

Sharp analysis of out-of-distribution error for “importance-weighted” estimators in the overparameterized regime

Kuo-Wei Lai[†] Vidya Muthukumar^{†,‡}

School of Electrical & Computer Engineering, Georgia Institute of Technology[†]
H. Milton School of Industrial & Systems Engineering, Georgia Institute of Technology[‡]

Abstract

Overparameterized models that achieve zero training error are observed to generalize well on average, but degrade in performance when faced with data that is under-represented in the training sample. In this work, we study an overparameterized Gaussian mixture model imbued with a *spurious feature*, and sharply analyze the *in-distribution* and *out-of-distribution* test error of a *cost-sensitive* interpolating solution that incorporates “importance weights”. Compared to recent work [1, 27], our analysis is sharp with matching upper and lower bounds, and significantly weakens required assumptions on data dimensionality. Our error characterizations also apply to any choice of importance weights and unveil a novel tradeoff between worst-case robustness to distribution shift and average accuracy as a function of the importance weight magnitude¹.

1 Introduction

Overparameterized models are ubiquitous in machine learning theory and practice today because of their state-of-the-art generalization guarantees (in the sense of low test error) even while perfectly fitting the training data [30, 7]. However, this “good generalization” property does not extend to test data that is distributed differently from training data, termed *out-of-distribution* (OOD) data [20, 21, 29]. A particularly acute scenario arises when the data is drawn as a mixture from multiple groups (each with a different distribution) and some groups are very under-represented in training data [2]. Under such models, the *worst-group generalization error* can be significantly degraded with respect to the average generalization error on all groups [1, 27, 21, 20].

The effect of distribution shift on generalization has been sharply characterized in a worst-case/minimax sense, e.g. [16, 12, 19]. However, distribution-shift arising from practical inequities in group representation is often much more structured. Of particular interest is distribution shift induced by *spurious correlations* present in the training dataset, which is a scenario that commonly arises on real-world image and language datasets². Here, the shifted distribution removes or reverses these correlations, degrading performance. Recent work [27, 1, 20] has examined the ramifications of training overparameterized models in the presence of spurious correlations and proposed several strategies to mitigate their harmful effects on worst-group error. For example, [27, 1, 11, 14] introduce novel loss functions with *adjustment weights* that upweight the importance of minority group examples. On the other hand, *group distributionally robust optimization* (DRO) techniques [20, 10, 13] adaptively optimize examples from the worst group during training. Despite these efforts, a tight characterization of the impact of overparameterization, number of examples and training loss function on OOD generalization remains missing, even in this special setting.

Our contributions: In this paper, we provide sharp matching upper and lower bounds on the group-wise generalization error of the overparameterized linear model on data drawn from the Gaussian Mixture Model (GMM) data distribution imbued with a spurious feature: this model was first introduced in [21]

¹A short version of this work will be presented at IEEE ISIT 2024.

²For example, the Waterbirds dataset [28] overwhelmingly features landbirds on land and waterbirds on water; in this dataset, the background of the image is a spurious feature.

and studied in [1]. We study, in particular, a form of *cost-sensitive minimum-norm interpolation* (cMNI) wherein minority examples are “upweighted” by an adjustment weight $1/\Delta_-$ (our bounds also carry over to the cost-sensitive support-vector-machine (SVM) via a corollary of [1]). Our results (Theorem 1 and Proposition 1) characterize the precise role of the number of majority and minority training examples, the model dimension, and the upweighting factor on both ID and OOD generalization. Our tight bounds uncover an interesting *robustness-accuracy tradeoff* (Table 2): while a smaller upweighting factor (larger Δ_-) results in higher average accuracy and worse robustness (i.e. higher worst-group error), a larger upweighting factor (larger Δ_+) results in improved robustness at the cost of average accuracy. We also demonstrate that explicit regularization in the form of a ridge penalty does not change our bounds except for universal constant factors, nor does it change the nature of this robustness-accuracy tradeoff.

At a technical level, to achieve matching upper and lower bounds we deviate significantly from the related analyses [1, 27], which study the regularization path of gradient descent on the training error objective directly, and require strong assumptions on the dimensionality of the data. We instead leverage techniques from benign overfitting in binary and multiclass classification [26, 25] and apply them to a careful and delicate analysis of the cost-sensitive estimators.

Organization of paper: The rest of this paper is organized as follows. Section 1.1 contextualizes our approach and results with the most closely related work. Section 2 presents our main results, Section 3 is overall proof, and Section 4 provides a brief conclusion. The proofs of technical and auxiliary lemmas are deferred to the appendices.

1.1 Related work

The problem of generalizing to *out-of-distribution* (OOD) data has been studied in a vast body of work through two lenses. First, the goal of *domain adaptation*, or generalizing to an unseen domain, has been studied from a minimax perspective in statistics, e.g. [12, 19] and learning theory, e.g. [16]. Such minimax bounds are often pessimistic and do not reflect real-world distribution shifts that may be more structured in nature. An alternative and increasingly popular model for distribution shift consists of a *mixture model* in which data comes from one of a finite set of pre-defined groups, each of which possesses a very different data distribution. Typically, some groups (designated as *minority groups*) are under-represented in the training data, and a natural type of distribution shift would entail the minority groups being more prevalent, or “upsampled”, in test data. Under this model, it suffices to minimize the *worst-group error* to achieve robustness to distribution shift. To model further structure, [21] introduced the Gaussian mixture model with a *spurious feature*. Here, the data is partitioned into two groups depending on the sign of the spurious feature. Such spurious features are often present in real-world datasets and induce *spurious correlations* between input and output due to imbalance in the representation of the two groups [21]. A related setting with *class imbalance* was also studied by [27]. For this setting, researchers have proposed specially designed *group-aware* loss functions to alleviate the impact of OOD data during training, as seen in [11, 15, 27, 3, 17]. Notably, [11] introduced the *vector-scaling* loss (VS-loss) and demonstrated both theoretical and empirical benefits in using this loss function for worst-group error [1]. The implicit bias of the VS-loss turns out to align with the cost-sensitive minimum-norm solution that we study in this work for high-dimensional data [1]. This makes our findings directly applicable to estimators derived using the VS-loss. Finally, while our work as well as the above discussion focused on unregularized empirical risk minimization with the gradient descent algorithm, numerous studies utilize adaptive/robust optimization methods [20, 10, 22]. These methods aim to minimize the worst-case error within predefined groups instead.

Compared to our work, the upper bound on worst-group error of [1] is more loose (in particular in its dependence on the number of minority examples); on the other hand, its assumptions on data dimensionality and signal strength are slightly less stringent. [27] studied the effect of importance weighting on an interpolating classifier trained with a *polynomial loss function*, and showed that a similar worst-class error scaling is possible to our Theorem 1 for a specific case of the importance weights. However, the assumptions on data dimensionality made in [27] are significantly more stringent. Moreover, neither of these papers provides a

matching lower bound on the worst-group risk or theoretically characterizes the impact of importance weight magnitude on ID and OOD error, nor do these papers characterize the impact of explicit ridge regularization. Simulations conducted in [27] do highlight a robustness-accuracy tradeoff as a function of the magnitude of the importance weight that we uncover mathematically in our work. See Table 1 for a brief contextualization of our results with these closely related papers.

Our approach is quite different from that taken in in [27, 1], which both apply the proof strategy proposed in [5] which derives bounds between gradient updates of data examples in gradient descent. While that approach is generally applicable to a variety of training loss functions and label noise models, it requires larger data dimension requirements that can be removed by other analyses even for ID model generalization [26, 18]. We instead take the approach of *benign overfitting in classification of binary/multiclass GMM* [26, 25, 4], which is known to provide sufficient and necessary conditions for ID generalization (i.e. in the absence of spurious features). The introduction of spurious features and unequal weights on training examples introduces new technical challenges in the analysis even for binary classification. We discuss in detail how we tackle these challenges in Appendix A.

Finally, we wish to mention that our in-depth contextualization above has focused on the impact of distribution shift in the spurious feature setup with distinct groups, and on classification error. A parallel question of the impact of covariate shift on regression error in the overparameterized regime has also received sizable recent interest; see, e.g. [24, 23, 8]. This work predominantly studies random-feature models and the impact of overparameterization on robustness.

1.2 Notation

We use lower-case boldface (e.g. $\mathbf{x}$) to denote vector notation and upper-case boldface (e.g. $\mathbf{X}$) to denote matrix notation. We use $\|\cdot\|_2$ to denote the ℓ_2 -norm of a vector and the operator norm of a matrix. We use $\lambda_k(\mathbf{X})$ to denote the k -th largest eigenvalue of matrix $\mathbf{X}$ for $\mathbf{X} \in \mathbb{R}^{n \times n}$ and $1 \leq k \leq n$. We use the shorthand notation $[n]$ to denote the set of natural numbers $\{1, \dots, n\}$. We use $C, C_1, C_2, \dots > 0$ to denote universal and finite constants independent of all problem parameters, that can change line to line.

Table 1: Comparison of our results with related work

	$\mathcal{R}_{wst}(\hat{\mathbf{w}})$	Assumed lower bound on d	Assumed lower bound on R_+
Wang et al. [27]	$\mathcal{O}(R_+^2 n_- / d)$	$\Omega(n^3 \log(n/\delta))$	$\Omega(n^2 \log(n/\delta))$
Behnia et al. [1]	$\mathcal{O}(R_+^2 / d)$	$\Omega(n^2 \log(n/\delta))$	$\Omega(\log(n/\delta))$
Our result ($\Delta_b = \frac{n_b}{n}$)	$\Theta(R_+^2 n_- / d)$	$\Omega(n^2 \log(n/\delta))$	$\Omega(n \log(n/\delta))$

2 Our group-wise error characterizations

Data distribution. We use the setup in [1], i.e. a binary classification problem with data possessing spurious features which incorrectly correlate with its label. We denote $\mathbf{x} \in \mathbb{R}^d$ as a *feature* and $y \in \{+1, -1\}$ as a *label* which is equal to 1 with probability π_{+1} and -1 with probability π_{-1} . Finally, we denote $a \in \{+1, -1\}$ as the *attribute* of the data, meaning that the case $a \neq y$ indicates that the example possesses spurious features. Accordingly, $b = y \times a \in \{+1, -1\}$ partitions the data into two groups distinguishing the presence of spurious features ($b = +1$) or not ($b = -1$). We consider a product distribution on the features, labels and attributes, i.e. $\mathcal{P} = \mathcal{P}_x \times \mathcal{P}_y \times \mathcal{P}_a$. We assume that the feature distribution is a *Gaussian mixture model* (GMM) such that $\mathbf{x}$ is a Gaussian vector conditioned on y and a with mean vector $\boldsymbol{\mu} = [y\boldsymbol{\mu}_c; a\boldsymbol{\mu}_s] \in \mathbb{R}^d$ and covariance is $\mathbf{I}_d$, where we denote $\boldsymbol{\mu}_c$ as the “correct” class mean vector and $\boldsymbol{\mu}_s$ is the “spurious” class

mean vector. We also denote $\bar{\boldsymbol{\mu}}_c := [\boldsymbol{\mu}_c; \mathbf{0}] \in \mathbb{R}^d$ and $\bar{\boldsymbol{\mu}}_s := [\mathbf{0}; \boldsymbol{\mu}_s] \in \mathbb{R}^d$. More formally, we have:

$$\mathbf{x}|y, a \sim \mathcal{N}\left(\begin{bmatrix} y\boldsymbol{\mu}_c \\ a\boldsymbol{\mu}_s \end{bmatrix}, \mathbf{I}_d\right), \quad (1)$$

$$\mathbf{x} = \begin{bmatrix} y\boldsymbol{\mu}_c \\ a\boldsymbol{\mu}_s \end{bmatrix} + \mathbf{z} = y\bar{\boldsymbol{\mu}}_c + a\bar{\boldsymbol{\mu}}_s + \mathbf{z}, \quad (2)$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. We also denote the mean vector for each group as $\boldsymbol{\mu}_b = y\boldsymbol{\mu} = [\boldsymbol{\mu}_c; b\boldsymbol{\mu}_s]$. Finally, we denote the *total signal strength* as $R_+ = \|\boldsymbol{\mu}\|_2^2 = \|\bar{\boldsymbol{\mu}}_c\|_2^2 + \|\bar{\boldsymbol{\mu}}_s\|_2^2$ and the *signal strength difference* as $R_- = \|\bar{\boldsymbol{\mu}}_c\|_2^2 - \|\bar{\boldsymbol{\mu}}_s\|_2^2$.

Training dataset. We consider a dataset $\{\mathbf{x}_i, y_i, a_i\}_{i=1}^n$ where each example is drawn i.i.d. from the distribution $\mathcal{P}$. We consider the number of *minority examples* (i.e. where $b_i = -1$) to be n_- and the number of *majority examples* (i.e. where $b_i = +1$) to be $n_+ = n - n_-$. We also denote $\mathbf{X} \in \mathbb{R}^{n \times d}$ as the training data matrix (where each row in $\mathbf{X}$ is a data point), $\mathbf{y} \in \mathbb{R}^n$ as the label vector, $\mathbf{a} \in \mathbb{R}^n$ as the attribute vector. The definition of $\mathbf{x}$ in Eq. (2) gives us

$$\mathbf{X} = \mathbf{y}\bar{\boldsymbol{\mu}}_c^\top + \mathbf{a}\bar{\boldsymbol{\mu}}_s^\top + \mathbf{Q}, \quad (3)$$

where $\mathbf{Q} \in \mathbb{R}^{n \times d}$ and each row in $\mathbf{Q}$ follows $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.

Assumptions on data. We first make some mild assumptions: a) we operate in the *overparameterized regime* ($d \geq n$) in which the dataset is linearly separable almost surely; b) the number of majority examples exceeds the minority examples ($n_+ \geq \frac{n}{2}$), and c) the strength of the correct features exceeds the strength of the spurious features ($R_- \geq 0$). Next, we introduce our stronger assumptions on the dataset that are required to sharply analyze the group-wise classification error. These assumptions are considerably weaker than those made in [27] and mostly identical to those made in [1], with the exception of Assumption 1(B) which is slightly stronger.

Assumption 1. For a target failure probability δ , there exists a large enough constant $C > 0$ such that

$$\begin{aligned} A) \quad n &\geq C \log(1/\delta), & C) \quad d &\geq CR_+n, \\ B) \quad \|\bar{\boldsymbol{\mu}}_c\|_2^2 &\geq Cn \log(n/\delta), & D) \quad d &\geq Cn^2 \log(n/\delta). \end{aligned}$$

Estimators. We deploy a linear model $\mathbf{w} \in \mathbb{R}^d$ that solves *empirical risk minimization (ERM) with ridge regularization*, i.e. $\min_{\mathbf{w} \in \mathbb{R}^d} \mathcal{R}(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^n \ell(y_i, \langle \mathbf{w}, \mathbf{x}_i \rangle) + \tau \|\mathbf{w}\|_2^2$, where ℓ is the loss function and $\tau \geq 0$. Note that when $\tau = 0$, we recover the special case of unregularized ERM, which will result in an interpolating solution. We begin our analysis with the squared loss with *adjusted labels* for this classification problem, $\ell_{\text{sq}}(y, z) = (\Delta^{-1}y - z)^2$, where $0 < \Delta \leq 1$ is the *adjustment weight* of the label (also commonly called *importance weight*). Typically, we adjust the weights Δ differently for majority and minority examples to mitigate spurious correlations. Accordingly, we denote Δ_+ and Δ_- as the adjustment weight for majority examples and minority examples. This results in the closed-form *cost-sensitive ridge estimator*

$$\hat{\mathbf{w}}_{\text{ridge}} := \mathbf{X}^\top \left(\mathbf{X}\mathbf{X}^\top + \tau \mathbf{I} \right)^{-1} \boldsymbol{\Delta}^{-1} \mathbf{y}, \quad (4)$$

where $\boldsymbol{\Delta} \in \mathbb{R}^{n \times n}$ is the diagonal matrix of adjustment weights, i.e. $\Delta_{i,i} = \Delta_{b_i}$. Since $d \geq n$ and we assume Gaussian covariates, $\mathbf{X}$ is almost surely full-rank. Therefore, in the special case where $\tau = 0$, solving the

unregularized ERM with gradient descent (initialized at 0) would yield the *cost-sensitive minimum-norm interpolation* (cMNI) of the adjusted labels, i.e.

$$\hat{\mathbf{w}}_{\text{cMNI}} := \arg \min_{\mathbf{w}} \|\mathbf{w}\|_2 \quad (5)$$

$$\text{s.t. } \Delta_{b_i} \langle \mathbf{w}, \mathbf{x}_i \rangle = y_i \text{ for all } i \in [n], \quad (6)$$

This solution turns out to be expressible in closed form as:

$$\hat{\mathbf{w}}_{\text{cMNI}} := \mathbf{X}^\top \left(\mathbf{X} \mathbf{X}^\top \right)^{-1} \Delta^{-1} \mathbf{y}. \quad (7)$$

Note that decreasing the adjustment weight Δ_{b_i} would induce a larger *margin* on the example $\mathbf{x}_i$ in Eq. (6), which could alleviate at least in part the harmful effect of spurious correlations. On the other hand, very low adjustment weights could also increase the variance of the cMNI estimator. As a final remark, we note that under Assumption 1, [1] shows that the cost-sensitive support-vector-machine (cSVM) solution (arising from gradient descent run on a logit-adjusted loss [11]) exactly coincides with the cMNI solution presented in Eq. (5). Therefore, our forthcoming characterization of the cMNI estimator also automatically applies to the cSVM estimator, which is more popular in practice [11].

Generalization error. For a fresh (test) sample $(\mathbf{x}, y, a)$, we can calculate the expected 0-1 error risk as $\mathcal{R}(\hat{\mathbf{w}}) = \mathbb{E} [\mathbb{1}(y \neq \hat{y})] = \Pr [y \hat{\mathbf{w}}^\top \mathbf{x} < 0]$. Further, we consider the worst-group error, defined below as

$$\mathcal{R}_{\text{wst}}(\hat{\mathbf{w}}) = \max_{b \in \pm 1} \mathcal{R}_b(\hat{\mathbf{w}}) = \max_{b \in \pm 1} \Pr [y \hat{\mathbf{w}}^\top \mathbf{x} < 0 | y \times a = b]. \quad (8)$$

Our main focus in this work is deriving tight upper and lower bounds for Eq. (8).

2.1 Error analysis results

We present our main result, an upper bound for the group-wise generalization error, in Theorem 1 below.

Theorem 1. *Under Assumption 1, the generalization error for each group $b \in \{+1, -1\}$ is upper bounded as*

$$\mathcal{R}_b(\hat{\mathbf{w}}) \leq \exp \left(-C_1 \frac{(\alpha_+ R_b^2 n_+ + \alpha_- R_{-b}^2 n_-)}{d} \right), \quad (9)$$

with probability at least $1 - \delta$, where we defined $\alpha_\pm := \frac{n_\pm / \Delta_\pm^2}{\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}} = \frac{n_\pm / \Delta_\pm^2}{n_\Delta}$ and we have $0 < \alpha_\pm < 1$.

