D_{\sqrt{\pi}\cdot\beta\cdot\alpha})^{-1}]\|\\
&\qquad + \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1}R_I^*\ddot{\not{H}}_{I,I}(\dot\Psi_I^*\dot\Psi_I)^{-1}\ddot{H}_{I,I}R_I(D_{\sqrt{\pi}\cdot\beta\cdot\alpha})^{-1}]\|\\
&\le [2\mathbb{P}(\mathcal{H}^c) + \mathbb{P}(\mathcal{G}^c)]\cdot\rho\underline{\alpha} + \|\mathbb{E}[D_{\sqrt{\pi}}^{-1}R_I^*\ddot{\not{H}}_{I,I}R_I(D_{\sqrt{\pi}})^{-1}]\|\cdot\|D_\alpha^{-1}\|\cdot\|D_\beta^{-1}\|\\
&\qquad + \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1}R_I^*\ddot{\not{H}}_{I,I}(\dot\Psi_I^*\dot\Psi_I)^{-1}\ddot{H}_{I,I}R_I(D_{\sqrt{\pi\cdot\beta}})^{-1}]\|\cdot\|D_\alpha^{-1}\|\cdot\|D_{\sqrt{\beta}}^{-1}\|. \qquad (75)
\end{aligned}$$

Now we have made all the necessary preparations to apply Corollary 5. Beginning with the first expectation, Corollary 5(a) gets us

$$\begin{aligned}
\|\mathbb{E}[D_{\sqrt{\pi}}^{-1}R_I^*\ddot{\not{H}}_{I,I}R_I(D_{\sqrt{\pi}})^{-1}]\| &\le 3\cdot\|D_{\sqrt{\pi}}\ddot{\not{H}}D_{\sqrt{\pi}}\|\\
&\le 3\cdot(\|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\varepsilon^2/2 + \|ZD_{\sqrt{\pi\cdot\beta}}\|)\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|, \qquad (76)
\end{aligned}$$

where the second inequality is proven in (121) in the appendix. For the second expectation in (75) we need the following bounds below, which are again addressed in the appendix, that is inequalities (124) and (127), yielding

$$\|\mathbb{E}[D_{\sqrt{\pi}}^{-1}R_I^*\ddot{\not{H}}_{I,I}\ddot{\not{H}}_{I,I}^*R_ID_{\sqrt{\pi}}^{-1}]\|^{\frac{1}{2}} \le (3\|ZD_{\sqrt{\pi\cdot\beta}}\| + 3\varepsilon\bar\beta^{1/2})\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|, \tag{77}$$

$$\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}R_I(D_{\sqrt{\pi\cdot\beta}})^{-1}]\|^{\frac{1}{2}} \le \sqrt{2}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|. \tag{78}$$

Using that $\|\mathbb{I}_S - \Psi_I^*\Psi_I\| \le \vartheta \le 1/4 < 1$ and the Neumann series results in

$$\begin{aligned}
\|(\dot\Psi_I^*\dot\Psi_I)^{-1}\| \le \underline{\beta}^{-1}\cdot\|(\Psi_I^*\Psi_I)^{-1}\| &= \underline{\beta}^{-1}\cdot\|\sum_{k\ge0}(\mathbb{I}_S - \Psi_I^*\Psi_I)^k\|\\
&\le \underline{\beta}^{-1}\cdot\sum_{k\ge0}\vartheta^k = \underline{\beta}^{-1}\cdot\frac{1}{1-\vartheta} \le \frac{4}{3}\cdot\underline{\beta}^{-1}. \qquad (79)
\end{aligned}$$

We apply Lemma 13 from the appendix to the second expectation in (75) and use the three bounds which we have established so far to get

$$\begin{aligned}
&\|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} \cdot (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \cdot \ddot{H}_{I,I} R_I (D_{\sqrt{\pi \cdot \beta}})^{-1}] \cdot \|D_{\sqrt{\beta}}^{-1}\| \\
&\quad \leq \|\mathbb{E}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} \ddot{H}_{I,I}^* R_I D_{\sqrt{\pi}}^{-1}\|^{\frac{1}{2}} \cdot 4/3 \cdot \underline{\beta}^{-3/2} \cdot \|\mathbb{E}[(D_{\sqrt{\pi \cdot \beta}})^{-1} R_I^* \ddot{H}_{I,I} \ddot{H}_{I,I} R_I (D_{\sqrt{\pi \cdot \beta}})^{-1}\|^{\frac{1}{2}} \\
&\quad \leq (3 \cdot \|Z D_{\sqrt{\pi \cdot \beta}}\| + 3 \cdot \varepsilon \bar{\beta}^{1/2}) \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\| \cdot 4/3 \cdot \underline{\beta}^{-3/2} \cdot \sqrt{2} \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\| \\
&\quad \leq 6 \cdot \underline{\beta}^{-3/2} \cdot (\|Z D_{\sqrt{\pi \cdot \beta}}\| + \varepsilon \underline{\beta}^{1/2} \gamma^{-1}) \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\|^2, \qquad (80)
\end{aligned}$$

where in the last inequality we used the fact that $(\bar{\beta}/\underline{\beta})^{1/2} \leq c_{\max}/c_{\min} = \gamma^{-1}$. Plugging (76) and (80) into (75) leads to

$$\begin{aligned}
\|\mathbb{E}[\hat{Y}]\| \leq [\mathbb{P}(\mathcal{H}^c) + \mathbb{P}(\mathcal{G}^c)] \cdot \rho \underline{\alpha} + 3 \cdot \underline{\alpha}^{-1} \underline{\beta}^{-1} \cdot (\|\Phi D_{\sqrt{\pi \cdot \beta}}\| \cdot \varepsilon^2/2 + \|Z D_{\sqrt{\pi \cdot \beta}}\|) \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\| \\
+ 6 \cdot \underline{\alpha}^{-1} \underline{\beta}^{-3/2} \cdot (\|Z D_{\sqrt{\pi \cdot \beta}}\| + \varepsilon \underline{\beta}^{1/2} \gamma^{-1}) \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\|^2 \qquad (81)
\end{aligned}$$

In order to apply the assumptions of Theorem 3 we consider the two cases, $\delta > \delta_\circ$ and $\delta \leq \delta_\circ$, separately. In the first regime (20) we have that $\max\{\|\Phi D_{\sqrt{\pi \cdot \beta}}\|, \|\Psi D_{\sqrt{\pi \cdot \beta}}\|\} \leq \underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/(4C)$, and hence $\|Z D_{\sqrt{\pi \cdot \beta}}\| \leq \underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/(2C)$. Furthermore we use that $\varepsilon \leq \sqrt{2}$ and $\nu \leq 1/3$. Using these conditions and the probability bound from Lemma 4 leads to

$$\begin{aligned}
\|\mathbb{E}[\hat{Y}]\| &\leq \underline{\alpha} \delta_*/32 + \underline{\alpha}^{-1} \underline{\beta}^{-1} \cdot 9 \cdot \left(\frac{\underline{\alpha} \gamma \nu \underline{\beta}^{1/2}}{4C} \right)^2 \\
&\qquad + 6 \cdot \underline{\alpha}^{-1} \underline{\beta}^{-3/2} \cdot \left(\frac{\underline{\beta}^{1/2}}{2C} + \sqrt{2} \gamma^{-1} \underline{\beta}^{1/2} \right) \cdot \left(\frac{\underline{\alpha} \gamma \nu \underline{\beta}^{1/2}}{4C} \right)^2 \\
&\leq \underline{\alpha} \delta_*/32 + 18/(16C) \cdot \underline{\alpha} \cdot \gamma \nu^2/C \leq \underline{\alpha} \cdot \Delta/16. \qquad (82)
\end{aligned}$$

Before we bound $\|\mathbb{E}[\hat{Y}]\|$ in the second regime (21), we note that for all $i \in [K]$ we get

$$\|\psi_i - \phi_i\| \leq \|\psi_i - \phi_i\| \cdot \beta_i^{1/2}/\underline{\beta}^{1/2} = \|(\psi_i - \phi_i)\beta_i^{1/2}\| \cdot \underline{\beta}^{-1/2}, \qquad (83)$$

and hence $\varepsilon \leq \delta$. For the second regime where $\delta \leq \delta_\circ$, we have as above $\varepsilon \leq \sqrt{2}$ and $\nu \leq 1/3$. We further use the assumption $\|\Phi D_{\sqrt{\pi \cdot \beta}}\| \leq \underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/(4C)$, while due to (63) and (10) we have $\|\Psi D_{\sqrt{\pi \cdot \beta}}\| \leq \underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/C$ and $\|Z D_{\sqrt{\pi \cdot \beta}}\| \leq \delta \underline{\beta}^{1/2}$. So we get

$$\begin{aligned}
\|\mathbb{E}[\hat{Y}]\| &\leq [\mathbb{P}(\mathcal{H}^c) + \mathbb{P}(\mathcal{G}^c)] \cdot \rho \underline{\alpha} + \underline{\alpha}^{-1} \underline{\beta}^{-1} \cdot 3(\|\Phi D_{\sqrt{\pi \cdot \beta}}\| \varepsilon^2/2 + \|Z D_{\sqrt{\pi \cdot \beta}}\|) \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\| \\
&\qquad + \underline{\alpha}^{-1} \cdot 6 \underline{\beta}^{-3/2} \cdot (\|Z D_{\sqrt{\pi \cdot \beta}}\| + \varepsilon \underline{\beta}^{1/2} \gamma^{-1}) \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\|^2 \\
&\leq \underline{\alpha} \delta_*/32 + \underline{\alpha}^{-1} \underline{\beta}^{-1} \cdot 3(\underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/(4C) \cdot \delta/\sqrt{2} + \delta \underline{\beta}^{1/2}) \cdot \underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/C \\
&\qquad + 6 \underline{\alpha}^{-1} \underline{\beta}^{-3/2} \cdot (\delta \underline{\beta}^{1/2} + \delta \underline{\beta}^{1/2} \gamma^{-1}) \cdot \left(\underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/C \right)^2 \\
&\leq \underline{\alpha} \delta_*/32 + 32/(16C) \cdot \delta \leq \Delta/16. \qquad (84)
\end{aligned}$$

Finally, we completed all necessary steps, in order to apply Bernstein's inequality 11. By choosing $t = m = \Delta/16$ and $r = \rho/2$ and performing a few simplifications we get the desired bound. ■

In Lemma 6 we obtained a better bound in comparison to its counterpart in [24]. In order to see this, we assume that the coefficient sequence c is i.i.d. We will denote the distance between the dictionary elements in [24] as

$$\hat{\delta}(\Psi,\Phi) := \max\left\{\|(\Psi-\Phi)D_{\sqrt{\pi}}\|\,,\|\Psi-\Phi\|_{2,1}\right\}.$$

Since the coefficient sequence c is i.i.d we obtain $\underline{\beta}^{1/2} = \|D_{\sqrt{\beta}}\|$ and thus $\delta = \hat{\delta}$. Hence we can compare Lemma 6 with [24]. The lemma in [24] yields

$$\|\Phi A(D_{\sqrt{\pi}\cdot\beta\cdot\alpha})^{-1} - \Phi D_{\sqrt{\pi}}\| \leq \frac{\underline{\alpha}\Delta}{8}.$$

If we want to bound the same via (19) and Lemma 6 we get

$$\begin{aligned}\|\Phi A(D_{\sqrt{\pi}\cdot\beta\cdot\alpha})^{-1} - \Phi D_{\sqrt{\pi}}\| &\leq \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\|(D_{\sqrt{\pi\cdot\beta}})^{-1}A(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \mathbb{I}\|\cdot\|D_{\sqrt{\beta}}^{-1}\| \\ &\leq \frac{\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}}{4C}\cdot\frac{\Delta}{8}\cdot\underline{\beta}^{-1/2} = \frac{\underline{\alpha}\gamma\nu}{4C}\cdot\frac{\Delta}{8}.\end{aligned}$$

We continue with the proof of the second inequality in 26.

