Appendix A. Full Statement of Main Theorem

We suppress some terms and constants in the statement of Theorem 2, so in this section we provide the full statement of our theorem.

Theorem 4 *Suppose that Assumptions 1-4 hold. Consider some* $\mathbf{x} \in \mathbb{R}^d$*. Assume*

$$p(\mathbf{x}) \geqslant \max\left\{2L_p bh, 8/(\omega_d b^d nh^d), \frac{4\pi^{d/2} R_K}{\Gamma(d/2) r_K^2} \cdot L_p h\right\}$$

and $d(\mathbf{x}, \partial\mathcal{S}) \geqslant bh$*. Define* $\Delta(\mathbf{x}) = |\sigma^2(\mathbf{x}) - \lambda|$*. Let* $\mathcal{R}(\mathbf{x})$ *be the Chow risk of Algorithm 1 and* ω_d *be volume of the unit ball in* $\mathbb{R}^d$*. If* $\sigma^2(\mathbf{x}) > \lambda$ *and*

$$\Delta(\mathbf{x}) \geqslant \frac{20R_K\sigma^2(\mathbf{x})}{ab^d\omega_d nh^d p(\mathbf{x})} + 5\mathsf{c}^{\sigma}_{cor.2}\frac{h^2}{p(\mathbf{x})}\left(1 + \frac{2R_K}{ab^d\omega_d nh^d p(\mathbf{x})}\right) - \frac{\sqrt{2}\lambda\|K\|_2 z_{1-\beta}}{\sqrt{nh^d\left(2p(\mathbf{x}) + L_p h\int_{\mathbb{R}^d}\|\mathbf{t}\|K(\mathbf{t})d\mu(\mathbf{t})\right)}},$$

$$\Delta(\mathbf{x}) \geqslant \frac{5L_f^2 h^2}{r_k^2}\log^2\frac{2e^2R_K}{ab^d\omega_d h^d p(\mathbf{x})} - \frac{\sqrt{2}\lambda\|K\|_2 z_{1-\beta}}{\sqrt{nh^d\left(2p(\mathbf{x}) + L_p h\int_{\mathbb{R}^d}\|\mathbf{t}\|K(\mathbf{t})d\mu(\mathbf{t})\right)}},$$

then the excess risk is at most

$$\mathbb{E}_{\mathcal{D}}[\mathcal{R}_\lambda(\mathbf{x}) - \mathcal{R}^*_\lambda(\mathbf{x})] \lesssim \left(\frac{\sigma^2(\mathbf{x}) + \frac{2L_{\sigma^2}h}{r_K}\log\frac{enR_K}{ar_K}}{nh^d p(\mathbf{x})} + \frac{h^4}{p^2(\mathbf{x})} + \{nh^{d-2}p(\mathbf{x})\}^{-1} + \Delta(\mathbf{x})\right)\cdot\mathbf{P}(\mathbf{x}),$$

$$\mathbf{P}(\mathbf{x}) \lesssim e^{-\Omega(nh^{d+2}p(\mathbf{x}))} + \exp\left(\frac{-\Omega(nh^d p(\mathbf{x})\cdot\delta(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2}h}{er_K}}\cdot\min\left\{\frac{\delta(\mathbf{x})/80}{\sigma^2(\mathbf{x}) + \mathsf{c}^{\sigma}_{cor.2}h}, \frac{1}{2}\right\}\right) + \exp\left(-\frac{\Omega(nh^{d-2}p(\mathbf{x})\delta^2(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2}h}{er_K}}\cdot\left([\mathsf{c}^f_{cor.2}]^2 + \frac{2L_f^2}{r_K^2}\log^2\frac{e^2R_K}{ab^d\omega_d h^d p(\mathbf{x})}\right)^{-1}\right),$$

$$\delta(\mathbf{x}) = \Delta(\mathbf{x}) + \lambda\|K\|_2 z_{1-\beta}\sqrt{\frac{2\omega_d^{-1}r_K^d/\log^d(n^2R_K)}{nh^d p(\mathbf{x})\cdot\left(1 + \frac{\log(n^2R_K)}{2br_K}\right)}}$$

for some constants $\mathsf{c}^f_{cor.2}$ *and* $\mathsf{c}^{\sigma}_{cor.2}$ *that do not depend on* $\mathbf{x}$ *as well as constants inside* $\lesssim$ *and* $\Omega(\cdot)$*. If, additionally, we have*

$$\Delta(\mathbf{x}) \geqslant \frac{20R_K\sigma^2(\mathbf{x})}{ab^d\omega_d nh^d p(\mathbf{x})} + 5\tilde{\mathsf{c}}^{\sigma}_{cor.2}h\left(1 + \frac{2R_K}{ab^d\omega_d nh^d p(\mathbf{x})}\right) - \frac{\sqrt{2}\lambda\|K\|_2 z_{1-\beta}}{\sqrt{nh^d\left(2p(\mathbf{x}) + L_p h\int_{\mathbb{R}^d}\|\mathbf{t}\|K(\mathbf{t})d\mu(\mathbf{t})\right)}},$$

where $\tilde{\mathsf{c}}^{\sigma}_{cor.2}$ *does not depend on* $\mathbf{x}, n, h$*, then*

$$\mathbf{P}(\mathbf{x}) \lesssim e^{-\Omega(nh^d p(\mathbf{x}))} + \exp\left(\frac{-\Omega(nh^d p(\mathbf{x}) \cdot \delta(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}} \cdot \min\left\{\frac{\delta(\mathbf{x})/80}{\sigma^2(\mathbf{x}) + \mathsf{c}^{\sigma}_{cor.2} h}, \frac{1}{2}\right\}\right) + \exp\left(-\frac{\Omega(nh^{d-2} p(\mathbf{x}) \delta^2(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}} \cdot \left([\mathsf{c}^{f}_{cor.2}]^2 + \frac{2L_f^2}{r_K^2} \log^2 \frac{e^2 R_K}{a b^d \omega_d h^d p(\mathbf{x})}\right)^{-1}\right)$$

If $\sigma^2(\mathbf{x}) \leqslant \lambda$ *and*

$$\Delta(\mathbf{x}) \geqslant \frac{20 R_K \sigma^2(\mathbf{x})}{a b^d \omega_d n h^d p(\mathbf{x})} + 5\mathsf{c}^{\sigma}_{cor.2} \frac{h^2}{p(\mathbf{x})}\left(1 + \frac{2R_K}{a b^d \omega_d n h^d p(\mathbf{x})}\right) + \frac{2\lambda (4\pi)^{-d/4} z_{1-\beta}}{\sqrt{a b^d \omega_d n h^d p(\mathbf{x})}},$$

$$\Delta(\mathbf{x}) \geqslant \frac{5 L_f^2 h^2}{r_k^2} \log^2 \frac{2e^2 R_K}{a b^d \omega_d h^d p(\mathbf{x})} + \frac{2\lambda (4\pi)^{-d/4} z_{1-\beta}}{\sqrt{a b^d \omega_d n h^d p(\mathbf{x})}},$$

then we have

$$\mathbb{E}_{\mathcal{D}}[\mathcal{R}(\mathbf{x}) - \mathcal{R}^*(\mathbf{x})] \lesssim \frac{\sigma^2(\mathbf{x}) + \frac{2L_{\sigma^2} h}{r_K} \log \frac{e n R_K}{a r_K}}{n h^d p(\mathbf{x})} + \frac{h^4}{p^2(\mathbf{x})} + \{n h^{d-2} p(\mathbf{x})\}^{-1} + \Delta(\mathbf{x}) \cdot \mathbf{P}'(\mathbf{x}),$$

$$\mathbf{P}'(\mathbf{x}) \lesssim e^{-\Omega(nh^{d+2} p(\mathbf{x}))} + \exp\left(\frac{-\Omega(nh^d p(\mathbf{x}) \cdot \delta'(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}} \cdot \min\left\{\frac{\delta'(\mathbf{x})/80}{\sigma^2(\mathbf{x}) + \mathsf{c}^{\sigma}_{cor.2} h}, \frac{1}{2}\right\}\right) + \exp\left(-\frac{\Omega(nh^{d-2} p(\mathbf{x}) \delta'^2(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}} \cdot \left([\mathsf{c}^{f}_{cor.2}]^2 + \frac{2L_f^2}{r_K^2} \log^2 \frac{e^2 R_K}{a b^d \omega_d h^d p(\mathbf{x})}\right)^{-1}\right),$$

$$\delta'(\mathbf{x}) = \Delta(\mathbf{x}) - \frac{2\lambda (4\pi)^{-d/4} z_{1-\beta}}{\sqrt{a b^d \omega_d n h^d p(\mathbf{x})}}.$$

where constants in $\lesssim$ *and* $\Omega(\cdot)$ *do not depend on* $\mathbf{x}$*. If, additionally, we have*

$$\Delta(\mathbf{x}) \geqslant \frac{20 R_K \sigma^2(\mathbf{x})}{a b^d \omega_d n h^d p(\mathbf{x})} + 5\tilde{\mathsf{c}}^{\sigma}_{cor.2} h \left(1 + \frac{2R_K}{a b^d \omega_d n h^d p(\mathbf{x})}\right) + \frac{2\lambda (4\pi)^{-d/4} z_{1-\beta}}{\sqrt{a b^d \omega_d n h^d p(\mathbf{x})}},$$

where $\tilde{\mathsf{c}}^{\sigma}_{cor.2}$ *does not depend on* $n, h, \mathbf{x}$*, then it holds that*

$$\mathbf{P}'(\mathbf{x}) \lesssim e^{-\Omega(nh^d p(\mathbf{x}))} + \exp\left(\frac{-\Omega(nh^d p(\mathbf{x}) \cdot \delta'(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}} \cdot \min\left\{\frac{\delta'(\mathbf{x})/80}{\sigma^2(\mathbf{x}) + \mathsf{c}^{\sigma}_{cor.2} h}, \frac{1}{2}\right\}\right) + \exp\left(-\frac{\Omega(nh^{d-2} p(\mathbf{x}) \delta'^2(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}} \cdot \left([\mathsf{c}^{f}_{cor.2}]^2 + \frac{2L_f^2}{r_K^2} \log^2 \frac{e^2 R_K}{a b^d \omega_d h^d p(\mathbf{x})}\right)^{-1}\right),$$

$$\delta'(\mathbf{x}) = \Delta(\mathbf{x}) - \frac{2\lambda (4\pi)^{-d/4} z_{1-\beta}}{\sqrt{a b^d \omega_d n h^d p(\mathbf{x})}}.$$

Finally, for any value of $\Delta(\mathbf{x})$*, we may bound*

$$\mathbb{E}_{\mathcal{D}}\mathcal{R}_{\lambda}(\mathbf{x}) - \mathcal{R}_{\lambda}^{*}(\mathbf{x}) \lesssim \frac{\sigma^2(\mathbf{x}) + \frac{2L_{\sigma^2}h}{r_K}\log\frac{enR_K}{ar_K}}{nh^d p(\mathbf{x})} + \frac{h^4}{p^2(\mathbf{x})} + \{nh^{d-2}p(\mathbf{x})\}^{-1} + \Delta(\mathbf{x}).$$

Appendix B. Proofs

B.1. Notation

Before we start our proves, we declare some notation we use:

- μ – Lebesgue's measure on $\mathbb{R}^d$;
- $\mathcal{B}_r(\mathbf{x})$ – the ball of the radius r and the center $\mathbf{x}$;
- $\boldsymbol{\omega}$ – weights of Nadaraya-Watson estimator, i.e.

$$\boldsymbol{\omega}_i = \frac{K\left(\frac{\mathbf{x}-X_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{\mathbf{x}-X_j}{h}\right)};$$

- $\boldsymbol{\sigma}^2$ – a vector that consists of $\sigma^2(X_i)$;
- $\mathbf{m}$ – a vector of means with respect to labels $Y_i \sim \mathcal{N}\left(f(X_i), \sigma^2(X_i)\right)$, i.e.

$$\mathbf{m}_i = f(X_i);$$

- $\mathbf{D}_{\mathbf{y}}$ – a diagonal matrix whose entries consists of vector $\mathbf{y}$'s elements;
- ω_d – the volume of a unit ball in $\mathbb{R}^d$.

