Supplementary material for:

Explaining the Explainer: A First Theoretical Analysis of LIME

In this supplementary material, we provide the proof of Theorem 3.1 of the main paper. It is a simplified version of Theorem 10.1. We first recall our setting in Section 7. Then, following Section 5 of the main paper, we study the covariance matrix in Section 8, and the right-hand side of the key equation (5.1) in Section 9. Finally, we state and prove Theorem 10.1 in Section 10. Some technical results (mainly Gaussian integrals computation) and external concentration results are collected in Section 11.

7 Setting

Let us recall briefly the main assumptions under which we prove Theorem 3.1. Recall that they are discussed in details in Section 2.2 of the main paper.

H1 (Linear f). The black-box model can be written $a^\top x + b$, with $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ fixed.

H2 (Gaussian sampling). The random variables $x_1, \ldots, x_n$ are i.i.d. $\mathcal{N}\left(\mu, \sigma^2 \mathrm{I}_d\right)$.

Also recall that, for any $1 \leqslant i \leqslant n$, we set the weights to

$$\pi_i := \exp\left(\frac{-\left\|x_i - \xi\right\|^2}{2\nu^2}\right) . \tag{7.1}$$

We will need the following scaling constant:

$$C_d := \left(\frac{\nu^2}{\nu^2 + \sigma^2}\right)^{d/2} \cdot \exp\left(\frac{-\left\|\xi - \mu\right\|^2}{2(\nu^2 + \sigma^2)}\right) , \tag{7.2}$$

which does not play any role in the final result. One can check that $C_d \to 1$ when $\nu \gg \sigma$, regardless of the dimension.

Finally, for any $1 \leqslant j \leqslant d$, recall that we defined

$$\alpha_j := \left[\frac{1}{2}\mathrm{erf}\left(\frac{x - \tilde{\mu}_j}{\tilde{\sigma}\sqrt{2}}\right)\right]_{q_{j-}}^{q_{j+}} , \tag{7.3}$$

and

$$\theta_j := \left[\frac{\tilde{\sigma}}{\sqrt{2\pi}} \exp\left(\frac{-(x - \tilde{\mu}_j)^2}{2\tilde{\sigma}^2}\right)\right]_{q_{j-}}^{q_{j+}} , \tag{7.4}$$

where $q_{j\pm}$ are the quantile boundaries of ξ_j. These coefficients are discussed in Section 5 of the main paper. Note that all the expected values are taken with respect to the randomness on the $x_1, \ldots, x_n$.

8 Covariance matrix

In this section, we state and prove the intermediate results used to control the covariance matrix $\widehat{\Sigma}$. The goal of this section is to obtain the control of $\left\|\widehat{\Sigma}^{-1} - \Sigma^{-1}\right\|_{\mathrm{op}}$ in probability. Intuitively, if this quantity is small enough, then we can inverse Eq. (5.1) and make very precise statements about $\widehat{\beta}$.

We first show that it is possible to compute the expected covariance matrix in closed form. Without this result, a concentration result would still hold, but it would be much harder to gain precise insights on the β_js.

Lemma 8.1 (Expected covariance matrix). *Under Assumption 2, the expected value of $\widehat{\Sigma}$ is given by*

$$\Sigma := C_d \begin{pmatrix} 1 & \alpha_1 & \cdots & \alpha_d \\ \alpha_1 & \alpha_1 & & \alpha_i\alpha_j \\ \vdots & & \ddots & \\ \alpha_d & \alpha_i\alpha_j & & \alpha_d \end{pmatrix} .$$

Proof. Elementary computations yield

$$\widehat{\Sigma} = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n \pi_i & \sum_{i=1}^n \pi_i z_{i1} & \cdots & \sum_{i=1}^n \pi_i z_{id} \\ \sum_{i=1}^n \pi_i z_{i1} & \sum_{i=1}^n \pi_i z_{i1} & & \sum_{i=1}^n \pi_i z_{ik} z_{i\ell} \\ \vdots & & \ddots & \\ \sum_{i=1}^n \pi_i z_{id} & \sum_{i=1}^n \pi_i z_{ik} z_{i\ell} & & \sum_{i=1}^n \pi_i z_{id} \end{pmatrix} .$$

Reading the coefficients of this matrix, we have essentially three computations to complete: $\mathbb{E}\left[\pi_i\right]$, $\mathbb{E}\left[\pi_i z_{ik}\right]$, and $\mathbb{E}\left[\pi_i z_{ik} z_{i\ell}\right]$.

Computation of $\mathbb{E}\left[\pi_i\right]$. Since the x_is are Gaussian (Assumption 2) and using the definition of the weights (Eq. (7.1)), we can write

$$\mathbb{E}\left[\pi_i\right] = \int_{\mathbb{R}^d} \exp\left(\frac{-\left\|x_i - \xi\right\|^2}{2\nu^2}\right) \exp\left(\frac{-\left\|x_i - \mu\right\|^2}{2\sigma^2}\right) \frac{\mathrm{d}x_{i1}\cdots x_{id}}{(2\pi\sigma^2)^{d/2}} .$$

By independence across coordinates, the last display amounts to

$$\prod_{j=1}^{d} \int_{-\infty}^{+\infty} \exp\left(\frac{-(x-\xi_j)^2}{2\nu^2} + \frac{-(x-\mu_j)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} .$$

We then apply Lemma 11.1 to each of the integrals within the product to obtain

$$\prod_{j=1}^{d} \frac{\nu}{\sqrt{\nu^2+\sigma^2}} \cdot \exp\left(\frac{-(\xi_j-\mu_j)^2}{2(\nu^2+\sigma^2)}\right) = \frac{\nu^d}{(\nu^2+\sigma^2)^{d/2}} \cdot \exp\left(\frac{-\left\|\xi-\mu\right\|^2}{2(\nu^2+\sigma^2)}\right) .$$

We recognize the definition of the scaling constant (Eq. (7.2)): we have proved that $\mathbb{E}\left[\pi_i\right] = C_d$.