Notice that Theorem 1 characterizes the group-wise error as a function of: a) the number of majority and minority examples n_+, n_- , b) the total and difference of signal strength R_+, R_- , c) the number of features d , and d) the parameter $\alpha_\pm$ which is a functional of the adjustment weights $\Delta_\pm$ (and n_+, n_-). The forthcoming Proposition 1 shows that this rate is sharp up to a universal constant factor in all of these problem parameters.

Proposition 1. *Under Assumption 1, the generalization error for each group $b \in \{+1, -1\}$ is lower bounded as*

$$\mathcal{R}_b(\hat{\mathbf{w}}) \geq C_2 \exp \left(-C_3 \frac{(\alpha_+ R_b^2 n_+ + \alpha_- R_{-b}^2 n_-)}{d} \right), \quad (10)$$

where $\alpha_\pm := \frac{n_\pm / \Delta_\pm^2}{\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}} = \frac{n_\pm / \Delta_\pm^2}{n_\Delta}$ and $0 < \alpha_\pm < 1$.

It is instructive to consider a few special cases of Theorem 1 and Proposition 1. First, if there is no spurious feature in the data (i.e. $\bar{\boldsymbol{\mu}}_s = \mathbf{0}$), it is easy to verify that we recover the *in-distribution* error rate for overparameterized binary GMM in [26]. Second, if the magnitude of the correct and spurious means are near-identical, i.e. $\|\bar{\boldsymbol{\mu}}_c\|_2 \approx \|\bar{\boldsymbol{\mu}}_s\|_2$ or $R_- \approx 0$, we obtain the rates $\mathcal{R}_b(\hat{\mathbf{w}}) \asymp \exp\left(-\frac{C\alpha_b R_+^2 n_b}{d}\right)$. Therefore, decreasing Δ_b , increasing total signal strength R_+ , or increasing n_b would improve the error rate. Second, we note that the ridge regularization parameter τ only affects the bound through its influence on the constants C_1 and C_3 , and $0 < C_1, C_3 < \infty$ for all values of τ . Therefore, the non-asymptotic dependence of the bound on the parameters d, n_b, R_b, α_b is, interestingly, unaffected by explicit ridge regularization. All the same, Figure 1 shows that the empirical worst-group error decreases with an increase in τ . This phenomenon is not captured by our proof technique, as we do not provide exact asymptotic characterizations of the test error³. Corollary 1 below discusses sufficient and necessary conditions under which we would obtain vanishing error, i.e. *consistency* for out-of-distribution generalization.

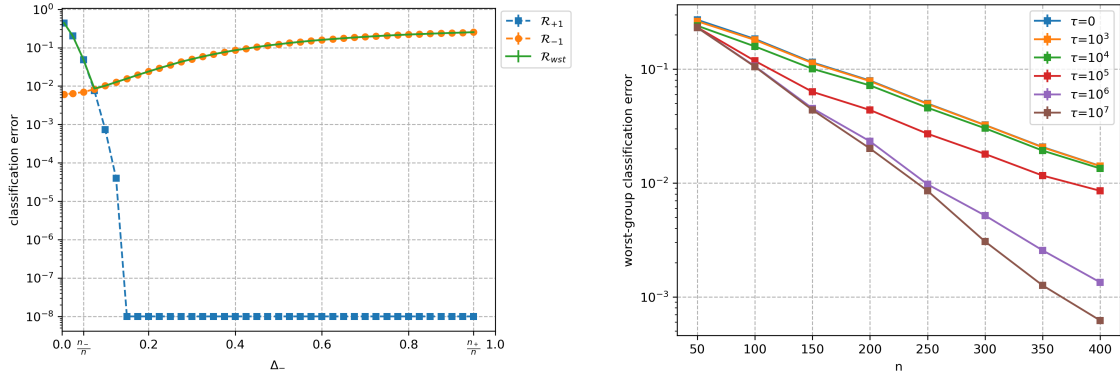


Figure 1: The left panel plots group-wise error as a function of Δ_- . We fix $d = 10^5$, $n = 200$, $n_- = 10$, $R_+ = d^{0.6}/4$, $\boldsymbol{\mu}_c = \boldsymbol{\mu}_s = \sqrt{R_+}/2\mathbf{e}_1$ (as in [1]), $\Delta_+ = \frac{n_+}{n}$ and make Δ_- decrease from $\frac{n_+}{n}$ to $\frac{n_-}{n}$. Observe that the worst-group error decreases when Δ_- decreases until $\Delta_- = \frac{n_-}{n}$, below which the worst-group becomes the majority group, whose error *increases* with decreased Δ_- . The right panel plots the worst-group error as a function of n fixing $d = 2n^2$, $n_- = 0.04n$, $R_+ = d^{0.6}/4$, $\boldsymbol{\mu}_c = \boldsymbol{\mu}_s = \sqrt{R_+}/2\mathbf{e}_1$, $\Delta_{\pm} = n_{\pm}/n$. Observe that increasing the ridge regularization parameter τ improves the worst-group error rate, but only up to a constant factor in the error exponent. These simulations were obtained by averaging over 10 trials, and error bars are small enough to not be visible.

Corollary 1. *Consider the case where $R_- = 0$. Then, under the conditions of Assumption 1, the generalization error for group b vanishes to 0 as $\frac{d}{n_b} \rightarrow \infty$ if and only if $R_+^2 = \Omega\left(\frac{d}{\alpha_b n_b}\right)$.*

Intuitively, the condition on $R_+^2 = \Omega\left(\frac{d}{\alpha_b n_b}\right)$ ensures that the total signal strength is sufficiently large, which reduces test error. Corollary 1 further shows how the adjustments $\Delta_{\pm}$, through their influence on the fractions α_b (and therefore the adjustments $\Delta_{\pm}$) can affect the condition on total signal strength for achieving vanishing group-wise test error. The implications of Corollary 1 are summarized in Table 2 for two extreme choices of adjustment weights: a) $\Delta_{\pm} = 1$ (i.e. no adjustments) and b) $\Delta_{\pm} = \frac{n_{\pm}}{n}$ (i.e. importance weighting). If we choose $\Delta_{\pm} = 1$, we get $\alpha_b = n_b/n$ and we need $R_+^2 = \Omega\left(\frac{dn}{n_b^2}\right)$ to have vanishing error. Then, since $n_+ \geq \frac{n}{2}$, the majority group generalization error vanishes *iff* $R_+^2 = \Omega\left(\frac{d}{n}\right)$, which matches the optimal rate of the case without spurious features [26]. However, the condition for vanishing minority-group

³We do expect that our non-asymptotic bounds would be tighter for higher values of τ , meaning that C_1 would *increase* with τ and C_3 would *decrease* with τ — but we do not conduct a detailed formal investigation of the exact impact of τ on the constants in our bounds in this paper.

error becomes much more stringent, and in the worst case yields $R_+^2 = \Omega(dn)$ when n_- is a constant. Thus, $\Delta_{\pm} = 1$ yields optimal majority-group error but highly suboptimal worst-group error. On the other hand, if we use $\Delta_{\pm} = n_{\pm}/n$, we get $\alpha_b \asymp n_-/n$ and can show that both the majority-group error and worst-group error would vanish *iff* $R_+^2 = \Omega\left(\frac{d}{n_-}\right)$. As a result, this choice improves worst-group error but worsens majority-group error. Thus Table 2 highlights an interesting tradeoff between *average accuracy* and *worst-group robustness* as a function of the adjustment weight magnitudes. We briefly explore this tradeoff empirically in Figure 2, which shows a sharp contrast in the rates of decrease of majority-group and minority-group error as a function of R_+^2 for two choices of the importance weights $\Delta_{\pm}$. This supports the scalings provided in Table 2.

Table 2: $\Delta_{\pm}$ changes the conditions for $\mathcal{R}_b \rightarrow 0$

	$\Delta_{\pm} = 1$	$\Delta_{\pm} = n_{\pm}/n$
$\mathcal{R}_b$	$\asymp \exp(-Cn_b^2 R_+^2/dn)$	$\asymp \exp(-Cn_b n_- R_+^2/dn)$
$\mathcal{R}_{+1}(\hat{\mathbf{w}}) \rightarrow 0$	$R_+^2 = \Omega(d/n)$	$R_+^2 = \Omega(d/n_-)$
$\mathcal{R}_{-1}(\hat{\mathbf{w}}) \rightarrow 0$	$R_+^2 = \Omega(dn/n_-^2)$	$R_+^2 = \Omega(d/n_-)$

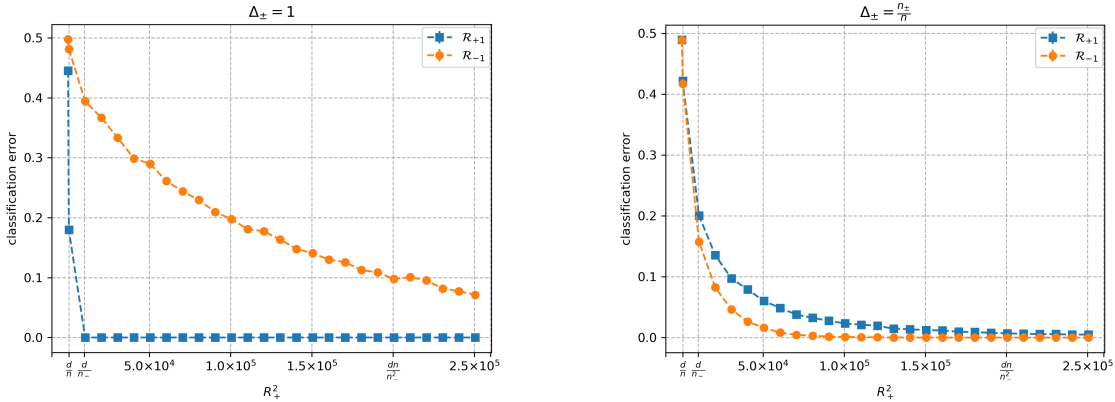


Figure 2: The left panel plots the majority-group and minority-group error as a function of the total signal strength squared (R_+^2) when the importance weights are set to be equal, i.e. $\Delta_{\pm} = 1$. We see that while $\mathcal{R}_{+1}(\hat{\mathbf{w}})$ approaches 0 when $R_+^2 = \frac{d}{n}$, $\mathcal{R}_{-1}(\hat{\mathbf{w}})$ is greater than 0.1 until $R_+^2 = \frac{dn}{n_-^2}$. The right panel also plots group-wise error for the alternative choice $\Delta_{\pm} = \frac{n_{\pm}}{n}$. In this case, as expected, both $\mathcal{R}_{+1}(\hat{\mathbf{w}})$ and $\mathcal{R}_{-1}(\hat{\mathbf{w}})$ decay to 0 at similar rates. For both plots, we fix $d = 10^5$, $n = 200$, $n_- = 10$, $\boldsymbol{\mu}_c = \boldsymbol{\mu}_s = \sqrt{R_+/2}\mathbf{e}_1$. These simulations were obtained by averaging over 10 trials, and error bars are small enough not to be visible.

3 Proof of Theorem 1

In this section, we present the proof of Theorem 1 in four steps. (Note that the lower bound in Proposition 1 follows via a matching sequence of steps and uses the matching exponential lower bound of the Q-function [6]; the full proof is in Appendix D.) We first present the upper bound of the worst group generalization error in analytical form. In step 2, we characterize the Gram matrix $\mathbf{X}\mathbf{X}^\top$ and its inverse recursively. Next, to simplify calculations, we define “primitive” quadratic terms and derive their bounds in step 3. Finally, we substitute these bounds on the primitive terms in the generalization error upper bound to complete the upper bound proof (Theorem 1). We note that steps 2 & 3 introduce new technical challenges, which we discuss in Appendix A.

Step 1. We restate [26, Lemma 1] below which upper bounds the worst group generalization error.

Lemma 1. *Consider a classification problem under GMM with a linear classifier $\hat{\mathbf{w}}$, the worst group risk is $\mathcal{R}_{wst}(\hat{\mathbf{w}}) = \max_{b \in \pm 1} Q\left(\frac{\hat{\mathbf{w}}^\top \boldsymbol{\mu}_b}{\sqrt{\hat{\mathbf{w}}^\top \hat{\mathbf{w}}}}\right)$. Specifically, if $\hat{\mathbf{w}}^\top \boldsymbol{\mu}_b > 0$, then*

$$\mathcal{R}_{wst}(\hat{\mathbf{w}}) \leq \max_{b \in \pm 1} \exp\left(-\frac{(\hat{\mathbf{w}}^\top \boldsymbol{\mu}_b)^2}{2\hat{\mathbf{w}}^\top \hat{\mathbf{w}}}\right). \quad (11)$$

According to Eq. (11), it is sufficient to derive a sharp lower bound on $\frac{(\hat{\mathbf{w}}^\top \boldsymbol{\mu}_b)^2}{2\hat{\mathbf{w}}^\top \hat{\mathbf{w}}}$ for each $b \in \pm 1$. Next, we substitute the closed form interpolation of the ridge estimator (Eq. (4)) into Eq. (11) to get

$$\frac{(\hat{\mathbf{w}}^\top \boldsymbol{\mu}_b)^2}{2\hat{\mathbf{w}}^\top \hat{\mathbf{w}}} = \frac{\left(\mathbf{y}^\top \boldsymbol{\Delta}^{-1} (\mathbf{X}\mathbf{X}^\top + \tau\mathbf{I})^{-1} \mathbf{X}\boldsymbol{\mu}_b\right)^2}{2\mathbf{y}^\top \boldsymbol{\Delta}^{-1} (\mathbf{X}\mathbf{X}^\top + \tau\mathbf{I})^{-1} \mathbf{X}\mathbf{X}^\top (\mathbf{X}\mathbf{X}^\top + \tau\mathbf{I})^{-1} \boldsymbol{\Delta}^{-1} \mathbf{y}}. \quad (12)$$

Step 2. In order to analyze and lower bound Eq. (12), we first analyze the Gram matrix $\mathbf{X}\mathbf{X}^\top$ and its inverse. Recalling the definition of $\mathbf{X}$ in Eq. (3), we define

$$\mathbf{M} := \mathbf{X}\mathbf{X}^\top + \tau\mathbf{I} = (\mathbf{v}_2\bar{\boldsymbol{\mu}}_2^\top + \mathbf{v}_1\bar{\boldsymbol{\mu}}_1^\top + \mathbf{Q})(\mathbf{v}_2\bar{\boldsymbol{\mu}}_2^\top + \mathbf{v}_1\bar{\boldsymbol{\mu}}_1^\top + \mathbf{Q})^\top + \tau\mathbf{I},$$

where we denote $\mathbf{v}_1 := \mathbf{a}$, $\mathbf{v}_2 := \mathbf{y}$, $\bar{\boldsymbol{\mu}}_1 := \bar{\boldsymbol{\mu}}_s$, $\bar{\boldsymbol{\mu}}_2 := \bar{\boldsymbol{\mu}}_c$ as shorthand. The presence of core and spurious mean vectors (along with the standard Wishart matrix $\mathbf{Q}\mathbf{Q}^\top$) in the expression for $\mathbf{X}\mathbf{X}^\top$ makes analyzing its inverse challenging, even for binary labels. Inspired by the idea in [25] (which was introduced for *multiclass* classification), we can decompose the data matrix $\mathbf{X}$ as

$$\mathbf{X}_0 := \mathbf{Q}, \quad \mathbf{X}_1 := \mathbf{v}_1\bar{\boldsymbol{\mu}}_1^\top + \mathbf{X}_0, \quad \mathbf{X} =: \mathbf{X}_2 = \mathbf{v}_2\bar{\boldsymbol{\mu}}_2^\top + \mathbf{X}_1.$$

Then, for $k = 1, 2$, we denote $\mathbf{M}_0 := \mathbf{X}_0\mathbf{X}_0^\top + \tau\mathbf{I} = \mathbf{Q}\mathbf{Q}^\top + \tau\mathbf{I}$ and

$$\begin{aligned} \mathbf{M}_k &:= \mathbf{X}_k\mathbf{X}_k^\top + \tau\mathbf{I} = \left(\mathbf{v}_k\bar{\boldsymbol{\mu}}_k^\top + \mathbf{X}_{k-1}\right) \left(\mathbf{v}_k\bar{\boldsymbol{\mu}}_k^\top + \mathbf{X}_{k-1}\right)^\top + \tau\mathbf{I} \\ &= \mathbf{M}_{k-1} + \underbrace{\begin{bmatrix} \|\bar{\boldsymbol{\mu}}_k\|_2 \mathbf{v}_k & \mathbf{d}_k & \mathbf{v}_k \end{bmatrix}}_{=: \mathbf{L}_k} \underbrace{\begin{bmatrix} \|\bar{\boldsymbol{\mu}}_k\|_2 \mathbf{v}_k^\top \\ \mathbf{v}_k^\top \\ \mathbf{d}_k^\top \end{bmatrix}}_{=: \mathbf{R}_k}, \end{aligned} \quad (13)$$

where $\mathbf{L}_k$, $\mathbf{R}_k$ are defined in the above display, and we define $\mathbf{d}_k := \mathbf{X}_{k-1}\bar{\boldsymbol{\mu}}_k = \mathbf{Q}\bar{\boldsymbol{\mu}}_k$. Next, as in [25], we apply the *recursive Woodbury identity* [9] to break the matrix inversion calculation into recursive steps as below:

$$\begin{aligned} \mathbf{M}_0^{-1} &= \left(\mathbf{X}_0\mathbf{X}_0^\top + \tau\mathbf{I}\right)^{-1} = \left(\mathbf{Q}\mathbf{Q}^\top + \tau\mathbf{I}\right)^{-1} \\ \mathbf{M}_k^{-1} &= \mathbf{M}_{k-1}^{-1} - \mathbf{M}_{k-1}^{-1} \mathbf{L}_k \mathbf{A}_k^{-1} \mathbf{R}_k \mathbf{M}_{k-1}^{-1}, \end{aligned} \quad (14)$$

where $\mathbf{A}_k := \mathbf{I} + \mathbf{R}_k \mathbf{M}_{k-1}^{-1} \mathbf{L}_k$ for $k = 1, 2$. As a result, we can represent the k th order inverse Gram matrix $\left(\mathbf{X}_k\mathbf{X}_k^\top + \tau\mathbf{I}\right)^{-1}$ as a constant multiple of the lower-order inverse Gram matrices. Ultimately, this shows that $\left(\mathbf{X}\mathbf{X}^\top + \tau\mathbf{I}\right)^{-1}$ is a constant multiple of the 0th order *inverse Wishart matrix* $\left(\mathbf{Q}\mathbf{Q}^\top + \tau\mathbf{I}\right)^{-1}$.