B.2. Proof of Proposition 1

Fix an estimator $\widehat{a}(\mathbf{x})$. Then the risk

$$\begin{aligned}\mathcal{R}_{\lambda}(\mathbf{x}) &= \mathbb{E}\left[(Y - \widehat{f}(X))^2\,\mathrm{I}\{\widehat{\alpha}(X) = 0\} \mid X = \mathbf{x}\right] + \lambda\,\mathrm{I}\{\widehat{\alpha}(\mathbf{x}) = 1\} \\ &= \mathbb{E}\left[(Y - \widehat{f}(X))^2 \mid X = \mathbf{x}\right]\mathrm{I}\{\widehat{\alpha}(\mathbf{x}) = 0\} + \lambda\,\mathrm{I}\{\widehat{\alpha}(\mathbf{x}) = 1\}\end{aligned}$$

attains the minimum for $\widehat{f}(\mathbf{x}) = \mathbb{E}[Y \mid X = \mathbf{x}]$. For such $\widehat{f}(\mathbf{x})$ we have

$$\mathcal{R}_{\lambda}(\mathbf{x}) = \sigma^2(\mathbf{x})\,\mathrm{I}\{\widehat{\alpha}(\mathbf{x}) = 0\} + \lambda\,\mathrm{I}\{\widehat{\alpha}(\mathbf{x}) = 1\}.$$

Clearly, $\alpha(\mathbf{x})$ is the optimal reject function.

B.3. Proof of Proposition 3

Consider two cases. If $\sigma^2(\mathbf{x}) \geqslant \lambda$, then $\mathcal{R}^*_\lambda(\mathbf{x}) = \lambda$. Thus,

$$\begin{aligned}\mathbb{E}_{\mathcal{D}}\mathcal{R}_\lambda(\mathbf{x}) - \mathcal{R}^*_\lambda(\mathbf{x}) &= \mathbb{E}\left[(Y - \widehat{f}(\mathbf{x}))^2\,\mathrm{I}\left\{\widehat{\alpha}(\mathbf{x}) = 0\right\} \mid X = \mathbf{x}\right] + \lambda\mathbb{P}\left(\widehat{\alpha}(\mathbf{x}) = 1\right) - \lambda \\ &= \mathbb{E}\left[\mathbb{E}\left[(Y - \widehat{f}(\mathbf{x}))^2 \mid X = \mathbf{x}\right]\mathrm{I}\left\{\widehat{\alpha}(\mathbf{x}) = 0\right\}\right] - \lambda\mathbb{P}\left(\widehat{\alpha}(\mathbf{x}) = 0\right).\end{aligned}$$

Then

$$\begin{aligned}&\mathbb{E}\left[(Y - \widehat{f}(\mathbf{x}))^2 \mid X = \mathbf{x}\right] = \mathbb{E}\left[(Y - f(\mathbf{x}))^2 \mid X = \mathbf{x}\right] \\ &\quad + 2\mathbb{E}\left[(Y - f(\mathbf{x}))(\widehat{f}(\mathbf{x}) - f(\mathbf{x})) \mid X = \mathbf{x}\right] + \mathbb{E}\left[(\widehat{f}(\mathbf{x}) - f(\mathbf{x}))^2 \mid X = \mathbf{x}\right] \\ &= \sigma^2(\mathbf{x}) + (\widehat{f}(\mathbf{x}) - f(\mathbf{x}))^2.\end{aligned}$$

Thus,

$$\mathbb{E}_{\mathcal{D}}\mathcal{R}_\lambda(\mathbf{x}) - \mathcal{R}^*_\lambda(\mathbf{x}) = \mathbb{E}_{\mathcal{D}}\left[(\widehat{f}(\mathbf{x}) - f(\mathbf{x}))^2\,\mathrm{I}\left\{\widehat{\alpha}(\mathbf{x}) = 0\right\}\right] + \Delta(\mathbf{x}) \cdot \mathbb{P}\left(\widehat{\alpha}(\mathbf{x}) = 0\right).$$

Since $\alpha(\mathbf{x}) = 1$, the proposition holds for the case $\sigma^2(\mathbf{x}) \geqslant \lambda$. The case $\sigma^2(\mathbf{x}) < \lambda$ can be checked analogously.

B.4. Weights bounding

Proposition 5 *Under Assumptions 3-4 it holds that*

$$\mathbb{P}\left(\sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right) \leqslant \mathbf{z}_{prop.1}\right) \leqslant \exp\left(-\mathbf{z}_{prop.1}/26\right),$$
$$\mathbf{z}_{prop.1} = ab^d\omega_d \cdot p(\mathbf{x}) \cdot nh^d/2$$

for any $\mathbf{x}$ *such that* $p(\mathbf{x}) > 2L_p bh$ *and* $d(\mathbf{x}, \partial S) \geqslant bh$.

Proof From Assumption 3, we have

$$\sum_{i=1}^n K\left(\frac{\mathbf{x} - X_i}{h}\right) \geqslant \sum_{i=1}^n a\,\mathrm{I}\left\{\|\mathbf{x} - X_i\| \leqslant hb\right\}. \tag{5}$$

The right-hand side is a sum of Bernoulli random variables multiplied by a. The probability of one indicator is

$$\mathbb{P}(\|\mathbf{x} - X_i\| \leqslant hb) = \int_{\mathcal{B}_{hb}(\mathbf{x})} p(\mathbf{y})d\mu(\mathbf{y}) \leqslant (p(\mathbf{x}) + L_p bh) \cdot \mu\left(\mathcal{B}_{bh}(0)\right)$$

since $p(\cdot)$ is L_p-Lipschitz according to Assumption 4. Thus,

$$\mathrm{Var}\,\mathrm{I}\left\{\|\mathbf{x} - X_i\| \leqslant bh\right\} \leqslant \mathbb{E}\,\mathrm{I}\left\{\|\mathbf{x} - X_i\| \leqslant bh\right\} \leqslant \omega_d \cdot (p(\mathbf{x}) + L_p bh) \cdot (bh)^d,$$

where ω_d is the measure of a unit ball in $\mathbb{R}^d$. Analogously, if $d(\mathbf{x}, \partial S) \geqslant bh$ we have

$$\mathbb{P}\left(\|\mathbf{x} - X_i\| \leqslant hb\right) \geqslant \omega_d \cdot (bh)^d \cdot \left(p(\mathbf{x}) - L_p bh\right).$$

Applying the Bernstein inequality, we obtain

$$\begin{aligned}
\mathbb{P}\bigg(\sum_{i=1}^n a\,\mathrm{I}\left\{\|\mathbf{x} - X_i\| \leqslant hb\right\} &\leqslant \frac{1}{2} an\mathbb{P}(\|\mathbf{x} - X_i\| \leqslant hb)\bigg) \\
&\leqslant \exp\left(-\frac{1}{8} \cdot \frac{a^2 n^2 (bh)^{2d} \omega_d^2 (p(\mathbf{x}) - L_p bh)^2}{an\omega_d \cdot (bh)^d \cdot (p(\mathbf{x}) + L_p bh) + \frac{1}{6} an\omega_d \cdot (p(\mathbf{x}) - L_p bh)(bh)^d}\right) \\
&\leqslant \exp\left(-\frac{1}{8} \cdot \frac{ab^d \omega_d \cdot (p(\mathbf{x}) - L_p bh) nh^d}{\frac{1}{6} + (p(\mathbf{x}) + L_p bh)/(p(\mathbf{x}) - L_p bh)}\right). \qquad (6)
\end{aligned}$$

Since $p(\mathbf{x}) \geqslant 2L_p bh$, we have

$$\begin{aligned}
&p(\mathbf{x}) - L_p bh \geqslant p(\mathbf{x})/2, \\
&\frac{p(\mathbf{x}) + L_p bh}{p(\mathbf{x}) - L_p bh} = 1 + \frac{2L_p bh}{p(\mathbf{x}) - L_p bh} \leqslant 3,
\end{aligned}$$

and we can bound (6) by $e^{-\mathbf{z}_{prop.1}/26}$. Combining it with (5), we obtain the proposition. ■

From the proposition, the following corollary follows:

Corollary 6 *Under Assumptions 3-4 it holds that*

$$\mathbb{P}\left(\max_i g(X_i)\boldsymbol{\omega}_i \geqslant g(\mathbf{x})\frac{R_K}{\mathbf{z}_{prop.1}} + \frac{LR_K h}{er_K \mathbf{z}_{prop.1}}\right) \leqslant \exp\left(-\mathbf{z}_{prop.1}/26\right)$$

simultaneously for L-Lipschitz g and any $\mathbf{x}$ such that $p(\mathbf{x}) > 2L_p bh$ and $d(\mathbf{x}, \partial S) \geqslant bh$.

Proof Since $g(\mathbf{x})$ is L-Lipschitz, we may state

$$g(X_i) \leqslant g(\mathbf{x}) + L\|X_i - \mathbf{x}\|.$$

Thus, $\max_i g(X_i)\boldsymbol{\omega}_i \leqslant g(\mathbf{x}) \max_i \boldsymbol{\omega}_i + L \max_i \|X_i - \mathbf{x}\|\boldsymbol{\omega}_i$. We may bound

$$\begin{aligned}
&\max_i \boldsymbol{\omega}_i \leqslant \frac{K\left(\frac{X_i - \mathbf{x}}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \leqslant \frac{R_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)}, \\
&\max_i \|X_i - \mathbf{x}\|\boldsymbol{\omega}_i \leqslant \frac{\max_i \|X_i - \mathbf{x}\| K\left(\frac{X_i - \mathbf{x}}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \\
&\leqslant \frac{R_K h \max_i \left\|\frac{X_i - \mathbf{x}}{h}\right\| e^{-r_K \frac{X_i - \mathbf{x}}{h}}}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \leqslant \frac{R_K h}{er_K \sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)}.
\end{aligned}$$

Thus,

$$\mathbb{P}\left(\max_i g(X_i)\boldsymbol{\omega}_i \geqslant g(\mathbf{x})\frac{2R_K}{\mathbf{z}_{prop.1}} + \frac{LR_K h}{er_K \mathbf{z}_{prop.1}}\right) \leqslant \mathbb{P}\left(\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right) \leqslant \mathbf{z}_{prop.1}\right)$$
$$\leqslant \exp\left(-\mathbf{z}_{prop.1}/26\right)$$

due to Proposition 5. ∎

B.5. Deviation of estimated noiseless mean

Before the main lemma of this section we introduce a simple auxiliary proposition:

Proposition 7 *Suppose that for a kernel $K\colon \mathbb{R}^d \to \mathbb{R}_+$ Assumption 3 holds. Then, we have*

$$\max_{\mathbf{t}\in\mathbb{R}^d} \|\mathbf{t}\|^m K(\mathbf{t}) \leqslant R_K \left(\frac{m}{r_K}\right)^m e^{-m},$$
$$\int_{\mathbf{t}\in\mathbb{R}^d} \|\mathbf{t}\|^m K^k(\mathbf{t}) d\mu(\mathbf{t}) \leqslant \frac{2\pi^{d/2} R_K^k \cdot m!}{\Gamma(d/2)(r_K k)^{m+1}}$$

for any non-negative integers k, m.