Computation of $\mathbb{E}\left[\pi_i z_{ik}\right]$. Since the x_is are Gaussian (Assumption 2) and using the definition of the weights (Eq. (7.1)),

$$\mathbb{E}\left[\pi_i\right] = \int_{\mathbb{R}^d} \exp\left(\frac{-\left\|x_i - \xi\right\|^2}{2\nu^2}\right) \exp\left(\frac{-\left\|x_i - \mu\right\|^2}{2\sigma^2}\right) \mathbf{1}_{\phi(x_i)_k = \phi(\xi)_k} \frac{\mathrm{d}x_{i1}\cdots x_{id}}{(2\pi\sigma^2)^{d/2}} .$$

By independence across coordinates, the last display amounts to

$$\int_{q_{k-}}^{q_{k+}} \exp\left(\frac{-(x-\xi_k)^2}{2\nu^2} + \frac{-(x-\mu_k)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \cdot \prod_{\substack{j=1 \\ j\neq k}}^{d} \int_{-\infty}^{+\infty} \exp\left(\frac{-(x-\xi_j)^2}{2\nu^2} + \frac{-(x-\mu_j)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} .$$

Using Lemma 11.1, we obtain

$$\frac{\nu^d}{(\nu^2+\sigma^2)^{d/2}} \cdot \exp\left(\frac{-\left\|\xi-\mu\right\|^2}{2(\nu^2+\sigma^2)}\right) \cdot \left[\frac{1}{2}\mathrm{erf}\left(\frac{\nu^2(x-\mu_k)+\sigma^2(x-\xi_k)}{\nu\sigma\sqrt{2(\nu^2+\sigma^2)}}\right)\right]_{q_{k-}}^{q_{k+}} .$$

We recognize the definition of the scaling constant (Eq. (7.2)) and of the α_k coefficient (Eq. (7.3)): we have proved that $\mathbb{E}\left[\pi_i z_{ik}\right] = C_d\alpha_k$.

Computation of $\mathbb{E}\left[\pi_i z_{ik} z_{i\ell}\right]$. Since the x_is are Gaussian (Assumption 2) and using the definition of the weights (Eq. (7.1)),

$$\mathbb{E}\left[\pi_i z_{ik} z_{i\ell}\right] = \int_{\mathbb{R}^d} \exp\left(\frac{-\left\|x_i - \xi\right\|^2}{2\nu^2}\right) \exp\left(\frac{-\left\|x_i - \mu\right\|^2}{2\sigma^2}\right) \mathbf{1}_{\phi(x_i)_k = \phi(\xi)_k} \mathbf{1}_{\phi(x_i)_\ell = \phi(\xi)_\ell} \frac{\mathrm{d}x_{i1} \cdots \mathrm{d}x_{id}}{(2\pi\sigma^2)^{d/2}} \, .$$

By independence across coordinates, the last display amounts to

$$\prod_{\substack{j=1 \\ j \neq k,\ell}}^{d} \int_{-\infty}^{+\infty} \exp\left(\frac{-(x-\xi_j)^2}{2\nu^2} + \frac{-(x-\mu_j)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \cdot \int_{q_{k-}}^{q_{k+}} \exp\left(\frac{-(x-\xi_k)^2}{2\nu^2} + \frac{-(x-\mu_k)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}}$$

$$\cdot \int_{q_{\ell-}}^{q_{\ell+}} \exp\left(\frac{-(x-\xi_\ell)^2}{2\nu^2} + \frac{-(x-\mu_\ell)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \, .$$

Using Lemma 11.1, we obtain

$$\frac{\nu^d}{(\nu^2+\sigma^2)^{d/2}} \cdot \exp\left(\frac{-\left\|\xi - \mu\right\|^2}{2(\nu^2+\sigma^2)}\right) \cdot \left[\frac{1}{2}\mathrm{erf}\left(\frac{\nu^2(x-\mu_k) + \sigma^2(x-\xi_k)}{\nu\sigma\sqrt{2(\nu^2+\sigma^2)}}\right)\right]_{q_{k-}}^{q_{k+}}$$

$$\cdot \left[\frac{1}{2}\mathrm{erf}\left(\frac{\nu^2(x-\mu_\ell) + \sigma^2(x-\xi_\ell)}{\nu\sigma\sqrt{2(\nu^2+\sigma^2)}}\right)\right]_{q_{\ell-}}^{q_{\ell+}} \, .$$

We recognize the definition of the scaling constant (Eq. (7.2)) and of the alphas (Eq. (7.3)): we have proved that $\mathbb{E}\left[\pi_i z_{ik} z_{i\ell}\right] = C_d \alpha_k \alpha_\ell$. □

As it turns out, we show that it is possible to invert Σ in closed-form, therefore simplifying tremendously our quest for control of $\left\|\hat{\Sigma}^{-1} - \Sigma^{-1}\right\|_{\mathrm{op}}$. Indeed, in most cases, even if concentration could be shown, one would not have a precise idea of the coefficients of Σ^{-1}.

Lemma 8.2 (Inverse of the covariance matrix). *If $\alpha_j \neq 0, 1$ for any $j \in \{1, \ldots, d\}$, then Σ is invertible, and*

$$\Sigma^{-1} = C_d^{-1} \begin{pmatrix} 1 + \sum_{j=1}^{d} \frac{\alpha_j}{1-\alpha_j} & \frac{-1}{1-\alpha_1} & \cdots & \frac{-1}{1-\alpha_d} \\ \frac{-1}{1-\alpha_1} & \frac{1}{\alpha_1(1-\alpha_1)} & & 0 \\ \vdots & & \ddots & \\ \frac{-1}{1-\alpha_d} & 0 & & \frac{1}{\alpha_d(1-\alpha_d)} \end{pmatrix} .$$

Proof. Define $\alpha \in \mathbb{R}^d$ the vector of the α_js. Set $A := 1$, $B := \alpha^\top$, $C := \alpha$, and

$$D := \begin{pmatrix} \alpha_1 & & \alpha_j \alpha_k \\ & \ddots & \\ \alpha_j \alpha_k & & \alpha_d \end{pmatrix} .$$

Then Σ is a block matrix that can be written $\Sigma = C_d \begin{bmatrix} A & B \\ C & D \end{bmatrix}$. We notice that

$$D - CA^{-1}B = \mathrm{Diag}\left(\alpha_1(1-\alpha_1), \ldots, \alpha_d(1-\alpha_d)\right) \, .$$

Note that, since erf is an increasing function, the α_js are always distinct from 0 and 1. Thus $D - CA^{-1}B$ is an invertible matrix, and we can use the block matrix inversion formula to obtain the claimed result. □

As a direct consequence of the computation of Σ^{-1}, we can control its largest eigenvalue.