Lemma 7 *Given the conditions stated in Theorem 3, we have*

$$\mathbb{P}\left(\|(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}B(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \mathbb{I}\| > \frac{\Delta}{4\gamma}\right) \leq 2K\exp\left(-\frac{N(\Delta/16)^2}{2\rho^2+\rho\Delta/16}\right).$$

Proof We closely adhere to the method used in the previous proof by rewriting the matrix $(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}B(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \mathbb{I}$ to $N^{-1}\sum_n \hat{Y}_n$, where the matrices $\hat{Y}_n$ are independent and $\|\hat{Y}_n\|$ as well as $\|\mathbb{E}[\hat{Y}_n]\|$ are bounded. Then we can apply the matrix Bernstein inequality and finish the proof. Again we take $\hat{I}_n$ to be the set determined by the thresholding and $\mathcal{B}(v) := \{I : \|\Psi_I^\dagger v\| \leq \kappa\|v\|\}$ to be the set of stable supports of v. We define

$$\begin{aligned}\hat{Y}_n &:= (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}R^*_{\hat{I}_n}\Psi^\dagger_{\hat{I}_n}y_ny_n^*(\Psi^\dagger_{\hat{I}_n})^*R_{\hat{I}_n}(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\mathbb{1}_{\mathcal{B}(y_n)}(\hat{I}_n) - \mathbb{I} \\ \text{and } Y_n &:= (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}R^*_{I_n}\Psi^\dagger_{I_n}y_ny_n^*(\Psi^\dagger_{I_n})^*R_{I_n}(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\mathbb{1}_{\mathcal{B}(y_n)}(I_n) - \operatorname{diag}(\mathbf{1}_{I_n})D_\pi^{-1},\end{aligned} \tag{85}$$

which coincide on $\mathcal{H}$ in expectation, since $I_n \in \mathcal{H}$ and $\mathbb{E}[\operatorname{diag}(\mathbf{1}_{I_n})D_\pi^{-1}] = \mathbb{I}$. We can find an upper limit chosen as

$$\max\{\|\hat{Y}_n\|,\|Y_n\|\} \leq \kappa^2S^2c_{\max}^2\|D_\beta^{-1}\|\|D_\alpha^{-2}\|\|D_\pi^{-1}\| + \|D_\pi^{-1}\| \leq 3\rho/4 \leq 3\gamma^{-1}\rho/4 =: r, \tag{86}$$

where we used that $\rho = 2\kappa^2S^2\gamma^{-2}\underline{\alpha}^{-2}\underline{\pi}^{-3/2}$. So following the same approach as in (69) and omitting the index n leads to

$$\|\mathbb{E}[\hat{Y}]\| \leq 2\rho\cdot\mathbb{P}(\mathcal{H}^c) + \rho\cdot\mathbb{P}(\mathcal{G}^c) + \|\mathbb{E}[1_{\mathcal{G}}(I)Y]\|. \tag{87}$$

For the upcoming steps we want to recall the notation $\dot{M} := MD_{\sqrt{\beta}}$ as well as $\ddot{M} := D_{\sqrt{\beta}}MD_{\sqrt{\beta}}$. With the same argument as in (70), where we used that $\mathcal{G} \subseteq \mathcal{B}(y)$, we take the expectation over (σ, c) and get

$$
\begin{aligned}
&\|\mathbb{E}[1_{\mathcal{G}}(I)Y]\| \\
&\quad = \|\mathbb{E}_I[\mathbb{1}_{\mathcal{G}}(I) \cdot \mathbb{E}_{\sigma,c}[(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} R_I^* \Psi_I^\dagger \Phi_I x_I x_I^* \Phi_I^* (\Psi_I^\dagger)^* R_I (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} - \operatorname{diag}(\mathbf{1}_I) D_\pi^{-1}]\| \\
&\quad = \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1} R_I^* (\Psi_I^\dagger \dot{\Phi}_I \dot{\Phi}_I^* (\Psi_I^\dagger)^* - (D_{\sqrt{\beta}\cdot\alpha})^2_{I,I}) R_I (D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}]\| \\
&\quad \leq \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi}\cdot\alpha})^{-1} R_I^* ((D_{\sqrt{\beta}}^{-1})_{I,I} \Psi_I^\dagger \dot{\Phi}_I \dot{\Phi}_I^* (\Psi_I^\dagger)^* (D_{\sqrt{\beta}}^{-1})_{I,I} - (D_\alpha)^2_{I,I}) R_I (D_{\sqrt{\pi}\cdot\alpha})^{-1}]\|. \qquad (88)
\end{aligned}
$$

With the help of (74a) we get the following identity

$$
\begin{aligned}
(D_{\sqrt{\beta}}^{-1})_{I,I} \Psi_I^\dagger \dot{\Phi}_I \dot{\Phi}_I^* (\Psi_I^\dagger)^* (D_{\sqrt{\beta}}^{-1})_{I,I} - (D_\alpha)^2_{I,I} &= (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \dot{\Psi}^* \dot{\Phi}_I \dot{\Phi}_I^* \dot{\Psi} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} - (D_\alpha)^2_{I,I} \\
&= (D_\alpha)_{I,I} \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} + (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \ddot{H}_{I,I}^* (D_\alpha)_{I,I} \\
&\quad + (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \ddot{H}_{I,I}^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1}. \qquad (89)
\end{aligned}
$$

We plug this into (88) and note that the first term is the transpose of the second term, which gets us

$$
\begin{aligned}
\|\mathbb{E}[1_{\mathcal{G}}(I)Y]\| &\leq 2 \cdot \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}]\| \cdot \|D_\alpha^{-1}\| \\
&\quad + \|D_\alpha^{-1}\|^2 \cdot \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \ddot{H}_{I,I}^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}]\|. \qquad (90)
\end{aligned}
$$

We begin by estimating the first term. For any I we have
$\|D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} R_I D_{\sqrt{\pi}}^{-1}\| \leq \rho\underline{\alpha}^2/2$. With the same procedure as in (75) and by using the identity $\mathbb{I}_S = \left(\dot{\Psi}_I^* \dot{\Psi}_I + \ddot{H}_{I,I}\right) (D_\beta)^{-1}_{I,I}$ as in (72), we obtain

$$
\begin{aligned}
&\|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}]\| \\
&\quad \leq \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} (\dot{\Psi}_I^* \dot{\Psi}_I + \ddot{H}_{I,I}) R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \cdot \|D_{\sqrt{\beta}}^{-1}\| \\
&\quad \leq \mathbb{P}(\mathcal{G}^c) \cdot \rho\underline{\alpha}^2/2 + \|\mathbb{E}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} R_I D_{\sqrt{\pi}}^{-1}]\| \cdot \|D_\beta^{-1}\| \\
&\qquad + \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} \cdot (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \cdot \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \cdot \|D_{\sqrt{\beta}}^{-1}\|. \qquad (91)
\end{aligned}
$$

Since we bounded the same expectations in (76) and (80), we get

$$
\begin{aligned}
&\|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}]\| \\
&\quad \leq \mathbb{P}(\mathcal{G}^c) \cdot \rho\underline{\alpha}^2/2 + 3 \cdot \underline{\beta}^{-1} \cdot (\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \cdot \varepsilon^2/2 + \|Z D_{\sqrt{\pi\cdot\beta}}\|) \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\| \\
&\qquad + 6 \cdot \underline{\beta}^{-3/2} \cdot (\|Z D_{\sqrt{\pi\cdot\beta}}\| + \varepsilon\underline{\beta}^{1/2}\gamma^{-1}) \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2.
\end{aligned}
$$

Similar to Lemma 6 we distinguish between the cases $\delta > \delta_\circ$ and $\delta \leq \delta_\circ$. In the first regime we use that $\max\{\|\Psi D_{\sqrt{\pi\cdot\beta}}\|, \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\} \leq \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)$ and thus $\|Z D_{\sqrt{\pi\cdot\beta}}\| \leq \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(2C)$. In the second regime, where $\delta \leq \delta_\circ$, we can use $\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \leq \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)$, while due to (63) and (10) we have $\|\Psi D_{\sqrt{\pi\cdot\beta}}\| \leq \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/C$ and $\|Z D_{\sqrt{\pi\cdot\beta}}\| \leq \delta\underline{\beta}^{1/2}$. Using that $\nu \leq 1/3$

from (58) yields

$$\delta > \delta_\circ : \quad \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{И}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}]\|$$

$$\leq \mathbb{P}(\mathcal{G}^c) \cdot \rho \underline{\alpha}^2/2 + \underline{\beta}^{-1} \cdot 9 \cdot \left(\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)\right)^2$$

$$+ 6 \cdot \underline{\beta}^{-3/2} \cdot (\underline{\beta}^{1/2}/(2C) + \sqrt{2}\gamma^{-1}\underline{\beta}^{1/2}) \cdot \left(\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)\right)^2$$

$$\leq \mathbb{P}(\mathcal{G}^c) \cdot \rho \underline{\alpha}^2/2 + 18/(16C) \cdot \underline{\alpha}^2 \cdot \gamma\nu^2/C$$

$$\leq \mathbb{P}(\mathcal{G}^c) \cdot \rho \underline{\alpha}^2/2 + 9/(8C) \cdot \underline{\alpha}^2 \cdot \delta_\circ,$$

$$\delta \leq \delta_\circ : \quad \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{И}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}]\|$$

$$\leq \mathbb{P}(\mathcal{G}^c) \cdot \rho \underline{\alpha}^2/2 + \underline{\beta}^{-1} \cdot 3 \cdot (\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C) \cdot \delta/\sqrt{2} + \delta\underline{\beta}^{1/2}) \cdot \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/C$$

$$+ 6 \cdot \underline{\beta}^{-3/2} \cdot (\delta\underline{\beta}^{1/2} + \delta\underline{\beta}^{1/2}\gamma^{-1}) \cdot \left(\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/C\right)^2$$

$$\leq \mathbb{P}(\mathcal{G}^c) \cdot \rho \underline{\alpha}^2/2 + 7/(4C) \cdot \underline{\alpha} \cdot \delta,$$

which we can summarise as

$$\|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{И}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}]\| \leq \mathbb{P}(\mathcal{G}^c) \cdot \frac{\rho\underline{\alpha}^2}{2} + \frac{7\underline{\alpha}}{4C} \cdot \min\{\delta_\circ, \delta\}. \tag{92}$$

For the second term in (90) we need to bound $\|\ddot{И}_{I,I}(\dot{\Psi}_I^* \dot{\Psi}_I)^{-1}\|$ in two different ways. Since $I \in \mathcal{G}$ and $И := (\Psi\mathcal{E} - Z)^*\Psi = (\Phi\mathcal{E} - ZD_\alpha)^*\Psi = (\Phi - \Psi D_\alpha)^*\Psi$ we obtain either

$$\|\ddot{И}_{I,I}(\dot{\Psi}_I^* \dot{\Psi}_I)^{-1}\| = \|(D_{\sqrt{\beta}})_{I,I}(\Phi - \Psi D_\alpha)_I^* \Psi_I (\Psi_I^* \Psi_I)^{-1} (D_{\sqrt{\beta}})_{I,I}^{-1}\|$$

$$= \|(D_{\sqrt{\beta}})_{I,I} \Phi_I^* (\Psi_I^\dagger)^* (D_{\sqrt{\beta}})_{I,I}^{-1} - (D_\alpha)_{I,I}\|$$

$$\leq \bar{\beta}^{1/2} \cdot \sqrt{\frac{1+\vartheta}{1-\vartheta}} \cdot \underline{\beta}^{-1/2} + \bar{\alpha} \leq \gamma^{-1}\sqrt{5/3} + 1 \leq \gamma^{-1} \cdot 7/3, \tag{93}$$

or since we have $\|Z_I D_{\sqrt{\beta}}\|^2 = \|Z_I D_\beta Z_I^*\| \leq 2 \cdot \delta^2 \underline{\beta} \log(nK\rho/\delta_*) = 2 \cdot \delta^2 \underline{\beta}/\nu^2$ on $\mathcal{G}$, we get

$$\|\ddot{И}_{I,I}(\dot{\Psi}_I^* \dot{\Psi}_I)^{-1}\| = \|(D_{\sqrt{\beta}})_{I,I}(\Psi\mathcal{E} - Z)_I^* \Psi_I (\Psi_I^* \Psi_I)^{-1} (D_{\sqrt{\beta}})_{I,I}^{-1}\|$$

$$\leq \|(D_{\sqrt{\beta}})_{I,I} \mathcal{E}_{I,I} (D_{\sqrt{\beta}})_{I,I}^{-1}\| + \|(D_{\sqrt{\beta}})_{I,I} Z_I^*\| \cdot \|\Psi_I^\dagger\| \cdot \|(D_{\sqrt{\beta}})_{I,I}^{-1}\|$$

$$\leq \delta/\sqrt{2} + \delta\underline{\beta}^{1/2}/\nu \cdot \sqrt{2 \cdot 4}/\sqrt{3} \cdot \underline{\beta}^{-1/2}$$

$$\leq 2 \cdot \delta/\nu. \tag{94}$$

We define $\Gamma := \min\{\gamma^{-1} \cdot 7/3, 2 \cdot \delta/\nu\}$ and use the identity (72) as above for the second term of (90), which enables us to apply Lemma 13 and Corollary 5(b):

$$
\begin{aligned}
&\|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \ddot{H}_{I,I}^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}]\| \\
&= \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi}\cdot\beta})^{-1} R_I^* (\ddot{H}_{I,I} + \dot{\Psi}_I^* \dot{\Psi}_I) \cdot (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \ddot{H}_{I,I}^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \cdot (\ddot{H}_{I,I}^* + \dot{\Psi}_I^* \dot{\Psi}_I) R_I D_{\sqrt{\pi}\cdot\beta}^{-1}]\| \\
&\leq \|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I}^* \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \cdot \|D_{\beta}^{-1}\| \\
&\qquad + 2\|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I}^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I) \cdot \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \cdot \|D_{\beta}^{-1}\| \\
&\qquad + \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I} \cdot (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \ddot{H}_{I,I}^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \cdot \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \cdot \|D_{\beta}^{-1}\| \\
&= \Big(\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I}^* \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\|^{1/2} \\
&\qquad\qquad + \Gamma \|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I} \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\|^{1/2} \Big)^2 \cdot \|D_{\beta}^{-1}\|. \qquad (95)
\end{aligned}
$$