Proof From Assumption 3 we have $K(\mathbf{t}) \leqslant R_K e^{-r_K\|\mathbf{t}\|}$. The first inequality is obtained via maximizing $\|\mathbf{t}\|^m R_K e^{-r_K\|\mathbf{t}\|}$, the second one from calculations

$$\int_{\mathbb{R}^d} \|\mathbf{t}\|^m K^k(\mathbf{t}) d\mu(\mathbf{t}) \leqslant \int_{\mathbb{R}^d} \|\mathbf{t}\|^m R_K^k e^{-r_K k\|\mathbf{t}\|} d\mu(\mathbf{t}) = R_K^k \cdot \mu(\mathbb{S}^{d-1}) \int_0^{+\infty} \rho^m e^{-r_K \rho k} d\rho$$
$$= R_k^k \cdot \frac{2\pi^{d/2}}{\Gamma(d/2)} \cdot \frac{\Gamma(m+1)}{(r_K k)^{m+1}} = \frac{2\pi^{d/2} R_K^k \cdot m!}{\Gamma(d/2)(r_K k)^{m+1}},$$

where $\mathbb{S}^{d-1}$ stands for a $(d-1)$-dimensional sphere. ∎

Proposition 8 *Assume that a kernel $K(\cdot)$ and $p(\cdot)$ satisfy Assumption 3 and Assumption 4 respectively. Let a function $g(\cdot)$ be twice differential with the Hessian bounded by H in the spectral norm and the gradient bounded by L. Finally, let $X_1, \ldots, X_n \sim p(\cdot)$ be identically independently distributed random variables. Then*

$$\mathbb{P}\left(\left|\sum_{i=1}^n \boldsymbol{\omega}_i g(X_i) - g(\mathbf{x})\right| \geqslant t\right) \leqslant 2\exp\left(-\frac{1}{2} \cdot \frac{nh^d p(\mathbf{x}) \cdot \mathtt{r}_1}{\mathtt{r}_2/\mathtt{r}_1 + \mathtt{r}_3/3}\right),$$

if $\mathtt{r}_1 > 0$ *where*

$$\mathtt{r}_1 = \left\{1 - \frac{L_p h}{p(\mathbf{x})} \cdot \frac{2\pi^{d/2} R_K}{\Gamma(d/2) r_K^2}\right\} \cdot t - \frac{h^2}{p(\mathbf{x})} \left\{\frac{2LL_p + HC_p}{2} \cdot \frac{4\pi^{d/2} R_K}{\Gamma(d/2) r_K^3}\right\}, \tag{7}$$

$$\mathtt{r}_2 = \left\{\frac{2\pi^{d/2} R_K^2}{\Gamma(d/2) r_k} + \frac{2hL_p}{p(\mathbf{x})} \cdot \frac{4\pi^{d/2} R_k}{\Gamma(d/2) r_k^2}\right\} \cdot t^2 + \left\{2L^2 \frac{\pi^{d/2} R_K^2}{2\Gamma(d/2) r_K^3} + \frac{2hL^2 L_p}{p(\mathbf{x})} \cdot \frac{4\pi^{d/2} R_K^2}{3\Gamma(d/2) r_K^4}\right\} \cdot h^2, \tag{8}$$

$$\mathtt{r}_3 = \frac{LhR_K}{er_K} + tR_K. \tag{9}$$

Proof We analyze the probability via bounding

$$\mathbb{P}\left(\left|\sum_{i=1}^n \boldsymbol{\omega}_i g(X_i) - g(\mathbf{x})\right| \geqslant t\right) \leqslant \mathbb{P}\left(\sum_{i=1}^n \boldsymbol{\omega}_i g(X_i) - g(\mathbf{x}) \geqslant t\right) + \mathbb{P}\left(\sum_{i=1}^n \boldsymbol{\omega}_i g(X_i) - g(\mathbf{x}) \leqslant -t\right).$$

We consider only the first term, the second one can be processed analogously. By rearranging terms, the problem reformulates as the bounding the probability of

$$\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)(g(X_i) - g(\mathbf{x}) - t) \geqslant 0.$$

The expectation of the above is

$$\mathbb{E}\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)(g(X_i) - g(\mathbf{x}) - t) = n\mathbb{E}\left(g(X_i) - g(\mathbf{x})\right) K\left(\frac{X_i - \mathbf{x}}{h}\right) - nt\cdot\mathbb{E}K\left(\frac{X_i - \mathbf{x}}{h}\right).$$

Meanwhile,

$$\begin{aligned}
&\left|\mathbb{E}\left(g(X_i) - g(\mathbf{x})\right) K\left(\frac{X_i - \mathbf{x}}{h}\right)\right| = \left|\int_{\mathbb{R}^d} (g(\mathbf{y}) - g(\mathbf{x})) K\left(\frac{\mathbf{y} - \mathbf{x}}{h}\right) p(\mathbf{y}) d\mu(\mathbf{y})\right| \\
&\leqslant \left|\int_{\mathbb{R}^d} \langle \nabla g(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle K\left(\frac{\mathbf{y} - \mathbf{x}}{h}\right) p(\mathbf{y}) d\mu(\mathbf{y})\right| + \frac{H}{2}\int_{\mathbb{R}^d} \|\mathbf{y} - \mathbf{x}\|^2 K\left(\frac{\mathbf{x} - \mathbf{y}}{h}\right) p(\mathbf{y}) d\mu(\mathbf{y})
\end{aligned}$$

since

$$g(\mathbf{x}) + \langle \nabla g(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle - \frac{H}{2}\|\mathbf{y} - \mathbf{x}\|^2 \leqslant g(\mathbf{y}) \leqslant g(\mathbf{x}) + \langle \nabla g(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \frac{H}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$

At the same time, we have

$$\langle \nabla g(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle K\left(\frac{\mathbf{y} - \mathbf{x}}{h}\right) p(\mathbf{y}) \leqslant p(\mathbf{x})\langle \nabla g(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle K\left(\frac{\mathbf{y} - \mathbf{x}}{h}\right) + LL_p\|\mathbf{y} - \mathbf{x}\|^2 K\left(\frac{\mathbf{y} - \mathbf{x}}{h}\right)$$

and

$$\langle \nabla g(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle K\left(\frac{\mathbf{y} - \mathbf{x}}{h}\right) p(\mathbf{y}) \geqslant p(\mathbf{x})\langle \nabla g(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle K\left(\frac{\mathbf{y} - \mathbf{x}}{h}\right) - LL_p\|\mathbf{y} - \mathbf{x}\|^2 K\left(\frac{\mathbf{y} - \mathbf{x}}{h}\right).$$

Bounding $p(\mathbf{y}) \leqslant C_p$, we obtain

$$\begin{aligned}\left|\mathbb{E}\left(g(X_i)-g(\mathbf{x})\right)K\left(\frac{X_i-\mathbf{x}}{h}\right)\right| \leqslant p(\mathbf{x})\int_{\mathbb{R}^d}\langle\nabla g(\mathbf{x}),\mathbf{y}-\mathbf{x}\rangle K\left(\frac{\mathbf{y}-\mathbf{x}}{h}\right)d\mu(\mathbf{y})\\ + LL_p\int_{\mathbb{R}^d}\|\mathbf{y}-\mathbf{x}\|^2K\left(\frac{\mathbf{y}-\mathbf{x}}{h}\right)d\mu(\mathbf{y})\\ + \frac{HC_p}{2}\int_{\mathbb{R}^d}\|\mathbf{y}-\mathbf{x}\|^2K\left(\frac{\mathbf{y}-\mathbf{x}}{h}\right)d\mu(\mathbf{y}).\end{aligned}$$

Using

$$\int_{\mathbb{R}^d}\langle\nabla g(\mathbf{x}),\mathbf{y}-\mathbf{x}\rangle K\left(\frac{\mathbf{y}-\mathbf{x}}{h}\right)d\mu(\mathbf{y}) = 0$$

from Assumption 3, we get

$$\left|\mathbb{E}\left(g(X_i)-g(\mathbf{x})\right)K\left(\frac{X_i-\mathbf{x}}{h}\right)\right| \leqslant LL_p\int_{\mathbb{R}^d}\|\mathbf{y}-\mathbf{x}\|^2K\left(\frac{\mathbf{y}-\mathbf{x}}{h}\right)d\mu(\mathbf{y}).$$

Finally, changing variables leads us to

$$\int_{\mathbb{R}^d}\|\mathbf{y}-\mathbf{x}\|^2K\left(\frac{\mathbf{y}-\mathbf{x}}{h}\right)d\mu(\mathbf{y}) = h^d\cdot h^2\int_{\mathbb{R}^d}\|\mathbf{t}\|^2K(\mathbf{t})d\mu(\mathbf{t}).$$

At the same time, we have

$$\begin{aligned}\mathbb{E}K\left(\frac{X_i-\mathbf{x}}{h}\right) &= h^d\int_{\mathbb{R}^d}p(\mathbf{x}+\mathbf{t}h)K(\mathbf{t})d\mu(\mathbf{t})\\ &\geqslant h^d\left(p(\mathbf{x})-L_ph\int_{\mathbb{R}^d}\|\mathbf{t}\|K(\mathbf{t})d\mu(\mathbf{t})\right).\end{aligned}$$

Consequently,

$$\left|\mathbb{E}\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)(g(X_i)-g(\mathbf{x})-t)\right| \geqslant nh^d\left(\mathsf{c}_t t-\mathsf{c}_{h^2}h^2\right), \tag{10}$$

where

$$\begin{aligned}\mathsf{c}_t &= p(\mathbf{x})-L_ph\int_{\mathbb{R}^d}\|\mathbf{t}\|K(\mathbf{t})d\mu(\mathbf{t}),\\ \mathsf{c}_{h^2} &= \frac{2LL_p+HC_p}{2}\int_{\mathbb{R}^d}\|\mathbf{t}\|^2K(\mathbf{t})\,d\mu(\mathbf{t}).\end{aligned}$$

Next, we bound

$$\begin{aligned}\operatorname{Var}K\left(\frac{X_i-\mathbf{x}}{h}\right)\{g(X_i)-g(\mathbf{x})-t\} &\leqslant \mathbb{E}K^2\left(\frac{X_i-\mathbf{x}}{h}\right)\{g(X_i)-g(\mathbf{x})-t\}^2\\ &\leqslant 2\mathbb{E}K^2\left(\frac{X_i-\mathbf{x}}{h}\right)\{g(X_i)-g(\mathbf{x})\}^2+2t^2\mathbb{E}K^2\left(\frac{X_i-\mathbf{x}}{h}\right)\\ &\leqslant 2L^2\mathbb{E}K^2\left(\frac{X_i-\mathbf{x}}{h}\right)\|X_i-\mathbf{x}\|^2+2t^2\mathbb{E}K^2\left(\frac{X_i-\mathbf{x}}{h}\right).\end{aligned}$$