Lemma 8.3 (Control of $\left\|\Sigma^{-1}\right\|_{\mathrm{op}}$). *We have the following bound on the operator norm of the inverse covariance matrix:*

$$\left\|\Sigma^{-1}\right\|_{op} \leqslant \frac{3dA_d}{C_d} \, ,$$

where $A_d := \max_{1 \leqslant j \leqslant d} \frac{1}{\alpha_j(1-\alpha_j)}$.

Proof. We control the operator norm of Σ^{-1} by its Frobenius norm: Namely,

$$\begin{aligned}\left\|\Sigma^{-1}\right\|_{\mathrm{op}}^2 &\leqslant \left\|\Sigma^{-1}\right\|_{\mathrm{F}}^2 \\ &= C_d^{-2}\left[\left(1+\sum\frac{\alpha_j}{1-\alpha_j}\right)^2+\sum\frac{1}{(1-\alpha_j)^2}+\sum\frac{1}{\alpha_j(1-\alpha_j)}\right] \\ \left\|\Sigma^{-1}\right\|_{\mathrm{op}}^2 &\leqslant 6C_d^{-2}d^2\left(\max\frac{1}{\alpha_j(1-\alpha_j)}\right)^2,\end{aligned}$$

where we used $\alpha_j \in (0,1)$ in the last step of the derivation. □

Remark 8.1. Better bounds can without doubt be obtained. A step in this direction is to notice that $S := C_d\Sigma^{-1}$ is an arrowhead matrix (O'Leary and Stewart, 1996). Thus the eigenvalues of S are solutions of the secular equation

$$1+\sum_{j=1}^d \frac{\alpha_j}{1-\alpha_j} - \lambda + \sum_{j=1}^d \frac{\alpha_j}{(1-\alpha_j)(1-\lambda\alpha_j(1-\alpha_j))} = 0\,.$$

Further study of this equation could yield an improved statement for Lemma 8.3.

We now show that the empirical covariance matrix concentrates around Σ. It is interesting to see that the non-linear nature of the new coordinates (the z_{ij}s) calls for complicated computations but allows us to use simple concentration tools since they are, in essence, Bernoulli random variables.

Lemma 8.4 (Concentration of the empirical covariance matrix). *Let $\hat{\Sigma}$ and Σ be defined as before. Then, for every $t > 0$,*

$$\mathbb{P}\left(\left\|\hat{\Sigma}-\Sigma\right\|_{op} \geqslant t\right) \leqslant 4d^2\exp(-2nt^2)\,.$$

Proof. Recall that $\|\cdot\|_{\mathrm{op}} \leqslant \|\cdot\|_{\mathrm{F}}$: it suffices to show the result for the Frobenius norm. Next, we notice that the summands appearing in the entries of $\hat{\Sigma}$, $X_i^{(1)} := \pi_i$, $X_i^{(2,k)} := \pi_i z_{ik}$, and $X_i^{(3,k,\ell)} := \pi_i z_{ik} z_{i\ell}$, are all bounded. Indeed, by the definition of the weights and the definition of the new features, they all take values in $[0,1]$. Moreover, for given k,ℓ, they are independent random variables. Thus we can apply Hoeffding's inequality (Theorem 11.1) to $X_i^{(1)}$, $X_i^{(2,k)}$, and $X_i^{(3,k,\ell)}$. For any given $t > 0$, we obtain

$$\begin{cases}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n(\pi_i-\mathbb{E}\left[\pi_i\right])\right| \geqslant t\right) \leqslant 2\exp(-2nt^2)\\ \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n(\pi_i z_{ik}-\mathbb{E}\left[\pi_i\right])\right| \geqslant t\right) \leqslant 2\exp(-2nt^2)\\ \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n(\pi_i z_{ik} z_{i\ell}-\mathbb{E}\left[\pi_i\right])\right| \geqslant t\right) \leqslant 2\exp(-2nt^2)\end{cases}$$

We conclude by a union bound on the $(d+1)^2 \leqslant 2d^2$ entries of the matrix. □

As a consequence of the two preceding lemmas, we can control the largest eigenvalue of Σ^{-1}.

Lemma 8.5 (Control of $\left\|\hat{\Sigma}^{-1}\right\|_{\mathbf{op}}$). *For every $t \in \left(0, \frac{C_d}{6dA_d}\right]$, with probability greater than $1-4d^2\exp(-2nt^2)$,*

$$\left\|\hat{\Sigma}^{-1}\right\|_{op} \leqslant \frac{6dA_d}{C_d}\,.$$

Proof. Let $t \in (0, C_d/(6dA_d)]$. According to Lemma 8.3, $\lambda_{\max}(\Sigma^{-1}) \leqslant 3dA_d/C_d$. We deduce that

$$\lambda_{\min}(\Sigma) \geqslant \frac{C_d}{3dA_d}\,.$$

Now let us use Lemma 8.4 with this t: there is an event Ω, which has probability greater than $1-4d^2\exp(-2nt^2)$, such that $\left\|\hat{\Sigma}-\Sigma\right\|_{\mathrm{op}} \leqslant t$. According to Weyl's inequality (Weyl, 1912), on this event,

$$\left|\lambda_{\min}(\hat{\Sigma})-\lambda_{\min}(\Sigma)\right| \leqslant \left\|\hat{\Sigma}-\Sigma\right\|_{\mathrm{op}} \leqslant t\,.$$

In particular,

$$\lambda_{\min}(\widehat{\Sigma}) \geqslant \lambda_{\min}(\Sigma) - t \geqslant \frac{C_d}{6dA_d}\,.$$

Finally, we deduce that

$$\left\|\widehat{\Sigma}^{-1}\right\|_{\mathrm{op}} \leqslant \frac{6dA_d}{C_d}\,.$$

□

We can now state and prove the main result of this section, controlling the operator norm of $\widehat{\Sigma} - \Sigma$ with high probability.