Looking at the first term of (95), we use a similar strategy as in (77) or rather in (121) and (123) by applying Corollary 5(b) and obtain

$$
\begin{aligned}
&\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I}^* \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \\
&\qquad \leq 9/2 \cdot \|D_{\sqrt{\beta}}^{-1}\|^2 \cdot \|D_{\sqrt{\pi}} \ddot{H}^* D_{\sqrt{\pi}}\|^2 + 3/2 \cdot \max_k \|e_k^* D_{\sqrt{\beta}}^{-1} \ddot{H}^* D_{\sqrt{\pi\cdot\beta}}\|^2 \\
&\qquad \leq 9/2 \cdot \underline{\beta}^{-1} \cdot (\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \cdot \varepsilon^2/2 + \|Z D_{\sqrt{\pi\cdot\beta}}\|)^2 \cdot \|\Psi D_{\sqrt{\pi}}\|^2 \\
&\qquad\qquad + 3/2 \cdot (\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \cdot \varepsilon^2/2 + \|Z D_{\sqrt{\pi\cdot\beta}}\|)^2 \\
&\qquad \leq (9/2 \cdot \underline{\beta}^{-1} \cdot \|\Psi_{\sqrt{\pi\cdot\beta}}\|^2 + 3/2) \cdot (\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \cdot \varepsilon^2/2 + \|Z D_{\sqrt{\pi\cdot\beta}}\|)^2. \qquad (96)
\end{aligned}
$$

Similar to above we observe the two cases $\delta > \delta_\circ$ and $\delta \leq \delta_\circ$ separately. Note that $\underline{\alpha} \geq 1 - \delta_\circ^2/2 \geq 1 - 1/(2C^2) \geq 17/18$ and that $\delta_\circ \leq 1/C$. This yields

$$
\begin{aligned}
\delta > \delta_\circ: \quad &\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I}^* \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \\
&\qquad \leq 2 \cdot \Big(\underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/(4C) + 2 \cdot \underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/(4C) \Big)^2 \leq 18 \cdot \Big(\underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/(4C) \Big)^2, \\
\delta \leq \delta_\circ: \quad &\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I}^* \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \\
&\qquad \leq 2 \cdot 18^2/17^2 \cdot \underline{\alpha}^2 \cdot \Big(\underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/(4C) \cdot \gamma \nu^2/(2C) \cdot \delta + \delta \underline{\beta}^{1/2} \Big)^2 \\
&\qquad \leq 3 \cdot \underline{\alpha}^2 \underline{\beta} \delta^2 \leq 3/C \cdot \underline{\alpha}^2 \underline{\beta} \delta.
\end{aligned}
$$

which concludes to

$$
\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I}^* \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \leq 3/C \cdot \underline{\alpha}^2 \underline{\beta} \cdot \min\{\delta_\circ, \delta\}. \qquad (97)
$$

With the bound we established on (78), we get

$$
\begin{aligned}
&\Big(\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I}^* \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\|^{1/2} \\
&\qquad\qquad + \Gamma \|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I} \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\|^{1/2} \Big)^2 \cdot \|D_{\beta}^{-1}\| \\
&\qquad \leq \Big(\sqrt{3/C} \cdot \underline{\alpha} \cdot \min\{\delta_\circ, \delta\}^{1/2} + \sqrt{2} \Gamma \underline{\beta}^{-1/2} \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\| \Big)^2. \qquad (98)
\end{aligned}
$$

We use the inequality above, (92) and the probability bound from Lemma 4 to bound

$$\begin{aligned}\|\mathbb{E}[\hat{Y}]\| \le \delta_*/16 + 2\underline{\alpha}^{-1} \cdot \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}] \\ + \underline{\alpha}^{-2} \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} \ddot{H}_{I,I}^* \ddot{H}_{I,I} (\dot{\Psi}_I^* \dot{\Psi}_I)^{-1} R_I D_{\sqrt{\pi}}^{-1}]\| \\ \le \delta_*/16 + 7/(4C) \cdot \min\{\delta_\circ, \delta\} \\ + \underline{\alpha}^{-2} \cdot \left(\sqrt{3/C} \cdot \underline{\alpha} \min\{\delta_\circ, \delta\}^{1/2} + \sqrt{2}\Gamma \underline{\beta}^{-1/2} \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\|\right)^2.\end{aligned}$$

Again we distinguish between the cases cases $\delta > \delta_\circ$ and $\delta \le \delta_\circ$, and note as before that $\nu \le 1/3$, which leads to

$$\begin{aligned}\delta > \delta_\circ: \quad & \|\mathbb{E}[\hat{Y}]\| \le \delta_*/16 + 7/(4C) \cdot \delta_\circ + (\sqrt{3/C} + \sqrt{2/C} \cdot 7/36)^2 \cdot \gamma^{-1}\delta_\circ \le \gamma^{-1}\Delta/6, \\ \delta \le \delta_\circ: \quad & \|\mathbb{E}[\hat{Y}]\| \le \delta_*/16 + 7/(4C) \cdot \delta + (\sqrt{3/C} + 4/C)^2 \cdot \delta \le \Delta/6. \end{aligned} \tag{99}$$

Now we fulfilled all preliminaries for Bernstein's inequality 11. By choosing $t = m = \gamma^{-1}\Delta/16$ and $r = \gamma^{-1}\rho/2$ and performing a few simplifications we get the desired bound. ■

If we compare the result of the Lemma 7 and its counterpart in [24] we notice that the obtained bound is slightly greater by the scale of γ^{-1}. The reason for this is that we needed to bound

$$\|D_{\sqrt{\beta}}\| \cdot \|D_{\sqrt{\beta}}^{-1}\| \le (\bar{\beta}/\underline{\beta})^{1/2} \le c_{\max}/c_{\min} = \gamma^{-1}.$$

If necessary Lemma 7 can also be bounded by $(\bar{\beta}/\underline{\beta})^{1/2} \cdot \Delta/8$.

Lemma 8 *Given the conditions stated in Theorem 3, we have*

$$\mathbb{P}\left(\|\Phi A(D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1} e_\ell - \Phi D_{\sqrt{\beta}} e_\ell\| > \underline{\beta}^{1/2} \cdot \Delta/8\right) \le 28 \exp\left(-\frac{N(\Delta/16)^2}{2\rho^2 + \rho\Delta/16}\right).$$

Proof Similar to the proofs of Lemma 6 and 7 we use Bernstein's inequality again, although this time we define $\hat{Y}_n$ and Y_n not as matrices, but as independent random vectors with fixed index ℓ, yielding

$$\begin{aligned}\hat{Y}_n &:= \left[y_n y_n^* (\Psi_{\hat{I}_n}^\dagger)^* R_{\hat{I}_n} (D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1} \cdot \mathbb{1}_{\mathcal{B}(y_n)}(\hat{I}_n) - D_{\sqrt{\beta}}\right] e_\ell, \\ Y_n &:= \left[y_n y_n^* (\Psi_{I_n}^\dagger)^* R_{I_n} (D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1} \cdot \mathbb{1}_{\mathcal{B}(y_n)}(I_n) - \operatorname{diag}(\mathbf{1}_{I_n}) D_\pi^{-1} D_{\sqrt{\beta}}\right] e_\ell. \end{aligned} \tag{100}$$

We can bound the ℓ_2-norm of $\hat{Y}_n$ and Y_n by

$$\begin{aligned}\max\left\{\|\hat{Y}_n\|, \|Y_n\|\right\} &\le \kappa S^2 c_{\max}^2 \|D_\pi^{-1}\| \|D_{\sqrt{\beta}}^{-1}\| \|D_\alpha^{-1}\| + S\|D_\pi^{-1}\| \\ &\le 3/4 \cdot \underline{\beta}^{1/2}\rho =: r.\end{aligned}$$

We follow the steps in (69) of Lemma 6 and obtain

$$\|\mathbb{E}[\hat{Y}]\| \le [2\mathbb{P}(\mathcal{H}^c) + \mathbb{P}(\mathcal{G}^c)] \cdot \rho + \|\mathbb{E}\left[\mathbb{1}_{\mathcal{G}}(I) Y\right]\| \le \delta_*/32 + \|\mathbb{E}\left[\mathbb{1}_{\mathcal{G}}(I) Y\right]\|. \tag{101}$$

We continue with recalling the expression from (74b)

$$\dot\Phi_I^*\dot\Psi_I(\dot\Psi_I^*\dot\Psi_I)^{-1} = (D_\alpha)_{I,I} + \ddot{H}_{I,I}(D_\beta^{-1})_{I,I} + \ddot{H}_{I,I}(\dot\Psi_I^*\dot\Psi_I)^{-1}\ddot{H}_{I,I}(D_\beta^{-1})_{I,I}.$$

Following the steps of (71) and substitute the expression above yields

$$\begin{aligned}\|\mathbb{E}\left[\mathbb{1}_{\mathcal{G}}(I)Y\right]\| &= \|\mathbb{E}_{\mathcal{G}}[\dot\Phi R_I^*(\dot\Phi_I^*\Psi_I^{\dagger *} - (D_\alpha)_{I,I})R_I e_\ell]\|/(\alpha_\ell\pi_\ell)\\ &\le \|\mathbb{E}_{\mathcal{G}}[\dot\Phi R_I^*\ddot{H}_{I,I}(\dot\Psi_I^*\dot\Psi_I)^{-1}R_I e_\ell]\|/(\underline{\alpha}\pi_\ell)\\ &\le \Big(\|\mathbb{E}_{\mathcal{G}}[\dot\Phi R_I^*\ddot{H}_{I,I}R_I e_\ell]\| + \|\mathbb{E}_{\mathcal{G}}[\dot\Phi R_I^*\ddot{H}_{I,I}(\dot\Psi_I^*\dot\Psi_I)^{-1}\ddot{H}_{I,I}R_I e_\ell]\|\Big)/(\underline{\alpha}\beta_\ell\pi_\ell).\end{aligned} \tag{102}$$

For the following step we want to use $\mathbb{I} = \mathbb{I}_{\ell^c} + e_\ell e_\ell^*$. Since $\ddot{H} = D_{\sqrt{\beta}}(\Phi\mathcal{E} - ZD_\alpha)^*\Psi D_{\sqrt{\beta}}$ has zero diagonal and for any matrix $M_{\ell\ell} = e_\ell M e_\ell$, we get that $e_\ell R_I^*\ddot{H}_{I,I}R_I e_\ell = 0$. So applying Corollary 5(c) to the first term results in

$$\begin{aligned}\|\mathbb{E}_{\mathcal{G}}[\dot\Phi R_I^*\ddot{H}_{I,I}R_I e_\ell]\| &= \|\mathbb{E}_{\mathcal{G}}\left[\Phi\mathbb{I}_{\ell^c}D_{\sqrt{\beta}}R_I^*R_I D_{\sqrt{\beta}}(\Phi\mathcal{E} - ZD_\alpha)^*\mathbb{1}_I(\ell)\right]\cdot\psi_\ell\beta_\ell^{1/2}\|\\ &\le \pi_\ell\cdot\|\Phi\mathbb{I}_{\ell^c}D_{\sqrt{\pi\cdot\beta}}\|\cdot\|(\Phi\mathcal{E} - ZD_\alpha)D_{\sqrt{\pi\cdot\beta}}\|\cdot\|\psi_\ell\|\cdot\beta_\ell^{1/2}\\ &\le \pi_\ell\cdot\|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot(\|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\varepsilon^2/2 + \|ZD_{\sqrt{\pi\cdot\beta}}\|)\cdot\beta_\ell^{1/2}.\end{aligned} \tag{103}$$

Recall that $H_{I,\ell} = R_I H e_\ell$ and $H_{\ell,I} = e_\ell^* H R_I^*$. For the second term in (102) we use the following bounds, which are addressed in Appendix A.2:

$$\begin{aligned}\|\mathbb{E}[\ddot{H}_{\ell,I}\ddot{H}_{\ell,I}^*\mathbb{1}_I(\ell)]\|^{1/2} &\le (\beta_\ell\pi_\ell)^{1/2}\cdot\varepsilon\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|,\\ \|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}\ddot{H}_{I,I}\ddot{H}_{I,I}^*\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\|^{1/2} &\le 3\cdot\min\{1,\delta\}\cdot(\pi_\ell\beta_\ell)^{1/2},\\ \|\mathbb{E}[\ddot{H}_{I,\ell}^*\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|^{1/2} &\le (\beta_\ell\pi_\ell)^{1/2}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|.\end{aligned}$$