Step 3. Inspired by [26, 25], we define several “primitive” quadratic terms involving $\left(\mathbf{X}_k\mathbf{X}_k^\top + \tau\mathbf{I}\right)^{-1}$ for $k = 0, 1, 2$. In Step 4 of the proof we will see that Eq. (12) can be characterized entirely as a function of these primitives.

Definition 1. For $i, j = 1, 2$ and order $k = 0, 1, 2$, we define the following basic primitives:

$$\begin{aligned} s_{i,j}^{(k)} &:= \mathbf{v}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_j, & t_{i,j}^{(k)} &:= \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{d}_j, & h_{i,j}^{(k)} &:= \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_j, \\ s_{u,u}^{(k)} &:= \mathbf{u}^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{u}, & s_{u,i}^{(k)} &:= \mathbf{u}^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i, & h_{i,u}^{(k)} &:= \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{u}, \end{aligned}$$

where we let $\mathbf{u} \in \mathbb{R}^n$ be an arbitrary unit vector, i.e. $\|\mathbf{u}\|_2 = 1$.

Definition 2. For $i, j = 1, 2$ and order $k = 0, 1, 2$, we define the following adjusted primitives:

$$\begin{aligned} s_{i\Delta,j}^{(k)} &:= \mathbf{v}_i^\top \Delta^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_j, & h_{i,j\Delta}^{(k)} &:= \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \Delta^{-1} \mathbf{v}_j, \\ s_{i\Delta,j\Delta}^{(k)} &:= \mathbf{v}_i^\top \Delta^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \Delta^{-1} \mathbf{v}_j, \\ o_{i\Delta,i\Delta}^{(k)} &:= \mathbf{v}_i^\top \Delta^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{X}_k \mathbf{X}_k^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \Delta^{-1} \mathbf{v}_i. \end{aligned}$$

Using the ideas in Step 2, these primitives consisting of quadratic forms on $\left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1}$ can be characterized sharply as quadratic forms on $\left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1}$. Then, we can apply sharp concentration inequalities on the inverse Wishart matrix $\left(\mathbf{Q} \mathbf{Q}^\top \right)^{-1}$ to characterize the primitives. These bounds are presented in Lemma 2 and Lemma 3 below; the proofs of these lemmas are contained in Appendix A. Henceforth, we define $\hat{C}_m = \frac{C_m+1}{C_m}$, $\check{C}_m = \frac{C_m-1}{C_m}$ for universal constants $C_m > 1$.

Lemma 2. We have the following upper and lower bounds for the basic primitives (Def. 1):

$$\begin{aligned} \frac{\check{C}_1 n}{d + \tau} &\leq s_{i,i}^{(k)} \leq \frac{\hat{C}_1 n}{d + \tau}, & t_{i,i}^{(k)} &\leq \frac{C_2 n \|\bar{\boldsymbol{\mu}}_i\|_2^2}{d + \tau} \\ \check{C}_3 \frac{n_+}{d + \tau} - \hat{C}_3 \frac{n_-}{d + \tau} &\leq s_{1,2}^{(k)} \leq \check{C}_3 \frac{n_+}{d + \tau} - \hat{C}_3 \frac{n_-}{d + \tau} \\ -\frac{C_4 n \|\bar{\boldsymbol{\mu}}_1\|_2 \|\bar{\boldsymbol{\mu}}_2\|_2}{d + \tau} &\leq t_{1,2}^{(k)} \leq \frac{C_4 n \|\bar{\boldsymbol{\mu}}_1\|_2 \|\bar{\boldsymbol{\mu}}_2\|_2}{d + \tau} \\ -\frac{C_5 n \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} &\leq h_{i,j}^{(k)} \leq \frac{C_5 n \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} \\ \frac{\check{C}_6}{d + \tau} &\leq s_{u,u}^{(k)} \leq \frac{\hat{C}_6}{d + \tau}, & -\frac{C_7 \sqrt{n}}{d + \tau} &\leq s_{u,i}^{(k)} \leq \frac{C_7 \sqrt{n}}{d + \tau} \\ -\frac{C_8 \sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} &\leq h_{i,u}^{(k)} \leq \frac{C_8 \sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} \end{aligned}$$

Lemma 3. We have the following upper and lower bounds for the adjusted primitives (Def. 2):

$$\begin{aligned} \frac{\left(\hat{C}_9 - \frac{1}{C_{10}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} &\leq s_{i\Delta,i}^{(k)} \leq \frac{\left(\hat{C}_9 + \frac{1}{C_{10}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} \\ \frac{\hat{C}_9 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) - \frac{1}{C_{10}} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} &\leq s_{2\Delta,1}^{(k)} \leq \frac{\hat{C}_9 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) + \frac{1}{C_{10}} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} \\ \frac{\check{C}_{11} n \Delta}{d + \tau} &\leq s_{i\Delta,i\Delta}^{(k)} \leq \frac{\hat{C}_{11} n \Delta}{d + \tau} \\ -\frac{C_{12} \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} &\leq h_{i,j\Delta}^{(k)} \leq \frac{C_{12} \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} \\ \frac{\check{C}_{13} n \Delta d}{(d + \tau)^2} &\leq o_{i\Delta,i\Delta}^{(k)} \leq \frac{\hat{C}_{13} n \Delta d}{(d + \tau)^2}, \end{aligned}$$

where we define $n_\Delta := \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}$.

Step 4. Finally, we can write Eq. (12) entirely as a function of the primitives in Step 3. We show the derivation for the case $b = +1$ ($b = -1$ follows by an identical set of calculations):

$$\begin{aligned}
\frac{(\hat{\mathbf{w}}^\top \boldsymbol{\mu}_{+1})^2}{2\hat{\mathbf{w}}^\top \hat{\mathbf{w}}} &= \frac{\left(\mathbf{y}^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}\mathbf{X}^\top + \tau\mathbf{I}\right)^{-1} \mathbf{X}\boldsymbol{\mu}_{+1}\right)^2}{2\mathbf{y}^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}\mathbf{X}^\top + \tau\mathbf{I}\right)^{-1} \mathbf{X}\mathbf{X}^\top \left(\mathbf{X}\mathbf{X}^\top + \tau\mathbf{I}\right)^{-1} \boldsymbol{\Delta}^{-1} \mathbf{y}} \\
&= \frac{\left(\mathbf{v}_2^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_2\mathbf{X}_2^\top + \tau\mathbf{I}\right)^{-1} (\mathbf{v}_2\bar{\boldsymbol{\mu}}_2^\top + \mathbf{v}_1\bar{\boldsymbol{\mu}}_1^\top + \mathbf{Q})(\bar{\boldsymbol{\mu}}_2 + \bar{\boldsymbol{\mu}}_1)\right)^2}{2\mathbf{v}_2^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_2\mathbf{X}_2^\top + \tau\mathbf{I}\right)^{-1} \mathbf{X}_2\mathbf{X}_2^\top \left(\mathbf{X}_2\mathbf{X}_2^\top + \tau\mathbf{I}\right)^{-1} \boldsymbol{\Delta}^{-1} \mathbf{v}_2} \\
&= \frac{\left(\|\bar{\boldsymbol{\mu}}_2\|_2^2 s_{2\Delta,2}^{(2)} + \|\bar{\boldsymbol{\mu}}_1\|_2^2 s_{2\Delta,1}^{(2)} + h_{2,2\Delta}^{(2)} + h_{1,2\Delta}^{(2)}\right)^2}{2o_{2\Delta,2\Delta}^{(2)}}. \tag{15}
\end{aligned}$$

Above, the second equality follows by the definition of $\boldsymbol{\mu}_{+1} = \bar{\boldsymbol{\mu}}_c + \bar{\boldsymbol{\mu}}_s$ and the definition of $\mathbf{X}$; the last equality substitutes the adjusted primitives from Def. 2. We also define the following auxiliary lemma that helps us identify lower-order terms that can be removed up to constant factors. The proof of Lemma 4 is deferred to Appendix B.

Lemma 4. *For some $C > C_1^2 > 0$, if Assumption 1(B) holds, or n_- is a factor of n such that $n_- = \alpha n$ for $0 < \alpha \leq 1/2$ is satisfied, we have $\frac{1}{2} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_2\|_2 \geq C_1 \sqrt{nn_\Delta}$.*

We proceed to introduce the primitive bounds from Lemma 3 into Eq. (15) (lower bounding the numerator terms and upper bounding the denominator terms) to get:

$$\begin{aligned}
(15) &\geq \frac{\left(\|\bar{\boldsymbol{\mu}}_2\|_2^2 \frac{(\hat{C}_9 - \frac{1}{C_{10}}) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d+\tau} + \|\bar{\boldsymbol{\mu}}_1\|_2^2 \frac{\hat{C}_9 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) - \frac{1}{C_{10}} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d+\tau} - \frac{C_{12}\sqrt{nn_\Delta} \|\bar{\boldsymbol{\mu}}_2\|_2}{d+\tau} - \frac{C_{12}\sqrt{nn_\Delta} \|\bar{\boldsymbol{\mu}}_1\|_2}{d+\tau} \right)^2}{\frac{\hat{C}_{13}n_\Delta d}{(d+\tau)^2}} \\
&= \frac{\left(\left(\hat{C}_9 \|\bar{\boldsymbol{\mu}}_2\|_2^2 - \frac{\|\bar{\boldsymbol{\mu}}_2\|_2^2 + \|\bar{\boldsymbol{\mu}}_1\|_2^2}{C_{10}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \hat{C}_9 \|\bar{\boldsymbol{\mu}}_1\|_2^2 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) - C_{12}\sqrt{nn_\Delta} (\|\bar{\boldsymbol{\mu}}_2\|_2 + \|\bar{\boldsymbol{\mu}}_1\|_2) \right)^2}{\hat{C}_{13}n_\Delta d}. \tag{16}
\end{aligned}$$

Next, since $R_- \geq 0$, i.e. $\|\bar{\boldsymbol{\mu}}_2\|_2 \geq \|\bar{\boldsymbol{\mu}}_1\|_2$, we get

$$\begin{aligned}
(16) &\geq \frac{\left(\left(\hat{C}_9 \|\bar{\boldsymbol{\mu}}_2\|_2^2 - \frac{2\|\bar{\boldsymbol{\mu}}_2\|_2^2}{C_{10}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \hat{C}_9 \|\bar{\boldsymbol{\mu}}_1\|_2^2 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) - 2C_{12}\sqrt{nn_\Delta} \|\bar{\boldsymbol{\mu}}_2\|_2 \right)^2}{\hat{C}_{13}n_\Delta d} \\
&\geq \frac{\left(\left(\hat{C}_9 - \frac{2}{C_{10}} - \frac{C_{12}}{C_{14}} \right) \|\bar{\boldsymbol{\mu}}_2\|_2^2 \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \hat{C}_9 \|\bar{\boldsymbol{\mu}}_1\|_2^2 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) \right)^2}{\hat{C}_{13}n_\Delta d} \\
&\geq \frac{\left(\hat{C}_{15} \|\bar{\boldsymbol{\mu}}_2\|_2^2 \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \hat{C}_{15} \|\bar{\boldsymbol{\mu}}_1\|_2^2 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) \right)^2}{\hat{C}_{13}n_\Delta d} \\
&= \frac{\left(\hat{C}_{15}R_+ \frac{n_+}{\Delta_+} + \hat{C}_{15}R_- \frac{n_-}{\Delta_-} \right)^2}{\hat{C}_{13}n_\Delta d} \geq \frac{\hat{C}_{15}^2 R_+^2 \frac{n_+^2}{\Delta_+^2} + \hat{C}_{15}^2 R_-^2 \frac{n_-^2}{\Delta_-^2}}{\hat{C}_{13}n_\Delta d}, \tag{17}
\end{aligned}$$

where the second inequality follows by Lemma 4 with a large enough constant $C_{14} > C_{12}$. Finally, we define $\alpha_\pm = \frac{n_\pm/\Delta_\pm^2}{n_\Delta}$, noting that $0 < \alpha_\pm < 1$ and $\alpha_+ + \alpha_- = 1$. Eq. (17) becomes

$$\frac{(\hat{\mathbf{w}}^\top \boldsymbol{\mu}_{+1})^2}{2\hat{\mathbf{w}}^\top \hat{\mathbf{w}}} \geq \frac{\hat{C}_{16} (\alpha_+ R_+^2 n_+ + \alpha_- R_-^2 n_-)}{d}.$$

Plugging the above into Lemma 1 completes the proof. $\square$

4 Conclusion

In this work, we provide a sharp rate on the worst-group error of cost-sensitive interpolating or ridge-regularized classifiers as well as SVM classifiers, and further show that this rate is tight up to universal constants. Our rate explicitly characterizes the role of the adjustment weights $\Delta_\pm$, through which we identify a new robustness-accuracy tradeoff for this class of estimators. An intriguing question that remains is whether this tradeoff is information-theoretically optimal under this data model in the overparameterized regime, or whether an estimator with an inductive bias that is fundamentally different from the ℓ_2 -norm might improve it.

Acknowledgements

We gratefully acknowledge the support of the NSF (through CAREER award CCF-2239151 and award IIS2212182), an Adobe Data Science Research Award, an Amazon Research Award and a Google Research Colabs award.

References

- [1] Tina Behnia, Ke Wang, and Christos Thrampoulidis. “On how to avoid exacerbating spurious correlations when models are overparameterized”. In: *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2022, pp. 121–126 (cit. on pp. 1–6).
- [2] Joy Buolamwini and Timnit Gebru. “Gender shades: Intersectional accuracy disparities in commercial gender classification”. In: *Conference on fairness, accountability and transparency*. PMLR. 2018, pp. 77–91 (cit. on p. 1).

- [3] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. “Learning imbalanced datasets with label-distribution-aware margin loss”. In: *Advances in neural information processing systems* 32 (2019) (cit. on p. 2).
- [4] Yuan Cao, Quanquan Gu, and Mikhail Belkin. “Risk bounds for over-parameterized maximum margin classification on sub-gaussian mixtures”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 8407–8418 (cit. on p. 3).
- [5] Niladri S Chatterji and Philip M Long. “Finite-sample analysis of interpolating linear classifiers in the overparameterized regime”. In: *The Journal of Machine Learning Research* 22.1 (2021), pp. 5721–5750 (cit. on p. 3).
- [6] Marco Chiani, Davide Dardari, and Marvin K Simon. “New exponential bounds and approximations for the computation of error probability in fading channels”. In: *IEEE Transactions on Wireless Communications* 2.4 (2003), pp. 840–845 (cit. on pp. 7, 40).
- [7] Zeyu Deng, Abba Kammoun, and Christos Thrampoulidis. “A model of double descent for high-dimensional binary linear classification”. In: *Information and Inference: A Journal of the IMA* 11.2 (2022), pp. 435–495 (cit. on p. 1).
- [8] Xingdong Feng, Xin He, Caixing Wang, Chao Wang, and Jingnan Zhang. “Towards a Unified Analysis of Kernel-based Methods Under Covariate Shift”. In: *Advances in Neural Information Processing Systems* 36 (2024) (cit. on p. 3).
- [9] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012 (cit. on p. 8).
- [10] Weihua Hu, Gang Niu, Issei Sato, and Masashi Sugiyama. “Does distributionally robust supervised learning give robust classifiers?” In: *International Conference on Machine Learning*. PMLR, 2018, pp. 2029–2037 (cit. on pp. 1, 2).
- [11] Ganesh Ramachandra Kini, Orestis Paraskevas, Samet Oymak, and Christos Thrampoulidis. “Label-imbalanced and group-sensitive classification under overparameterization”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 18970–18983 (cit. on pp. 1, 2, 5).
- [12] Samory Kpotufe and Guillaume Martinet. “Marginal singularity and the benefits of labels in covariate-shift”. In: *The Annals of Statistics* 49.6 (2021), pp. 3299–3323 (cit. on pp. 1, 2).
- [13] Daniel Kuhn, Peyman Mohajerin Esfahani, Viet Anh Nguyen, and Soroosh Shafieezadeh-Abadeh. “Wasserstein distributionally robust optimization: Theory and applications in machine learning”. In: *Operations research & management science in the age of analytics*. Informs, 2019, pp. 130–166 (cit. on p. 1).
- [14] Kuo-Wei Lai and Vidya Muthukumar. “General loss functions lead to (approximate) interpolation in high dimensions”. In: *arXiv preprint arXiv:2303.07475* (2023) (cit. on p. 1).
- [15] Mingchen Li, Xuechen Zhang, Christos Thrampoulidis, Jiasi Chen, and Samet Oymak. “Autobalance: Optimized loss functions for imbalanced data”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 3163–3177 (cit. on p. 2).
- [16] Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. “Domain adaptation with multiple sources”. In: *Advances in neural information processing systems* 21 (2008) (cit. on pp. 1, 2).
- [17] Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. “Long-tail learning via logit adjustment”. In: *arXiv preprint arXiv:2007.07314* (2020) (cit. on p. 2).
- [18] Vidya Muthukumar, Adhyayan Narang, Vignesh Subramanian, Mikhail Belkin, Daniel Hsu, and Anant Sahai. “Classification vs regression in overparameterized regimes: Does the loss function matter?” In: *The Journal of Machine Learning Research* 22.1 (2021), pp. 10104–10172 (cit. on pp. 3, 14).
- [19] Reese Pathak, Cong Ma, and Martin Wainwright. “A new similarity measure for covariate shift with applications to nonparametric regression”. In: *International Conference on Machine Learning*. PMLR, 2022, pp. 17517–17530 (cit. on pp. 1, 2).

- [20] Shiori Sagawa, Pang Wei Koh, Tatsunori B Hashimoto, and Percy Liang. “Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization”. In: *arXiv preprint arXiv:1911.08731* (2019) (cit. on pp. 1, 2).
- [21] Shiori Sagawa, Aditi Raghunathan, Pang Wei Koh, and Percy Liang. “An investigation of why overparameterization exacerbates spurious correlations”. In: *International Conference on Machine Learning*. PMLR. 2020, pp. 8346–8356 (cit. on pp. 1, 2).
- [22] Dvir Samuel and Gal Chechik. “Distributional robustness loss for long-tail learning”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 9495–9504 (cit. on p. 2).
- [23] Nilesh Tripuraneni, Ben Adlam, and Jeffrey Pennington. “Covariate shift in high-dimensional random feature regression”. In: *arXiv preprint arXiv:2111.08234* (2021) (cit. on p. 3).
- [24] Nilesh Tripuraneni, Ben Adlam, and Jeffrey Pennington. “Overparameterization improves robustness to covariate shift in high dimensions”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 13883–13897 (cit. on p. 3).
- [25] Ke Wang, Vidya Muthukumar, and Christos Thrampoulidis. “Benign overfitting in multiclass classification: All roads lead to interpolation”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 24164–24179 (cit. on pp. 2, 3, 8, 14, 15).
- [26] Ke Wang and Christos Thrampoulidis. “Binary classification of gaussian mixtures: Abundance of support vectors, benign overfitting, and regularization”. In: *SIAM Journal on Mathematics of Data Science* 4.1 (2022), pp. 260–284 (cit. on pp. 2, 3, 6, 8, 16).
- [27] Ke Alexander Wang, Niladri S Chatterji, Saminul Haque, and Tatsunori Hashimoto. “Is importance weighting incompatible with interpolating classifiers?” In: *arXiv preprint arXiv:2112.12986* (2021) (cit. on pp. 1–4).
- [28] Peter Welinder, Steve Branson, Takeshi Mita, Catherine Wah, Florian Schroff, Serge Belongie, and Pietro Perona. “Caltech-UCSD birds 200”. In: (2010) (cit. on p. 1).
- [29] John R Zech, Marcus A Badgeley, Manway Liu, Anthony B Costa, Joseph J Titano, and Eric Karl Oermann. “Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study”. In: *PLoS medicine* 15.11 (2018), e1002683 (cit. on p. 1).
- [30] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. “Understanding deep learning (still) requires rethinking generalization”. In: *Communications of the ACM* 64.3 (2021), pp. 107–115 (cit. on p. 1).