Proposition 8.1 (Control of $\left\|\widehat{\Sigma}^{-1} - \Sigma^{-1}\right\|_{\mathbf{op}}$). *For every $t \in \left(0, \frac{3dA_d}{C_d}\right]$, we have*

$$\mathbb{P}\left(\left\|\widehat{\Sigma}^{-1} - \Sigma^{-1}\right\|_{op} \geqslant t\right) \leqslant 8d^2 \exp\left(\frac{-C_d^4 nt^2}{162 d^4 A_d^4}\right)\,.$$

Remark 8.2. Proposition 8.1 is the key tool to invert Eq. (5.1) and gain precise control over $\widehat{\beta}$. In the regime that we consider, the dimension d as well as the number of bins p are *fixed*, and d, C_d, and A_d are essentially numerical constants. We did not optimize these constant with respect to d, since the main message is to consider the behavior for a large number of new examples ($n \to +\infty$).

Proof. We notice that, assuming that $\widehat{\Sigma}$ is invertible, $\widehat{\Sigma}^{-1} - \Sigma^{-1} = \widehat{\Sigma}^{-1}(\Sigma - \widehat{\Sigma})\Sigma^{-1}$. Since $\left\|\cdot\right\|_{\mathrm{op}}$ is sub-multiplicative, we just have to control each term individually. Lemma 8.3 gives us

$$\left\|\Sigma^{-1}\right\|_{\mathrm{op}} \leqslant \frac{3dA_d}{C_d}\,.$$

Next, set $t_1 := \frac{C_d^2 t}{18 d^2 A_d^2}$. According to Lemma 8.4, with probability greater than $1 - 4d^2\exp(-2nt_1^2)$,

$$\left\|\widehat{\Sigma} - \Sigma\right\|_{\mathrm{op}} \leqslant t_1\,.$$

Finally, set $t_2 := t_1$. It is easy to check that $t_2 \leqslant C_d/(6dA_d)$. Thus we can use Lemma 8.5: with probability greater than $1 - 4d^2\exp(-2nt_1^2)$,

$$\left\|\widehat{\Sigma}^{-1}\right\|_{\mathrm{op}} \leqslant \frac{6dA_d}{C_d}\,.$$

By the union bound, with probability greater than $1 - 8d^2 \exp\left(\frac{-C_d^4 nt^2}{162 d^4 A_d^4}\right)$,

$$\begin{aligned}\left\|\widehat{\Sigma}^{-1} - \widehat{\Sigma}\right\|_{\mathrm{op}} &\leqslant \left\|\Sigma^{-1}\right\|_{\mathrm{op}} \cdot \left\|\widehat{\Sigma} - \Sigma\right\|_{\mathrm{op}} \cdot \left\|\widehat{\Sigma}^{-1}\right\|_{\mathrm{op}} \\ &\leqslant \frac{3dA_d}{C_d} \cdot t_1 \cdot \frac{6dA_d}{C_d} = t\,.\end{aligned}$$

□

9 Right-hand side of Eq. (5.1)

In this section, we state and prove the results in relation to $\widehat{\Gamma}$. We begin with the computation of Γ, the expected value of $\widehat{\Gamma}$.

Lemma 9.1 (Computation of Γ). *Under Assumption 2 and 1, the expected value of $\widehat{\Gamma}$ is given by*

$$\Gamma = C_d \begin{pmatrix} f(\tilde{\mu}) \\ \alpha_1 f(\tilde{\mu}) - a_1\theta_1 \\ \vdots \\ \alpha_d f(\tilde{\mu}) - a_d\theta_d \end{pmatrix},$$

where the θ_js are defined by

$$\theta_j := \left[\frac{\tilde{\sigma}}{\sqrt{2\pi}} \exp\left(\frac{-(x-\tilde{\mu}_j)^2}{2\tilde{\sigma}^2} \right) \right]_{q_{j-}}^{q_{j+}} .$$

Proof. Recall that $\widehat{\Gamma} = \frac{1}{n} Z^\top \Pi f(x)$. Elementary computations yield

$$\widehat{\Gamma} = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n \pi_i f(x_i) \\ \sum_{i=1}^n \pi_i z_{i1} f(x_i) \\ \vdots \\ \sum_{i=1}^n \pi_i z_{id} f(x_i) \end{pmatrix} .$$

Given the expression of $\widehat{\Gamma}$, we have essentially two computations to manage: $\mathbb{E}\left[\pi_i f(x_i)\right]$ and $\mathbb{E}\left[\pi_i z_{ij} f(x_i)\right]$.

Computation of $\mathbb{E}\left[\pi_i f(x_i)\right]$. Under Assumption 1, by linearity of the integral,

$$\mathbb{E}\left[\pi_i f(x_i)\right] = \mathbb{E}\left[\pi_i(a^\top x_i + b)\right] = b\mathbb{E}\left[\pi_i\right] + \sum_{j=1}^d a_j \mathbb{E}\left[\pi_i x_{ij}\right] . \tag{9.1}$$

Now we have already seen in the proof of Lemma 8.1 that $\mathbb{E}\left[\pi_i\right] = C_d$. Thus we can focus on the computation of $\mathbb{E}\left[\pi_i x_{ij}\right]$ for fixed i, j. Under Assumption 2, we have

$$\mathbb{E}\left[\pi_i x_{ij}\right] = \int_{\mathbb{R}^d} x_j \cdot \exp\left(\frac{-\left\|x-\xi\right\|^2}{2\nu^2} + \frac{-\left\|x-\mu\right\|^2}{2\sigma^2} \right) \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_d}{(2\pi\sigma^2)^{d/2}} .$$