As previously,

$$\mathbb{E}K^2\left(\frac{X_i - \mathbf{x}}{h}\right) = \int_{\mathbb{R}^d} K\left(\frac{\mathbf{y}-\mathbf{x}}{h}\right) p(\mathbf{y}) d\mu(\mathbf{y})$$
$$\leqslant h^d p(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mu(\mathbf{t}) + h^{d+1} L_p \int_{\mathbb{R}^d} \|\mathbf{t}\| K(\mathbf{t}) d\mu(\mathbf{t})$$

and

$$\mathbb{E}K^2\left(\frac{X_i - \mathbf{x}}{h}\right)\|X_i - \mathbf{x}\|^2 \leqslant h^{d+2} p(\mathbf{x}) \int_{\mathbb{R}^d} \|\mathbf{t}\|^2 K^2(\mathbf{t}) d\mu(\mathbf{t}) + h^{d+3} L_p \int_{\mathbb{R}^d} \|\mathbf{t}\|^3 K^2(\mathbf{t}) d\mu(\mathbf{t}).$$

Consequently,

$$\operatorname{Var} K\left(\frac{X_i - \mathbf{x}}{h}\right)\{g(X_i) - g(\mathbf{x}) - t\} \leqslant h^d \left\{\mathsf{c}'_t t^2 + \mathsf{c}'_{h^2} h^2\right\} \tag{11}$$

where

$$\mathsf{c}'_t = 2p(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mu(\mathbf{t}) + 2hL_p \int_{\mathbb{R}^d} \|\mathbf{t}\| K(\mathbf{t}) d\mu(\mathbf{t}),$$
$$\mathsf{c}'_{h^2} = 2L^2 p(\mathbf{x}) \int_{\mathbb{R}^d} \|\mathbf{t}\|^2 K^2(\mathbf{t}) d\mu(\mathbf{t}) + 2hL^2 L_p \int_{\mathbb{R}^d} \|\mathbf{t}\|^3 K^2(\mathbf{t}) d\mu(\mathbf{t}).$$

Finally,

$$\left| K\left(\frac{X_1 - \mathbf{x}}{h}\right)(g(X_i) - g(\mathbf{x}) - t)\right| \leqslant K\left(\frac{X_1 - \mathbf{x}}{h}\right)(|g(X_i) - g(\mathbf{x})| + t)$$
$$\leqslant Lh \max_{\mathbf{t}\in\mathbb{R}^d} \|\mathbf{t}\| K(\mathbf{t}) + t \max_{\mathbf{t}\in\mathbb{R}^d} K(\mathbf{t}). \tag{12}$$

Define

$$\xi_i = K\left(\frac{X_i - \mathbf{x}}{h}\right)(g(X_i) - g(\mathbf{x}) - t),$$

then the probability from the statement can be bounded as

$$\mathbb{P}\left(\sum_{i=1}^n \xi_i \geqslant 0\right) = \mathbb{P}\left(\sum_{i=1}^n \xi_i - \sum_{i=1}^n \mathbb{E}\xi_i \geqslant -\sum_{i=1}^n \mathbb{E}\xi_i\right).$$

If $\mathsf{c}_t t > \mathsf{c}_{h^2} h^2$ then $\mathbb{E}\xi_i$ is negative due to inequality (10), and, whence, the above can be bounded via the Bernstein inequality:

$$\mathbb{P}\left(\sum_{i=1}^n \xi_i - \sum_{i=1}^n \mathbb{E}\xi_i \geqslant -\sum_{i=1}^n \mathbb{E}\xi_i\right) \leqslant \exp\left(-\frac{1}{2}\cdot\frac{(n\mathbb{E}\xi_1)^2}{n\operatorname{Var}\xi_1 + n|\mathbb{E}\xi_1|/3 \cdot \operatorname{ess\,sup}_{X_1}|\xi_1|}\right). \tag{13}$$

Substituting bounds (10), (11), (12) instead of $\mathbb{E}\xi_1$, $\operatorname{Var}\xi_1$ and $\operatorname{ess\,sup}_{X_1}|\xi_1|$ respectively, we obtain

$$(13) \leqslant \exp\left(-\frac{1}{2}\cdot\frac{n^2 h^{2d}(\mathsf{c}_t t - \mathsf{c}_{h^2} h^2)^2}{nh^d\{\mathsf{c}'_t t^2 + \mathsf{c}'_{h^2} h^2\} + \frac{nh^d}{3}(\mathsf{c}_t t - \mathsf{c}_{h^2} h^2)\{Lh \max_{\mathbf{t}\in\mathbb{R}^d}\|\mathbf{t}\|K(\mathbf{t}) + t\max_{\mathbf{t}\in\mathbb{R}^d} K(\mathbf{t})\}}\right)$$
$$\leqslant \exp\left(-\frac{1}{2}\cdot\frac{nh^d p(\mathbf{x})\cdot \mathtt{r}_1}{\mathtt{r}_2/\mathtt{r}_1 + \mathtt{r}_3/3}\right),$$

where

$$\mathtt{r}_1 = \frac{\mathtt{c}_t t - \mathtt{c}_{h^2} h^2}{p(\mathbf{x})}, \quad \mathtt{r}_2 = \frac{\mathtt{c}'_t t^2 + \mathtt{c}'_{h^2} h^2}{p(\mathbf{x})}, \quad \mathtt{r}_3 = Lh \max_{\mathbf{t} \in \mathbb{R}^d} \|\mathbf{t}\| K(\mathbf{t}) + t \max_{\mathbf{t} \in \mathbb{R}^d} K(\mathbf{t}).$$

Replacing integrals and maxima in the above with their bounds from Proposition 7, we obtain the statement of the proposition. ■

In most of the cases, it is sufficient to use the simplified version of the proposition.

Corollary 9 *Assume that a kernel $K(\cdot)$ and $p(\cdot)$ satisfy Assumption 3 and Assumption 4 respectively. Let a function $g(\cdot)$ be twice differential with the Hessian bounded by H in the spectral norm and the gradient bounded by L. Finally, let $X_1, \ldots, X_n \sim p(\cdot)$ be identically independently distributed random variables. If*

1. $p(\mathbf{x}) \geqslant 2L_p h \cdot \max\{b, \frac{2\pi^{d/2} R_K}{\Gamma(d/2) r_K^2}\}$,
2. *and*

$$t > \frac{\mathtt{c}_{cor.2} h^2}{p(\mathbf{x})}, \quad \mathtt{c}_{cor.2} = \left\{ (2LL_p + HC_p) \frac{4\pi^{d/2} R_K}{\Gamma(d/2) r_K^3} \right\},$$

then

$$\mathbb{P}\left(\left| \sum_{i=1}^n \boldsymbol{\omega}_i g(X_i) - g(\mathbf{x}) \right| \geqslant t \right) \leqslant 2 \exp\left(-Cnh^{d+2} p(\mathbf{x}) \right),$$

for some constant C that does not depend on $\mathbf{x}$. Additionally, if $t > \tilde{\mathtt{c}}_{cor.2} h$, where

$$\tilde{\mathtt{c}}_{cor.2} = \frac{1}{2L_p b} \left\{ (2LL_p + HC_p) \frac{4\pi^{d/2} R_K}{\Gamma(d/2) r_K^3} \right\},$$

when

$$\mathbb{P}\left(\left| \sum_{i=1}^n \boldsymbol{\omega}_i g(X_i) - g(\mathbf{x}) \right| \geqslant t \right) \leqslant 2 \exp\left(-C' nh^d p(\mathbf{x}) \right),$$

for some constant C' that does not depend on $\mathbf{x}$.

Proof Consider the definition (7) of $\mathtt{r}_1$ in Proposition 8. First, we analyze the coefficient of t. Since $p(\mathbf{x}) \geqslant 2L_p h \cdot \frac{2\pi^{d/2} R_K}{\Gamma(d/2) r_K^2}$, we have:

$$1 - \frac{L_p h}{p(\mathbf{x})} \cdot \frac{2\pi^{d/2} R_K}{\Gamma(d/2) r_K^2} \geqslant \frac{1}{2}.$$

Thus, we may state that $\mathtt{r}_1 \geqslant \frac{1}{2}(t^2 - \mathtt{c}_1 h^2 / p(\mathbf{x}))$ where $\mathtt{c}_1$ does not depend on $\mathbf{x}$.

Next, we bound coefficients of $\mathtt{r}_2$ defined by (8). The first condition ensures that

$$\frac{2\pi^{d/2}R_K^2}{\Gamma(d/2)r_k} + \frac{2hL_p}{p(\mathbf{x})}\cdot\frac{4\pi^{d/2}R_k}{\Gamma(d/2)r_k^2} \leqslant \frac{2\pi^{d/2}R_K^2}{\Gamma(d/2)r_k}\left(1+\frac{2b}{r_K}\right),$$
$$2L^2\frac{\pi^{d/2}R_K^2}{2\Gamma(d/2)r_K^3} + \frac{2hL^2L_p}{p(\mathbf{x})}\cdot\frac{4\pi^{d/2}R_K^2}{3\Gamma(d/2)r_K^4} \leqslant L^2\frac{\pi^{d/2}R_K^2}{\Gamma(d/2)r_K^3}\left(1+\frac{4b}{3r_K}\right).$$

Consequently, $\mathtt{r}_2$ is at most $\mathtt{c}_2 t^2 + \mathtt{c}_3 h^2$ for two constants $\mathtt{c}_2$ and $\mathtt{c}_3$ that do not depend on $\mathbf{x}$. Clearly, $\mathtt{r}_3 = \mathtt{c}_4 t + \mathtt{c}_5 h$ where $\mathtt{c}_4$ and $\mathtt{c}_5$ do not depend on $\mathbf{x}$ too. Bounding $\mathtt{r}_3 \leqslant t$, we obtain

$$\frac{\mathtt{r}_1}{\mathtt{r}_2/\mathtt{r}_1 + \mathtt{r}_3/3} \geqslant \frac{1}{4}\frac{(t - \mathtt{c}_1 h^2/p(\mathbf{x}))^2}{\mathtt{c}_2 t^2 + \mathtt{c}_3 h^2 + (\mathtt{c}_4 t^2 + \mathtt{c}_5 th)/3}.$$

The third condition of the corollary guarantees that $t \geqslant 2\mathtt{c}_1 h^2/p(\mathbf{x})$, so the right-hand side of the above can not be zero. It is bounded below by

$$\frac{1}{16}\frac{t^2}{\mathtt{c}_2 t^2 + \mathtt{c}_3 h^2 + (\mathtt{c}_4 t^2 + \mathtt{c}_5 th)/3} \geqslant Ch^2,$$

since $t \geqslant 2\mathtt{c}_1 h^2/p(\mathbf{x})$ and h is bounded by $C_p/(2L_p b)$. Applying Proposition 8, we obtain the first part of the corollary. If $t > 2\tilde{\mathtt{c}}_{cor.2} h$, then we have

$$\frac{1}{16}\frac{t^2}{\mathtt{c}_2 t^2 + \mathtt{c}_3 h^2 + (\mathtt{c}_4 t^2 + \mathtt{c}_5 th)/3} \geqslant C',$$

and the second part of the corollary follows. ■

While the above provides probabilistic bound, we also requires deterministic bound:

Proposition 10 *Suppose a function g is L-Lipschitz and for a kernel $K(\cdot)$ Assumption 3 holds. Then*

$$\sum_{i=1}^n |g(X_i) - g(\mathbf{x})|^s \boldsymbol{\omega}_i \leqslant \frac{2L^s h^s}{r_K^s}\left(s \vee \log\frac{nR_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)}\right)^s.$$

Proof The proof is straightforward. We start with

$$\sum_{i=1}^n |g(X_i) - g(\mathbf{x})|^s \boldsymbol{\omega}_i \leqslant L^s \sum_{i=1}^n \|X_i - \mathbf{x}\|^s \boldsymbol{\omega}_i = \frac{L^s \sum_{i=1}^n \|X_i - \mathbf{x}\|^s K\left(\frac{X_i - \mathbf{x}}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)}.$$

Then impose some parameter $t_0 \geqslant \frac{s}{r_K}$. Such a restriction guarantees that $t^s e^{-r_K t} \leqslant t_0^s e^{-R_K t_0}$ for any $t \geqslant t_0$. Consider

$$\sum_{i=1}^n \|X_i - \mathbf{x}\|^s K\left(\frac{X_i - \mathbf{x}}{h}\right) = \sum_{i | X_i \in \mathcal{B}_{ht_0}(\mathbf{x})} \|X_i - \mathbf{x}\|^s K\left(\frac{X_i - \mathbf{x}}{h}\right) + \sum_{i | X_i \notin \mathcal{B}_{ht_0}(\mathbf{x})} \|X_i - \mathbf{x}\|^s K\left(\frac{X_i - \mathbf{x}}{h}\right)$$
$$\leqslant h^s t_0^s \sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right) + nR_k h^s t_0^s e^{-r_K t_0} = h^s t_0^s \left(\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right) + nR_k e^{-r_K t_0}\right).$$

If $\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right) > nR_K e^{-s}$, set $t_0 = s/r_K$. Then

$$\frac{L^s \sum_{i=1}^n \|X_i-\mathbf{x}\|^s K\left(\frac{X_i-\mathbf{x}}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)} \leqslant \frac{L^s h^s t_0^s \cdot 2\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)} \leqslant 2L^s h^s s^s / r_K^s.$$

Otherwise choose t_0 such that

$$nR_K e^{-r_K t_0} = \sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right).$$

Then

$$\frac{L^s \sum_{i=1}^n \|X_i-\mathbf{x}\|^s K\left(\frac{X_i-\mathbf{x}}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)} \leqslant 2L^s h^s t_0^s$$

for

$$t_0 = \frac{1}{r_K}\log\frac{nR_K}{\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)}.$$

Thus, the statement holds. ■

B.6. Estimation of variance

In this section we establish the concentration properties of the estimator

$$\begin{aligned}\widehat{\sigma}_n^2(\mathbf{x}) &= \mathbf{Y}^{\mathrm{T}}\mathbf{D}_{\boldsymbol{\omega}}\mathbf{Y} - \mathbf{Y}^{\mathrm{T}}\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}\mathbf{Y} \\ &= (\mathbf{Y}-\mathbf{m})^{\mathrm{T}}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\mathbf{Y}-\mathbf{m}) \qquad (14)\\ &\quad + 2(\mathbf{Y}-\mathbf{m})^{\mathrm{T}}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\mathbf{m} \qquad (15)\\ &\quad + \mathbf{m}^{\mathrm{T}}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\mathbf{m}. \qquad (16)\end{aligned}$$

We estimate each term separately. First, we bound deviations of term 14. Define n independent random variables $Z_i \sim \mathcal{N}(0,1)$. Further, we will show that

$$(\mathbf{Y}-\mathbf{m})^{\mathrm{T}}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\mathbf{Y}-\mathbf{m}) \overset{d}{=} \sum_{i=1}^n \lambda_i(\Sigma) Z_i^2,$$

where $\Sigma = \mathbf{D}_{\boldsymbol{\sigma}^2}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})$. The following proposition allows to establish large deviations inequality:

Proposition 11 *For any* $\mathbf{x}$ *and* $t > 0$*, we have*

$$\mathbb{P}\left(\left|(\mathbf{Y}-\mathbf{m})^{\mathrm{T}}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\mathbf{Y}-\mathbf{m}) - \mathrm{Tr}(\Sigma)\right| \geqslant 16\max\{\sqrt{\mathrm{Tr}(\Sigma^2)t}, \|\Sigma\|t\}\right) \leqslant 2e^{-t},$$

where $\Sigma = \mathbf{D}_{\boldsymbol{\sigma}^2}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})$*.*

Proof Since term (14) is a quadratic form of Gaussian vector, it admits the representation

$$(14) = \sum_{i=1}^{n} \lambda_i(\Sigma') Z_i^2,$$

where Z_i are independent random variables from standard normal distribution and

$$\Sigma' = \mathbf{D}_{\boldsymbol{\sigma}^2}^{1/2}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\mathbf{D}_{\boldsymbol{\sigma}^2}^{1/2}.$$

Meanwhile, the non-zero part of the spectrum of arbitrary matrix product $\mathbf{AB}$ coincides with the one of $\mathbf{BA}$. Consequently, the mean of (14) is equal to $\sum_{i=1}^{n} \lambda_i(\Sigma') = \mathrm{Tr}(\Sigma)$, $\max_i |\lambda_i(\Sigma')| = \max_i |\lambda_i(\Sigma)| = \|\Sigma\|$ and $\mathrm{Tr}((\Sigma')^2) = \mathrm{Tr}(\Sigma^2)$. Thus, applying standard inequality for sum of sub-exponential random variables (see, for example, (Wainwright, 2019)), we have:

$$\mathbb{P}\left(\left|(\mathbf{Y}-\mathbf{m})^{\mathrm{T}}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\mathbf{Y}-\mathbf{m}) - \mathrm{Tr}(\Sigma)\right| \geqslant r \mid X\right) \leqslant 2 \cdot \begin{cases} e^{-\frac{r^2}{16\,\mathrm{Tr}(\Sigma^2)}}, & \text{if } 0 \leqslant r \leqslant \frac{\mathrm{Tr}(\Sigma^2)}{\|\Sigma\|}, \\ e^{-\frac{r}{16\|\Sigma\|}} & \text{if } r > \frac{\mathrm{Tr}(\Sigma^2)}{\|\Sigma\|}. \end{cases}$$

Substituting r with $16\max\{\sqrt{\mathrm{Tr}(\Sigma^2)t}, \|\Sigma\|t\}$, we obtain the statement of the proposition. ■

The above propositions allows to establish precise large deviation bounds for term 14. Let $\mathsf{c}_{cor.2}^{f}, \mathsf{c}_{cor.2}^{\sigma}$ be constants obtained from Corollary 9 applied for functions $\sigma^2(\cdot)$ and $f(\cdot)$ as $\mathsf{c}_{cor.2}$:

$$\mathsf{c}_{cor.2}^{f} = \frac{1}{L_p b}\left\{(2L_f L_p + H_f C_p)\frac{2\pi^{d/2} R_K}{\Gamma(d/2) r_K^3}\right\},$$

$$\mathsf{c}_{cor.2}^{\sigma} = \frac{1}{L_p b}\left\{(2L_f L_p + H_{\sigma^2} C_p)\frac{2\pi^{d/2} R_K}{\Gamma(d/2) r_K^3}\right\}.$$

Analogously, we define constants $\tilde{\mathsf{c}}_{cor.2}^{f}, \tilde{\mathsf{c}}_{cor.2}^{\sigma}$.

Lemma 12 *Suppose Assumptions 3-4 hold. Assume that*

1. *for the density $p(\mathbf{x})$ we have $p(\mathbf{x}) \geqslant 2L_p h \cdot \max\{b, \frac{2\pi^{d/2} R_K}{\Gamma(d/2) r_K^2}\}$,*
2. *the Euclidean distance $d(\mathbf{x}, \partial\mathcal{S})$ from $\mathbf{x}$ to the bound of $\mathcal{S}$ is at least bh,*
3. *for a function $\delta(\mathbf{x})$ we have:*

$$\delta(\mathbf{x}) \geqslant \frac{10 R_K \sigma^2(\mathbf{x})}{\mathsf{z}_{prop.1}} + \frac{5\mathsf{c}_{cor.2}^{\sigma} h^2}{p(\mathbf{x})}\left(1 + \frac{R_K}{\mathsf{z}_{prop.1}}\right).$$

When

$$\mathbb{P}\left(\left|(\mathbf{Y}-\mathbf{m})^{\mathrm{T}}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\mathbf{Y}-\mathbf{m}) - \sigma^2(\mathbf{x})\right| \geqslant \frac{2\delta(\mathbf{x})}{5}\right)$$

$$\leqslant O(1) \cdot e^{-\Omega(nh^{d+2}p(\mathbf{x}))} + O(1) \cdot \exp\left(\frac{-\Omega(nh^d p(\mathbf{x}) \cdot \delta(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}} \cdot \min\left\{\frac{\delta(\mathbf{x})/80}{\sigma^2(\mathbf{x}) + \frac{\mathsf{c}_{cor.2}^{\sigma} h^2}{p(\mathbf{x})}}, \frac{1}{2}\right\}\right).$$

where constants inside Ω *do not depend on* $\mathbf{x}$*. Additionally, if*

$$\delta(\mathbf{x}) \geqslant \frac{10 R_K \sigma^2(\mathbf{x})}{\mathsf{z}_{prop.1}} + \tilde{\mathsf{c}}^{\sigma}_{cor.2} h \left(1 + \frac{R_K}{\mathsf{z}_{prop.1}}\right),$$

then

$$\mathbb{P}\left(\left| (\mathbf{Y} - \mathbf{m})^{\mathrm{T}} (\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\mathbf{Y} - \mathbf{m}) - \sigma^2(\mathbf{x}) \right| \geqslant \frac{2\delta(\mathbf{x})}{5} \right)$$
$$\leqslant O(1) \cdot e^{-\Omega(nh^d p(\mathbf{x}))} + O(1) \cdot \exp\left(\frac{-\Omega(nh^d p(\mathbf{x}) \cdot \delta(\mathbf{x}))}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}} \cdot \min\left\{ \frac{\delta(\mathbf{x})/80}{\sigma^2(\mathbf{x}) + \tilde{\mathsf{c}}^{\sigma}_{cor.2} h}, \frac{1}{2} \right\} \right).$$

Proof First, we bound

$$\mathbb{P}\left(\left| \mathrm{Tr}(\Sigma) - \sigma^2(\mathbf{x}) \right| \geqslant \frac{\delta(\mathbf{x})}{5} \right).$$