A Proofs of primitive lemmas (Lemmas 2 and 3)

In this section, we present the proofs of bounds on the primitives. There are two essential techniques that we use throughout these proofs. First, in a manner similar to [25], we use the recursive Woodbury identity to characterize the Gram matrix inverse $(\mathbf{X}\mathbf{X}^\top + \tau\mathbf{I})^{-1}$ (and, thereby, the second-order primitives) in terms of the zero-th order primitives involving the inverse Wishart matrix $(\mathbf{Q}\mathbf{Q}^\top + \tau\mathbf{I})^{-1}$. We note that, unlike in [25] (which assumed orthogonal label vectors across classes), the label vector and the attribute vector *are not* orthogonal here; therefore, a new technique is required to characterize the second-order primitives in terms of zeroth order primitives up to multiplicative constants. Second, due to the presence of the label vector $\mathbf{y}$ and attribute vector $\mathbf{a}$, we are required to bound several “cross terms” in the primitives, for which naive variational characterizations of the inverse Wishart matrix $(\mathbf{Q}\mathbf{Q}^\top + \tau\mathbf{I})^{-1}$ will not suffice. The analysis uses concentration bounds on quadratic forms of inverse Wishart matrices that are sharp even in their multiplicative constant, facilitating a tight analysis of these cross terms. Finally, special care is taken to sharply characterize the role of the label adjustment matrix $\mathbf{\Delta}^{-1}$ in the analysis.

To begin with, we recall the definition of $\mathbf{X}_k\mathbf{X}_k^\top$ and $(\mathbf{X}_k\mathbf{X}_k^\top + \tau\mathbf{I})^{-1}$ in Eq. (13) and Eq. (14) for $k = 0, 1, 2$. Meanwhile, we write $\mathbf{A}_k$ in terms of the primitives as below:

$$\mathbf{A}_k = \begin{bmatrix} 1 + \|\bar{\boldsymbol{\mu}}_k\|_2^2 s_{k,k}^{(k-1)} & \|\bar{\boldsymbol{\mu}}_k\|_2 h_{k,k}^{(k-1)} & \|\bar{\boldsymbol{\mu}}_k\|_2 s_{k,k}^{(k-1)} \\ \|\bar{\boldsymbol{\mu}}_k\|_2 s_{k,k}^{(k-1)} & 1 + h_{k,k}^{(k-1)} & s_{k,k}^{(k-1)} \\ \|\bar{\boldsymbol{\mu}}_k\|_2 h_{k,k}^{(k-1)} & t_{k,k}^{(k-1)} & 1 + h_{k,k}^{(k-1)} \end{bmatrix},$$

its determinant as

$$\det(\mathbf{A}_k) = s_{k,k}^{(k-1)} \left(\|\bar{\boldsymbol{\mu}}_k\|_2^2 - t_{k,k}^{(k-1)} \right) + \left(h_{k,k}^{(k-1)} + 1 \right)^2, \quad (18)$$

and its adjugate matrix $\text{adj}(\mathbf{A}_k)$ in column format as

$$\begin{aligned} \text{adj}(\mathbf{A}_k)_1 &= \begin{bmatrix} \left(h_{k,k}^{(k-1)} + 1 \right)^2 - s_{k,k}^{(k-1)} t_{k,k}^{(k-1)} \\ - \|\bar{\boldsymbol{\mu}}_k\|_2 s_{k,k}^{(k-1)} \\ \|\bar{\boldsymbol{\mu}}_k\|_2 \left(s_{k,k}^{(k-1)} t_{k,k}^{(k-1)} - h_{k,k}^{(k-1)} - h_{k,k}^{(k-1)2} \right) \end{bmatrix}, \\ \text{adj}(\mathbf{A}_k)_2 &= \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_k\|_2 \left(s_{k,k}^{(k-1)} t_{k,k}^{(k-1)} - h_{k,k}^{(k-1)} - h_{k,k}^{(k-1)2} \right) \\ h_{k,k}^{(k-1)} + 1 + \|\bar{\boldsymbol{\mu}}_k\|_2^2 s_{k,k}^{(k-1)} \\ \|\bar{\boldsymbol{\mu}}_k\|_2^2 h_{k,k}^{(k-1)2} - t_{k,k}^{(k-1)} \left(1 + \|\bar{\boldsymbol{\mu}}_k\|_2^2 s_{k,k}^{(k-1)} \right) \end{bmatrix}, \\ \text{adj}(\mathbf{A}_k)_3 &= \begin{bmatrix} - \|\bar{\boldsymbol{\mu}}_k\|_2 s_{k,k}^{(k-1)} \\ - s_{k,k}^{(k-1)} \\ h_{k,k}^{(k-1)} + 1 + \|\bar{\boldsymbol{\mu}}_k\|_2^2 s_{k,k}^{(k-1)} \end{bmatrix}. \end{aligned}$$

Next, we will need the following inverse Wishart concentration lemma from [18].

Lemma 5 ([18], Lemma 21). *Let $\mathbf{A} \sim \text{Wishart}(d, \mathbf{I}_n)$ and $d'(n) := d - n + 1$. For any vector $\mathbf{u} \in S^{n-1}$ and any $t > 0$, we have*

$$\Pr \left[\frac{1}{\mathbf{u}^\top \mathbf{A}^{-1} \mathbf{u}} > d'(n) + \sqrt{2t \cdot d'(n)} + 2t \right] \leq e^{-t} \quad (19)$$

$$\Pr \left[\frac{1}{\mathbf{u}^\top \mathbf{A}^{-1} \mathbf{u}} < d'(n) - \sqrt{2t \cdot d'(n)} \right] \leq e^{-t}, \quad (20)$$

provided that $d'(n) > 2\max\{t, 1\}$.

The following corollary further characterizes the eigenvalues of an inverse ridge-regularized inverse Wishart matrix.

Corollary 2. *Let $\mathbf{A} \sim \text{Wishart}(d, \mathbf{I}_n)$ and $d'(n) := d - n + 1$. For any $t > 0$, and $\tau \geq 0$, we have*

$$\frac{1}{d'(n) + \sqrt{2t \cdot d'(n)} + 2t + \tau} \leq \lambda_n \left((\mathbf{A} + \tau \mathbf{I})^{-1} \right) \leq \lambda_1 \left((\mathbf{A} + \tau \mathbf{I})^{-1} \right) \leq \frac{1}{d'(n) - \sqrt{2t \cdot d'(n)} + \tau}, \quad (21)$$

with probability at least $1 - e^{-t}$, provided that $d'(n) > 2\max\{t, 1\}$.

Proof. We first show the upper bound

$$\lambda_1 \left((\mathbf{A} + \tau \mathbf{I})^{-1} \right) = \frac{1}{\lambda_n(\mathbf{A} + \tau \mathbf{I})} = \frac{1}{\lambda_n(\mathbf{A}) + \tau} = \frac{1}{\frac{1}{\lambda_1(\mathbf{A}^{-1})} + \tau} \leq \frac{1}{d'(n) - \sqrt{2t \cdot d'(n)} + \tau},$$

where the inequality follows by Eq. (20) with probability at least $1 - e^{-t}$. Similarly, for the lower bound, we have

$$\lambda_n \left((\mathbf{A} + \tau \mathbf{I})^{-1} \right) = \frac{1}{\lambda_1(\mathbf{A} + \tau \mathbf{I})} = \frac{1}{\lambda_1(\mathbf{A}) + \tau} = \frac{1}{\frac{1}{\lambda_n(\mathbf{A}^{-1})} + \tau} \geq \frac{1}{d'(n) + \sqrt{2t \cdot d'(n)} + 2t + \tau},$$

where the inequality follows by Eq. (19) with probability at least $1 - e^{-t}$. □

A.1 Proof for Lemma 2 (without adjustment weight)

Proof. (Lemma 2) We prove the lemma by induction. We first show the bounds for the order $k = 0$ (base case). Then, we assume that the bounds are true for order $k = k$ and we show the bounds for order $k = k + 1$. The proof for the base case $k = 0$ resembles the primitive proofs in [25]. Note that the indexing of constants in the proof below does not match the indexing of constants in the main statement of Lemma 2.

Bounds for order $k = 0$.

We have $\bullet s_{i,i}^{(0)} = \mathbf{v}_i^\top \left(\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i = \mathbf{v}_i^\top \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i$ for $i = 1, 2$. Based on Corollary 2, let $t = \sqrt{2 \log(n/\delta)}$. Then, we have with probability at least $1 - \frac{\delta^2}{n^2}$,

$$s_{i,i}^{(0)} \leq \frac{\|\mathbf{v}_i\|_2^2}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \leq \frac{n}{d'(n) \left(1 - \frac{2}{\sqrt{C_n}} \right) + \tau} = \frac{n \left(\frac{\sqrt{C_n}}{\sqrt{C_n} - 2} \right)}{d'(n) + \left(\frac{\sqrt{C_n}}{\sqrt{C_n} - 2} \right) \tau} \leq \left(\frac{C_1 + 1}{C_1} \right) \frac{n}{d + \tau},$$

where the second inequality follows by Assumption 1(D) such that $\frac{1}{C_n^2} \geq \frac{\log(n/\delta)}{d'(n)}$. Similarly,

$$s_{i,i}^{(0)} \geq \frac{\|\mathbf{v}_i\|_2^2}{d'(n) + \sqrt{4 \log(n/\delta) \cdot d'(n)} + 4 \log(n/\delta) + \tau} \geq \frac{n}{d'(n) \left(1 + \frac{2}{\sqrt{C_n}} + \frac{4}{C_n} \right) + \tau} \geq \left(\frac{C_1 - 1}{C_1} \right) \frac{n}{d + \tau}.$$

$\bullet s_{1,2}^{(0)} = \mathbf{v}_1^\top \left(\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_2 = \mathbf{y}^\top \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{a}$. Note that this is a “cross-term” primitive. We apply the parallelogram law to get

$$s_{1,2}^{(0)} = \mathbf{y}^\top \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{a} = \frac{1}{4} \left(\mathbf{v}_+^\top \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_+ - \mathbf{v}_-^\top \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_- \right),$$

where we denote $\mathbf{v}_\pm := \mathbf{y} \pm \mathbf{a}$. It is easy to verify that $\|\mathbf{v}_+\|_2^2 = 4n_+$ and $\|\mathbf{v}_-\|_2^2 = 4n_-$. Hence, again applying Corollary 2 gives us

$$\begin{aligned}
s_{1,2}^{(0)} &\leq \frac{1}{4} \left(\frac{4n_+}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} - \frac{4n_-}{d'(n) + \sqrt{4 \log(n/\delta) \cdot d'(n)} + 4 \log(n/\delta) + \tau} \right) \\
&\leq \frac{n_+}{d'(n) \left(1 - \frac{2}{\sqrt{Cn^2}}\right) + \tau} - \frac{n_-}{d'(n) \left(1 + \frac{2}{\sqrt{Cn^2}} + \frac{4}{Cn^2}\right) + \tau} \\
&\leq \left(\frac{C_2+1}{C_2}\right) \frac{n_+}{d+\tau} - \left(\frac{C_2-1}{C_2}\right) \frac{n_-}{d+\tau} =: \hat{C}_3 \frac{n_+}{d+\tau} - \check{C}_3 \frac{n_-}{d+\tau}. \\
s_{1,2}^{(0)} &\geq \frac{1}{4} \left(\frac{4n_+}{d'(n) + \sqrt{4 \log(n/\delta) \cdot d'(n)} + 4 \log(n/\delta) + \tau} - \frac{4n_-}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \right) \\
&\geq \frac{n_+}{d'(n) \left(1 + \frac{2}{\sqrt{Cn^2}} + \frac{4}{Cn^2}\right) + \tau} - \frac{n_-}{d'(n) \left(1 - \frac{2}{\sqrt{Cn^2}}\right) + \tau} \\
&\geq \left(\frac{C_2-1}{C_2}\right) \frac{n_+}{d+\tau} - \left(\frac{C_2+1}{C_2}\right) \frac{n_-}{d+\tau} =: \check{C}_3 \frac{n_+}{d+\tau} - \hat{C}_3 \frac{n_-}{d+\tau}.
\end{aligned}$$

• $t_{i,i}^{(0)} = \mathbf{d}_i^\top (\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I})^{-1} \mathbf{d}_i = \mathbf{d}_i^\top (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \mathbf{d}_i$, for $i = 1, 2$. Note that the lemma in [26][Corollary 8.1] gives $\|\mathbf{d}_i\|_2 \leq C\sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2$, we have

$$t_{i,i}^{(0)} \leq \frac{\|\mathbf{d}_i\|_2^2}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \leq \frac{Cn \|\bar{\boldsymbol{\mu}}_i\|_2^2}{d'(n) \left(1 - \frac{2}{\sqrt{Cn^2}}\right) + \tau} \leq \frac{C_3 n \|\bar{\boldsymbol{\mu}}_i\|_2^2}{d+\tau}.$$

• $t_{1,2}^{(0)} = \mathbf{d}_1^\top (\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I})^{-1} \mathbf{d}_2 = \mathbf{d}_1^\top (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \mathbf{d}_2$. By the sub-multiplicative property of the matrix norm, we get

$$\begin{aligned}
t_{1,2}^{(0)} &\leq \|\mathbf{d}_1\|_2 \|\mathbf{d}_2\|_2 \left\| (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \right\|_2 \leq \frac{Cn \|\bar{\boldsymbol{\mu}}_1\|_2 \|\bar{\boldsymbol{\mu}}_2\|_2}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \leq \frac{C_3 n \|\bar{\boldsymbol{\mu}}_1\|_2 \|\bar{\boldsymbol{\mu}}_2\|_2}{d+\tau}. \\
t_{1,2}^{(0)} &\geq -\|\mathbf{d}_1\|_2 \|\mathbf{d}_2\|_2 \left\| (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \right\|_2 \geq \frac{-C_3 n \|\bar{\boldsymbol{\mu}}_1\|_2 \|\bar{\boldsymbol{\mu}}_2\|_2}{d+\tau}.
\end{aligned}$$

• $h_{i,j}^{(0)} = \mathbf{d}_i^\top (\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I})^{-1} \mathbf{v}_j = \mathbf{d}_i^\top (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \mathbf{v}_j$, for $i, j = 1, 2$. Hence,

$$\begin{aligned}
h_{i,j}^{(0)} &\leq \|\mathbf{d}_i\|_2 \|\mathbf{v}_j\|_2 \left\| (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \right\|_2 \leq \frac{Cn \|\bar{\boldsymbol{\mu}}_i\|_2}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \leq \frac{C_4 n \|\bar{\boldsymbol{\mu}}_i\|_2}{d+\tau}. \\
h_{i,j}^{(0)} &\geq -\|\mathbf{d}_i\|_2 \|\mathbf{v}_j\|_2 \left\| (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \right\|_2 \geq \frac{-C_4 n \|\bar{\boldsymbol{\mu}}_i\|_2}{d+\tau}.
\end{aligned}$$

• $s_{u,u}^{(0)} = \mathbf{u}^\top (\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I})^{-1} \mathbf{u} = \mathbf{u}^\top (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \mathbf{u}$. By Corollary 2 and setting $t = \sqrt{2 \log(n/\delta)}$, we have with probability at least $1 - \frac{\delta^2}{n^2}$,

$$s_{u,u}^{(0)} \leq \frac{1}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \leq \frac{1}{d'(n) \left(1 - \frac{2}{\sqrt{C_n}}\right) + \tau} \leq \left(\frac{C_5 + 1}{C_5}\right) \frac{1}{d + \tau},$$

$$s_{u,u}^{(0)} \geq \frac{1}{d'(n) + \sqrt{4 \log(n/\delta) \cdot d'(n)} + 4 \log(1/\delta) + \tau} \geq \frac{n}{d'(n) \left(1 + \frac{2}{\sqrt{C_n}} + \frac{4}{C_n}\right) + \tau} \geq \left(\frac{C_5 - 1}{C_5}\right) \frac{1}{d + \tau}.$$

• $s_{u,i}^{(0)} = \mathbf{u}^\top (\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I})^{-1} \mathbf{u} = \mathbf{u}^\top (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \mathbf{u}$ for $i = 1, 2$. Again, by Corollary 2, we have

$$s_{u,i}^{(0)} \leq \|\mathbf{u}\|_2 \|\mathbf{v}_i\|_2 \left\| (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \right\|_2 \leq \frac{C \sqrt{n}}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \leq \frac{C_6 \sqrt{n}}{d + \tau}.$$

$$s_{u,i}^{(0)} \geq -\|\mathbf{u}\|_2 \|\mathbf{v}_i\|_2 \left\| (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \right\|_2 \geq \frac{-C_6 \sqrt{n}}{d + \tau}.$$

• $h_{i,u}^{(0)} = \mathbf{d}_i^\top (\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I})^{-1} \mathbf{u} = \mathbf{d}_i^\top (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \mathbf{u}$, for $i = 1, 2$. Hence,

$$h_{i,u}^{(0)} \leq \|\mathbf{d}_i\|_2 \|\mathbf{u}\|_2 \left\| (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \right\|_2 \leq \frac{C \sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \leq \frac{C_7 \sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau}.$$

$$h_{i,u}^{(0)} \geq -\|\mathbf{d}_i\|_2 \|\mathbf{u}\|_2 \left\| (\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I})^{-1} \right\|_2 \geq \frac{-C_7 \sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau}.$$

Thus, we have characterized all of the zeroth-order primitives. Next, we assume that the bounds for order $k = k$ are true, and we show the bounds for order $k = k + 1$. The key idea is that the higher order primitives are a constant fraction of the previous-order primitives. The following lemma, whose proof is deferred to Appendix B, details the recursive calculation that shows this.