By independence, the last display amounts to

$$\int_{-\infty}^{+\infty} x \cdot \exp\left(\frac{-(x-\xi_j)^2}{2\nu^2} + \frac{-(x-\mu_j)^2}{2\sigma^2} \right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \cdot \prod_{k\neq j} \int_{-\infty}^{+\infty} \exp\left(\frac{-(x-\xi_k)^2}{2\nu^2} + \frac{-(x-\mu_k)^2}{2\sigma^2} \right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} .$$

A straightforward application of Lemmas 11.1 and 11.2 yields

$$\mathbb{E}\left[\pi_i x_{ij}\right] = C_d \cdot \frac{\nu^2 \mu_j + \sigma^2 \xi_j}{\nu^2 + \sigma^2} .$$

Back to Eq. (9.1), we have shown that

$$\mathbb{E}\left[\pi_i f(x_i)\right] = C_d b + \sum_{j=1}^d a_j \cdot C_d \frac{\nu^2 \mu_j + \sigma^2 \xi_j}{\nu^2 + \sigma^2} = C_d f(\tilde{\mu}) .$$

Computation of $\mathbb{E}\left[\pi_i z_{ij} f(x_i)\right]$. Under Assumption 1, by linearity of the integral,

$$\mathbb{E}\left[\pi_i z_{ij} f(x_i)\right] = b\mathbb{E}\left[\pi_i z_{ij}\right] + \sum_{k=1}^d a_k \cdot \mathbb{E}\left[\pi_i z_{ij} x_{ik}\right] . \tag{9.2}$$

We have already computed $\mathbb{E}\left[\pi_i z_{ij}\right]$ in the proof of Lemma 8.1 and found that

$$\mathbb{E}\left[\pi_i z_{ij}\right] = C_d \alpha_j .$$

Regarding the computation of $\mathbb{E}\left[\pi_i z_{ij} x_{ik}\right]$, there are essentially two cases to consider depending whether $k = \ell$ or not. Let us first consider the case $k = j$. Then we obtain

$$\mathbb{E}\left[\pi_i z_{ij} x_{ik}\right] = \int_{\mathbb{R}^d} x_j \exp\left(\frac{-\left\|x-\xi\right\|^2}{2\nu^2} + \frac{-\left\|x-\mu\right\|^2}{2\sigma^2} \right) \mathbf{1}_{\phi(x)_j = \phi(\xi)_j} \frac{\mathrm{d}x_1 \cdots \mathrm{d}x_d}{(2\pi\sigma^2)^{d/2}} .$$

By independence, the last display amounts to

$$\int_{q_{j-}}^{q_{j+}} x \cdot \exp\left(\frac{-(x-\xi_j)^2}{2\nu^2} + \frac{-(x-\mu_j)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \cdot \prod_{k\neq j} \int_{-\infty}^{+\infty} \exp\left(\frac{-(x-\xi_k)^2}{2\nu^2} + \frac{-(x-\mu_k)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \,.$$

According to Lemma 11.2 and the definition of α_j and θ_j (Eqs. (7.3) and (7.3)), we have

$$\mathbb{E}\left[\pi_i z_{ij} x_{ij}\right] = C_d \frac{\sigma^2 \xi_j + \nu^2 \mu_j}{\nu^2 + \sigma^2} \alpha_j - C_d \theta_j \,.$$

Now if $k \neq j$, by independence, $\mathbb{E}\left[\pi_i z_{ij} x_{ik}\right]$ splits in three parts:

$$\begin{aligned}\mathbb{E}\left[\pi_i z_{ij} x_{ik}\right] = \int_{-\infty}^{+\infty} x \cdot \exp\left(\frac{-(x-\xi_k)^2}{2\nu^2} + \frac{-(x-\mu_k)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \cdot \int_{q_{j-}}^{q_{j+}} \exp\left(\frac{-(x-\xi_j)^2}{2\nu^2} + \frac{-(x-\mu_j)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \cdot \\ \cdot \prod_{\ell \neq j,k} \int_{-\infty}^{+\infty} \exp\left(\frac{-(x-\xi_k)^2}{2\nu^2} + \frac{-(x-\mu_k)^2}{2\sigma^2}\right) \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \,.\end{aligned}$$

Lemma 11.1 and 11.2 yield

$$\mathbb{E}\left[\pi_i z_{ij} x_{ik}\right] = C_d \cdot \frac{\sigma^2 \xi_k + \nu^2 \mu_k}{\nu^2 + \sigma^2} \cdot \alpha_j \,.$$

In definitive, plugging these results into Eq. (9.2) gives

$$\begin{aligned}\mathbb{E}\left[\pi_i z_{ij} f(x_i)\right] &= C_d \alpha_j b + a_j \left(C_d \frac{\sigma^2 \xi_j + \nu^2 \mu_j}{\nu^2 + \sigma^2} \alpha_j - C_d \theta_j\right) + \sum_{k \neq j} a_k \cdot C_d \frac{\sigma^2 \xi_k + \nu^2 \mu_k}{\nu^2 + \sigma^2} \alpha_j \\ &= C_d \alpha_j f(\tilde{\mu}) - C_d a_j \theta_j \,.\end{aligned}$$

□

As a consequence of Lemma 9.1, we can control $\lVert\Gamma\rVert$.

Lemma 9.2 (Control of $\lVert\Gamma\rVert$). *Under Assumptions 2 and 1, it holds that*

$$\lVert\Gamma\rVert^2 \leqslant C_d^2 \left(3 d f(\tilde{\mu})^2 + d \tilde{\sigma}^2 \lVert\nabla f\rVert^2\right) \,.$$

Proof. According to Lemma 9.1, we have

$$\lVert\Gamma\rVert^2 = C_d^2 \left(f(\tilde{\mu})^2 + \sum_{j=1}^{d} (\alpha_j f(\tilde{\mu}) - a_j \theta_j)^2\right) \,.$$

Successively using $(x-y)^2 \leqslant 2(x^2+y^2)$, $\alpha_j \in [0,1]$ and $\theta_j \in [-\tilde{\sigma}/\sqrt{2\pi}, \tilde{\sigma}/\sqrt{2\pi}]$, we write

$$\begin{aligned}\lVert\Gamma\rVert^2 &\leqslant C_d^2 \left(f(\tilde{\mu})^2 + \sum_{j=1}^{d} 2(\alpha_j^2 f(\tilde{\mu})^2 + a_j^2 \theta_j^2)\right) \\ &\leqslant C_d^2 \left(3 d f(\tilde{\mu})^2 + d \tilde{\sigma}^2 \lVert a\rVert^2\right) \,,\end{aligned}$$

which concludes the proof. □

Finally, we conclude this section with a concentration result for $\widehat{\Gamma}$.