By the definition of Σ, we have $\mathrm{Tr}(\Sigma) = \sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i - \boldsymbol{\omega}^{\mathrm{T}} \mathbf{D}_{\boldsymbol{\sigma}^2} \boldsymbol{\omega}$. Meanwhile, due to Corollary 6, we have

$$\boldsymbol{\omega}^{\mathrm{T}} \mathbf{D}_{\boldsymbol{\sigma}^2} \boldsymbol{\omega} = \sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i^2 \leqslant \frac{2R_K}{\mathsf{z}_{prop.1}} \left\{ \sigma^2(\mathbf{x}) + \sum_{i=1} \sigma^2(X_i)\boldsymbol{\omega}_i - \sigma^2(\mathbf{x}) \right\}$$

with probability $1 - e^{-\mathsf{z}_{prop.1}/26} = 1 - e^{-\Omega(nh^d p(\mathbf{x}))}$. Under this event we have

$$\left| \mathrm{Tr}(\Sigma) - \sigma^2(\mathbf{x}) \right| \geqslant \frac{\delta(\mathbf{x})}{5} \implies \left(1 + \frac{R_K}{\mathsf{z}_{prop.1}}\right) \left| \sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i - \sigma^2(\mathbf{x}) \right| \geqslant \frac{\delta(\mathbf{x})}{5} - \frac{2R_K \sigma^2(\mathbf{x})}{\mathsf{z}_{prop.1}}.$$

If $\left(\frac{\delta(\mathbf{x})}{5} - \frac{2R_K\sigma^2(\mathbf{x})}{\mathsf{z}_{prop.1}} \right) / (1 + \frac{R_K}{\mathsf{z}_{prop.1}}) \geqslant \mathsf{c}^{\sigma}_{cor.2} h^2 / p(\mathbf{x})$, we may apply Corollary 9. Thus, condition 3 of the lemma ensures that

$$\mathbb{P}\left(\left| \mathrm{Tr}(\Sigma) - \sigma^2(\mathbf{x}) \right| \geqslant \frac{\delta(\mathbf{x})}{5} \right) = \exp\left(-\Omega(nh^{d+2} p(\mathbf{x})) \right). \tag{17}$$

Next, we will bound

$$\mathbb{P}\left(\left| (\mathbf{Y} - \mathbf{m})^{\mathrm{T}} (\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\mathbf{Y} - \mathbf{m}) - \mathrm{Tr}(\Sigma) \right| \geqslant \frac{\delta(\mathbf{x})}{5} \right)$$

using Proposition 11. First,

$$\|\Sigma\| = \|\mathbf{D}_{\boldsymbol{\sigma}^2} (\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\| \leqslant \max_i \left\{ \sigma^2(X_i)\boldsymbol{\omega}_i + \sum_j \sigma^2(X_i)\boldsymbol{\omega}_i\boldsymbol{\omega}_j \right\}$$

due to Gershgorin's circle theorem. Thus, we have

$$
\begin{aligned}
\|\Sigma\| &\leqslant 2 \max_i \sigma^2(X_i)\boldsymbol{\omega}_i \\
&\leqslant 2 \max_i \sigma^2(\mathbf{x})\boldsymbol{\omega}_i + 2 \max_i |\sigma^2(X_i) - \sigma^2(\mathbf{x})|\boldsymbol{\omega}_i \\
&\leqslant 2\sigma^2(\mathbf{x}) \max_i \boldsymbol{\omega}_i + 2L_{\sigma^2} \frac{\max_i \|X_i - \mathbf{x}\| K\left(\frac{X_i - \mathbf{x}}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \\
&\leqslant 2\sigma^2(\mathbf{x}) \frac{2R_K}{\mathsf{z}_{prop.1}} + \frac{2L_{\sigma^2} R_K h}{r_K e \mathsf{z}_{prop.1}}
\end{aligned}
$$

with probability $1 - e^{-Cnh^d p(\mathbf{x})}$. At the same time

$$
\begin{aligned}
\operatorname{Tr}(\Sigma^2) &= \sum_{i=1}^n \sigma^4(X_i)\boldsymbol{\omega}_i^2 - 2\boldsymbol{\omega}^{\mathrm{T}} \mathbf{D}_{\boldsymbol{\sigma}^2} \mathbf{D}_{\boldsymbol{\omega}} \mathbf{D}_{\boldsymbol{\sigma}^2} \boldsymbol{\omega} + (\boldsymbol{\omega}^{\mathrm{T}} \mathbf{D}_{\boldsymbol{\sigma}^2} \boldsymbol{\omega})^2 \\
&= \sum_{i=1}^n \sigma^4(X_i)\boldsymbol{\omega}_i^2 - 2\sum_{i=1}^n \sigma^4(X_i)\boldsymbol{\omega}_i^3 + \left(\sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i^2\right)^2 \\
&\leqslant \sum_{i=1}^n \sigma^4(X_i)\boldsymbol{\omega}_i^2 + \left(\sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i^2\right)^2 \leqslant 2\sum_{i=1}^n \sigma^4(X_i)\boldsymbol{\omega}_i^2,
\end{aligned}
$$

since

$$
\frac{\sum_{i=1}^n \sigma^4(X_i)\boldsymbol{\omega}_i^2}{\sum_{i=1}^n \boldsymbol{\omega}_i^2} \geqslant \left(\frac{\sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i^2}{\sum_{i=1}^n \boldsymbol{\omega}_i^2}\right)^2,
$$

or, rearranging terms,

$$
\sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i^2 \geqslant \sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i^2 \cdot \sum_{i=1}^n \boldsymbol{\omega}_i^2 \geqslant \left(\sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i^2\right)^2.
$$

As previously, bounding

$$
\max_{i \in [n]} \sigma^2(\mathbf{x})\boldsymbol{\omega}_i \leqslant \sigma^2(\mathbf{x}) \frac{2R_K}{\mathsf{z}_{prop.1}} + \frac{L_{\sigma^2} R_K h}{r_K e \mathsf{z}_{prop.1}},
$$

we obtain

$$
\begin{aligned}
\operatorname{Tr}(\Sigma^2) &\leqslant 2\left\{\sigma^2(\mathbf{x}) \frac{2R_K}{\mathsf{z}_{prop.1}} + \frac{L_{\sigma^2} R_K h}{r_K e \mathsf{z}_{prop.1}}\right\} \sum_{i=1}^n \sigma^2(X_i)\boldsymbol{\omega}_i \\
&\overset{\text{Cor. } 9}{\leqslant} \frac{2R_K}{\mathsf{z}_{prop.1}} \left\{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}\right\} \left\{\sigma^2(\mathbf{x}) + \frac{\mathsf{c}^{\sigma}_{cor.2} h^2}{p(\mathbf{x})}\right\}
\end{aligned}
$$

with probability $1 - e^{-Cnh^{d+2} p(\mathbf{x})}$. Choose t such that $16\sqrt{\operatorname{Tr}(\Sigma^2) t} \leqslant \delta(\mathbf{x})/5$ and $16\|\Sigma\| t \leqslant \delta(\mathbf{x})/5$, e.g.

$$
t = \frac{\delta(\mathbf{x})}{160 R_K} \cdot \frac{\mathsf{z}_{prop.1}}{\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}} \cdot \min\left\{\frac{\delta(\mathbf{x})/80}{\sigma^2(\mathbf{x}) + \frac{\mathsf{c}^{\sigma}_{cor.2} h^2}{p(\mathbf{x})}}, \frac{1}{2}\right\}.
$$

Thus, from Proposition 11 we infer

$$\mathbb{P}\left(\left| (\mathbf{Y} - \mathbf{m})^{\mathrm{T}} (\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\mathbf{Y} - \mathbf{m}) - \mathrm{Tr}(\Sigma) \right| \geqslant \frac{\delta(\mathbf{x})}{5} \right) \leqslant 2e^{-t} + O(1) \cdot e^{-\Omega(nh^{d+2} p(\mathbf{x}))}.$$

Combining the above with (17) finalizes the proof of the first part of the lemma. The second part can be obtained analogously using the second part of Corollary 9. ■

Next, we analyze term (15).

Lemma 13 *Suppose Assumptions 1-4 hold,* $p(\mathbf{x}) \geqslant 2L_p h \cdot \max\{b, \frac{2\pi^{d/2} R_K}{\Gamma(d/2) r_K^2}\}$ *and* $d(\mathbf{x}, \partial\mathcal{S}) \geqslant bh$. *Then it holds that*

$$\mathbb{P}\left(\left| (\mathbf{m} - \mathbf{Y})^{\mathrm{T}} (\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega})^{\mathrm{T}} \mathbf{m} \right| \geqslant \frac{\delta(\mathbf{x})}{5} \right) \leqslant O(1) \cdot \exp\left(-\Omega(nh^d p(\mathbf{x})) \right)$$
$$+ 2\exp\left(-\frac{\delta^2(\mathbf{x})}{50\left(\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}\right)} \cdot \frac{\Omega(nh^{d-2} p(\mathbf{x}))}{2R_K} \cdot \left([\tilde{\mathsf{c}}^f_{cor.2}]^2 + \frac{2L_f^2}{r_K^2} \log^2 \frac{e^2 n R_K}{\mathsf{z}_{prop.1}} \right)^{-1} \right).$$

Proof Notice, that $(\mathbf{m} - \mathbf{Y})^{\mathrm{T}} (\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega})^{\mathrm{T}} \mathbf{m} \mid X$ is distributed as $\mathcal{N}(0, v^2)$ for

$$\begin{aligned} v^2 &= \sum_{i=1}^n \mathbf{m}_i^2 \sigma^2(X_i) \boldsymbol{\omega}_i^2 - 2(\mathbf{m}^{\mathrm{T}} \boldsymbol{\omega}) \cdot \boldsymbol{\omega}^{\mathrm{T}} \mathbf{D}_{\boldsymbol{\sigma}^2} \mathbf{D}_{\boldsymbol{\omega}} \mathbf{m} + (\mathbf{m}^{\mathrm{T}} \boldsymbol{\omega})^2 (\boldsymbol{\omega}^{\mathrm{T}} \mathbf{D}_{\boldsymbol{\sigma}^2} \boldsymbol{\omega}) \\ &= \sum_{i=1}^n (\mathbf{m}_i - \mathbf{m}^{\mathrm{T}} \boldsymbol{\omega})^2 \sigma^2(X_i) \boldsymbol{\omega}_i^2. \end{aligned}$$

Hence,

$$\mathbb{P}\left(\left| (\mathbf{m} - \mathbf{Y})^{\mathrm{T}} (\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega})^{\mathrm{T}} \mathbf{m} \right| \geqslant rv \mid X \right) \leqslant 2\exp\left(-\frac{r^2}{2} \right). \tag{18}$$

We can remove conditioning on X for any fixed positive r. Then, bound

$$\begin{aligned} v^2 &\leqslant (\max_i \sigma^2(X_i) \boldsymbol{\omega}_i) \cdot \sum_{i=1}^n \left(\mathbf{m}_i - \mathbf{m}^{\mathrm{T}} \boldsymbol{\omega} \right)^2 \boldsymbol{\omega}_i \\ &\leqslant 2\left(\max_i \sigma^2(X_i) \boldsymbol{\omega}_i \right) \cdot \left[(\mathbf{m}^{\mathrm{T}} \boldsymbol{\omega} - f(\mathbf{x}))^2 + \sum_{i=1}^n (\mathbf{m}_i - f(\mathbf{x}))^2 \boldsymbol{\omega}_i \right]. \end{aligned}$$