Lemma 6. *For a function $f_{A_k} : \mathbb{R}^4 \rightarrow \mathbb{R}$ such that*

$$\begin{aligned} f_{A_k} \begin{pmatrix} x_a, x_b, \\ x_c, x_d \end{pmatrix} &:= [\|\bar{\boldsymbol{\mu}}_k\|_2 x_a, x_b, x_a] \text{adj}(\mathbf{A}_k) \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_k\|_2 x_c \\ x_c \\ x_d \end{bmatrix} \\ &= \left(\|\bar{\boldsymbol{\mu}}_k\|_2^2 - t_{k,k}^{(k-1)} \right) x_a x_c + \left(1 + h_{k,k}^{(k-1)} \right) (x_a x_d + x_b x_c) - s_{k,k}^{(k-1)} x_b x_d, \end{aligned}$$

we have the absolute value upper bound

$$\left| f_{A_k} \begin{pmatrix} x_a, x_b, \\ x_c, x_d \end{pmatrix} \right| \leq \|\bar{\boldsymbol{\mu}}_k\|_2^2 |x_a| |x_c| + |s_{k,k}^{(k-1)}| |x_b| |x_d| + \left(1 + \frac{1}{C_1} \right) (|x_a| |x_d| + |x_b| |x_c|).$$

The following lemma provides an upper and lower bound for $\det(\mathbf{A}_k)$. The proof of this lemma can be found in Appendix B.

Lemma 7. *For $k = 1, 2$, we have*

$$\check{C}_1 \leq \det(\mathbf{A}_k) \leq \hat{C}_1,$$

for some $C_1 > 1$ and $\hat{C}_m := \frac{C_m+1}{C_m}$ and $\check{C}_m := \frac{C_m-1}{C_m}$.

We can now complete the proof of Lemma 2. We only present the calculation of primitive $s_{i,i}^{(k+1)}$ due to page limitations; the proofs of the other primitives follow the same idea and are presented in Appendix C.

Bounds for order $k = k + 1$

• $s_{i,i}^{(k+1)}$ for $i = 1, 2$.

$$\begin{aligned}
s_{i,i}^{(k+1)} &= \mathbf{v}_i^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i \\
&= \mathbf{v}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i - \mathbf{v}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i \\
&= s_{i,i}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)}, s_{i,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{i,k+1}^{(k)} \\ s_{i,k+1}^{(k)} \\ h_{k+1,i}^{(k)} \end{bmatrix} \\
&= \frac{s_{i,i}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)}, s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
s_{i,i}^{(k+1)} &\leq \frac{\left| s_{i,i}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)}, s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| s_{i,i}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{i,k+1}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{i,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| + \left| h_{k+1,i}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right) + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{n}{d + \tau} + \frac{C_3 + 1}{C_3} \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2}{(d + \tau)^2} + \frac{C_4 + 1}{C_4} \frac{C_5 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n^2}{(d + \tau)^2} + \frac{C_6 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^3}{(d + \tau)^3} \right) \\
&\leq \frac{C_7 + 1}{C_7} \frac{n}{d + \tau},
\end{aligned}$$

where in the third inequality we apply lower bound of $\det(\mathbf{A}_{k+1})$ in Lemma 7 and primitive bounds in order k , and in the last inequality, we apply Assumption 1(C) such that $d \geq CR_+ n$. Similarly,

$$\begin{aligned}
s_{i,i}^{(k+1)} &\geq \frac{s_{i,i}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)}, s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) s_{i,i}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{i,k+1}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{i,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| + \left| h_{k+1,i}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right) + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 - 1}{C_2} \frac{n}{d + \tau} - \frac{C_3 + 1}{C_3} \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2}{(d + \tau)^2} - \frac{C_4 + 1}{C_4} \frac{C_5 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n^2}{(d + \tau)^2} - \frac{C_6 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^3}{(d + \tau)^3} \right) \\
&\geq \frac{C_7 - 1}{C_7} \frac{n}{d + \tau}.
\end{aligned}$$

This completes the proof of the lemma. $\square$

A.2 Proof for Lemma 3 (with adjustment weights)

Proof. (Lemma 3) Before we start the proof of Lemma 3, we need the following auxiliary lemma, whose proof we defer to Appendix B.

Lemma 8. For $\frac{1}{n} \leq \Delta_{\pm} \leq 1$ and $\Delta_+ \geq \Delta_-$, we have

$$\tilde{C} \left[\frac{1}{\sqrt{Cn}} \left(n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \right) \right] \leq \frac{1}{2} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right).$$

for some $C > 144$, and $\tilde{C} > 2$.

Similar to the proof of Lemma 2, we prove by induction.

Bounds for base case $k = 0$:

• $s_{i_{\Delta}, i}^{(0)} = \mathbf{v}_i^\top \Delta^{-1} \left(\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i = \mathbf{v}_i^\top \Delta^{-1} \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i$, for $i = 1, 2$. Applying the parallelogram law gives

$$\begin{aligned} s_{i_{\Delta}, i}^{(0)} &= \frac{1}{4} \left(\left(\Delta^{-1} \mathbf{v}_i + \mathbf{v}_i \right)^\top \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \left(\Delta^{-1} \mathbf{v}_i + \mathbf{v}_i \right) - \left(\Delta^{-1} \mathbf{v}_i - \mathbf{v}_i \right)^\top \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \left(\Delta^{-1} \mathbf{v}_i - \mathbf{v}_i \right) \right). \end{aligned}$$

Next, we have

$$\begin{aligned} \left\| \Delta^{-1} \mathbf{v}_i + \mathbf{v}_i \right\|_2^2 &= \left(1 + \frac{1}{\Delta_+} \right)^2 n_+ + \left(1 + \frac{1}{\Delta_-} \right)^2 n_- = n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}, \\ \left\| \Delta^{-1} \mathbf{v}_i - \mathbf{v}_i \right\|_2^2 &= \left(1 - \frac{1}{\Delta_+} \right)^2 n_+ + \left(1 - \frac{1}{\Delta_-} \right)^2 n_- = n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}. \end{aligned}$$

Applying Corollary 2 with $t = 2 \log(n/\delta)$, we have, with probability at least $1 - \frac{\delta^2}{n^2}$,

$$\begin{aligned} s_{i_{\Delta}, i}^{(0)} &\leq \frac{1}{4} \left(\frac{n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) - \sqrt{4 \log(n/\delta)} \cdot d'(n) + \tau} - \frac{n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) + \sqrt{4 \log(n/\delta)} \cdot d'(n) + 4 \log(n/\delta) + \tau} \right) \\ &\leq \frac{1}{4} \left(\frac{n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) \left(1 - \frac{2}{\sqrt{Cn}} \right) + \tau} - \frac{n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) \left(1 + \frac{2}{\sqrt{Cn}} + \frac{4}{Cn^2} \right) + \tau} \right) \\ &\leq \frac{\left(1 + \frac{2}{Cn^2} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \left(\frac{1}{\sqrt{Cn}} + \frac{1}{Cn^2} \right) \left(n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \right)}{d'(n) \left(1 - \frac{2}{\sqrt{Cn}} \right) \left(1 + \frac{2}{\sqrt{Cn}} + \frac{4}{Cn^2} \right) + \tau} \\ &\leq \frac{\left(1 + \frac{2}{Cn^2} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \frac{2}{\sqrt{Cn}} \left(n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \right)}{d + \tau} \\ &\leq \frac{\left(1 + \frac{2}{Cn^2} + \frac{1}{\tilde{C}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} = \frac{\left(\frac{C_1+1}{C_1} + \frac{1}{\tilde{C}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau}, \end{aligned}$$

where the second inequality follows by Assumption 1(D) such that $d'(n) > Cn^2 \log(n/\delta)$, and the last

inequality follows by Lemma 8. Next, we have the lower bound:

$$\begin{aligned}
s_{i_{\Delta},i}^{(0)} &\geq \frac{1}{4} \left(\frac{n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) + \sqrt{4 \log(n/\delta) \cdot d'(n)} + 4 \log(n/\delta) + \tau} - \frac{n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \right) \\
&\geq \frac{1}{4} \left(\frac{n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) \left(1 + \frac{2}{\sqrt{Cn}} + \frac{4}{Cn^2}\right) + \tau} - \frac{n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) \left(1 - \frac{2}{\sqrt{Cn}}\right) + \tau} \right) \\
&\geq \frac{\left(1 + \frac{2}{Cn^2}\right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-}\right) - \left(\frac{1}{\sqrt{Cn}} + \frac{1}{Cn^2}\right) \left(n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}\right)}{d'(n) \left(1 - \frac{2}{\sqrt{Cn}}\right) \left(1 + \frac{2}{\sqrt{Cn}} + \frac{4}{Cn^2}\right) + \tau} \\
&\geq \frac{\left(1 + \frac{2}{Cn^2}\right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-}\right) - \frac{2}{\sqrt{Cn}} \left(n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}\right)}{d + \tau} \\
&\geq \frac{\left(1 + \frac{2}{Cn^2} - \frac{1}{\bar{C}}\right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-}\right)}{d + \tau} = \frac{\left(\frac{C_1-1}{C_1} - \frac{1}{\bar{C}}\right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-}\right)}{d + \tau},
\end{aligned}$$

where the second inequality follows by Assumption 1(D) such that $d'(n) > Cn^2 \log(n/\delta)$, and the fourth inequality follows by Lemma 8.

• $s_{2_{\Delta},1}^{(0)} = \mathbf{v}_2^\top \Delta^{-1} \left(\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_1 = \mathbf{y}^\top \Delta^{-1} \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{a}$. Applying the parallelogram law gives

$$s_{2_{\Delta},1}^{(0)} = \frac{1}{4} \left(\left(\Delta^{-1} \mathbf{y} + \mathbf{a} \right)^\top \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \left(\Delta^{-1} \mathbf{y} + \mathbf{a} \right) - \left(\Delta^{-1} \mathbf{y} - \mathbf{a} \right)^\top \left(\mathbf{Q} \mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \left(\Delta^{-1} \mathbf{y} - \mathbf{a} \right) \right).$$

Next, we have

$$\begin{aligned}
\left\| \Delta^{-1} \mathbf{y} + \mathbf{a} \right\|_2^2 &= \left(1 + \frac{1}{\Delta_+}\right)^2 n_+ + \left(1 - \frac{1}{\Delta_-}\right)^2 n_- = n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}. \\
\left\| \Delta^{-1} \mathbf{y} - \mathbf{a} \right\|_2^2 &= \left(1 - \frac{1}{\Delta_+}\right)^2 n_+ + \left(1 + \frac{1}{\Delta_-}\right)^2 n_- = n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}.
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
s_{2_{\Delta},1}^{(0)} &\leq \frac{1}{4} \left(\frac{n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} - \frac{n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) + \sqrt{4 \log(n/\delta) \cdot d'(n)} + 4 \log(n/\delta) + \tau} \right) \\
&\leq \frac{1}{4} \left(\frac{n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) \left(1 - \frac{2}{\sqrt{Cn}}\right) + \tau} - \frac{n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) \left(1 + \frac{2}{\sqrt{Cn}} + \frac{4}{Cn^2}\right) + \tau} \right) \\
&\leq \frac{\left(1 + \frac{2}{Cn^2}\right) \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-}\right) + \left(\frac{1}{\sqrt{Cn}} + \frac{1}{Cn^2}\right) \left(n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}\right)}{d'(n) \left(1 - \frac{2}{\sqrt{Cn}}\right) \left(1 + \frac{2}{\sqrt{Cn}} + \frac{4}{Cn^2}\right) + \tau} \\
&\leq \frac{\left(1 + \frac{2}{Cn^2}\right) \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-}\right) + \frac{2}{\sqrt{Cn}} \left(n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}\right)}{d + \tau} \\
&\leq \frac{\left(1 + \frac{2}{Cn^2}\right) \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-}\right) + \frac{1}{\bar{C}} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-}\right)}{d + \tau} = \frac{\left(\frac{C_1+1}{C_1}\right) \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-}\right) + \frac{1}{\bar{C}} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-}\right)}{d + \tau}.
\end{aligned}$$

$$\begin{aligned}
s_{2\Delta,1}^{(0)} &\geq \frac{1}{4} \left(\frac{n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) + \sqrt{4 \log(n/\delta) \cdot d'(n)} + 4 \log(n/\delta) + \tau} - \frac{n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \right) \\
&\geq \frac{1}{4} \left(\frac{n + \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} - \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) \left(1 + \frac{2}{\sqrt{Cn}} + \frac{4}{Cn^2}\right) + \tau} - \frac{n - \frac{2n_+}{\Delta_+} + \frac{n_+}{\Delta_+^2} + \frac{2n_-}{\Delta_-} + \frac{n_-}{\Delta_-^2}}{d'(n) \left(1 - \frac{2}{\sqrt{Cn}}\right) + \tau} \right) \\
&\geq \frac{\left(1 + \frac{2}{Cn^2}\right) \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-}\right) - \left(\frac{1}{\sqrt{Cn}} + \frac{1}{Cn^2}\right) \left(n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}\right)}{d'(n) \left(1 - \frac{2}{\sqrt{Cn}}\right) \left(1 + \frac{2}{\sqrt{Cn}} + \frac{4}{Cn^2}\right) + \tau} \\
&\geq \frac{\left(1 + \frac{2}{Cn^2}\right) \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-}\right) - \frac{2}{\sqrt{Cn}} \left(n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}\right)}{d + \tau} \\
&\geq \frac{\left(1 + \frac{2}{Cn^2}\right) \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-}\right) - \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-}\right)}{d + \tau} = \frac{\left(\frac{C_1+1}{C_1}\right) \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-}\right) - \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-}\right)}{d + \tau}.
\end{aligned}$$

• $s_{i\Delta,i\Delta}^{(0)}$ for $i = 1, 2$,

$$s_{i\Delta,i\Delta}^{(0)} = \mathbf{v}_i^\top \mathbf{\Delta}^{-1} \left(\mathbf{Q}\mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{\Delta}^{-1} \mathbf{v}_i.$$

Based on Corollary 2, let $t = \sqrt{2 \log(n/\delta)}$. The variational characterization of eigenvalues gives us, with probability at least $1 - \frac{\delta^2}{n^2}$,

$$\begin{aligned}
s_{i\Delta,i\Delta}^{(0)} &\leq \frac{\|\mathbf{\Delta}_{-1} \mathbf{v}_i\|_2^2}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \leq \left(\frac{C_1 + 1}{C_1} \right) \frac{\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}}{d + \tau} \\
s_{i\Delta,i\Delta}^{(0)} &\geq \frac{\|\mathbf{\Delta}_{-1} \mathbf{v}_i\|_2^2}{d'(n) + \sqrt{4 \log(n/\delta) \cdot d'(n)} + 4 \log(n/\delta) + \tau} \geq \left(\frac{C_1 - 1}{C_1} \right) \frac{\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2}}{d + \tau}.
\end{aligned}$$

• $h_{i,j\Delta}^{(0)} = \mathbf{d}_i^\top \left(\mathbf{X}_0 \mathbf{X}_0^\top + \tau \mathbf{I} \right)^{-1} \mathbf{\Delta}^{-1} \mathbf{v}_j = \mathbf{d}_i^\top \left(\mathbf{Q}\mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{\Delta}^{-1} \mathbf{v}_j$ for $i, j = 1, 2$. Hence,

$$\begin{aligned}
h_{i,j\Delta}^{(0)} &\leq \|\mathbf{d}_i\|_2 \|\mathbf{\Delta}_{-1} \mathbf{v}_j\|_2 \left\| \left(\mathbf{Q}\mathbf{Q}^\top + \tau \mathbf{I} \right)^{-1} \right\|_2 \leq \frac{C \sqrt{n \left(\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \right)} \|\bar{\boldsymbol{\mu}}_i\|_2}{d'(n) - \sqrt{4 \log(n/\delta) \cdot d'(n)} + \tau} \leq \frac{C_1 \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau}. \\
h_{i,j\Delta}^{(0)} &\geq \frac{-C_1 \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau}.
\end{aligned}$$

• $o_{i\Delta,i\Delta}^{(k)} = \mathbf{v}_i^\top \mathbf{\Delta}^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{X}_k \mathbf{X}_k^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{\Delta}^{-1} \mathbf{v}_i$, for $i = 1, 2$. We start with the upper bound

$$\begin{aligned}
o_{i\Delta,i\Delta}^{(k)} &= \mathbf{v}_i^\top \mathbf{\Delta}^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{X}_k \mathbf{X}_k^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{\Delta}^{-1} \mathbf{v}_i \\
&\leq \lambda_1 \left(\mathbf{X}_k \mathbf{X}_k^\top \right) \left\| \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{\Delta}^{-1} \mathbf{v}_i \right\|_2^2 \\
&\leq \lambda_1 \left(\mathbf{X}_k \mathbf{X}_k^\top \right) \left\| \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \right\|_2^2 \left\| \mathbf{\Delta}^{-1} \mathbf{v}_i \right\|_2^2,
\end{aligned} \tag{22}$$

where the first inequality follows by the variational characterization of eigenvalues, and the second inequality applies the sub-multiplicative property of the matrix norm. Next, we focus on analyzing the eigenvalues of $\mathbf{X}_k \mathbf{X}_k^\top$ and $\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I}$. We have

$$\begin{aligned}
(22) &= \lambda_1 \left(\mathbf{X}_k \mathbf{X}_k^\top \right) \lambda_1 \left(\left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \right)^2 \left\| \Delta^{-1} \mathbf{v}_i \right\|_2^2 \\
&= \frac{\lambda_1 \left(\left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \right)^2}{\lambda_n \left(\left(\mathbf{X}_k \mathbf{X}_k^\top \right)^{-1} \right)} \left\| \Delta^{-1} \mathbf{v}_i \right\|_2^2 \leq \frac{\left(\hat{C}_1 / (d + \tau) \right)^2}{\check{C}_2 / d} \left(\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \right) = \frac{\hat{C}_3 n_\Delta d}{(d + \tau)^2},
\end{aligned}$$

where the inequality follows by applying the primitive bounds of $s_{u,u}^{(k)}$ in Lemma 2. Similarly, for the lower bound we have

$$\begin{aligned}
o_{i_\Delta, i_\Delta}^{(k)} &= \mathbf{v}_i^\top \Delta^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{X}_k \mathbf{X}_k^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \Delta^{-1} \mathbf{v}_i \\
&\geq \lambda_n \left(\left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{X}_k \mathbf{X}_k^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \right) \left\| \Delta^{-1} \mathbf{v}_i \right\|_2^2 \\
&= \frac{\left\| \Delta^{-1} \mathbf{v}_i \right\|_2^2}{\lambda_1 \left(\left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right) \left(\mathbf{X}_k \mathbf{X}_k^\top \right)^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right) \right)} \\
&\geq \frac{\left\| \Delta^{-1} \mathbf{v}_i \right\|_2^2}{\lambda_1 \left(\left(\mathbf{X}_k \mathbf{X}_k^\top \right)^{-1} \right) \lambda_1 \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^2}. \tag{23}
\end{aligned}$$

Next, we have

$$(23) = \frac{\lambda_n \left(\left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \right)^2}{\lambda_1 \left(\left(\mathbf{X}_k \mathbf{X}_k^\top \right)^{-1} \right)} \left\| \Delta^{-1} \mathbf{v}_i \right\|_2^2 \geq \frac{\left(\check{C}_1 / (d + \tau) \right)^2}{\hat{C}_2 / d} \left(\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \right) = \frac{\check{C}_3 n_\Delta d}{(d + \tau)^2}.$$

Finally, we show the inductive step from $k \rightarrow k+1$. The calculation is identical to the one that we did for the proof of Lemma 2, since we reuse the $f_{\mathbf{A}_{K+1}}$ function in Lemma 6, and only the input variables change. (The full proof of the inductive step is provided in Appendix C for completeness.) This completes the proof of the lemma. $\square$

B Proofs of auxiliary lemmas (Lemmas 4, 6, 7 and 8)

In this section, we provide the proofs of auxiliary technical lemmas used in Appendix A.