Lemma 9.3 (Concentration of $\lVert\widehat{\Gamma}\rVert$). *Under Assumptions 2 and 1, for any $t > 0$, we have*

$$\mathbb{P}\left(\lVert\widehat{\Gamma} - \Gamma\rVert > t\right) \leqslant 4d \exp\left(\frac{-nt^2}{2 \lVert\nabla f\rVert^2 \sigma^2}\right) \,.$$

Proof. Since the x_i are Gaussian with variance σ^2 (Assumption 2), the random variable $a^\top x_i + b$ is Gaussian with variance $\left\|a\right\|^2 \sigma^2$, and the $X_i^{(1)} := \pi_i x_i$ are sub-Gaussian with parameter $\left\|a\right\|^2 \sigma^2$. They are also independent, thus we can apply Theorem 11.2 to the $X_i^{(1)}$:

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \pi_i f(x_i) - \mathbb{E}\left[\pi_i f(x_i)\right] \right| > t \right) \leqslant 2 \exp\left(\frac{-nt^2}{2\left\|a\right\|^2 \sigma^2} \right) .$$

Furthermore, the z_{ij} are $\{0,1\}$-valued. Thus the random variables $X_i^{(j)} := \pi_i z_{ij} f(x_i)$ are also sub-Gaussian with parameter $\left\|a\right\|^2 \sigma^2$. We use Hoeffding's inequality (Theorem 11.2) again, to obtain, for any j,

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \pi_i z_{ij} f(x_i) - \mathbb{E}\left[\pi_i z_{ij} f(x_i)\right] \right| > t \right) \leqslant 2 \exp\left(\frac{-nt^2}{2\left\|a\right\|^2 \sigma^2} \right) .$$

By the union bound,

$$\mathbb{P}\left(\left\| \widehat{\Gamma} - \Gamma \right\| > t \right) \leqslant 2(d+1) \exp\left(\frac{-nt^2}{2\left\|a\right\|^2 \sigma^2} \right) .$$

We deduce the result since $d \geqslant 1$. □

10 Proof of the main result

In this section, we state and prove our main result, Theorem 10.1. It is a more precise version than Theorem 3.1 in the main paper.

Theorem 10.1 (Concentration of $\widehat{\beta}$). *Let $\eta \in (0,1)$ and $\varepsilon > 0$. Take*

$$n \geqslant \max\left(\frac{288\left\|\nabla f\right\|^2 \sigma^2 d^2 A_d^2}{\varepsilon^2 C_d^2} \log \frac{12d}{\eta}, \frac{18d^2 A_d^2}{C_d^2} \log \frac{24d^2}{\eta}, \frac{648 d^5 A_d^4 (3f(\tilde{\mu})^2 + \tilde{\sigma}^2 \left\|\nabla f\right\|^2)}{C_d^2 \varepsilon^2} \log \frac{24d^2}{\eta} \right) .$$

Then, under assumptions 2 and 1,

$$\left\| \widehat{\beta} - \Sigma^{-1}\Gamma \right\| \leqslant \varepsilon ,$$

with probability greater than $1-\eta$.

Proof. The main idea of the proof is to notice that

$$\begin{aligned} \left\| \widehat{\beta} - \Sigma^{-1}\Gamma \right\| &= \left\| \widehat{\Sigma}^{-1}\widehat{\Gamma} - \Sigma^{-1}\Gamma \right\| \\ &\leqslant \left\| \widehat{\Sigma}^{-1}(\widehat{\Gamma} - \Gamma) \right\| + \left\| (\widehat{\Sigma}^{-1} - \Sigma^{-1})\Gamma \right\| , \end{aligned}$$

and then to control these two terms using the results of Section 8 and 9.

Control of $\left\| \widehat{\Sigma}^{-1}(\widehat{\Gamma} - \Gamma) \right\|$. We use the upper bound $\left\| \widehat{\Sigma}^{-1}(\widehat{\Gamma} - \Gamma) \right\| \leqslant \left\| \widehat{\Sigma}^{-1} \right\|_{\mathrm{op}} \cdot \left\| \widehat{\Gamma} - \Gamma \right\|$. We then achieve control of the operator norm of the empirical covariance matrix in probability with Lemma 8.5, and control of the norm of $\widehat{\Gamma} - \Gamma$ in probability with Lemma 9.3. Set

$$t_1 := \frac{C_d}{6dA_d} \quad \text{and} \quad n_1 := \frac{18d^2}{C_d^2} \log \frac{12d^2}{\eta} .$$

According to Lemma 8.5, for any $n \geqslant n_1$, there is an event Ω_1^n which has probability greater than $1 - 4d^2 \exp(-2nt_1^2)$ such that

$$\left\| \widehat{\Sigma}^{-1} \right\|_{\mathrm{op}} \leqslant \frac{6dA_d}{C_d}$$

on this event. It is easy to check that $4d^2 \exp(-2n_1 t_1^2) = \eta/3$, thus Ω_1^n has probability greater than $1 - \eta/3$. Now set

$$t_2 := \frac{\varepsilon C_d}{12dA_d} \quad \text{and} \quad n_2 := \frac{288\left\|a\right\|^2 \sigma^2 d^2 A_d^2}{\varepsilon^2 C_d^2} \log \frac{12d}{\eta} .$$

According to Lemma 9.3, for any $n \geqslant n_2$, there exists an event Ω_2^n which has probability greater than $1 - 4d \exp\left(\frac{-nt_2^2}{2\|a\|^2\sigma^2}\right)$ such that $\left\|\hat{\Gamma} - \Gamma\right\| \leqslant t_2$ on that event. One can check that

$$4d \exp\left(\frac{-n_2 t_2^2}{2\left\|a\right\|^2 \sigma^2}\right) = \frac{\eta}{3}\,,$$

thus Ω_2^n has probability greater than $1 - \eta/3$. On the event $\Omega_1^n \cap \Omega_2^n$, we have

$$\left\|\hat{\Sigma}^{-1}(\hat{\Gamma} - \Gamma)\right\| \leqslant \left\|\hat{\Sigma}^{-1}\right\|_{\mathrm{op}} \cdot \left\|\hat{\Gamma} - \Gamma\right\| \leqslant \frac{6dA_d}{C_d} \cdot t_2 \leqslant \frac{\varepsilon}{2}\,,$$

by definition of t_2.