We take the decomposition $\mathbb{I} = \mathbb{I}_{\ell^c} + e_\ell e_\ell^*$ again and employ Lemma 13 on both terms to obtain

$$\begin{aligned}&\|\mathbb{E}_{\mathcal{G}}[\dot\Phi R_I^*\ddot{H}_{I,I}(\dot\Psi_I^*\dot\Psi_I)^{-1}\ddot{H}_{I,I}R_I e_\ell]\|\\ &\le \|\phi_\ell\beta_\ell^{1/2}\|\cdot\|\mathbb{E}_{\mathcal{G}}[\ddot{H}_{\ell,I}\cdot(\dot\Psi_I^*\dot\Psi_I)^{-1}\cdot\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|\\ &\qquad + \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\|D_{\sqrt{\beta}}\|\cdot\|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\cdot(\dot\Psi_I^*\dot\Psi_I)^{-1}\cdot\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|\\ &\le \beta_\ell^{1/2}\cdot\|\mathbb{E}[\ddot{H}_{\ell,I}\ddot{H}_{\ell,I}^*\mathbb{1}_I(\ell)]\|^{1/2}\cdot\|(\dot\Psi_I^*\dot\Psi_I)^{-1}\|\cdot\|\mathbb{E}[\ddot{H}_{I,\ell}^*\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|^{1/2} + \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\bar\beta^{1/2}\times\\ &\quad\times\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\|^{1/2}\cdot\|(\dot\Psi_I^*\dot\Psi_I)^{-1}\|\cdot\|\mathbb{E}[\ddot{H}_{I,\ell}^*\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|^{1/2}\\ &\le \Big(\|\mathbb{E}[\ddot{H}_{\ell,I}\ddot{H}_{\ell,I}^*\mathbb{1}_I(\ell)]\|^{1/2} + \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\|^{1/2}\Big)\times\\ &\qquad\qquad\times 4/3\cdot\bar\beta^{1/2}\underline{\beta}^{-1}\cdot\|\mathbb{E}[\ddot{H}_{I,\ell}^*\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|^{1/2}\\ &\le \beta_\ell\pi_\ell\cdot\min\{\sqrt{2},\delta\}\cdot 4/3\cdot\underline{\beta}^{-1/2}\gamma^{-1}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|\cdot(\|\Psi D_{\sqrt{\pi\cdot\beta}}\| + 3\cdot\|\Phi D_{\sqrt{\pi\cdot\beta}}\|),\end{aligned} \tag{104}$$

where in the last inequality we used that $(\bar\beta/\underline{\beta})^{1/2} \le c_{\max}/c_{\min} = \gamma^{-1}$ and that $\varepsilon \le \min\{\sqrt{2},\delta\}$. If we substitute these bounds into (102) and (101) we get

$$\begin{aligned}\|\mathbb{E}[\hat{Y}]\| \le \underline{\beta}^{1/2}\cdot\delta_*/32 + \underline{\alpha}^{-1}\beta_\ell^{-1/2}\cdot\|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot(\|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\varepsilon^2/2 + \|ZD_{\sqrt{\pi\cdot\beta}}\|)\\ + 4/3\cdot\underline{\alpha}^{-1}\cdot\min\{\sqrt{2},\delta\}\cdot\underline{\beta}^{-1/2}\gamma^{-1}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|\cdot(\|\Psi D_{\sqrt{\pi\cdot\beta}}\| + 3\cdot\|\Phi D_{\sqrt{\pi\cdot\beta}}\|).\end{aligned}$$

Similar to Lemma 6 and 7 we distinguish between the cases $\delta > \delta_\circ$ and $\delta \leq \delta_\circ$. In the first regime we use that $\max\{\|\Psi D_{\sqrt{\pi\cdot\beta}}\|, \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\} \leq \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)$ and thus $\|ZD_{\sqrt{\pi\cdot\beta}}\| \leq \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(2C)$, which yields

$$\begin{aligned}\|\mathbb{E}[\hat{Y}]\| &\leq \underline{\beta}^{1/2}\cdot\frac{\delta_*}{32} + \underline{\beta}^{1/2}\cdot\frac{\gamma\nu^2}{C}\cdot\frac{\underline{\alpha}}{16C}\cdot(\gamma+2\gamma+4/3\cdot\sqrt{2}\cdot(1+3))\\ &\leq \underline{\beta}^{1/2}\cdot\frac{\delta_*}{32} + \underline{\beta}^{1/2}\cdot\delta_\circ\cdot\frac{11\underline{\alpha}}{16C} \leq \underline{\beta}^{1/2}\cdot\frac{\Delta}{16}.\end{aligned} \tag{105}$$

In the second regime, where $\delta \leq \delta_\circ$, we can use $\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \leq \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)$, while due to (63) and (10) we have $\|\Psi D_{\sqrt{\pi\cdot\beta}}\| \leq \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/C$ and $\|ZD_{\sqrt{\pi\cdot\beta}}\| \leq \delta\underline{\beta}^{1/2}$ resp., which results in

$$\begin{aligned}\|\mathbb{E}[\hat{Y}]\| &\leq \underline{\beta}^{1/2}\cdot\frac{\delta_*}{32} + \underline{\beta}^{1/2}\cdot\delta\cdot\frac{\gamma\nu}{C}\cdot\left(\frac{\underline{\alpha}\gamma\nu}{16C}\cdot 1/\sqrt{2}+1+4/3\cdot\left(\frac{\underline{\alpha}\nu}{C}+\frac{3\underline{\alpha}\nu}{4C}\right)\right)\\ &\leq \underline{\beta}^{1/2}\cdot\frac{\delta_*}{32} + \underline{\beta}^{1/2}\cdot\delta\cdot\frac{18\gamma\nu}{16C} \leq \underline{\beta}^{1/2}\cdot\frac{\Delta}{16}.\end{aligned} \tag{106}$$

Now we fulfilled all preliminaries for Bernstein's inequality 11. By choosing $t = m = \underline{\beta}^{1/2}\cdot\Delta/16$ and $r = 3/4\cdot\underline{\beta}^{1/2}\cdot\rho$ and performing a few simplifications we get the desired bound. ■

Lemma 9 *Given the conditions stated in Theorem 3 for* $\Lambda := \max\left\{\frac{\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}}{4C}, \|\Psi D_{\sqrt{\pi\cdot\beta}}\|\right\}$, *we get*

$$\mathbb{P}(\Lambda\cdot\|\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}B(D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1}e_\ell\| > \underline{\beta}^{1/2}\cdot\frac{3\Delta}{16}) \leq 28\exp\left(-\frac{N(\Delta/16)^2}{2\rho^2+\rho\Delta/16}\right).$$

Proof As in Lemma 8 we use the vector Bernstein inequality and express $(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}B(D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1}e_\ell$ as a sum of independent random vectors $\hat{Y}_n$. Concretely we set

$$\hat{Y}_n := \Lambda\cdot\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}R^*_{\hat{I}_n}\Psi^\dagger_{\hat{I}_n}y_ny_n^*\Psi^{\dagger *}_{\hat{I}_n}R_{\hat{I}_n}(D_{\sqrt{\beta}\cdot\pi\cdot\alpha})^{-1}\mathbb{I}_{\mathcal{B}(y_n)}(\hat{I}_n)e_\ell. \tag{107}$$

Analogously we define Y_n , but using the correct support I_n instead of $\hat{I}_n$. We can bound the ℓ_2-norm of these vectors by

$$\max\left\{\|\hat{Y}_n\|, \|Y_n\|\right\} \leq \Lambda\kappa^2S^2c^2_{\max}\|D_\alpha^{-2}\|\|D_\beta^{-1}\|\|D_\pi^{-3/2}\| \leq 3/4\cdot\underline{\beta}^{1/2}\rho =: r. \tag{108}$$

Apart from a few slight modifications we structure the proof as in Lemma 7. Similar to (88) we obtain

$$\Lambda^{-1}\cdot\|\mathbb{E}[\hat{Y}]\| \leq \frac{\delta_*}{32} + \|\mathbb{E}_{\mathcal{G}}[(D^{-1}_{\sqrt{\pi}\cdot\alpha})\mathbb{I}_{\ell^c}R_I^*(D^{-1}_{\sqrt{\beta}})_{I,I}\Psi_I^\dagger\dot{\Phi}_I\dot{\Phi}_I^*(\Psi_I^\dagger)^*(D^{-1}_{\sqrt{\beta}})_{I,I}R_Ie_\ell]\|/(\pi_\ell\alpha_\ell). \tag{109}$$

Note that $\mathbb{I}_{\ell^c}De_\ell = 0$ for every diagonal matrix D, so we can rewrite the expression above as in (89) and get

$$\begin{aligned}&(D^{-1}_{\sqrt{\pi}\cdot\alpha})\mathbb{I}_{\ell^c}(D^{-1}_{\sqrt{\beta}})_{I,I}\Psi_I^\dagger\dot{\Phi}_I\dot{\Phi}_I^*(\Psi_I^\dagger)^*(D^{-1}_{\sqrt{\beta}})_{I,I}R_Ie_\ell\\ &\quad= (D^{-1}_{\sqrt{\pi}\cdot\alpha})\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\dot{\Psi}_I^*\dot{\Phi}_I\dot{\Phi}_I^*\dot{\Psi}_I(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_Ie_\ell\\ &\quad= (D^{-1}_{\sqrt{\pi}\cdot\alpha})\mathbb{I}_{\ell^c}R_I^*(D_\alpha)_{I,I}\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_Ie_\ell + (D^{-1}_{\sqrt{\pi}\cdot\alpha})\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}^*_{I,I}(D_\alpha)_{I,I}R_Ie_\ell\\ &\qquad\qquad + \mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}^*_{I,I}\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_Ie_\ell.\end{aligned} \tag{110}$$

We estimate the expectation of these three terms separately, beginning with the first one. We use the decomposition of the identity $\mathbb{I}_S = \left(\ddot{H}_{I,I} + \dot{\Psi}_I^*\dot{\Psi}_I\right)(D_\beta)_{I,I}^{-1}$ as in (72). By replacing $\dot{\Phi}$ with $(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}$ we can take the bounds we obtained in (103) and (104), and get

$$\begin{aligned}
&\mathbb{E}_\mathcal{G}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_Ie_\ell]\| \\
&\le \left(\|\mathbb{E}_\mathcal{G}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}R_Ie_\ell]\| + \|D_{\sqrt{\beta}}\|\cdot\|\mathbb{E}_\mathcal{G}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|\right)\cdot\beta_\ell^{-1} \\
&\le \pi_\ell\beta_\ell^{-1/2}\cdot\left(\|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\varepsilon^2/2 + \|ZD_{\sqrt{\pi\cdot\beta}}\| + 4\cdot\min\{1,\delta\}\cdot\beta_\ell^{1/2}\underline{\beta}^{-1/2}\gamma^{-1}\cdot\|\Psi_{\sqrt{\pi\cdot\beta}}\|\right), \quad (111)
\end{aligned}$$

We continue by distinguishing between the both regimes $\delta > \delta_\circ$ and $\delta \le \delta_\circ$. In the first regime we use that $\max\{\|\Psi D_{\sqrt{\pi\cdot\beta}}\|, \|\Phi D_{\sqrt{\pi\cdot\beta}}\|\} \le \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)$ and thus $\|ZD_{\sqrt{\pi\cdot\beta}}\| \le \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(2C)$, while in the second regime, where $\delta \le \delta_\circ$, we can use $\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \le \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)$, while due to (63) and (10) we have $\|\Psi D_{\sqrt{\pi\cdot\beta}}\| \le \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/C$ and $\|ZD_{\sqrt{\pi\cdot\beta}}\| \le \delta\underline{\beta}^{1/2}$. This yields

$$\begin{aligned}
\delta > \delta_\circ:\quad &\|\mathbb{E}_\mathcal{G}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_Ie_\ell]\| \\
&\qquad\le \pi_\ell\beta_\ell^{-1/2}\cdot\left(3\cdot\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C) + 4\cdot\beta_\ell^{1/2}\cdot\underline{\alpha}\nu/(4C)\right) \\
&\qquad\le 1/20\cdot\pi_\ell\underline{\alpha}\gamma^{-1}\nu^{-1}\cdot\delta_\circ, \quad (112)\\
\delta \le \delta_\circ:\quad &\mathbb{E}_\mathcal{G}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_Ie_\ell]\| \\
&\qquad\le \pi_\ell\beta_\ell^{-1/2}\cdot\left(\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C)\cdot\delta\delta_\circ/2 + \delta\underline{\beta}^{1/2} + 4\cdot\delta\beta_\ell^{1/2}\cdot\underline{\alpha}\nu/C\right) \\
&\qquad\le 11/10\cdot\pi_\ell\cdot\delta. \quad (113)
\end{aligned}$$