B.1 Proof of Lemma 4

We have

$$\begin{aligned}
\frac{1}{2} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_c\|_2 &\geq \frac{1}{2} \sqrt{\frac{n_+^2}{\Delta_+^2} + \frac{n_-^2}{\Delta_-^2}} \|\bar{\boldsymbol{\mu}}_c\|_2 \\
&= \sqrt{\frac{n_+}{\Delta_+^2} \left(\frac{n_+ \|\bar{\boldsymbol{\mu}}_c\|_2^2}{4} \right) + \frac{n_-}{\Delta_-^2} \left(\frac{n_- \|\bar{\boldsymbol{\mu}}_c\|_2^2}{4} \right)} \\
&\geq \sqrt{\left(\frac{n_- \|\bar{\boldsymbol{\mu}}_c\|_2^2}{4} \right) \left(\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \right)} \\
&\geq \sqrt{\left(\frac{n_- C n \log(n/\delta)}{4} \right) \left(\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \right)} \\
&= C_1 \sqrt{n \left(\frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \right)} = C_1 \sqrt{nn_\Delta},
\end{aligned}$$

where the second inequality follows by $\Delta_- \leq \Delta_+$, the third inequality follows via the assumptions made in the lemma, and the last equality follows by the choice $C_1 = \sqrt{\frac{C n_- \log(n/\delta)}{4}} > 1$. On the other hand, if $n_- = \alpha n$, we can pick $C_1 = \sqrt{\alpha} \frac{\|\bar{\boldsymbol{\mu}}_c\|_2}{2}$. This completes the proof of the lemma. $\square$

B.2 Proof of Lemma 6

According to the definition of $f_{\mathbf{A}_k}$, we can compute the function value based on the definition of adjugate matrix of $\mathbf{A}_k$, by a series of algebraic steps. To upper bound $|f_{\mathbf{A}_k}|$, we apply the bounds on the $(k-1)$ -th order primitives to get

$$\begin{aligned}
\left| f_{\mathbf{A}_k} \begin{pmatrix} x_a, x_b, \\ x_c, x_d \end{pmatrix} \right| &= \left| \left(\|\bar{\boldsymbol{\mu}}_k\|_2^2 - t_{k,k}^{(k-1)} \right) x_a x_c + \left(1 + h_{k,k}^{(k-1)} \right) (x_a x_d + x_b x_c) - s_{k,k}^{(k-1)} x_b x_d \right| \\
&\leq \left| \|\bar{\boldsymbol{\mu}}_k\|_2^2 - t_{k,k}^{(k-1)} \right| |x_a| |x_c| + \left| 1 + h_{k,k}^{(k-1)} \right| (|x_a| |x_d| + |x_b| |x_c|) + \left| s_{k,k}^{(k-1)} \right| |x_b| |x_d| \\
&\leq \|\bar{\boldsymbol{\mu}}_k\|_2^2 |x_a| |x_c| + \left| s_{k,k}^{(k-1)} \right| |x_b| |x_d| + \left(1 + \frac{C_1 n \|\bar{\boldsymbol{\mu}}_k\|_2}{d + \tau} \right) (|x_a| |x_d| + |x_b| |x_c|) \\
&\leq \|\bar{\boldsymbol{\mu}}_k\|_2^2 |x_a| |x_c| + \left(1 + \frac{1}{C_2} \right) (|x_a| |x_d| + |x_b| |x_c|) + \left| s_{k,k}^{(k-1)} \right| |x_b| |x_d|,
\end{aligned}$$

where we introduce the primitive bounds in the second inequality, and apply Assumption 1(C) in the last inequality. $\square$

B.3 Proof of Lemma 7

According to the definition of $\det(\mathbf{A}_k)$ in Eq. (18), we have

$$\det(\mathbf{A}_k) = s_{k,k}^{(k-1)} \left(\|\bar{\boldsymbol{\mu}}_k\|_2^2 - t_{k,k}^{(k-1)} \right) + \left(h_{k,k}^{(k-1)} + 1 \right)^2.$$

Thus, we introduce the upper bound on the $(k-1)$ th order primitive to yield

$$\begin{aligned}\det(\mathbf{A}_k) &\leq \left(\frac{C_1+1}{C_1}\right) \frac{n}{d+\tau} \|\bar{\boldsymbol{\mu}}_k\|_2^2 + \left(\frac{C_2 n \|\bar{\boldsymbol{\mu}}_k\|_2}{d+\tau} + 1\right)^2 \\ &\leq \left(\frac{C_1+1}{C_1}\right) \frac{n}{d+\tau} \|\bar{\boldsymbol{\mu}}_k\|_2^2 + \left(\frac{1}{C_3} + 1\right)^2 \\ &\leq \left(\frac{C_4+1}{C_4}\right) \left(\frac{n}{d+\tau} \|\bar{\boldsymbol{\mu}}_k\|_2^2 + \frac{C_4+1}{C_4}\right) \leq \frac{C_5+1}{C_5},\end{aligned}$$

where the second and third inequality follow by applying Assumption 1(C), i.e. $\frac{R_+ n}{d} \leq \frac{1}{C}$. Similarly,

$$\det(\mathbf{A}_k) \geq \left(\frac{C_1-1}{C_1}\right)^2 \frac{n}{d+\tau} \|\bar{\boldsymbol{\mu}}_k\|_2^2 + \left(\frac{-C_2 n \|\bar{\boldsymbol{\mu}}_k\|_2}{d+\tau} + 1\right)^2 \geq \left(\frac{C_3-1}{C_3}\right).$$

This completes the proof of the lemma. $\square$

B.4 Proof of Lemma 8

Since $\Delta_{\pm} \leq 1$, we have

$$\begin{aligned}\frac{1 + \frac{1}{\Delta_{\pm}^2}}{2} &\leq \frac{1}{\Delta_{\pm}} + \frac{1}{2} \left(\frac{1}{\Delta_{\pm}^2} - 1 \right) \\ \iff 1 + \frac{1}{\Delta_{\pm}^2} &\leq \frac{1}{2\Delta_{\pm}} \left(4 + 2 \left(\frac{1}{\Delta_{\pm}} - \Delta_{\pm} \right) \right) \\ \iff n_{\pm} + \frac{n_{\pm}}{\Delta_{\pm}^2} &\leq \frac{n_{\pm}}{2\Delta_{\pm}} \left(4 + 2 \left(\frac{1}{\Delta_{\pm}} - \Delta_{\pm} \right) \right).\end{aligned}$$

Then, we have

$$\begin{aligned}n_+ + \frac{n_+}{\Delta_+^2} + n_- + \frac{n_-}{\Delta_-^2} &= n + \frac{n_+}{\Delta_+^2} + \frac{n_-}{\Delta_-^2} \\ &\leq \frac{n_+}{2\Delta_+} \left(4 + 2 \left(\frac{1}{\Delta_+} - \Delta_+ \right) \right) + \frac{n_-}{2\Delta_-} \left(4 + 2 \left(\frac{1}{\Delta_-} - \Delta_- \right) \right) \\ &\leq \left(\frac{n_+}{2\Delta_+} + \frac{n_-}{2\Delta_-} \right) \left(4 + 2 \left(\frac{1}{\Delta_-} - \Delta_- \right) \right),\end{aligned}$$

where the last inequality comes from $\Delta_- \leq \Delta_+$. Hence, it is sufficient to show $\frac{\sqrt{C}n}{C} \geq \left(4 + 2 \left(\frac{1}{\Delta_-} - \Delta_- \right) \right)$.

To this end, we have

$$\left(4 + 2 \left(\frac{1}{\Delta_-} - \Delta_- \right) \right)^2 \leq \left(4 + 2 \left(\frac{1}{\Delta_-} \right) \right)^2 = 16 + \frac{4}{\Delta_-^2} + \frac{16}{\Delta_-} \leq 16 + 4n^2 + 16n \leq \frac{Cn^2}{\tilde{C}^2}$$

for some $\frac{C}{\tilde{C}^2} \geq 36$ since $\frac{1}{n} \leq \Delta_- \leq 1$. This completes the proof of the lemma. $\square$

C Omitted calculations in proofs of Lemmas 2 and 3

In this section, we provide the detailed calculations of the inductive step, i.e. going from $k \rightarrow k+1$, in the proofs of Lemmas 2 and 3.

C.1 Inductive step for Lemma 2 (without adjustment weight)

- $s_{i,i}^{(k+1)}$ for $i = 1, 2$: We have

$$\begin{aligned}
s_{i,i}^{(k+1)} &= \mathbf{v}_i^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i \\
&= \mathbf{v}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i - \mathbf{v}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i \\
&= s_{i,i}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)}, s_{i,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{i,k+1}^{(k)} \\ s_{i,k+1}^{(k)} \\ h_{k+1,i}^{(k)} \end{bmatrix} \\
&= \frac{s_{i,i}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)}, s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
s_{i,i}^{(k+1)} &\leq \frac{\left| s_{i,i}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)}, s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| s_{i,i}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{i,k+1}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{i,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| + \left| h_{k+1,i}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right) + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{n}{d + \tau} + \frac{C_3 + 1}{C_3} \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2}{(d + \tau)^2} + \frac{C_4 + 1}{C_4} \frac{C_5 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n^2}{(d + \tau)^2} + \frac{C_6 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^3}{(d + \tau)^3} \right) \\
&\leq \frac{C_7 + 1}{C_7} \frac{n}{d + \tau},
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds, and in the last inequality, we apply Assumption 1(C), i.e. $d \geq CR_+ n$. Similarly,

$$\begin{aligned}
s_{i,i}^{(k+1)} &\geq \frac{s_{i,i}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)}, s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) s_{i,i}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{i,k+1}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{i,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| + \left| h_{k+1,i}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right) + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 - 1}{C_2} \frac{n}{d + \tau} - \frac{C_3 + 1}{C_3} \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2}{(d + \tau)^2} - \frac{C_4 + 1}{C_4} \frac{C_5 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n^2}{(d + \tau)^2} - \frac{C_6 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^3}{(d + \tau)^3} \right) \\
&\geq \frac{C_7 - 1}{C_7} \frac{n}{d + \tau}.
\end{aligned}$$

• $s_{1,2}^{(k+1)}$: We have

$$\begin{aligned}
s_{1,2}^{(k+1)} &= \mathbf{v}_1^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_2 \\
&= \mathbf{v}_1^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_2 - \mathbf{v}_1^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_2 \\
&= s_{1,2}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{1,k+1}^{(k)}, h_{k+1,1}^{(k)}, s_{1,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{2,k+1}^{(k)} \\ s_{2,k+1}^{(k)} \\ h_{k+1,2}^{(k)} \end{bmatrix} \\
&= \frac{s_{1,2}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(s_{1,k+1}^{(k)}, h_{k+1,1}^{(k)}, s_{2,k+1}^{(k)}, h_{k+1,2}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
s_{1,2}^{(k+1)} &\leq \frac{\left| s_{1,2}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(s_{1,k+1}^{(k)}, h_{k+1,1}^{(k)}, s_{2,k+1}^{(k)}, h_{k+1,2}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| s_{1,2}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{1,k+1}^{(k)} \right| \left| s_{2,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,1}^{(k)} \right| \left| h_{k+1,2}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{1,k+1}^{(k)} \right| \left| h_{k+1,2}^{(k)} \right| + \left| h_{k+1,1}^{(k)} \right| \left| s_{2,k+1}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \left(\frac{C_3 + 1}{C_3} \frac{n_+}{d + \tau} - \frac{C_3 - 1}{C_3} \frac{n_-}{d + \tau} \right) \right. \\
&\quad \left. + \frac{C_4 + 1}{C_4} \left(\frac{C_5 + 1}{C_5} \frac{n_+}{d + \tau} - \frac{C_5 - 1}{C_5} \frac{n_-}{d + \tau} \right) \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} + \frac{C_6 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^3}{(d + \tau)^3} \right. \\
&\quad \left. + \frac{C_7 + 1}{C_7} \left(\left(\frac{C_8 + 1}{C_8} \frac{n_+}{d + \tau} - \frac{C_8 - 1}{C_8} \frac{n_-}{d + \tau} \right) \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n}{d + \tau} + \frac{C_{10} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n^2}{(d + \tau)^2} \right) \right) \\
&\leq \frac{C_{11} + 1}{C_{11}} \frac{n_+}{d + \tau} - \frac{C_{11} - 1}{C_{11}} \frac{n_-}{d + \tau},
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds, and in the last inequality, we apply Assumption 1(C), i.e. $d \geq CR_+n$. Similarly,

$$\begin{aligned}
s_{1,2}^{(k+1)} &\geq \frac{s_{1,2}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(s_{1,k+1}^{(k)}, h_{k+1,1}^{(k)}, s_{2,k+1}^{(k)}, h_{k+1,2}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) s_{1,2}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{1,k+1}^{(k)} \right| \left| s_{2,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,1}^{(k)} \right| \left| h_{k+1,2}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{1,k+1}^{(k)} \right| \left| h_{k+1,2}^{(k)} \right| + \left| h_{k+1,1}^{(k)} \right| \left| s_{2,k+1}^{(k)} \right| \right) \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 - 1}{C_2} \left(\frac{C_3 - 1}{C_3} \frac{n_+}{d + \tau} - \frac{C_3 + 1}{C_3} \frac{n_-}{d + \tau} \right) \right. \\
&\quad - \frac{C_4 + 1}{C_4} \left(\frac{C_5 + 1}{C_5} \frac{n_+}{d + \tau} - \frac{C_5 - 1}{C_5} \frac{n_-}{d + \tau} \right) \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} - \frac{C_6 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^3}{(d + \tau)^3} \\
&\quad \left. - \frac{C_7 + 1}{C_7} \left(\left(\frac{C_8 + 1}{C_8} \frac{n_+}{d + \tau} - \frac{C_8 - 1}{C_8} \frac{n_-}{d + \tau} \right) \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n}{d + \tau} + \frac{C_{10} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n^2}{(d + \tau)^2} \right) \right) \\
&\geq \frac{C_{11} - 1}{C_{11}} \frac{n_+}{d + \tau} - \frac{C_{11} + 1}{C_{11}} \frac{n_-}{d + \tau}.
\end{aligned}$$

• $t_{i,i}^{(k+1)}$ for $i = 1, 2$: We have

$$\begin{aligned}
t_{i,i}^{(k+1)} &= \mathbf{d}_i^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{d}_i \\
&= \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{d}_i - \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{d}_i \\
&= t_{i,i}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, h_{i,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 h_{i,k+1}^{(k)} \\ h_{i,k+1}^{(k)} \\ t_{i,k+1}^{(k)} \end{bmatrix} \\
&= \frac{t_{i,i}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
t_{i,i}^{(k+1)} &\leq \frac{\left| t_{i,i}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| t_{i,i}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| h_{i,k+1}^{(k)} \right| \left| h_{i,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| t_{i,k+1}^{(k)} \right| \left| t_{i,k+1}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| h_{i,k+1}^{(k)} \right| \left| t_{i,k+1}^{(k)} \right| + \left| t_{i,k+1}^{(k)} \right| \left| h_{i,k+1}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{C_3 n \|\bar{\boldsymbol{\mu}}_i\|_2^2}{d + \tau} + \frac{C_4 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_i\|_2^2 n^2}{(d + \tau)^2} \right. \\
&\quad \left. + \frac{C_5 \|\bar{\boldsymbol{\mu}}_i\|_2^2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^3}{(d + \tau)^3} + \frac{C_6 \|\bar{\boldsymbol{\mu}}_i\|_2^2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2}{(d + \tau)^2} \right) \\
&\leq \frac{C_7 n \|\bar{\boldsymbol{\mu}}_i\|_2^2}{d + \tau},
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds, and in the last inequality, we apply Assumption 1(C), i.e. $d \geq CR_+ n$.