Control of $\left\|(\hat{\Sigma}^{-1} - \Sigma^{-1})\Gamma\right\|$. We use the upper bound $\left\|(\hat{\Sigma}^{-1} - \Sigma^{-1})\Gamma\right\| \leqslant \left\|\hat{\Sigma}^{-1} - \Sigma^{-1}\right\|_{\mathrm{op}} \cdot \|\Gamma\|$. We then achieve control of $\left\|\hat{\Sigma}^{-1} - \Sigma^{-1}\right\|_{\mathrm{op}}$ in probability with Proposition 8.1, whereas we can bound the norm of Γ almost surely with Lemma 9.2. If $\|\Gamma\| = 0$, then there is nothing to prove. Otherwise, set

$$t_3 := \min\left(\frac{\varepsilon}{2\left\|\Gamma\right\|}, \frac{3dA_d}{C_d}\right),\; n_3 := \frac{18d^2A_d^2}{C_d^2}\log\frac{24d^2}{\eta}, \quad \text{and} \quad n_4 := \frac{648d^5A_d^4(3f(\tilde{\mu})^2 + \tilde{\sigma}^2\left\|a\right\|^2)}{C_d^2\varepsilon^2}\log\frac{24d^2}{\eta}\,.$$

According to Proposition 8.1, for any $n \geqslant \max(n_3, n_4)$, there is an event Ω_3^n which has probability greater than $1 - 8d^2 \exp\left(\frac{-C_d^3 n t_3^2}{162d^2A_d^4}\right)$ such that

$$\left\|\hat{\Sigma}^{-1} - \Sigma^{-1}\right\|_{\mathrm{op}} \leqslant t_3$$

on this event. With the help of Lemma 9.2, one can check that

$$\max\left(8d^2\exp\left(\frac{-C_d^3 n_3 t_3^2}{162d^2A_d^4}\right), 8d^2\exp\left(\frac{-C_d^3 n_4 t_3^2}{162d^2A_d^4}\right)\right) \leqslant \frac{\eta}{3}\,.$$

Therefore, Ω_3^n has probability greater than $\eta/3$ and, on this event,

$$\left\|(\hat{\Sigma}^{-1} - \Sigma^{-1})\Gamma\right\| \leqslant \left\|\hat{\Sigma}^{-1} - \Sigma^{-1}\right\|_{\mathrm{op}} \cdot \|\Gamma\| \leqslant t_3 \cdot \|\Gamma\| \leqslant \frac{\varepsilon}{2}\,.$$

Conclusion. Set $n \geqslant \max(n_i, i = 1\ldots 4)$. Define $\Omega^n := \Omega_1^n \cap \Omega_2^n \cap \Omega_3^n$, where the Ω_i^n are defined as before. According to the previous reasoning, on the event Ω^n,

$$\begin{aligned}\left\|\hat{\beta} - \Sigma^{-1}\Gamma\right\| &= \left\|\hat{\Sigma}^{-1}\hat{\Gamma} - \Sigma^{-1}\Gamma\right\| \\ &\leqslant \left\|\hat{\Sigma}^{-1}(\hat{\Gamma} - \Gamma)\right\| + \left\|(\hat{\Sigma}^{-1} - \Sigma^{-1})\Gamma\right\| \\ &\leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.\end{aligned}$$

Moreover, the union bound gives $\mathbb{P}\left(\Omega^n\right) \geqslant 1 - \eta$. We conclude by noticing that n_1 is always smaller than n_3, thus we just have to require $n \geqslant \max(n_2, n_3, n_4)$, as in the statement of our result. □

11 Technical lemmas

11.1 Gaussian integrals

In this section, we collect some Gaussian integral computations that are needed in our derivations. We provide succinct proof, since essentially any modern computer algebra system will provide these formulas. Our first result is for zero-th order Gaussian integral.

Lemma 11.1 (Gaussian integral, 0-th order). *Let ξ, μ be real numbers, and ν, σ be positive real numbers. Then, it holds that*

$$\int \exp\left(\frac{-(x-\xi)^2}{2\nu^2}+\frac{-(x-\mu)^2}{2\sigma^2}\right)\frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} = \frac{\nu}{\sqrt{\nu^2+\sigma^2}}\cdot\exp\left(\frac{-(\xi-\mu)^2}{2(\nu^2+\sigma^2)}\right)\cdot\frac{1}{2}\mathrm{erf}\left(\frac{\nu^2(x-\mu)+\sigma^2(x-\xi)}{\nu\sigma\sqrt{2(\nu^2+\sigma^2)}}\right).$$

In particular,

$$\int_{-\infty}^{+\infty} \exp\left(\frac{-(x-\xi)^2}{2\nu^2}+\frac{-(x-\mu)^2}{2\sigma^2}\right)\frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} = \frac{\nu}{\sqrt{\nu^2+\sigma^2}}\cdot\exp\left(\frac{-(\xi-\mu)^2}{2(\nu^2+\sigma^2)}\right).$$

Proof. For any reals a, b, and c, it holds that

$$\int \mathrm{e}^{-ax^2+bx+c}\,\mathrm{d}x = \sqrt{\frac{\pi}{a}}\cdot\mathrm{e}^{\frac{b^2}{4a}+c}\cdot\frac{1}{2}\mathrm{erf}\left(\frac{2ax-b}{2\sqrt{a}}\right).$$

We apply this formula with $a = \frac{1}{2\nu^2}+\frac{1}{2\sigma^2}$, $b = \frac{\xi}{\nu^2}+\frac{\mu}{\sigma^2}$, and $c = -\left(\frac{\xi^2}{2\nu^2}+\frac{\mu^2}{\sigma^2}\right)$. We then notice that $b^2/(4a)+c = \frac{-(\xi-\mu)^2}{2(\nu^2+\sigma^2)}$ and

$$\frac{2ax-b}{2\sqrt{a}} = \frac{\nu^2(x-\mu)+\sigma^2(x-\xi)}{\nu\sigma\sqrt{2(\nu^2+\sigma^2)}}.$$

□

Remark 11.1. We often replace $\frac{\nu^2(x-\mu)+\sigma^2(x-\xi)}{\nu\sigma\sqrt{2(\nu^2+\sigma^2)}}$ by the more readable $(x-\tilde{\mu})/(\tilde{\sigma}\sqrt{2})$ in the main text of the paper.