For the second term we employ that $\ddot{H}_{I,I}^*R_Ie_\ell = R_I\mathbb{I}_{\ell^c}\ddot{H}^*e_\ell\mathbb{1}_I(\ell)$, which holds true since $\ddot{H}$ has zero diagonal. Similar to above we use $\mathbb{I}_S = (D_\beta)_{I,I}^{-1}\left(\ddot{H}_{I,I} + \dot{\Psi}_I^*\dot{\Psi}_I\right)$ in order to apply Corollary 5(c) to the first term and Lemma 13 to the second one:

$$\begin{aligned}
&\|\mathbb{E}_\mathcal{G}[(D_{\sqrt{\pi}\cdot\alpha})^{-1}\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*(D_\alpha)_{I,I}R_Ie_\ell]\| \\
&\le \|\mathbb{E}_\mathcal{G}[(D_{\sqrt{\pi}\cdot\alpha})^{-1}]\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_I\mathbb{I}_{\ell^c}D_{\sqrt{\pi}}^{-1}\mathbb{1}_I(\ell)]\|\cdot\|D_{\sqrt{\pi}}\ddot{H}^*e_\ell\|\cdot\alpha_\ell \\
&\le \underline{\alpha}^{-1}\cdot\|z_\ell\beta_\ell^{1/2}\|\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|\cdot\Big(\underline{\beta}^{-1}\cdot\|\mathbb{E}_\mathcal{G}\left[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*R_I\mathbb{I}_{\ell^c}D_{\sqrt{\pi}}^{-1}\mathbb{1}_I(\ell)\right]\| \\
&\qquad\qquad + \underline{\beta}^{-1/2}\cdot\|\mathbb{E}_\mathcal{G}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\cdot(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\cdot\mathbb{I}_{\ell^c}R_ID_{\sqrt{\pi}}^{-1}\mathbb{1}_I(\ell)]\|\Big) \\
&\le \underline{\alpha}^{-1}\underline{\beta}^{-1}\cdot\|z_\ell\beta_\ell^{1/2}\|\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|\times \\
&\qquad\times(\pi_\ell + 4/3\cdot\underline{\beta}^{-1/2}\pi_\ell^{1/2}\cdot\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\|^{1/2}). \quad (114)
\end{aligned}$$

As above we discuss the two cases $\delta > \delta_\circ$ and $\delta \le \delta_\circ$ separately. In the appendix in (137) and (138) we show that

$$\begin{aligned}
\delta > \delta_\circ:\quad &\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\| \le 9\cdot\pi_\ell\cdot(\underline{\alpha}\gamma\nu\beta_\ell^{1/2}/C)^2, \\
\delta \le \delta_\circ:\quad &\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\| \le 9\cdot\pi_\ell\cdot(2\gamma\nu\bar{\beta}^{1/2}/C)^2.
\end{aligned}$$

With the bound above and the fact that $(\bar{\beta}/\underline{\beta})^{1/2} \leq \gamma^{-1}$, we get

$$\delta > \delta_\circ : \quad \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi}\cdot\alpha})^{-1}\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*(D_\alpha)_{I,I}R_Ie_\ell]\| \leq \varepsilon\cdot\nu/(4C)\cdot(\pi_\ell + 4\cdot\underline{\beta}^{-1/2}\pi_\ell\cdot\underline{\alpha}\gamma\nu\beta_\ell^{1/2}/C) \leq 1/20\cdot\pi_\ell\gamma^{-1}\nu^{-1}\cdot\delta_\circ, \quad (115)$$

$$\delta \leq \delta_\circ : \quad \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi}\cdot\alpha})^{-1}\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*(D_\alpha)_{I,I}R_Ie_\ell]\| \leq \delta\cdot\gamma\nu/C\cdot(\pi_\ell + 6\cdot\underline{\beta}^{-1/2}\pi_\ell\cdot\gamma\nu\bar{\beta}^{1/2}/C) \leq 2/C\cdot\pi_\ell\cdot\delta. \quad (116)$$

To the third term of (110) we apply the identity $\mathbb{I}_S = \left(\ddot{H}_{I,I} + \dot{\Psi}_I^*\dot{\Psi}_I\right)(D_\beta)_{I,I}^{-1}$ on the left-hand and on the right-hand side following the same procedure as in (95), which leads to

$$\begin{aligned}
&\|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi}\cdot\alpha})^{-1}\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_Ie_\ell\mathbb{1}_I(\ell)]\| \\
&\leq \underline{\alpha}^{-1}\underline{\beta}^{-1/2}\cdot\|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}\cdot\alpha})^{-1}\mathbb{I}_{\ell^c}R_I^*(\ddot{H}_{I,I} + \dot{\Psi}_I^*\dot{\Psi}_I)\cdot(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\times \\
&\qquad\times(\ddot{H}_{I,I}^* + \dot{\Psi}_I^*\dot{\Psi}_I)R_Ie_\ell\mathbb{1}_I(\ell)\|\cdot\beta_\ell^{-1} \\
&\leq \underline{\alpha}^{-1}\underline{\beta}^{-1/2}\beta_\ell^{-1}\cdot\big(\|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}^*\ddot{H}_{I,I}R_Ie_\ell]\| \\
&\qquad + \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}R_Ie_\ell]\| \\
&\qquad + \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}R_Ie_\ell]\| \\
&\qquad + \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}R_Ie_\ell]\|\big).
\end{aligned} \quad (117)$$

Now we have made all the necessary preparations to bound $\|\mathbb{E}[\hat{Y}]\|$, since

$$\begin{aligned}
\|\mathbb{E}[\hat{Y}]\| &\leq \Lambda\cdot\delta_*/32 + \Lambda\cdot\pi_\ell^{-1}\cdot\underline{\alpha}^{-1}\times \\
&\times\Big(\|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_Ie_\ell]\| + \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*R_Ie_\ell]\| \\
&\quad + \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_Ie_\ell\mathbb{1}_I(\ell)]\|\Big).
\end{aligned} \quad (118)$$

Again we distinguish between the two cases $\delta > \delta_\circ$ and $\delta \leq \delta_\circ$. In the appendix we have further bounded (117), which we plug with the bounds of (112), (113), (115) and (116) into (118), yielding

$$\begin{aligned}
\delta > \delta_\circ : \quad \|\mathbb{E}[\hat{Y}]\| &\leq \underline{\beta}^{1/2}\cdot\delta_*/32 + \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/C\cdot\underline{\alpha}^{-1}\cdot\big(1/20\cdot\gamma^{-1}\nu^{-1}\cdot\delta_\circ \\
&\qquad + 1/20\cdot\gamma^{-1}\nu^{-1}\cdot\delta_\circ + 5/2\cdot\gamma^{-1}\nu^{-1}\cdot\delta_\circ\big) \\
&\leq \underline{\beta}^{1/2}\cdot\delta_*/32 + 26/(10C)\cdot\underline{\beta}^{1/2}\cdot\delta_\circ, && (119) \\
\delta \leq \delta_\circ : \quad \|\mathbb{E}[\hat{Y}]\| &\leq \underline{\beta}^{1/2}\cdot\delta_*/32 + \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/C\cdot\underline{\alpha}^{-1}\cdot(11/10\cdot\delta + 2/C\cdot\delta + 11/10\cdot\delta) \\
&\leq \underline{\beta}^{1/2}\cdot\delta_*/32 + 23/(10C)\cdot\underline{\beta}^{1/2}\cdot\delta \leq \underline{\beta}^{1/2}\cdot\Delta/16. && (120)
\end{aligned}$$

Now we fulfilled all preliminaries for Bernstein's inequality 11. By choosing $t = m = \underline{\beta}^{1/2}\cdot\Delta/16$ and $r = 3/4\cdot\underline{\beta}^{1/2}\cdot\rho$ and performing a few simplifications we get the desired bound.

■

With this we have proved the four inequalities which are needed for the main theorem. In the following chapter we summarize our results of this thesis and provide an outlook for further related research work.

6. Discussion

In this thesis we showed convergence of the MOD and ODL algorithm for data models with non-uniform distribution of the supports of sparse coefficients in combination with non-homogeneous distribution of the coefficient amplitudes. As part of our future work we want to consider thresholding in a similar manner as in [19] to improve the bound on the probability that thresholding recovers the correct support. We also want to generalise the convergence result of the ITKrM algorithm in [21] to the signal model in Definition 2. There it should not even be necessary to change the conditions of [21]. Since D_β does not get trapped between two non-diagonal matrices and therefore can still be canceled out by scaling with D_β^{-1}.

Appendix A. Appendix

In this section we detail some calculations used in the proofs of Lemma 6 to 9 and restate some theorems and lemmas which are essential for the thesis.

A.1 Auxiliary calculations for Lemma 6 and Lemma 7

We start by elaborating on two inequalities which help to bound $\|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{И}_{I,I} \ddot{И}_{I,I}^* R_I D_{\sqrt{\pi}}^{-1}]\|^{\frac{1}{2}}$ in (77). We recall that $И = (\Phi\mathcal{E} - ZD_\alpha)^*\Psi = (\Psi\mathcal{E} - Z)^*\Psi$ with $\mathcal{E}_{kk} = \langle \psi_k, z_k \rangle$ and $\ddot{И} = D_{\sqrt{\beta}} И D_{\sqrt{\beta}}$. Then, since $\mathbb{I} - \psi_k^*\psi_k$ is an orthogonal projection, we obtain

$$
\begin{aligned}
\|D_{\sqrt{\pi}} \ddot{И} D_{\sqrt{\pi}}\| &= \|D_{\sqrt{\pi}} \ddot{И}^* D_{\sqrt{\pi}}\| \\
&\leq \|(\Phi\mathcal{E} - ZD_\alpha) D_{\sqrt{\pi\cdot\beta}}\| \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\| = (\|\Phi D_{\sqrt{\pi\cdot\beta}} \mathcal{E} - Z D_{\sqrt{\pi\cdot\beta}} D_\alpha\|) \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\| \\
&\leq (\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \cdot \varepsilon^2/2 + \|Z D_{\sqrt{\pi\cdot\beta}}\|) \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\|, \qquad (121)\\
\|e_k^* \ddot{И} D_{\sqrt{\pi}}\| &\leq \|(\psi_k \langle \psi_k, z_k \rangle - z_k)\beta_k^{1/2}\| \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\| \\
&= \|(\psi_k \psi_k^* - \mathbb{I}) z_k \beta_k^{1/2}\| \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\| \leq \|z_k \beta_k^{1/2}\| \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\| \\
&\leq \min\{\varepsilon \bar{\beta}^{1/2}, \delta \underline{\beta}^{1/2}\} \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\|, \qquad (122)\\
\|e_k^* D_{\sqrt{\beta}}^{-1} \ddot{И}^* D_{\sqrt{\pi}}\| &\leq \|\psi\| \cdot \|(\Phi\mathcal{E} - ZD_\alpha) D_{\sqrt{\pi\cdot\beta}}\| \leq \|\Phi D_{\sqrt{\pi\cdot\beta}}\| \cdot \varepsilon^2/2 + \|Z D_{\sqrt{\pi\cdot\beta}}\|. \qquad (123)
\end{aligned}
$$

We continue by using the fact that $\|\Phi D_{\sqrt{\pi\cdot\beta}}\| \leq \underline{\alpha}\gamma\nu^2\underline{\beta}^{1/2}/(4C) \leq \underline{\beta}^{1/2} \leq \bar{\beta}^{1/2}$ and $\varepsilon \leq \min\{\sqrt{2}, \delta\}$, which we showed in (83), to obtain a bound for (77) by using Corollary 5, meaning

$$
\begin{aligned}
&2 \cdot \|\mathbb{E}[D_{\sqrt{\pi}}^{-1} R_I^* \ddot{И}_{I,I} \ddot{И}_{I,I}^* R_I D_{\sqrt{\pi}}^{-1}]\| \\
&\leq 9 \cdot \|D_{\sqrt{\pi}} \ddot{И} D_{\sqrt{\pi}}\|^2 + 3 \cdot \max_k \|e_k^* \ddot{И} D_{\sqrt{\pi}}\|^2 \\
&\leq \Big(9 \cdot \|\Phi D_{\sqrt{\pi\cdot\beta}}\|^2 \cdot \varepsilon^4/4 + 9 \cdot \|\Phi D_{\sqrt{\pi\cdot\beta}}\| \cdot \varepsilon^2 \cdot \|Z D_{\sqrt{\pi\cdot\beta}}\| + 9 \cdot \|Z D_{\sqrt{\pi\cdot\beta}}\|^2 \\
&\qquad + 3 \cdot \min\{\varepsilon \bar{\beta}^{1/2}, \delta \underline{\beta}^{1/2}\}^2\Big) \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2 \\
&\leq \Big(9/2 \cdot \min\{\varepsilon \bar{\beta}^{1/2}, \delta \underline{\beta}^{1/2}\}^2 + 18 \cdot \underline{\beta}^{1/2}\varepsilon \cdot \|Z D_{\sqrt{\pi\cdot\beta}}\| + 9 \cdot \|Z D_{\sqrt{\pi\cdot\beta}}\|^2 \\
&\qquad + 3 \cdot \min\{\varepsilon \bar{\beta}^{1/2}, \delta \underline{\beta}^{1/2}\}^2\Big) \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2 \\
&\leq (3 \cdot \|Z D_{\sqrt{\pi\cdot\beta}}\| + 3 \cdot \min\{\varepsilon \bar{\beta}^{1/2}, \delta \underline{\beta}^{1/2}\})^2 \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2 \\
&\leq (3 \cdot \|Z D_{\sqrt{\pi\cdot\beta}}\| + 3 \cdot \varepsilon \bar{\beta}^{1/2})^2 \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2. \qquad (124)
\end{aligned}
$$