• $t_{1,2}^{(k+1)}$: We have

$$\begin{aligned}
t_{1,2}^{(k+1)} &= \mathbf{d}_1^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{d}_2 \\
&= \mathbf{d}_1^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{d}_2 - \mathbf{d}_1^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{d}_2 \\
&= t_{1,2}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 h_{1,k+1}^{(k)}, t_{1,k+1}^{(k)}, h_{1,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 h_{2,k+1}^{(k)} \\ h_{2,k+1}^{(k)} \\ t_{2,k+1}^{(k)} \end{bmatrix} \\
&= \frac{t_{1,2}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(h_{1,k+1}^{(k)}, t_{1,k+1}^{(k)}, h_{2,k+1}^{(k)}, t_{2,k+1}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
t_{1,2}^{(k+1)} &\leq \frac{\left| t_{1,2}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(h_{1,k+1}^{(k)}, t_{1,k+1}^{(k)}, h_{2,k+1}^{(k)}, t_{2,k+1}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| t_{1,2}^{(k)} \right| + \left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2^2 \left| h_{1,k+1}^{(k)} \right| \left| h_{2,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| t_{1,k+1}^{(k)} \right| \left| t_{2,k+1}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| h_{1,k+1}^{(k)} \right| \left| t_{2,k+1}^{(k)} \right| + \left| t_{1,k+1}^{(k)} \right| \left| h_{2,k+1}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{C_3 n \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2}{d + \tau} + \frac{C_4 \left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2^2 \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2 n^2}{(d + \tau)^2} \right. \\
&\quad \left. + \frac{C_5 \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2^2 n^3}{(d + \tau)^3} + \frac{C_6 \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2 n^2}{(d + \tau)^2} \right) \\
&\leq \frac{C_7 n \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2}{d + \tau},
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds, and in the last inequality, we apply Assumption 1(C), i.e. $d \geq CR_+ n$. Similarly,

$$\begin{aligned}
t_{1,2}^{(k+1)} &\geq \frac{t_{1,2}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(h_{1,k+1}^{(k)}, t_{1,k+1}^{(k)}, h_{2,k+1}^{(k)}, t_{2,k+1}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) t_{1,2}^{(k)} - \left(\left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2^2 \left| h_{1,k+1}^{(k)} \right| \left| h_{2,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| t_{1,k+1}^{(k)} \right| \left| t_{2,k+1}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| h_{1,k+1}^{(k)} \right| \left| t_{2,k+1}^{(k)} \right| + \left| t_{1,k+1}^{(k)} \right| \left| h_{2,k+1}^{(k)} \right| \right) \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(- \frac{C_2 + 1}{C_2} \frac{C_3 n \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2}{d + \tau} - \frac{C_4 \left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2^2 \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2 n^2}{(d + \tau)^2} \right. \\
&\quad \left. - \frac{C_5 \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2^2 n^3}{(d + \tau)^3} - \frac{C_6 \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2 n^2}{(d + \tau)^2} \right) \\
&\geq - \frac{C_7 n \left\| \bar{\boldsymbol{\mu}}_1 \right\|_2 \left\| \bar{\boldsymbol{\mu}}_2 \right\|_2}{d + \tau}.
\end{aligned}$$

• $h_{i,j}^{(k+1)} = \mathbf{d}_i^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_j$, for $i, j = 1, 2$: We have

$$\begin{aligned}
h_{i,j}^{(k+1)} &= \mathbf{d}_i^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_j \\
&= \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_j - \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_j \\
&= h_{i,j}^{(k)} - \left[\left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2 h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, h_{i,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \left\| \bar{\boldsymbol{\mu}}_{k+1} \right\|_2 s_{j,k+1}^{(k)} \\ s_{j,k+1}^{(k)} \\ h_{k+1,j}^{(k)} \end{bmatrix} \\
&= \frac{h_{i,j}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, s_{j,k+1}^{(k)}, h_{k+1,j}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
h_{i,j}^{(k+1)} &\leq \frac{\left| h_{i,j}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, s_{j,k+1}^{(k)}, h_{k+1,j}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| h_{i,j}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| h_{i,k+1}^{(k)} \right| \left| s_{j,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| t_{i,k+1}^{(k)} \right| \left| h_{k+1,j}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| h_{i,k+1}^{(k)} \right| \left| h_{k+1,j}^{(k)} \right| + \left| t_{i,k+1}^{(k)} \right| \left| s_{j,k+1}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{C_3 n \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} + \frac{C_4 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_i\|_2 n^2}{(d + \tau)^2} \right. \\
&\quad \left. + \frac{C_5 \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^3}{(d + \tau)^3} + \frac{C_6 \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n^2}{(d + \tau)^2} \right) \\
&\leq \frac{C_7 n \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau},
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds, and in the last inequality, we apply Assumption 1(C), i.e. $d \geq CR_+ n$. Similarly,

$$\begin{aligned}
h_{i,j}^{(k+1)} &\geq \frac{h_{i,j}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, s_{j,k+1}^{(k)}, h_{k+1,j}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) h_{i,j}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| h_{i,k+1}^{(k)} \right| \left| s_{j,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| t_{i,k+1}^{(k)} \right| \left| h_{k+1,j}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| h_{i,k+1}^{(k)} \right| \left| h_{k+1,j}^{(k)} \right| + \left| t_{i,k+1}^{(k)} \right| \left| s_{j,k+1}^{(k)} \right| \right) \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(- \frac{C_2 + 1}{C_2} \frac{C_3 n \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} - \frac{C_4 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_i\|_2 n^2}{(d + \tau)^2} \right. \\
&\quad \left. - \frac{C_5 \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^3}{(d + \tau)^3} - \frac{C_6 \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n^2}{(d + \tau)^2} \right) \\
&\geq - \frac{C_7 n \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau}.
\end{aligned}$$

• $s_{u,u}^{(k+1)}$: We have

$$\begin{aligned}
s_{u,u}^{(k+1)} &= \mathbf{u}^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{u} \\
&= \mathbf{u}^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{u} - \mathbf{u}^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{u} \\
&= s_{u,u}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)}, s_{u,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{u,k+1}^{(k)} \\ s_{u,k+1}^{(k)} \\ h_{k+1,u}^{(k)} \end{bmatrix} \\
&= \frac{s_{u,u}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)}, s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
s_{u,u}^{(k+1)} &\leq \frac{\left| s_{u,u}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)}, s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| s_{u,u}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{u,k+1}^{(k)} \right| \left| s_{u,k+1}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{u,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| + \left| h_{k+1,u}^{(k)} \right| \left| s_{u,k+1}^{(k)} \right| \right) + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2+1}{C_2} \frac{1}{d+\tau} + \frac{C_3 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{(d+\tau)^2} + \frac{C_4 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n}{(d+\tau)^2} + \frac{C_5 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2}{(d+\tau)^3} \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_6+1}{C_6} \frac{1}{d+\tau} \right) \leq \frac{C_7+1}{C_7} \frac{1}{d+\tau},
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds, and in the last inequality, we apply Assumption 1(C) such that $d \geq CR_+n$. Similarly,

$$\begin{aligned}
s_{u,u}^{(k+1)} &\geq \frac{s_{u,u}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)}, s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) s_{u,u}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{u,k+1}^{(k)} \right| \left| s_{u,k+1}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{u,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| + \left| h_{k+1,u}^{(k)} \right| \left| s_{u,k+1}^{(k)} \right| \right) + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2-1}{C_2} \frac{1}{d+\tau} - \frac{C_3 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{(d+\tau)^2} - \frac{C_4 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n}{(d+\tau)^2} - \frac{C_5 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2}{(d+\tau)^3} \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_6-1}{C_6} \frac{1}{d+\tau} \right) = \frac{C_7-1}{C_7} \frac{1}{d+\tau}.
\end{aligned}$$

• $s_{u,i}^{(k+1)}$ for $i = 1, 2$: We have

$$\begin{aligned}
s_{u,i}^{(k+1)} &= \mathbf{u}^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i \\
&= \mathbf{u}^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i - \mathbf{u}^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i \\
&= s_{u,i}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)}, s_{u,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{i,k+1}^{(k)} \\ s_{i,k+1}^{(k)} \\ h_{k+1,i}^{(k)} \end{bmatrix} \\
&= \frac{s_{u,i}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)}, s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
s_{u,i}^{(k+1)} &\leq \frac{\left| s_{u,i}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)}, s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| s_{u,i}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{u,k+1}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{u,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| + \left| h_{k+1,u}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{C_3 \sqrt{n}}{d + \tau} + \frac{C_4 + 1}{C_4} \frac{C_5 \sqrt{n}}{d + \tau} \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} + \frac{C_6 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2 \sqrt{n}}{(d + \tau)^3} \right. \\
&\quad \left. + \frac{C_7 + 1}{C_7} \left(\frac{C_8 \sqrt{n}}{d + \tau} \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n}{d + \tau} + \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n \sqrt{n}}{(d + \tau)^2} \right) \right) \\
&\leq \frac{C_{10} \sqrt{n}}{d + \tau},
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds, and in the last inequality, we apply Assumption 1(C) such that $d \geq CR_+ n$. Similarly,

$$\begin{aligned}
s_{u,i}^{(k+1)} &\geq \frac{s_{u,i}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)}, s_{i,k+1}^{(k)}, h_{k+1,i}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) s_{u,i}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{u,k+1}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{u,k+1}^{(k)} \right| \left| h_{k+1,i}^{(k)} \right| + \left| h_{k+1,u}^{(k)} \right| \left| s_{i,k+1}^{(k)} \right| \right) \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(- \frac{C_2 - 1}{C_2} \frac{C_3 \sqrt{n}}{d + \tau} - \frac{C_4 + 1}{C_4} \frac{C_5 \sqrt{n}}{d + \tau} \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} - \frac{C_6 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2 \sqrt{n}}{(d + \tau)^3} \right. \\
&\quad \left. - \frac{C_7 + 1}{C_7} \left(\frac{C_8 \sqrt{n}}{d + \tau} \frac{\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n}{d + \tau} + \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n \sqrt{n}}{(d + \tau)^2} \right) \right) \\
&\geq - \frac{C_{10} \sqrt{n}}{d + \tau}.
\end{aligned}$$

• $h_{i,u}^{(k+1)} = \mathbf{d}_i^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{u}$, for $i = 1, 2$: We have

$$\begin{aligned}
h_{i,u}^{(k+1)} &= \mathbf{d}_i^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{u} \\
&= \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{u} - \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{u} \\
&= h_{i,u}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, h_{i,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{u,k+1}^{(k)} \\ s_{u,k+1}^{(k)} \\ h_{k+1,u}^{(k)} \end{bmatrix} \\
&= \frac{h_{i,u}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
h_{i,u}^{(k+1)} &\leq \frac{\left| h_{i,u}^{(k)} \det(\mathbf{A}_{k+1}) + f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| h_{i,u}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| h_{i,k+1}^{(k)} \right| \left| s_{u,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| t_{i,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| h_{i,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| + \left| t_{i,k+1}^{(k)} \right| \left| s_{u,k+1}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{C_3 \sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} + \frac{C_4 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_i\|_2 n \sqrt{n}}{(d + \tau)^2} \right. \\
&\quad \left. + \frac{C_5 \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2 \sqrt{n}}{(d + \tau)^3} + \frac{C_6 \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n \sqrt{n}}{(d + \tau)^2} \right) \\
&\leq \frac{C_7 \sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau},
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds, and in the last inequality, we apply Assumption 1(C) such that $d \geq CR_+ n$. Similarly,

$$\begin{aligned}
h_{i,u}^{(k+1)} &\geq \frac{h_{i,u}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, s_{u,k+1}^{(k)}, h_{k+1,u}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) h_{i,u}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| h_{i,k+1}^{(k)} \right| \left| s_{u,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| t_{i,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| h_{i,k+1}^{(k)} \right| \left| h_{k+1,u}^{(k)} \right| + \left| t_{i,k+1}^{(k)} \right| \left| s_{u,k+1}^{(k)} \right| \right) \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(- \frac{C_2 + 1}{C_2} \frac{C_3 \sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} - \frac{C_4 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_i\|_2 n \sqrt{n}}{(d + \tau)^2} \right. \\
&\quad \left. - \frac{C_5 \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2 \sqrt{n}}{(d + \tau)^3} - \frac{C_6 \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 n \sqrt{n}}{(d + \tau)^2} \right) \\
&\geq - \frac{C_7 \sqrt{n} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau}.
\end{aligned}$$

C.2 Inductive step for Lemma 3 (with adjustment weight)

- $s_{i_\Delta, i}^{(k+1)}$ for $i = 1, 2$: We have

$$\begin{aligned}
s_{i_\Delta, i}^{(k+1)} &= \mathbf{v}_i^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i \\
&= \mathbf{v}_i^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i - \mathbf{v}_i^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_i \\
&= s_{i_\Delta, i}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{i_\Delta, k+1}^{(k)}, h_{k+1, i_\Delta}^{(k)}, s_{i_\Delta, k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{i, k+1}^{(k)} \\ s_{i, k+1}^{(k)} \\ h_{k+1, i}^{(k)} \end{bmatrix} \\
&= \frac{s_{i_\Delta, i}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(s_{i_\Delta, k+1}^{(k)}, h_{k+1, i_\Delta}^{(k)}, s_{i, k+1}^{(k)}, h_{k+1, i}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
s_{i_\Delta, i}^{(k+1)} &\leq \frac{\left| s_{i_\Delta, i}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(s_{i_\Delta, k+1}^{(k)}, h_{k+1, i_\Delta}^{(k)}, s_{i, k+1}^{(k)}, h_{k+1, i}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| s_{i_\Delta, i}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{i_\Delta, k+1}^{(k)} \right| \left| s_{i, k+1}^{(k)} \right| + \left| s_{k+1, k+1}^{(k)} \right| \left| h_{k+1, i_\Delta}^{(k)} \right| \left| h_{k+1, i}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{i_\Delta, k+1}^{(k)} \right| \left| h_{k+1, i}^{(k)} \right| + \left| h_{k+1, i_\Delta}^{(k)} \right| \left| s_{i, k+1}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{\left(\hat{C}_3 + \frac{1}{C_4} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} \right. \\
&\quad + \frac{C_5 + 1}{C_5} \frac{\left(\hat{C}_6 + \frac{1}{C_7} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} + \frac{C_8 + 1}{C_8} \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2 \sqrt{nn\Delta}}{(d + \tau)^3} \\
&\quad \left. + \frac{C_{10} + 1}{C_{10}} \left(\frac{\left(\hat{C}_{11} + \frac{1}{C_{12}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) C_{13} n \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{d + \tau} + \frac{C_{14} n \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} \right) \right),
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds. Next, Lemma 4 tells us that $\sqrt{nn\Delta} \leq \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_2\|_2$. We then have

$$\begin{aligned}
s_{i_\Delta, i}^{(k+1)} &\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{\left(\hat{C}_3 + \frac{1}{C_4} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} \right. \\
&\quad + \frac{C_5 + 1}{C_5} \frac{\left(\hat{C}_6 + \frac{1}{C_7} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} + \frac{C_8 + 1}{C_8} \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_2\|_2 n^2 \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{(d + \tau)^3} \\
&\quad \left. + \frac{C_{10} + 1}{C_{10}} \left(\frac{\left(\hat{C}_{11} + \frac{1}{C_{12}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) C_{13} n \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{d + \tau} + \frac{C_{14} n \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_2\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} \right) \right) \\
&\leq \frac{\left(\hat{C}_{15} + \frac{1}{C_{16}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau},
\end{aligned}$$

where in the last inequality, we apply Assumption 1(C), i.e. $d \geq CR_+n$. Similarly,

$$\begin{aligned}
s_{i_\Delta, i}^{(k+1)} &\geq \frac{s_{i_\Delta, i}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(s_{i_\Delta, k+1}^{(k)}, h_{k+1, i_\Delta}^{(k)}, s_{i, k+1}^{(k)}, h_{k+1, i}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) s_{i_\Delta, i}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{i_\Delta, k+1}^{(k)} \right| \left| s_{i, k+1}^{(k)} \right| + \left| s_{k+1, k+1}^{(k)} \right| \left| h_{k+1, i_\Delta}^{(k)} \right| \left| h_{k+1, i}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{i_\Delta, k+1}^{(k)} \right| \left| h_{k+1, i}^{(k)} \right| + \left| h_{k+1, i_\Delta}^{(k)} \right| \left| s_{i, k+1}^{(k)} \right| \right) \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 - 1}{C_2} \frac{\left(\hat{C}_3 - \frac{1}{C_4} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} \right. \\
&\quad - \frac{C_5 + 1}{C_5} \frac{\left(\hat{C}_6 + \frac{1}{C_7} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} - \frac{C_8 + 1}{C_8} \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_2\|_2 n^2 \frac{1}{C}}{(d + \tau)^3} \\
&\quad \left. - \frac{C_{10} + 1}{C_{10}} \left(\frac{\left(\hat{C}_{11} + \frac{1}{C_{12}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) C_{13} n \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{d + \tau} - \frac{C_{14} n \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_2\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} \right) \right) \\
&\geq \frac{\left(\hat{C}_{15} - \frac{1}{C_{16}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau}.
\end{aligned}$$

• $s_{2\Delta, 1}^{(k+1)}$: We have

$$\begin{aligned}
s_{2\Delta, 1}^{(k+1)} &= \mathbf{v}_2^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_1 \\
&= \mathbf{v}_2^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_1 - \mathbf{v}_2^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{v}_1 \\
&= s_{2\Delta, 1}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{2\Delta, k+1}^{(k)}, h_{k+1, 2\Delta}^{(k)}, s_{2\Delta, k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{1, k+1}^{(k)} \\ s_{1, k+1}^{(k)} \\ h_{k+1, 1}^{(k)} \end{bmatrix} \\
&= \frac{s_{2\Delta, 1}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(s_{2\Delta, k+1}^{(k)}, h_{k+1, 2\Delta}^{(k)}, s_{1, k+1}^{(k)}, h_{k+1, 1}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
s_{2\Delta,1}^{(k+1)} &\leq \frac{\left| s_{2\Delta,1}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(s_{2\Delta,k+1}^{(k)}, h_{k+1,2\Delta}^{(k)}, s_{1,k+1}^{(k)}, h_{k+1,1}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) s_{2\Delta,1}^{(k)} + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{2\Delta,k+1}^{(k)} \right| \left| s_{1,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,2\Delta}^{(k)} \right| \left| h_{k+1,1}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{2\Delta,k+1}^{(k)} \right| \left| h_{k+1,1}^{(k)} \right| + \left| h_{k+1,2\Delta}^{(k)} \right| \left| s_{1,k+1}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{\hat{C}_3 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) + \frac{1}{C_4} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} \right. \\
&\quad + \frac{C_5 + 1}{C_5} \frac{\left(\hat{C}_6 + \frac{1}{C_7} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} + \frac{C_8 + 1}{C_8} \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2 \sqrt{nn\Delta}}{(d + \tau)^3} \\
&\quad \left. + \frac{C_{10} + 1}{C_{10}} \left(\frac{\left(\hat{C}_{11} + \frac{1}{C_{12}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) C_{13} n \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{d + \tau} + \frac{C_{14} n \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} \right) \right)
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds. Next, Lemma 4 tells us that $\sqrt{nn\Delta} \leq \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_2\|_2$. We then have

$$\begin{aligned}
s_{2\Delta,1}^{(k+1)} &\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{\hat{C}_3 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) + \frac{1}{C_4} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} \right. \\
&\quad + \frac{C_5 + 1}{C_5} \frac{\left(\hat{C}_6 + \frac{1}{C_7} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} + \frac{C_8 + 1}{C_8} \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_2\|_2 n^2 \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{(d + \tau)^3} \\
&\quad + \frac{C_{10} + 1}{C_{10}} \left(\frac{\left(\hat{C}_{11} + \frac{1}{C_{12}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) C_{13} n \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{d + \tau} + \frac{C_{14} n \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_2\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} \right) \Bigg) \\
&\leq \frac{\hat{C}_{15} \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) + \frac{1}{C_{16}} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau},
\end{aligned}$$

where in the last inequality, we apply Assumption 1(C), i.e. $d \geq CR_+n$. Similarly,

$$\begin{aligned}
s_{2\Delta,1}^{(k+1)} &\geq \frac{s_{2\Delta,1}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(s_{2\Delta,k+1}^{(k)}, h_{k+1,2\Delta}^{(k)}, s_{1,k+1}^{(k)}, h_{k+1,1}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) s_{2\Delta,1}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{2\Delta,k+1}^{(k)} \right| \left| s_{1,k+1}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| h_{k+1,2\Delta}^{(k)} \right| \left| h_{k+1,1}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{2\Delta,k+1}^{(k)} \right| \left| h_{k+1,1}^{(k)} \right| + \left| h_{k+1,2\Delta}^{(k)} \right| \left| s_{1,k+1}^{(k)} \right| \right) \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 - 1}{C_2} \frac{\hat{C}_3 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) - \frac{1}{C_4} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau} \right. \\
&\quad - \frac{C_5 + 1}{C_5} \frac{\left(\hat{C}_6 + \frac{1}{C_7} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n}{d + \tau} - \frac{C_8 + 1}{C_8} \frac{C_9 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_2\|_2 n^2 \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{(d + \tau)^3} \\
&\quad \left. - \frac{C_{10} + 1}{C_{10}} \left(\frac{\left(\hat{C}_{11} + \frac{1}{C_{12}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) C_{13} n \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{d + \tau} - \frac{C_{14} n \frac{1}{C} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_2\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} \right) \right) \\
&\geq \frac{\hat{C}_{15} \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) - \frac{1}{C_{16}} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d + \tau}.
\end{aligned}$$