Applying Corollary 6 and Corollary 9, we obtain

$$\begin{aligned} \max_i \sigma^2(X_i) \boldsymbol{\omega}_i &\leqslant \frac{R_K}{\mathsf{z}_{prop.1}} \left(\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K} \right), \\ \left(\mathbf{m}^{\mathrm{T}} \boldsymbol{\omega} - f(\mathbf{x}) \right)^2 &\leqslant (\tilde{\mathsf{c}}^f_{cor.2} h)^2 \end{aligned}$$

with probability $1 - e^{-\Omega\left(nh^d p(\mathbf{x})\right)}$. We bound the sum $\sum_{i=1}^n \left(\mathbf{m}_i - f(\mathbf{x})\right)^2 \boldsymbol{\omega}_i$ using Proposition 10. Thus, with probability $1 - e^{-\Omega(nh^d p(\mathbf{x}))}$ we obtain

$$\begin{aligned}\sum_{i=1}^n \left(\mathbf{m}_i - f(\mathbf{x})\right)^2 \boldsymbol{\omega}_i &\leqslant \frac{2L_f^2 h^2}{r_K^2}\left(2 \vee \log \frac{nR_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)}\right)^2 \\ &\leqslant \frac{2L_f^2 h^2}{r_K^2} \log^2 \frac{e^2 nR_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \\ &\leqslant \frac{2L_f^2 h^2}{r_K^2} \log^2 \frac{e^2 nR_K}{\mathsf{z}_{prop.1}}.\end{aligned}$$

Consequently, with probability $1 - O(1) \cdot e^{-\Omega(nh^d p(\mathbf{x}))}$ we have

$$v \leqslant \sqrt{\frac{2R_K h^2}{\mathsf{z}_{prop.1}}\left(\sigma^2(\mathbf{x}) + \frac{L_{\sigma^2} h}{e r_K}\right)\left([\tilde{\mathsf{c}}^f_{cor.2}]^2 + \frac{2L_f^2}{r_K^2} \log^2 \frac{e^2 nR_K}{\mathsf{z}_{prop.1}}\right)}.$$

Choose r such that $rv \leqslant \frac{\delta(\mathbf{x})}{5}$ and apply (18) to finalize the proof. ■

Finally, we analyze term (16).

Lemma 14 *Suppose $p(\mathbf{x}) \geqslant 2L_p bh$ and $d(\mathbf{x}, \partial\mathcal{S}) \geqslant bh$. Then under Assumptions 1-3, we have*

$$0 \leqslant \mathbf{m}^{\mathrm{T}}(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\mathbf{m} \leqslant \frac{L_f^2 h^2}{r_K^2} \log^2 \frac{e^2 nR_K}{\mathsf{z}_{prop.1}}$$

with probability $1 - e^{-\Omega\left(nh^d p(\mathbf{x})\right)}$ where a constant in $\Omega(\cdot)$ does not depend on $\mathbf{x}$.

Proof From the Gershgorin circle theorem, we infer that $\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}$ is non-negative defined. Then notice, that

$$\begin{aligned}&\mathbf{m}^{\mathrm{T}}\left(\mathbf{D}_{\boldsymbol{\omega}} - \boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}\right)\mathbf{m} = \sum_{i=1}^n \mathbf{m}_i^2 \boldsymbol{\omega}_i - (\mathbf{m}^{\mathrm{T}}\boldsymbol{\omega})^2 \\ &= \sum_{i=1}^n (\mathbf{m}_i^2 - f(\mathbf{x})^2)\boldsymbol{\omega}_i + (f(\mathbf{x}) - \mathbf{m}^{\mathrm{T}}\boldsymbol{\omega})(f(\mathbf{x}) + \mathbf{m}^{\mathrm{T}}\boldsymbol{\omega}) \\ &= \sum_{i=1}^n \left(\mathbf{m}_i - f(\mathbf{x})\right)^2 \boldsymbol{\omega}_i + 2f(\mathbf{x})\sum_{i=1}^n \left(\mathbf{m}_i - f(\mathbf{x})\right)\boldsymbol{\omega}_i - (f(\mathbf{x}) - \mathbf{m}^{\mathrm{T}}\boldsymbol{\omega})^2 + 2f(\mathbf{x})(f(\mathbf{x}) - \mathbf{m}^{\mathrm{T}}\boldsymbol{\omega}) \\ &= \sum_{i=1}^n \left(\mathbf{m}_i - f(\mathbf{x})\right)^2 \boldsymbol{\omega}_i - (f(\mathbf{x}) - \mathbf{m}^{\mathrm{T}}\boldsymbol{\omega})^2 \leqslant \sum_{i=1}^n \left(\mathbf{m}_i - f(\mathbf{x})\right)^2 \boldsymbol{\omega}_i.\end{aligned}$$

The sum can be bounded via Proposition 10. Thus,

$$\mathbf{m}^{\mathrm{T}}\left(\mathbf{D}_{\boldsymbol{\omega}}-\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}\right)\mathbf{m}\leqslant\frac{L_f^2h^2}{r_K^2}\log^2\frac{e^2nR_K}{\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)}.$$

Application of Proposition 5 finalizes the proof. ∎

We summarize this section with the following corollary.

Corollary 15 *Suppose Assumptions 1-3 hold. Assume that*

1. *the probability density* $p(\mathbf{x})$ *is at least* $2L_p h\cdot\max\{b,\frac{2\pi^{d/2}R_K}{\Gamma(d/2)r_K^2}\}$,
2. *the distance* $d(\mathbf{x},\partial\mathcal{S})\geqslant bh$,
3. *for a function* $\delta(\mathbf{x})$ *we have*

$$\delta(\mathbf{x})\geqslant\frac{10R_K\sigma^2(\mathbf{x})}{\mathsf{z}_{prop.1}}+5\frac{\mathsf{c}^{\sigma}_{cor.2}h^2}{p(\mathbf{x})}\left(1+\frac{R_K}{\mathsf{z}_{prop.1}}\right),$$
$$\delta(\mathbf{x})\geqslant\frac{5L_f^2h^2}{r_K^2}\log^2\frac{e^2nR_K}{\mathsf{z}_{prop.1}}.$$

Then

$$\begin{aligned}\mathbb{P}\left(|\widehat{\sigma}_n^2(\mathbf{x})-\sigma^2(\mathbf{x})|\geqslant\delta(\mathbf{x})\right)&\lesssim e^{-\Omega(nh^{d+2}p(\mathbf{x}))}\\&+\exp\left(\frac{-\Omega(nh^dp(\mathbf{x})\cdot\delta(\mathbf{x}))}{\sigma^2(\mathbf{x})+\frac{L_{\sigma^2}h}{er_K}}\cdot\min\left\{\frac{\delta(\mathbf{x})/80}{\sigma^2(\mathbf{x})+\mathsf{c}^{\sigma}_{cor.2}h},\frac{1}{2}\right\}\right)\\&+\exp\left(-\frac{\delta^2(\mathbf{x})}{\sigma^2(\mathbf{x})+\frac{L_{\sigma^2}h}{er_K}}\cdot\frac{\Omega(nh^{d-2}p(\mathbf{x}))}{2R_K}\cdot\left([\mathsf{c}^f_{cor.2}]^2+\frac{2L_f^2}{r_K^2}\log^2\frac{e^2nR_K}{\mathsf{z}_{prop.1}}\right)^{-1}\right).\end{aligned}$$

Additionally, if we have

$$\delta(\mathbf{x})\geqslant\frac{10R_K\sigma^2(\mathbf{x})}{\mathsf{z}_{prop.1}}+5\tilde{\mathsf{c}}^{\sigma}_{cor.2}h\left(1+\frac{R_K}{\mathsf{z}_{prop.1}}\right),$$

then it holds that

$$\begin{aligned}\mathbb{P}\left(|\widehat{\sigma}_n^2(\mathbf{x})-\sigma^2(\mathbf{x})|\geqslant\delta(\mathbf{x})\right)&\lesssim e^{-\Omega(nh^{d}p(\mathbf{x}))}\\&+\exp\left(\frac{-\Omega(nh^dp(\mathbf{x})\cdot\delta(\mathbf{x}))}{\sigma^2(\mathbf{x})+\frac{L_{\sigma^2}h}{er_K}}\cdot\min\left\{\frac{\delta(\mathbf{x})/80}{\sigma^2(\mathbf{x})+\mathsf{c}^{\sigma}_{cor.2}h},\frac{1}{2}\right\}\right)\\&+\exp\left(-\frac{\delta^2(\mathbf{x})}{\sigma^2(\mathbf{x})+\frac{L_{\sigma^2}h}{er_K}}\cdot\frac{\Omega(nh^{d-2}p(\mathbf{x}))}{2R_K}\cdot\left([\mathsf{c}^f_{cor.2}]^2+\frac{2L_f^2}{r_K^2}\log^2\frac{e^2nR_K}{\mathsf{z}_{prop.1}}\right)^{-1}\right).\end{aligned}$$

Proof We decompose $\widehat{\sigma}_n^2(\mathbf{x})$ as $\widehat{\sigma}_n^2(\mathbf{x}) = (14) + (15) + (16)$. Then

$$\mathbb{P}\left(|\widehat{\sigma}_n^2(\mathbf{x}) - \sigma^2(\mathbf{x})| \geqslant \delta(\mathbf{x})\right) \leqslant$$
$$\leqslant \mathbb{P}\left(|(14) - \sigma^2(\mathbf{x})| \geqslant \frac{2\delta(\mathbf{x})}{5}\right) + \mathbb{P}\left(2|(15)| \geqslant \frac{2\delta(\mathbf{x})}{5}\right) + \mathbb{P}\left(|(16)| \geqslant \frac{\delta(\mathbf{x})}{5}\right).$$

The first term and the second term can be bounded via Lemma 12 and Lemma 13 respectively. For the third term, we use Lemma 14 and the third condition of the corollary. ∎

B.7. Bias-Variance tradeoff for L_p-risk

Proposition 16 *For a binomial random variable $B(n, q)$ it holds that*

$$\mathbb{E}\frac{1}{(r + B(n,q))^r} \leqslant \frac{1}{(n+1)^r q^r}.$$

Proof For a binomial random variable $B(n, q)$ it holds

$$\mathbb{E}\frac{1}{(r + B(n,q))^r} = \sum_{k=0}^{n} \frac{1}{(r+k)^r}\binom{n}{k} q^k (1-q)^{n-k}$$
$$\leqslant \sum_{k=0}^{n}\prod_{j=1}^{r}\frac{1}{(j+k)}\binom{n}{k} q^k(1-q)^{n-k} = \prod_{j=1}^{r}\frac{1}{(n+j)}\sum_{k=0}^{n}\binom{n+r}{k+r}q^k(1-q)^{n-k}$$
$$\leqslant \frac{1}{(n+1)^r q^r}\sum_{k=0}^{n}\binom{n+r}{k+r}q^{k+r}(1-q)^{n-k} = \frac{1}{(n+1)^r q^r}\sum_{k'=r}^{n+r}\binom{n+r}{k'}q^{k'}(1-q)^{n+r-k'}$$
$$\leqslant \frac{1}{(n+1)^r q^r}(q + (1-q))^{n+r} = \frac{1}{(n+1)^r q^r}.$$