Next we want to establish a bound for $\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1} R_I^* \ddot{H}_{I,I} \ddot{H}_{I,I} R_I (D_{\sqrt{\pi\cdot\beta}})^{-1}]\|^{\frac{1}{2}}$ in (78). We recall that $\ddot{H} = D_{\sqrt{\beta}} \Psi^* \Psi D_{\sqrt{\beta}} - D_\beta$ and note that $D_{\sqrt{\beta}} \Psi^* \Psi D_{\sqrt{\beta}}$ is a positive semidefinite matrix which yields

$$
\begin{aligned}
\|D_{\sqrt{\pi}} D_{\sqrt{\beta}}^{-1} \ddot{H} D_{\sqrt{\pi}}\| &= \|D_{\sqrt{\beta}}^{-1}\| \cdot \|D_{\sqrt{\pi\cdot\beta}} \Psi^* \Psi D_{\sqrt{\pi\cdot\beta}} - \operatorname{diag}(D_{\sqrt{\pi\cdot\beta}} \Psi^* \Psi D_{\sqrt{\pi\cdot\beta}})\| \\
&\leq \underline{\beta}^{-1/2} \cdot \|D_{\sqrt{\pi\cdot\beta}} \Psi^* \Psi D_{\sqrt{\pi\cdot\beta}}\| = \underline{\beta}^{-1/2} \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2, \qquad (125)\\
\|e_k^* D_{\sqrt{\beta}}^{-1} \ddot{H} D_{\sqrt{\pi}}\| &= \|(e_k^* - \psi_k^* \Psi) D_{\sqrt{\pi\cdot\beta}}\| \leq \|\psi_k^* \Psi D_{\sqrt{\pi\cdot\beta}}\| \leq \|\Psi D_{\sqrt{\pi\cdot\beta}}\|. \qquad (126)
\end{aligned}
$$

We apply Corollary 5(b) to (78) and use that $\|\Psi D_{\sqrt{\pi\cdot\beta}}\| \leq \underline{\alpha}\gamma\nu^2\underline{\beta}^{1/2}/C \leq \underline{\beta}^{1/2}/3$ holds true in both regimes, which results in

$$\begin{aligned}
&\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}R_I(D_{\sqrt{\pi\cdot\beta}})^{-1}]\| \\
&\qquad = \|\mathbb{E}[D_{\sqrt{\pi}}^{-1}R_I^*(D_{\sqrt{\beta}}^{-1}\ddot{H})_{I,I}(\ddot{H}D_{\sqrt{\beta}})_{I,I}^{-1}R_I D_{\sqrt{\pi}}^{-1}]\| \\
&\qquad \leq \frac{9}{2}\cdot\|D_{\sqrt{\pi}}D_{\sqrt{\beta}}^{-1}\ddot{H}D_{\sqrt{\pi}}\|^2 + \frac{3}{2}\cdot\max_k\|e_k^*D_{\sqrt{\beta}}^{-1}\ddot{H}D_{\sqrt{\pi}}\|^2 \\
&\qquad \leq \frac{9}{2}\cdot\underline{\beta}^{-1}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|^4 + \frac{3}{2}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2 \\
&\qquad \leq \frac{1}{2}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2 + \frac{3}{2}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2 = 2\|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2. \qquad (127)
\end{aligned}$$

A.2 Auxiliary calculations for Lemma 8 and Lemma 9

We begin with some fundamental bounds, which are used in both Lemmas. Therefore we note that, $\ddot{\mathit{H}}_{\ell,I} = e_\ell^*\ddot{\mathit{H}}R_I^*$ and $\ddot{H}_{I,\ell} = R_I\ddot{H}e_\ell$, and apply Corollary 5(c) by setting $W = V = e_\ell^*\ddot{\mathit{H}}$ for the first inequality and $W = V = \ddot{H}e_\ell$ for the second one

$$\begin{aligned}
\|\mathbb{E}[\ddot{\mathit{H}}_{\ell,I}\ddot{\mathit{H}}_{\ell,I}^*\mathbb{1}_I(\ell)]\| &\leq \pi_\ell\cdot(\|e_\ell^*\ddot{\mathit{H}}D_{\sqrt{\pi}}\|^2 + \|\ddot{\mathit{H}}_{\ell\ell}\|^2) \\
&= \pi_\ell\beta_\ell\cdot\|e_\ell^*(\Psi\mathcal{E} - Z)^*\Psi D_{\sqrt{\pi\cdot\beta}}\|^2 \\
&\leq \pi_\ell\beta_\ell\cdot\|z_\ell\|^2\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2 \leq \pi_\ell\beta_\ell\varepsilon^2\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2, \qquad (128)\\
\|\mathbb{E}[\ddot{\mathit{H}}_{I,\ell}^*\ddot{\mathit{H}}_{I,\ell}\mathbb{1}_I(\ell)]\| &\leq \pi_\ell\cdot(\|e_\ell^*\ddot{\mathit{H}}^*D_{\sqrt{\pi}}\|^2 + \|\ddot{\mathit{H}}_{\ell\ell}^*\|^2) \\
&= \pi_\ell\beta_\ell\cdot\|e_\ell^*\Psi^*(\Phi\mathcal{E} - ZD_\alpha)D_{\sqrt{\pi\cdot\beta}}\|^2 \\
&\leq \pi_\ell\beta_\ell\cdot(\|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\varepsilon^2/2 + \|ZD_{\sqrt{\pi\cdot\beta}}\|)^2, \qquad (129)\\
\|\mathbb{E}[\ddot{H}_{I,\ell}^*\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\| &\leq \pi_\ell\cdot(\|e_\ell^*\ddot{H}D_{\sqrt{\pi}}\|^2 + \|\ddot{H}_{\ell\ell}\|^2) \\
&\leq \pi_\ell\cdot\|\beta_\ell^{1/2}\psi_\ell\|^2\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2 \leq \pi_\ell\beta_\ell\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2. \qquad (130)
\end{aligned}$$

Next we want to bound $\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{\mathit{H}}_{I,I}^*\ddot{\mathit{H}}_{I,I}R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\|$. Hence we need the following calculations. We recall that $\dot{\mathit{H}} = \mathit{H}D_{\sqrt{\beta}}$ and obtain

$$\begin{aligned}
\max_k(D_{\sqrt{\beta}}^{-1}\ddot{\mathit{H}}_{k\ell}^*)^2 &= \max_k(e_k^*\Psi^*(\Psi\mathcal{E} - Z)D_{\sqrt{\beta}}e_\ell)^2 \leq \max_k(\|\psi_k\|\cdot\|\psi_\ell\langle\psi_\ell, z_\ell\rangle - z_\ell\|\cdot\beta_\ell^{1/2})^2 \\
&\leq \|(\psi_\ell\psi_\ell^* - \mathbb{I})z_\ell\|^2\cdot\beta_\ell \leq \|z_\ell\|^2\cdot\beta_\ell \leq \varepsilon^2\beta_\ell \leq \min\{2, \delta^2\}\cdot\beta_\ell. \qquad (131)
\end{aligned}$$

Following a similar approach as in (121) and (122) we get

$$\begin{aligned}
3 \cdot \|D_{\sqrt{\beta}}^{-1} D_{\sqrt{\pi}} \ddot{H}^* e_\ell\|^2 &\le 3 \cdot \|(D_{\sqrt{\pi \cdot \beta}})^{-1}\|^2 \cdot \|\Psi D_{\sqrt{\beta}}\|^2 \cdot \|\psi_\ell \langle \psi_\ell, z_\ell\rangle - z_\ell\|^2 \cdot \beta_\ell \\
&\le 3 \cdot \underline{\beta}^{-1} \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\|^2 \cdot \|(\psi_\ell \psi_\ell^* - \mathbb{I}) z_\ell\|^2 \cdot \beta_\ell \\
&\le 3 \cdot \underline{\beta}^{-1} \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\|^2 \cdot \|z_\ell\|^2 \cdot \beta_\ell \\
&\le 1/C \cdot \min\{1, \delta^2\} \cdot \beta_\ell, \qquad (132) \\
9/2 \cdot \|D_{\sqrt{\beta}}^{-1} D_{\sqrt{\pi}} \ddot{H}^* D_{\sqrt{\pi}}\|^2 &\le 9/2 \cdot \|D_{\sqrt{\beta}}^{-1}\|^2 \cdot \|D_{\sqrt{\pi}} \ddot{H} D_{\sqrt{\pi}}\|^2 \\
&\le 9/2 \cdot \underline{\beta}^{-1} \cdot (\|\Phi D_{\sqrt{\pi \cdot \beta}}\| \cdot \varepsilon^2/2 + \|Z D_{\sqrt{\pi \cdot \beta}}\|)^2 \cdot \|\Psi D_{\sqrt{\pi \cdot \beta}}\|^2 \\
&\le 1/C \cdot \min\{1, \delta^2\} \cdot \beta_\ell, \qquad (133) \\
3/2 \cdot \max_k \|e_k^* D_{\sqrt{\beta}}^{-1} \ddot{H}^* D_{\sqrt{\pi}}\|^2 &\le 3/2 \cdot \|\psi_k\|^2 \cdot \|(\Phi \mathcal{E} - Z D_\alpha) D_{\sqrt{\pi \cdot \beta}}\|^2 \\
&\le 3/2 \cdot (\|\Phi D_{\sqrt{\pi \cdot \beta}}\| \cdot \varepsilon^2/2 + \|Z D_{\sqrt{\pi \cdot \beta}}\|)^2 \\
&\le \min\{1/C, 2 \cdot \delta^2\} \cdot \beta_\ell. \qquad (134)
\end{aligned}$$

By using Corollary 5(d) we get

$$\begin{aligned}
&\|\mathbb{E}[(D_{\sqrt{\pi \cdot \beta}})^{-1} \mathbb{I}_{\ell^c} R_I^* \ddot{H}_{I,I}^* \ddot{H}_{I,I} R_I \mathbb{I}_{\ell^c} (D_{\sqrt{\pi \cdot \beta}})^{-1} \mathbb{1}_I(\ell)]\| \\
&\quad \le 3/2 \cdot \pi_\ell \cdot \big(3 \cdot \|D_{\sqrt{\beta}}^{-1} D_{\sqrt{\pi}} \ddot{H}^* e_\ell\|^2 + \max_k (D_{\sqrt{\beta}}^{-1} \ddot{H}_{k\ell}^*)^2 \\
&\qquad\qquad + 9/2 \cdot \|D_{\sqrt{\beta}}^{-1} D_{\sqrt{\pi}} \ddot{H}^* D_{\sqrt{\pi}}\|^2 + 3/2 \cdot \max_k \|e_k^* D_{\sqrt{\beta}}^{-1} \ddot{H}^* D_{\sqrt{\pi}}\|^2\big) \\
&\quad \le 9 \cdot \min\{4/11, \delta^2\} \cdot \pi_\ell \beta_\ell. \qquad (135)
\end{aligned}$$