• $s_{i\Delta,i\Delta}^{(k+1)}$ for $i = 1, 2$: We have

$$\begin{aligned}
s_{i\Delta,i\Delta}^{(k+1)} &= \mathbf{v}_i^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \boldsymbol{\Delta}^{-1} \mathbf{v}_i \\
&= \mathbf{v}_i^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \boldsymbol{\Delta}^{-1} \mathbf{v}_i \\
&\quad - \mathbf{v}_i^\top \boldsymbol{\Delta}^{-1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \boldsymbol{\Delta}^{-1} \mathbf{v}_i \\
&= s_{i\Delta,i\Delta}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{i\Delta,k+1}^{(k)}, h_{k+1,i\Delta}^{(k)}, s_{i\Delta,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{i\Delta,k+1}^{(k)} \\ s_{i\Delta,k+1}^{(k)} \\ h_{k+1,i\Delta}^{(k)} \end{bmatrix} \\
&= \frac{s_{i\Delta,i\Delta}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(s_{i\Delta,k+1}^{(k)}, h_{k+1,i\Delta}^{(k)}, s_{i\Delta,k+1}^{(k)}, h_{k+1,i\Delta}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
s_{i_\Delta, i_\Delta}^{(k+1)} &\leq \frac{\left| s_{i_\Delta, i_\Delta}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(s_{i_\Delta, k+1}^{(k)}, h_{k+1, i_\Delta}^{(k)}, s_{i_\Delta, k+1}^{(k)}, h_{k+1, i_\Delta}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| s_{i_\Delta, i_\Delta}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{i_\Delta, k+1}^{(k)} \right| \left| s_{i_\Delta, k+1}^{(k)} \right| + \left| s_{k+1, k+1}^{(k)} \right| \left| h_{k+1, i_\Delta}^{(k)} \right| \left| h_{k+1, i_\Delta}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{i_\Delta, k+1}^{(k)} \right| \left| h_{k+1, i_\Delta}^{(k)} \right| + \left| h_{k+1, i_\Delta}^{(k)} \right| \left| s_{i_\Delta, k+1}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{\hat{C}_3 n_\Delta}{d + \tau} + \frac{\left(\hat{C}_4 + \frac{1}{C_5} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)^2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2}{(d + \tau)^2} \right. \\
&\quad \left. + \frac{C_6 + 1}{C_6} \frac{C_7 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2 n_\Delta}{(d + \tau)^3} + \frac{C_8 + 1}{C_8} \frac{2 \left(\hat{C}_9 + \frac{1}{C_{10}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) C_{11} \sqrt{nn_\Delta} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{d + \tau} \right)
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds. Next, by the Cauchy-Schwarz inequality, we have $\left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)^2 \leq 2 \left(\frac{n_+^2}{\Delta_+^2} + \frac{n_-^2}{\Delta_-^2} \right) \leq 2nn_\Delta$. Also, Lemma 4 tells us that $\sqrt{nn_\Delta} \leq \frac{1}{\bar{C}} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_2\|_2$. We then have

$$\begin{aligned}
s_{i_\Delta, i_\Delta}^{(k+1)} &\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{\hat{C}_3 n_\Delta}{d + \tau} + \frac{\left(\hat{C}_4 + \frac{1}{C_5} \right) 2nn_\Delta \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2}{(d + \tau)^2} + \frac{C_6 + 1}{C_6} \frac{C_7 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2 n_\Delta}{(d + \tau)^3} \right. \\
&\quad \left. + \frac{C_8 + 1}{C_8} \frac{2 \left(\hat{C}_9 + \frac{1}{C_{10}} \right) \frac{2}{\bar{C}} nn_\Delta \|\bar{\boldsymbol{\mu}}_2\|_2 C_{11} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{d + \tau} \right) \\
&\leq \frac{\hat{C}_{12} n_\Delta}{d + \tau},
\end{aligned}$$

where in the last inequality, we apply Assumption 1(C), i.e. $d \geq CR_+ n$. Similarly,

$$\begin{aligned}
s_{i_\Delta, i_\Delta}^{(k+1)} &\geq \frac{s_{i_\Delta, i_\Delta}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(s_{i_\Delta, k+1}^{(k)}, h_{k+1, i_\Delta}^{(k)}, s_{i_\Delta, k+1}^{(k)}, h_{k+1, i_\Delta}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) s_{i_\Delta, i_\Delta}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| s_{i_\Delta, k+1}^{(k)} \right| \left| s_{i_\Delta, k+1}^{(k)} \right| + \left| s_{k+1, k+1}^{(k)} \right| \left| h_{k+1, i_\Delta}^{(k)} \right| \left| h_{k+1, i_\Delta}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| s_{i_\Delta, k+1}^{(k)} \right| \left| h_{k+1, i_\Delta}^{(k)} \right| + \left| h_{k+1, i_\Delta}^{(k)} \right| \left| s_{i_\Delta, k+1}^{(k)} \right| \right) \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{\check{C}_3 n_\Delta}{d + \tau} - \frac{\left(\hat{C}_4 + \frac{1}{C_5} \right) 2nn_\Delta \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2}{(d + \tau)^2} \right. \\
&\quad \left. - \frac{C_6 + 1}{C_6} \frac{C_7 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 n^2 n_\Delta}{(d + \tau)^3} - \frac{C_8 + 1}{C_8} \frac{2 \left(\hat{C}_9 + \frac{1}{C_{10}} \right) \frac{2}{\bar{C}} nn_\Delta \|\bar{\boldsymbol{\mu}}_2\|_2 C_{11} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{d + \tau} \right) \\
&\geq \frac{\check{C}_{12} n_\Delta}{d + \tau}.
\end{aligned}$$

• $h_{i,j\Delta}^{(k+1)}$ for $i, j = 1, 2$: We have

$$\begin{aligned}
h_{i,j\Delta}^{(k+1)} &= \mathbf{d}_i^\top \left(\mathbf{X}_{k+1} \mathbf{X}_{k+1}^\top + \tau \mathbf{I} \right)^{-1} \boldsymbol{\Delta}^{-1} \mathbf{v}_j \\
&= \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \boldsymbol{\Delta}^{-1} \mathbf{v}_j - \mathbf{d}_i^\top \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \mathbf{L}_{k+1} \mathbf{A}_{k+1}^{-1} \mathbf{R}_{k+1} \left(\mathbf{X}_k \mathbf{X}_k^\top + \tau \mathbf{I} \right)^{-1} \boldsymbol{\Delta}^{-1} \mathbf{v}_j \\
&= h_{i,j\Delta}^{(k)} - \left[\|\bar{\boldsymbol{\mu}}_{k+1}\|_2 h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, h_{i,k+1}^{(k)} \right] \times \frac{\text{adj}(\mathbf{A}_{k+1})}{\det(\mathbf{A}_{k+1})} \begin{bmatrix} \|\bar{\boldsymbol{\mu}}_{k+1}\|_2 s_{k+1,j\Delta}^{(k)} \\ s_{k+1,j\Delta}^{(k)} \\ h_{k+1,j\Delta}^{(k)} \end{bmatrix} \\
&= \frac{h_{i,j\Delta}^{(k)} \det(\mathbf{A}_{k+1}) - f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, s_{k+1,j\Delta}^{(k)}, h_{k+1,j\Delta}^{(k)} \right)}{\det(\mathbf{A}_{k+1})}.
\end{aligned}$$

Thus, we can apply Lemma 6 to get

$$\begin{aligned}
h_{i,j\Delta}^{(k+1)} &\leq \frac{\left| h_{i,j\Delta}^{(k)} \det(\mathbf{A}_{k+1}) \right| + \left| f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, s_{k+1,j\Delta}^{(k)}, h_{k+1,j\Delta}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) \left| h_{i,j\Delta}^{(k)} \right| + \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| h_{i,k+1}^{(k)} \right| \left| s_{k+1,j\Delta}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| t_{i,k+1}^{(k)} \right| \left| h_{k+1,j\Delta}^{(k)} \right| \right. \\
&\quad \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| h_{i,k+1}^{(k)} \right| \left| h_{k+1,j\Delta}^{(k)} \right| + \left| t_{i,k+1}^{(k)} \right| \left| s_{k+1,j\Delta}^{(k)} \right| \right) \right) \\
&\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{C_3 \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} \right. \\
&\quad + \frac{\left(\hat{C}_4 + \frac{1}{C_5} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 C_6 n \|\bar{\boldsymbol{\mu}}_i\|_2}{(d + \tau)^2} + \frac{C_7 + 1}{C_7} \frac{C_8 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_i\|_2 n^2 \sqrt{nn\Delta}}{(d + \tau)^3} \\
&\quad \left. + \frac{C_9 + 1}{C_9} \left(\frac{C_{10} n \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} + \frac{C_{11} \left(\hat{C}_{12} + \frac{1}{C_{13}} \right) n \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} \right) \right),
\end{aligned}$$

where in the third inequality we apply the lower bound on $\det(\mathbf{A}_{k+1})$ (Lemma 7) and the k -th order primitive bounds. Next, by the Cauchy-Schwarz inequality, we have $\left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) \leq \sqrt{2 \left(\frac{n_+^2}{\Delta_+^2} + \frac{n_-^2}{\Delta_-^2} \right)} \leq \sqrt{2nn\Delta}$. We then have

$$\begin{aligned}
h_{i,j\Delta}^{(k+1)} &\leq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\frac{C_2 + 1}{C_2} \frac{C_3 \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} \right. \\
&\quad + \frac{\left(\hat{C}_4 + \frac{1}{C_5} \right) C_6 n \sqrt{2nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2}{(d + \tau)^2} + \frac{C_7 + 1}{C_7} \frac{C_8 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_i\|_2 n^2 \sqrt{nn\Delta}}{(d + \tau)^3} \\
&\quad \left. + \frac{C_9 + 1}{C_9} \left(\frac{C_{10} n \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} + \frac{C_{11} \left(\hat{C}_{12} + \frac{1}{C_{13}} \right) n \sqrt{2nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} \right) \right) \\
&\leq \frac{C_{12} \sqrt{nn\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau},
\end{aligned}$$

where in the last inequality, we apply Assumption 1(C), i.e. $d \geq CR_+n$. Similarly,

$$\begin{aligned}
h_{i,j_\Delta}^{(k+1)} &\geq \frac{h_{i,j_\Delta}^{(k)} \det(\mathbf{A}_{k+1}) - \left| f_{\mathbf{A}_{k+1}} \left(h_{i,k+1}^{(k)}, t_{i,k+1}^{(k)}, s_{k+1,j_\Delta}^{(k)}, h_{k+1,j_\Delta}^{(k)} \right) \right|}{\det(\mathbf{A}_{k+1})} \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(\det(\mathbf{A}_{k+1}) h_{i,j_\Delta}^{(k)} - \left(\|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \left| h_{i,k+1}^{(k)} \right| \left| s_{k+1,j_\Delta}^{(k)} \right| + \left| s_{k+1,k+1}^{(k)} \right| \left| t_{i,k+1}^{(k)} \right| \left| h_{k+1,j_\Delta}^{(k)} \right| \right. \right. \\
&\quad \left. \left. + \left(1 + \frac{1}{C_1} \right) \left(\left| h_{i,k+1}^{(k)} \right| \left| h_{k+1,j_\Delta}^{(k)} \right| + \left| t_{i,k+1}^{(k)} \right| \left| s_{k+1,j_\Delta}^{(k)} \right| \right) \right) \right) \\
&\geq \frac{1}{\det(\mathbf{A}_{k+1})} \left(-\frac{C_2 - 1}{C_2} \frac{C_3 \sqrt{nn_\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau} \right. \\
&\quad - \frac{\left(\hat{C}_4 + \frac{1}{C_5} \right) C_6 n \sqrt{2nn_\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2}{(d + \tau)^2} - \frac{C_7 + 1}{C_7} \frac{C_8 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2^2 \|\bar{\boldsymbol{\mu}}_i\|_2 n^2 \sqrt{nn_\Delta}}{(d + \tau)^3} \\
&\quad \left. - \frac{C_9 + 1}{C_9} \left(\frac{C_{10} n \sqrt{nn_\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} - \frac{C_{11} \left(\hat{C}_{12} + \frac{1}{C_{13}} \right) n \sqrt{2nn_\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2 \|\bar{\boldsymbol{\mu}}_{k+1}\|_2}{(d + \tau)^2} \right) \right) \\
&\geq -\frac{C_{12} \sqrt{nn_\Delta} \|\bar{\boldsymbol{\mu}}_i\|_2}{d + \tau}.
\end{aligned}$$

D Proof of Proposition 1

In this section, we provide the proof of Proposition 1. The proof steps mostly resemble those in the proof of Theorem 1. We show the derivation for the case $b = +1$, noting that the case $b = -1$ follows by an identical set of calculations.

First of all, we use the matching exponential lower bound on the Q-function [6], which yields

$$\mathcal{R}_{+1}(\hat{\mathbf{w}}) = Q \left(\frac{\hat{\mathbf{w}}^\top \boldsymbol{\mu}_{+1}}{\sqrt{\hat{\mathbf{w}}^\top \hat{\mathbf{w}}}} \right) \geq C \exp \left(-\frac{(\hat{\mathbf{w}}^\top \boldsymbol{\mu}_{+1})^2}{2\hat{\mathbf{w}}^\top \hat{\mathbf{w}}} \right). \quad (24)$$

Therefore, it is sufficient to show an upper bound on $\frac{(\hat{\mathbf{w}}^\top \boldsymbol{\mu}_{+1})^2}{2\hat{\mathbf{w}}^\top \hat{\mathbf{w}}}$ to lower bound Eq. (24). We first reproduce the equivalent expression in terms of adjusted primitives (Eq. (15)), and then apply the bounds on these

primitives (provided in Lemma 3) to get the following upper bound:

$$\begin{aligned}
& \frac{(\hat{\mathbf{w}}^\top \boldsymbol{\mu}_{+1})^2}{2\hat{\mathbf{w}}^\top \hat{\mathbf{w}}} \\
&= \frac{\left(\|\bar{\boldsymbol{\mu}}_2\|_2^2 s_{2\Delta,2}^{(2)} + \|\bar{\boldsymbol{\mu}}_1\|_2^2 s_{2\Delta,1}^{(2)} + h_{2,2\Delta}^{(2)} + h_{1,2\Delta}^{(2)} \right)^2}{2o_{2\Delta,2\Delta}^{(2)}} \\
&\leq \frac{\left(\|\bar{\boldsymbol{\mu}}_2\|_2^2 \frac{(\hat{C}_9 + \frac{1}{\hat{C}_{10}}) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d+\tau} + \|\bar{\boldsymbol{\mu}}_1\|_2^2 \frac{\hat{C}_9 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) + \frac{1}{\hat{C}_{10}} \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right)}{d+\tau} + \frac{C_{12}\sqrt{nn_\Delta}\|\bar{\boldsymbol{\mu}}_2\|_2}{d+\tau} + \frac{C_{12}\sqrt{nn_\Delta}\|\bar{\boldsymbol{\mu}}_1\|_2}{d+\tau} \right)^2}{\frac{\check{C}_{13}n_\Delta d}{(d+\tau)^2}} \\
&= \frac{\left(\left(\hat{C}_9 \|\bar{\boldsymbol{\mu}}_2\|_2^2 + \frac{\|\bar{\boldsymbol{\mu}}_2\|_2^2 + \|\bar{\boldsymbol{\mu}}_1\|_2^2}{\hat{C}_{10}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \hat{C}_9 \|\bar{\boldsymbol{\mu}}_1\|_2^2 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) + C_{12}\sqrt{nn_\Delta} (\|\bar{\boldsymbol{\mu}}_2\|_2 + \|\bar{\boldsymbol{\mu}}_1\|_2) \right)^2}{\check{C}_{13}n_\Delta d}. \tag{25}
\end{aligned}$$

Next, since $R_- \geq 0$, i.e. $\|\bar{\boldsymbol{\mu}}_2\|_2 \geq \|\bar{\boldsymbol{\mu}}_1\|_2$, we have

$$\begin{aligned}
(25) &\leq \frac{\left(\left(\hat{C}_9 \|\bar{\boldsymbol{\mu}}_2\|_2^2 + \frac{2\|\bar{\boldsymbol{\mu}}_2\|_2^2}{\hat{C}_{10}} \right) \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \hat{C}_9 \|\bar{\boldsymbol{\mu}}_1\|_2^2 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) + 2C_{12}\sqrt{nn_\Delta} \|\bar{\boldsymbol{\mu}}_2\|_2 \right)^2}{\check{C}_{13}n_\Delta d} \\
&\leq \frac{\left(\left(\hat{C}_9 + \frac{2}{\hat{C}_{10}} + \frac{C_{12}}{\hat{C}_{14}} \right) \|\bar{\boldsymbol{\mu}}_2\|_2^2 \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \hat{C}_9 \|\bar{\boldsymbol{\mu}}_1\|_2^2 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) \right)^2}{\check{C}_{13}n_\Delta d} \\
&= \frac{\left(\hat{C}_{15} \|\bar{\boldsymbol{\mu}}_2\|_2^2 \left(\frac{n_+}{\Delta_+} + \frac{n_-}{\Delta_-} \right) + \hat{C}_{15} \|\bar{\boldsymbol{\mu}}_1\|_2^2 \left(\frac{n_+}{\Delta_+} - \frac{n_-}{\Delta_-} \right) \right)^2}{\check{C}_{13}n_\Delta d} \\
&= \frac{\left(\hat{C}_{15}R_+ \frac{n_+}{\Delta_+} + \hat{C}_{15}R_- \frac{n_-}{\Delta_-} \right)^2}{\check{C}_{13}n_\Delta d} \leq \frac{2\hat{C}_{15}^2 R_+^2 \frac{n_+^2}{\Delta_+^2} + 2\hat{C}_{15}^2 R_-^2 \frac{n_-^2}{\Delta_-^2}}{\check{C}_{13}n_\Delta d}, \tag{26}
\end{aligned}$$

where the second inequality follows from Lemma 4 applied with a large enough constant $C_{14} > C_{12}$. Finally, we define $\alpha_\pm = \frac{n_\pm/\Delta_\pm^2}{n_\Delta}$, noting that $0 < \alpha_\pm < 1$ and $\alpha_+ + \alpha_- = 1$. Then, Eq. (26) becomes

$$\frac{(\hat{\mathbf{w}}^\top \boldsymbol{\mu}_{+1})^2}{2\hat{\mathbf{w}}^\top \hat{\mathbf{w}}} \leq \frac{C_{16} (\alpha_+ R_+^2 n_+ + \alpha_- R_-^2 n_-)}{d}.$$

Plugging the above into Eq. (24) completes the proof. $\square$