Since f is assumed to be linear in most of the paper, we need first order computations as well:

Lemma 11.2 (Gaussian integral, 1st order). *Let ξ, μ be real numbers, and ν, σ be positive numbers. Then it holds that*

$$\int x\cdot\exp\left(\frac{-(x-\xi)^2}{2\nu^2}+\frac{-(x-\mu)^2}{2\sigma^2}\right)\frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} = \frac{\nu}{\sqrt{\nu^2+\sigma^2}}\cdot\exp\left(\frac{-(\xi-\mu)^2}{2(\nu^2+\sigma^2)}\right)\cdot$$
$$\left[\frac{\sigma^2\xi+\nu^2\mu}{\nu^2+\sigma^2}\cdot\frac{1}{2}\mathrm{erf}\left(\frac{\nu^2(x-\mu)+\sigma^2(x-\xi)}{\nu\sigma\sqrt{2(\nu^2+\sigma^2)}}\right)-\frac{\nu\sigma}{\sqrt{2\pi}\sqrt{\nu^2+\sigma^2}}\cdot\exp\left(-\left(\frac{\nu^2(x-\mu)+\sigma^2(x-\xi)}{\nu\sigma\sqrt{2(\nu^2+\sigma^2)}}\right)^2\right)\right].$$

In particular,

$$\int_{-\infty}^{+\infty} x\cdot\exp\left(\frac{-(x-\xi)^2}{2\nu^2}+\frac{-(x-\mu)^2}{2\sigma^2}\right)\frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} = \frac{\sigma^2\xi+\nu^2\mu}{\nu^2+\sigma^2}\cdot\frac{\nu}{\sqrt{\nu^2+\sigma^2}}\cdot\exp\left(\frac{-(\xi-\mu)^2}{2(\nu^2+\sigma^2)}\right).$$

Proof. For any a, b, c with $a > 0$, it holds that

$$\int x\cdot\mathrm{e}^{-ax^2+bx+c}\,\mathrm{d}x = \frac{\sqrt{\pi}b}{4a^{3/2}}\mathrm{e}^{b^2/(4a)+c}\mathrm{erf}\left(\frac{2ax-b}{2\sqrt{a}}\right)-\frac{1}{2a}\mathrm{e}^{-ax^2+bx+c}.$$

□

Finally we want to mention the following result.

Lemma 11.3 (Gaussian integral, 2nd order). *Let ξ, μ be real numbers, and ν, σ be positive real numbers. Then, it holds that*

$$\int_{-\infty}^{+\infty} x^2\cdot\exp\left(\frac{-(x-\xi)^2}{2\nu^2}+\frac{-(x-\mu)^2}{2\sigma^2}\right)\frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} = \frac{(\sigma^2\xi+\nu^2\mu)^2+\nu^2\sigma^2(\nu^2+\sigma^2)}{(\nu^2+\sigma^2)^2}\cdot\frac{\nu}{\sqrt{\nu^2+\sigma^2}}\cdot\exp\left(\frac{-(\xi-\mu)^2}{2(\nu^2+\sigma^2)}\right).$$

Remark 11.2. As a consequence of Lemma 11.3, it would be possible to further our analysis by adding second degree terms to f. Indeed, quantities depending on $\|x_i - \xi\|$, which would have to be computed to extend the proofs of Lemmas 9.1 and 9.3, can be computed with this lemma. For instance, one can show that

$$\mathbb{E}\left[\pi_i \left\|x_i - \xi\right\|^2\right] = C_d \cdot \left[\frac{\nu^4}{(\nu^2+\sigma^2)^2} \left\|\xi - \mu\right\|^2 + \frac{\nu^2\sigma^2 d}{\nu^2+\sigma^2}\right] .$$

Proof. We use the fact that

$$\int x^2 \cdot \mathrm{e}^{-ax^2+bx+c} \,\mathrm{d}x = \frac{\sqrt{\pi}(2a+b^2)}{8a^{5/2}} \mathrm{e}^{\frac{b^2}{4a}+c} \cdot \operatorname{erf}\left(\frac{2ax-b}{2\sqrt{a}}\right) - \frac{ax+b}{4a^2} \cdot \mathrm{e}^{-ax^2+bx+c} .$$

□

11.2 Concentration results

In this section we collect some concentration results used throughout our proofs. Note that we use rather use the two-sided version of these results.

Theorem 11.1 (Hoeffding's inequality). *Let $X_1, \ldots, X_n$ be independent random variables such that X_i takes its values in $[a_i, b_i]$ almost surely for all $i \leqslant n$. Then for every $t > 0$,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}\left[X_i\right]) \geqslant t\right) \leqslant \exp\left(\frac{-2t^2n^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) .$$

Proof. This is Theorem 2.8 in Boucheron et al. (2013) in our notation. □

Theorem 11.2 (Hoeffding's inequality for sub-Gaussian random variables). *Let $X_1, \ldots, X_n$ be independent random variables such that X_i is sub-Gaussian with parameter $s^2 > 0$. Then, for every $t > 0$,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}\left[X_i\right] > t\right) \leqslant \exp\left(\frac{-nt^2}{2s^2}\right) .$$

Proof. This is Proposition 2.1 in Wainwright (2019). □