We take a similar approach for $\|\mathbb{E}[(D_{\sqrt{\pi \cdot \beta}})^{-1} \mathbb{I}_{\ell^c} R_I^* \ddot{H}_{I,I} \ddot{H}_{I,I}^* R_I \mathbb{I}_{\ell^c} (D_{\sqrt{\pi \cdot \beta}})^{-1} \mathbb{1}_I(\ell)]\|$. We distinguish between the two cases $\delta > \delta_\circ$ and $\delta \le \delta_\circ$. While in the first regime we simply get that

$$\max_k (D_{\sqrt{\beta}}^{-1} \ddot{H})_{k\ell}^2 \le \max_{i \ne j} |\langle \psi_i, \psi_j \rangle \beta_j^{1/2}|^2 \le \mu_{\sqrt{\beta}}(\Psi)^2 \le (\underline{\alpha} \gamma \nu^2 \underline{\beta}^{1/2}/C)^2,$$

the second regime takes further attention. Following a similar approach as in (64) leads to

$$\begin{aligned}
\max_k (D_{\sqrt{\beta}}^{-1} \ddot{H})_{k\ell}^2 &\le \underline{\beta}^{-1} \cdot \max_{i \ne j} |\beta_i^{1/2} \langle \psi_i, \psi_j \rangle \beta_j^{1/2}|^2 \\
&= \underline{\beta}^{-1} \cdot \max_{i \ne j} \Big(|\beta_i^{1/2} \langle \phi_i, \phi_j \rangle \beta_j^{1/2}| + |\beta_i^{1/2} \langle \phi_i, \psi_j - \phi_j \rangle \beta_j^{1/2}| \\
&\qquad\qquad + |\beta_i^{1/2} \langle \psi_i - \phi_i, \psi_j \rangle \beta_j^{1/2}|\Big)^2 \\
&\le \underline{\beta}^{-1} \cdot (\bar{\beta}^{1/2} \mu_{\sqrt{\beta}}(\Phi) + 2\bar{\beta}^{1/2} \max_k \|(\psi_k - \phi_k) \beta_k^{1/2}\|)^2 \\
&\le \underline{\beta}^{-1} \cdot (\bar{\beta}^{1/2} \mu_{\sqrt{\beta}}(\Phi) + 2\bar{\beta}^{1/2} \underline{\beta}^{1/2} \cdot \delta)^2 \\
&\le \underline{\beta}^{-1} \cdot (\bar{\beta}^{1/2} \underline{\alpha} \gamma \nu^2 \underline{\beta}^{1/2}/C + \bar{\beta}^{1/2} \underline{\beta}^{1/2} \gamma \nu^2/C)^2 \le (2\bar{\beta}^{1/2} \gamma \nu^2/C)^2.
\end{aligned}$$

Recalling that in both regimes $\|\Psi D_{\sqrt{\beta}}\| \le \underline{\alpha} \gamma \nu \underline{\beta}^{1/2}/C$ we bound the following

$$\begin{aligned}
\|D_{\sqrt{\pi}} \dot{H} e_\ell\|^2 = \|D_{\sqrt{\beta}}^{-1} D_{\sqrt{\pi}} \ddot{H} e_\ell\|^2 &\le \|D_{\sqrt{\beta}}\|^2 \cdot \|D_{\sqrt{\beta}} \Psi^* \Psi D_{\sqrt{\beta}} e_\ell\|^2 \\
&\le \underline{\beta}^{-1} \cdot \|\Psi D_{\sqrt{\beta}}\|^2 \cdot \|\psi\|^2 \cdot \beta_\ell \le (\underline{\alpha} \gamma \nu/C)^2 \cdot \beta_\ell. \qquad (136)
\end{aligned}$$

Using $\|D_{\sqrt{\pi}}\dot{H}D_{\sqrt{\pi}}\| \leq \underline{\beta}^{-1/2} \cdot \|\Psi D_{\sqrt{\pi\cdot\beta}}\|^2$ and $\|e_k^*\dot{H}D_{\sqrt{\pi}}\| \leq \|\Psi D_{\sqrt{\pi\cdot\beta}}\|$ from (125) and (126) resp. and Corollary 5(d), we obtain for the first regime that

$$
\begin{aligned}
&\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\| \\
&\quad\leq 3/2\cdot\pi_\ell\cdot\Big(3\cdot\|D_{\sqrt{\pi}}\dot{H}e_\ell\|^2 + \max_k (D_{\sqrt{\beta}}^{-1}\ddot{H})_{k\ell}^2 \\
&\qquad\qquad\qquad\qquad + 9/2\cdot\|D_{\sqrt{\pi}}\dot{H}D_{\sqrt{\pi}}\|^2 + 3/2\cdot\max_k\|e_k^*\dot{H}D_{\sqrt{\pi}}\|^2\Big) \\
&\quad\leq \frac{3}{2}\cdot\pi_\ell\cdot\left(3\cdot\left(\frac{\underline{\alpha}\gamma\nu}{C}\right)^2\cdot\beta_\ell + \left(\frac{\underline{\alpha}\gamma\nu^2\underline{\beta}^{1/2}}{C}\right)^2 + \frac{9}{2}\cdot\underline{\beta}\cdot\left(\frac{\underline{\alpha}\gamma\nu}{4C}\right)^4 + \frac{3}{2}\cdot\left(\frac{\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}}{4C}\right)^2\right) \\
&\quad\leq 9\cdot\pi_\ell\cdot(\underline{\alpha}\gamma\nu\beta_\ell^{1/2}/C)^2, \qquad (137)
\end{aligned}
$$

and for the second regime that

$$
\|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\| \leq 9\cdot\pi_\ell\cdot(2\gamma\nu\bar{\beta}^{1/2}/C)^2. \qquad (138)
$$

Next we establish a bound for (117). Therefore we mainly use the calculations from (129)-(138) and Lemma 13. Recall that $\|\ddot{\mathcal{H}}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\| \leq \min\{\gamma^{-1}\cdot 7/3, 2\delta/\nu\} = \Gamma$.
For the first term in (117) we use the inequalities of (129) and (135) after applying Lemma 13, yielding

$$
\begin{aligned}
&\|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{\mathcal{H}}_{I,I}^*\cdot\ddot{\mathcal{H}}_{I,I}R_Ie_\ell]\| \\
&\qquad\leq \|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{\mathcal{H}}_{I,I}^*\ddot{\mathcal{H}}_{I,I}R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}}\cdot\|\mathbb{E}[\ddot{\mathcal{H}}_{I,\ell}^*\ddot{\mathcal{H}}_{I,\ell}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}} \\
&\qquad\leq 3\cdot\pi_\ell\beta_\ell\cdot\min\{2/\sqrt{11},\delta\}\cdot(\|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\varepsilon^2/2 + \|ZD_{\sqrt{\pi\cdot\beta}}\|).
\end{aligned}
$$

Continuing in both regimes separately and using that $\underline{\alpha} \geq 1-\delta_\circ^2/2 \geq 1-1/(2C^2) \geq 17/18$ results in

$$
\begin{aligned}
\delta > \delta_\circ: \quad &\|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{\mathcal{H}}_{I,I}^*\cdot\ddot{\mathcal{H}}_{I,I}R_Ie_\ell]\| \\
&\qquad\leq 6/\sqrt{11}\cdot\pi_\ell\beta_\ell\cdot\frac{3\cdot\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}}{4C} \leq 14/10\cdot\pi_\ell\beta_\ell\cdot\underline{\alpha}\underline{\beta}^{1/2}\nu^{-1}\cdot\delta_\circ, \qquad (139) \\
\delta \leq \delta_\circ: \quad &\|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{\mathcal{H}}_{I,I}^*\cdot\ddot{\mathcal{H}}_{I,I}R_Ie_\ell]\| \\
&\qquad\leq 3\cdot\underline{\alpha}\cdot 18/17\cdot\pi_\ell\beta_\ell\cdot\frac{\gamma\nu^2}{C}\cdot\delta\underline{\beta}^{1/2}\cdot(1/C+1) \\
&\qquad\leq 4/C\cdot\pi_\ell\beta_\ell\cdot\underline{\alpha}\underline{\beta}^{1/2}\cdot\delta. \qquad (140)
\end{aligned}
$$

For the second term we use the inequalities of (130) and (135) to obtain

$$
\begin{aligned}
&\|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{\mathcal{H}}_{I,I}^*\cdot\ddot{\mathcal{H}}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\cdot\ddot{H}_{I,I}R_Ie_\ell]\| \\
&\quad\leq \|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{\mathcal{H}}_{I,I}^*\ddot{\mathcal{H}}_{I,I}R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}}\cdot\Gamma\cdot\|\mathbb{E}[\ddot{H}_{I,\ell}^*\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}} \\
&\quad\leq 3\cdot\min\{1,\delta\}\cdot(\pi_\ell\beta_\ell)^{1/2}\cdot\Gamma\cdot(\pi_\ell\beta_\ell)^{1/2}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\| \\
&\quad\leq 3\cdot\min\{1,\delta\}\cdot\pi_\ell\beta_\ell\cdot\Gamma\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|.
\end{aligned}
$$

Continuing in both regimes separately leads to

$$
\begin{aligned}
\delta > \delta_\circ: \quad & \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}^* \cdot \ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\cdot \ddot{H}_{I,I}R_I e_\ell]\| \\
& \leq 3\cdot \pi_\ell\beta_\ell\cdot\gamma^{-1}\cdot 7/3\cdot \underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C) \\
& \leq 7/8\cdot\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell\cdot\gamma^{-1}\nu^{-1}\cdot\gamma\nu^2/C \leq 7/8\cdot\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell\gamma^{-1}\nu^{-1}\cdot\delta_\circ, \quad (141)\\
\delta \leq \delta_\circ: \quad & \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}^* \cdot \ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\cdot \ddot{H}_{I,I}R_I e_\ell]\| \\
& \leq 3\cdot\pi_\ell\beta_\ell\delta\cdot 2\delta\nu^{-1}\cdot\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/C \\
& \leq 3/C\cdot\pi_\ell\cdot\gamma\nu^2/C\cdot 2\delta\cdot\beta_\ell\cdot\underline{\alpha}\gamma\underline{\beta}^{1/2} \leq 3/C^2\cdot\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell\gamma\cdot\delta. \quad (142)
\end{aligned}
$$

For the next term we need the inequalities of (129), (137) and (138), yielding

$$
\begin{aligned}
& \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\cdot(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\cdot\ddot{H}_{I,I}R_I e_\ell]\| \\
& \quad \leq \|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}}\cdot\Gamma\cdot\|\mathbb{E}[\ddot{H}_{I,\ell}^*\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}} \\
& \quad \leq \|\mathbb{E}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*R_I\mathbb{I}_{\ell^c}(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}}\times \\
& \qquad\qquad \times\Gamma\cdot(\beta_\ell\pi_\ell)^{1/2}\cdot(\|\Phi D_{\sqrt{\pi\cdot\beta}}\|\cdot\varepsilon^2/2+\|ZD_{\sqrt{\pi\cdot\beta}}\|).
\end{aligned}
$$

Continuing in both regimes separately results in

$$
\begin{aligned}
\delta > \delta_\circ: \quad & \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\cdot(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\cdot\ddot{H}_{I,I}R_I e_\ell]\| \\
& \leq 3\cdot\pi_\ell^{1/2}\cdot\underline{\alpha}\gamma\nu\beta_\ell^{1/2}/C\cdot\gamma^{-1}\cdot 7/3\cdot(\beta_\ell\pi_\ell)^{1/2}\cdot 3\cdot\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C) \\
& \leq 11/(2C^2)\cdot\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell\cdot\delta_\circ, \quad (143)\\
\delta \leq \delta_\circ: \quad & \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\cdot(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\cdot\ddot{H}_{I,I}R_I e_\ell]\| \\
& \leq 6\cdot\pi_\ell^{1/2}\cdot\gamma\nu\bar{\beta}^{1/2}/C\cdot 2\delta\nu^{-1}\cdot(\beta_\ell\pi_\ell)^{1/2}\cdot\gamma\nu^2/C\cdot\underline{\beta}^{1/2}\cdot(1/C+1) \\
& \leq 13/C^2\cdot\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell\gamma^2\nu^3\cdot\delta \leq 13/C^2\cdot\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell\cdot\delta. \quad (144)
\end{aligned}
$$

For the last term we employ the inequalities of (130), (137) and (138) after applying Lemma 13, yielding

$$
\begin{aligned}
& \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\cdot(\Psi_I^*\Psi_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\Psi_I^*\Psi_I)^{-1}\cdot\ddot{H}_{I,I}R_I e_\ell]\| \\
& \quad \leq \|\mathbb{E}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*R_I\mathbb{I}_{\ell^c}D_{\sqrt{\pi}}^{-1}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}}\cdot\Gamma^2\cdot\|\mathbb{E}[\ddot{H}_{I,\ell}^*\ddot{H}_{I,\ell}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}} \\
& \quad \leq \|\mathbb{E}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\ddot{H}_{I,I}^*R_I\mathbb{I}_{\ell^c}D_{\sqrt{\pi}}^{-1}\mathbb{1}_I(\ell)]\|^{\frac{1}{2}}\cdot\Gamma^2\cdot(\beta_\ell\pi_\ell)^{1/2}\cdot\|\Psi D_{\sqrt{\pi\cdot\beta}}\|.
\end{aligned}
$$

Continuing in both regimes separately leads to

$$
\begin{aligned}
\delta > \delta_\circ: \quad & \|\mathbb{E}_{\mathcal{G}}[D_{\sqrt{\pi}}^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\cdot(\Psi_I^*\Psi_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\Psi_I^*\Psi_I)^{-1}\cdot\ddot{H}_{I,I}R_I e_\ell]\| \\
& \leq 3\cdot\pi_\ell^{1/2}\cdot\underline{\alpha}\gamma\nu\beta_\ell^{1/2}/C\cdot\gamma^{-2}\cdot 49/9\cdot(\beta_\ell\pi_\ell)^{1/2}\cdot\underline{\alpha}\gamma\nu\underline{\beta}^{1/2}/(4C) \\
& \leq 8/C\cdot\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell\gamma^{-1}\cdot\gamma\nu^2/C \leq 8/C\cdot\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell\gamma^{-1}\cdot\delta_\circ, \quad (145)\\
\delta \leq \delta_\circ: \quad & \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi\cdot\beta}})^{-1}\mathbb{I}_{\ell^c}R_I^*\ddot{H}_{I,I}\cdot(\Psi_I^*\Psi_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\Psi_I^*\Psi_I)^{-1}\cdot\ddot{H}_{I,I}R_I e_\ell]\| \\
& \leq 6\cdot\pi_\ell^{1/2}\cdot\gamma\nu\bar{\beta}^{1/2}/C\cdot 4\delta^2\nu^{-1}\cdot\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell^{1/2}\gamma\nu/C \\
& \leq 24/C^2\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\nu\beta_\ell\cdot\gamma\nu^2/C\cdot\delta \leq 12/C^3\cdot\pi_\ell\underline{\alpha}\underline{\beta}^{1/2}\beta_\ell\cdot\delta. \quad (146)
\end{aligned}
$$

Substituting the bounds(139)-(146) into (117) yields

$$\begin{aligned}
\delta > \delta_\circ : \quad & \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi}\cdot\alpha})^{-1}\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_I e_\ell \mathbb{1}_I(\ell)]\| \\
& \leq \pi_\ell \cdot \big(14/10 \cdot \gamma^{-1}\nu^{-1} \cdot \delta_\circ + 7/8 \cdot \gamma^{-1}\nu^{-1} \cdot \delta_\circ \\
& \qquad + 11/(2C^2) \cdot \delta_\circ + 8/C \cdot \gamma^{-1} \cdot \delta_\circ\big) \\
& \leq 5/2 \cdot \pi_\ell \gamma^{-1}\nu^{-1} \cdot \delta_\circ, \qquad (147) \\
\delta \leq \delta_\circ : \quad & \|\mathbb{E}_{\mathcal{G}}[(D_{\sqrt{\pi}\cdot\alpha})^{-1}\mathbb{I}_{\ell^c}R_I^*(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}\ddot{H}_{I,I}^*\ddot{H}_{I,I}(\dot{\Psi}_I^*\dot{\Psi}_I)^{-1}R_I e_\ell \mathbb{1}_I(\ell)]\| \\
& \leq \pi_\ell \cdot \big(4/C \cdot \delta + 3/C^2 \cdot \delta + 13/C^2 \cdot \delta + 12/C^3 \cdot \delta\big) \\
& \leq 11/10 \cdot \pi_\ell \cdot \delta. \qquad (148)
\end{aligned}$$

Next we want to restate several well-known lemmas and theorems, which we use throughout this thesis.

A.3 Matrix bounds

In this chapter we state useful matrix bounds on which we heavily rely in this thesis.

Theorem 10 (Matrix Chernoff inequality [33]) *Let $X_1, \dots, X_N$ be independent random positive semi-definite matrices taking values in $\mathbb{R}^{d\times d}$. Assume that for all $n \in \{1, \dots, N\}$, $\|X_n\| \leq \eta$ a.s. and $\left\|\sum_{n=1}^N \mathbb{E}[X_n]\right\| \leq \mu_{\max}$. Then, for all $t \geq e\mu_{\max}$,*

$$\mathbb{P}\left(\|\sum_{n=1}^N X_n\| \geq t\right) \leq K\left(\frac{e\mu_{\max}}{t}\right)^{\frac{t}{\eta}}.$$

Theorem 11 (Matrix resp. vector Bernstein inequality [33], [18]) *Consider a sequence $Y_1, \dots, Y_N$ of independent, random matrices (resp. vectors) with dimension $d \times K$ (resp. d). Assume that each random matrix (resp. vector) satisfies*

$$\|Y_n\| \leq r \quad \text{a.s.} \quad \text{and} \quad \|\mathbb{E}[Y_n]\| \leq m.$$

Then, for all $t > 0$,

$$\mathbb{P}\left(\|\frac{1}{N}\sum_{n=1}^N Y_n\| \geq m + t\right) \leq \kappa \exp\left(\frac{-Nt^2}{2r^2 + (r+m)t}\right),$$

where $\kappa = d + K$ for the matrix Bernstein inequality and $\kappa = 28$ for the vector Bernstein inequality.

Theorem 12 (Operator norm of a random submatrix [22]) *Let Ψ be a dictionary and assume $I \subseteq \mathbb{K}$ is chosen according to the rejective sampling model with probabilities $p_1, \dots, p_K$ such that $\sum_{i=1}^K p_i = S$. Further, let D_p denote the diagonal matrix with the vector p on its diagonal. Then*

$$\mathbb{P}\left(\|\Psi_I^*\Psi_I - \mathbb{I}\| > \vartheta\right) \leq 216K \exp\left(-\min\left\{\frac{\vartheta^2}{4e^2\|\Psi D_p \Psi^*\|}, \frac{\vartheta}{2\mu(\Psi)}\right\}\right).$$

Lemma 13 (Sums of products of matrices [8], [14]) *Let $A_n \in \mathbb{R}^{d_1 \times d_2}, B_n \in \mathbb{R}^{d_2 \times d_3}$, $C_n \in \mathbb{R}^{d_3 \times d_4}$. Then*

$$\left\| \sum_{n=1}^{N} A_n B_n C_n \right\| \leq \left\| \sum_{n=1}^{N} A_n A_n^* \right\|^{\frac{1}{2}} \max_n \|B_n\| \left\| \sum_{n=1}^{N} C_n^* C_n \right\|^{\frac{1}{2}}.$$

Theorem 14 ([22], [24]) *Let $\mathbb{P}_B$ be the probability measure corresponding to the Poisson sampling model with weights $p_i < 1$ and $\mathbb{P}_S$ be the probability measure corresponding to the associated rejective sampling model with parameter S, $\mathbb{P}_S(I) = \mathbb{P}_B(I \mid |I| = S)$, as in Definition 1. Further, denote by $\mathbb{E}_S$ the expectation with respect to $\mathbb{P}_S$ and by π_S the vector of first order inclusion probabilities of level S, meaning $\pi_S(i) = \mathbb{P}_S(i \in I)$ or equivalently $\pi_S = \mathbb{E}_S(\mathbf{1}_I)$. We have*

$$\begin{aligned}(1 - \|p\|_\infty) \cdot p_i \leq \pi_S(i) \leq 2 \cdot p_i, \quad &\text{if} \quad \sum_k p_k = S, \\ \pi_{S-1}(i) \leq \pi_S(i),& \\ \mathbb{P}_S(\{i,j\} \subseteq I) \leq \pi_S(i) \cdot \pi_S(j), \quad &\text{if } i \neq j.\end{aligned}$$

Further, defining for $L \subseteq [K]$ with $|L| < S$ the set $\mathcal{L} = \{I \subseteq [K] : L \subseteq I\}$, we have

$$\mathbb{E}_S \left[\mathbf{1}_{I \setminus L} \mathbf{1}_{I \setminus L}^* \cdot \mathbb{1}_{\mathcal{L}}(I) \right] \cdot \prod_{\ell \in L} [1 - \pi_S(\ell)] \preceq \mathbb{E}_{S-|L|} [\mathbf{1}_I \mathbf{1}_I^*] \cdot \prod_{\ell \in L} \pi_S(\ell).$$

Finally, if $\pi := \pi_S$ satisfies $\|\pi\|_\infty < 1$, then for any $K \times K$ matrix A we have

$$\|A \odot \mathbb{E}[\mathbf{1}_I \mathbf{1}_I^*]\| \leq \frac{1 + \|\pi\|_\infty}{(1 - \|\pi\|_\infty)^2} \cdot \|D_\pi [A - \mathrm{diag}(A)] D_\pi\| + \|\mathrm{diag}(A) D_\pi\|.$$

References

[1] A. Agarwal, A. Anandkumar, P. Jain, and P. Netrapalli. Learning sparsely used overcomplete dictionaries via alternating minimization. *SIAM Journal on Optimization*, 26(4):2775–2799, 2016.

[2] A. Agarwal, A. Anandkumar, and P. Netrapalli. A clustering approach to learning sparsely used overcomplete dictionaries. *IEEE Transactions on Information Theory*, 63(1):575–592, 2017.

[3] M. Aharon, M. Elad, and A.M. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing.*, 54(11):4311–4322, November 2006.

[4] S. Arora, R. Ge, and A. Moitra. New algorithms for learning incoherent and overcomplete dictionaries. In *COLT 2014 (arXiv:1308.6273)*, 2014.

[5] S. Arora, R. Ge, T. Ma, and A. Moitra. Simple, efficient, and neural algorithms for sparse coding. In *COLT 2015 (arXiv:1503.00778)*, 2015.

[6] B. Barak, J.A. Kelner, and D. Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In *STOC 2015 (arXiv:1407.1543)*, 2015.

[7] N. Chatterji and P. Bartlett. Alternating minimization for dictionary learning with random initialization. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017.

[8] W. Dai and Y. Ye. A characterization on singular value inequalities of matrices. *Journal of Function Spaces*, 2020:1–4, 02 2020. doi: 10.1155/2020/1657381.

[9] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407 – 499, 2004.

[10] K. Engan, S.O. Aase, and J.H. Husoy. Method of optimal directions for frame design. In *ICASSP99*, volume 5, pages 2443 – 2446, 1999.

[11] D.J. Field and B.A. Olshausen. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[12] R. Gribonval and K. Schnass. Dictionary identifiability - sparse matrix-factorisation via l_1-minimisation. *IEEE Transactions on Information Theory*, 56(7):3523–3539, July 2010.

[13] R. Gribonval, R. Jenatton, and F. Bach. Sparse and spurious: dictionary learning with noise and outliers. *IEEE Transactions on Information Theory*, 61(11):6298–6319, 2015.

[14] O. Hirzallah and F. Kittaneh. Inequalities for sums and direct sums of Hilbert space operators. *Linear Algebra and its Applications*, 424(1):71–82, 2007.

[15] K. Kreutz-Delgado, J.F. Murray, B.D. Rao, K. Engan, T. Lee, and T.J. Sejnowski. Dictionary learning algorithms for sparse representation. *Neural Computations*, 15(2): 349–396, 2003.

[16] M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Computations*, 12(2):337–365, 2000.

[17] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11:19–60, 2010.

[18] S. Minsker. On some extensions of Bernstein's inequality for self-adjoint operators. *Statistics and Probability Letters*, 127:111–119, 2017.

[19] M.C. Pali and K. Schnass. Dictionary learning – from local towards global and adaptive. *Information and Inference: A Journal of the IMA*, 12(3), 2023.

[20] Q. Qu, Y. Zhai, X. Li, Y. Zhang, and Z. Zhu. Geometric analysis of nonconvex optimization landscapes for overcomplete learning. In *International Conference on Learning Representations*, 2020.

[21] S. Ruetz. Adapted variable density subsampling for compressed sensing. *arXiv:2206.13796*, 2022.

[22] S. Ruetz and K. Schnass. Submatrices with non-uniformly selected random supports and insights into sparse approximation. *SIAM Journal on Matrix Analysis and Applications*, 42(3):1268–1289, 2021.

[23] S. Ruetz and K. Schnass. Non-asymptotic bounds for inclusion probabilities in rejective sampling. *arXiv:2212.09391*, 2022.

[24] S. Ruetz and K. Schnass. Convergence of alternating minimisation algorithms for dictionary learning. *arXiv:2304.01768*, 2023.

[25] C. Rusu and B. Dumitrescu. Stagewise K-SVD to design efficient dictionaries for sparse representations. *IEEE Signal Processing Letters*, 19(10):631–634, 2012.

[26] K. Schnass. On the identifiability of overcomplete dictionaries via the minimisation principle underlying K-SVD. *Applied and Computational Harmonic Analysis*, 37(3): 464–491, 2014.

[27] K. Schnass. Local identification of overcomplete dictionaries. *Journal of Machine Learning Research*, 16(Jun):1211–1242, 2015.

[28] K. Schnass. Convergence radius and sample complexity of ITKM algorithms for dictionary learning. *Applied and Computational Harmonic Analysis*, 45(1):22–58, 2018.

[29] K. Skretting and K. Engan. Recursive least squares dictionary learning algorithm. *IEEE Transactions on Signal Processing*, 58(4):2121–2130, 2010.

[30] D. Spielman, H. Wang, and J. Wright. Exact recovery of sparsely-used dictionaries. In *COLT 2012 (arXiv:1206.5882)*, 2012.

[31] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere I: Overview and geometric picture. *IEEE Transactions on Information Theory*, 63(2):853–884, 2017.

[32] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere II: Recovery by Riemannian trust-region method. *IEEE Transactions on Information Theory*, 63 (2):885–915, 2017.

[33] J. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.