∎

Lemma 17 *Suppose that Assumptions 3,4 hold and $p(\mathbf{x}) \geqslant 2L_p h \cdot \max\{b, \frac{2\pi^{d/2}R_K}{\Gamma(d/2)r_K^2}\}$, $d(\mathbf{x}, \partial S) \geqslant bh$. Then*

$$\mathbb{E}\left[\frac{1}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ra}{nh^d}\right\}\right]^r \leqslant \left(\frac{a\omega_d b^d}{4} nh^d p(\mathbf{x})\right)^{-r}.$$

Proof By the definition of $\widehat{p}_n(\mathbf{x})$: $\widehat{p}_n(\mathbf{x}) \geqslant ar/(nh^d)$ implies

$$\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right) \geqslant ar.$$

In particular, it means that

$$\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right) \geqslant \frac{ar + \sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)}{2}.$$

Thus,

$$\frac{1}{\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)} \leqslant \frac{2}{ar+\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)} \leqslant \frac{2}{ar+a\sum_{i=1}^n \mathrm{I}\left\{X_i \in \mathcal{B}_{hb}(\mathbf{x})\right\}} \tag{19}$$

due to Assumption 3. The sum $\sum_{i=1}^n \mathrm{I}\left\{X_i \in \mathcal{B}_{hb}(\mathbf{x})\right\}$ is a binomial random variable. Due to Proposition 16, we have

$$\mathbb{E}\frac{1}{\left(r+\sum_{i=1}^n \mathrm{I}\left\{X_i \in \mathcal{B}_{hb}(\mathbf{x})\right\}\right)^r} \leqslant \left\{(n+1)\mathbb{P}(X_1 \in \mathcal{B}_{bh}(\mathbf{x}))\right\}^{-r}.$$

Using (19) and the above, we obtain

$$\mathbb{E}\left[\frac{1}{\sum_{i=1}^n K\left(\frac{X_i-\mathbf{x}}{h}\right)}\mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ra}{nh^d}\right\}\right]^r \leqslant \left\{\frac{a}{2}(n+1)\mathbb{P}(X_1 \in \mathcal{B}_{bh}(\mathbf{x}))\right\}^{-r}. \tag{20}$$

Finally,

$$\int_{\mathcal{B}_{bh}(\mathbf{x})} p(\mathbf{y})d\mu(\mathbf{y}) \geqslant (p(\mathbf{x})-L_p bh)\int_{\mathcal{B}_{bh}(\mathbf{x})} d\mu(\mathbf{y}) = \mu(\mathcal{B}_{bh})\left(p(\mathbf{x})-L_p bh\right).$$

Since $p(\mathbf{x}) \geqslant 2L_p bh$, we have $p(\mathbf{x}) - L_p bh \geqslant p(\mathbf{x})/2$, and, thus,

$$\mathbb{P}(X_1 \in \mathcal{B}_{bh}(\mathbf{x})) \geqslant \frac{p(\mathbf{x})\omega_d b^d}{2}h^d.$$

Combining the above with (20), we infer the statement. ■

Lemma 18 *Suppose Assumptions 1-4 hold and assume that* $p(\mathbf{x}) \geqslant 2L_p bh$ *and* $d(\mathbf{x}, \partial S) \geqslant bh$. *Define*

$$\mathsf{c}_t = 1 - \frac{L_p h}{p(\mathbf{x})}\cdot\frac{2\pi^{d/2}R_K}{\Gamma(d/2)r_K^2},$$

$$\mathsf{c}_h = \frac{2L_f L_p + H_f C_p}{2}\cdot\frac{4\pi^{d/2}R_K}{\Gamma(d/2)r_K^3},$$

$$\mathsf{c}'_t = \frac{2\pi^{d/2}R_K^2}{\Gamma(d/2)r_k} + \frac{2hL_p}{p(\mathbf{x})}\cdot\frac{4\pi^{d/2}R_k}{\Gamma(d/2)r_k^2},$$

$$\mathsf{c}'_h = L_f^2\frac{\pi^{d/2}R_K^2}{\Gamma(d/2)r_K^3} + \frac{2hL_f^2 L_p}{p(\mathbf{x})}\cdot\frac{4\pi^{d/2}R_K^2}{3\Gamma(d/2)r_K^4},$$

Assume $\mathsf{c}_t > 0$. *Then*

$$\mathbb{E}\left|\widehat{f}_n(\mathbf{x}) - f(\mathbf{x})\right|^r \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} \leqslant \frac{\Gamma\left(\frac{r+1}{2}\right)}{2\sqrt{\pi}}\left(\frac{32R_K\left\{\sigma^2(\mathbf{x}) + \frac{2L_\sigma h}{r_K}\log\frac{enR_K}{ar}\right\}}{a\omega_d b^d\cdot nh^d p(\mathbf{x})}\right)^{r/2}$$

$$+\left(\frac{4h^2}{p(\mathbf{x})}\cdot\frac{\mathsf{c}_h}{\mathsf{c}_t}\right)^r + \frac{r2^{3r/2}\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}\cdot h^r \times \left(\frac{\left\{\mathsf{c}'_t + \frac{\mathsf{c}_t R_K}{3}\right\}\cdot 4^{1/r}\frac{L_f^2}{r_K^2}\log^2\frac{e^r nR_K}{ar} + \mathsf{c}'_h + \frac{R_K}{3}\frac{2^{1/r}L_f^2}{er_K^2}\log\frac{e^r nR_K}{ar}}{nh^d p(\mathbf{x})\mathsf{c}_t^2}\right)^{r/2}.$$

Proof First, we have

$$\mathbb{E}\left|\widehat{f}_n(\mathbf{x}) - f(\mathbf{x})\right|^r \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{n}\right\} \leqslant 2^{r-1}\mathbb{E}\left|\widehat{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{f}_n(\mathbf{x}) \mid X]\right|^r \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} + 2^{r-1}\mathbb{E}\left|\mathbb{E}[\widehat{f}_n(\mathbf{x}) \mid X] - f(\mathbf{x})\right|^r \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\}. \tag{21}$$

We analyze each term separately. For the first term, we have

$$\mathbb{E}\left|\widehat{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{f}_n(\mathbf{x}) \mid X]\right|^r \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} = \mathbb{E}\left\{\mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} \cdot \mathbb{E}\left[\left|\widehat{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{f}_n(\mathbf{x}) \mid X]\right|^r \mid X\right]\right\}.$$

Conditioned on X, the random variable $\widehat{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{f}_n(\mathbf{x}) \mid X]$ is Gaussian with zero mean and variance $v^2 = \sum_{i=1}^n \boldsymbol{\omega}_i^2 \sigma^2(X_i)$. Thus,

$$\mathbb{E}\left|\widehat{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{f}_n(\mathbf{x}) \mid X]\right|^r \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} = 2^{r/2}\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}\mathbb{E}v^r \,\mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\}. \tag{22}$$

We bound

$$\begin{aligned} v^2 = \sum_{i=1}^n \boldsymbol{\omega}_i^2 \sigma^2(X_i) &\leqslant (\max_i \boldsymbol{\omega}_i) \sum_{i=1}^n \boldsymbol{\omega}_i \sigma^2(X_i) \\ &\leqslant \frac{R_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \left\{\sigma^2(\mathbf{x}) + \sum_{i=1}^n \boldsymbol{\omega}_i |\sigma^2(X_i) - \sigma^2(\mathbf{x})|\right\}. \end{aligned}$$

From Proposition 10, we have

$$v^2 \leqslant \frac{R_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \left\{\sigma^2(\mathbf{x}) + \frac{2L_\sigma h}{r_K} \log \frac{enR_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)}\right\}.$$

Substituting the above into (22), we obtain

$$\begin{aligned} (22) \leqslant\; & 2^{r/2}\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}\mathbb{E}\,\mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} \times \\ & \times \left(\frac{R_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \left\{\sigma^2(\mathbf{x}) + \frac{2L_\sigma h}{r_K} \log \frac{enR_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)}\right\}\right)^{r/2} \\ \leqslant\; & 2^{r/2}\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}\mathbb{E}\,\mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} \times \\ & \times \left(\frac{R_K}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)} \left\{\sigma^2(\mathbf{x}) + \frac{2L_\sigma h}{r_K} \log \frac{enR_K}{ar}\right\}\right)^{r/2} \\ \leqslant\; & 2^{r/2}\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}} \left(R_K\left\{\sigma^2(\mathbf{x}) + \frac{2L_\sigma h}{r_K} \log \frac{enR_K}{ar}\right\}\right)^{r/2} \mathbb{E}\left(\frac{\mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\}}{\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)}\right)^{r/2}. \end{aligned}$$

Applying Lemma 17 to bound the expectation, we obtain the variance term:

$$\mathbb{E}\left|\widehat{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{f}_n(\mathbf{x}) \mid X]\right|^r \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} \leqslant \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}} \left(\frac{8R_K\left\{\sigma^2(\mathbf{x}) + \frac{2L_\sigma h}{r_K}\log\frac{enR_K}{ar}\right\}}{a\omega_d b^d \cdot nh^d p(\mathbf{x})}\right)^{r/2}. \tag{23}$$

Analysis of the bias term is much simpler. We have

$$\begin{aligned}\mathbb{E}&\left|\mathbb{E}[\widehat{f}_n(\mathbf{x}) \mid X] - f(\mathbf{x})\right|^r \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} \\ &= \int_0^\infty \mathbb{P}\left(\left|\sum_{i=1}^n \boldsymbol{\omega}_i f(X_i) - f(\mathbf{x})\right|^r \geqslant t \text{ and } \widehat{p}_n(\mathbf{x}) \geqslant \frac{ra}{nh^d}\right) dt.\end{aligned}$$

Since $\left|\sum_{i=1}^n \boldsymbol{\omega}_i f(X_i) - f(\mathbf{x})\right|^r \leqslant \sum_{i=1}^n |f(X_i) - f(\mathbf{x})|^r \boldsymbol{\omega}_i$ from the Jensen inequality, we obtain the following:

$$\left|\sum_{i=1}^n \boldsymbol{\omega}_i f(X_i) - f(\mathbf{x})\right|^r \leqslant \frac{2L^r h^r}{r_K^r}\log^r \frac{e^r nR_K}{\sum_{i=1}^n K\left(\frac{\mathbf{x}-X_i}{h}\right)}$$

via Proposition 10. If $\widehat{p}_n(\mathbf{x}) \geqslant \frac{ra}{nh^d}$, the sum $\sum_{i=1}^n K\left(\frac{\mathbf{x}-X_i}{h}\right)$ is at least ar, and, consequently,

$$\begin{aligned}\mathbb{E}&\left|\mathbb{E}[\widehat{f}_n(\mathbf{x}) \mid X] - f(\mathbf{x})\right|^r \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{ar}{nh^d}\right\} \\ &\leqslant \int_0^{\frac{2L^r h^r}{r_K^r}\log^r\frac{e^r nR_K}{ar}} \mathbb{P}\left(\left|\sum_{i=1}^n \boldsymbol{\omega}_i f(X_i) - f(\mathbf{x})\right|^r \geqslant t\right) dt.\end{aligned} \tag{24}$$

Define

$$t_0 = \left(\frac{h^2}{p(\mathbf{x})}\cdot\frac{\mathsf{c}_h}{\mathsf{c}_t}\right)^r, \quad t_{\max} = \frac{2L^r h^r}{r_K^r}\log^r\frac{e^r nR_K}{ar}.$$

According to Lemma 8, for $t \geqslant t_0$ we have

$$\mathbb{P}\left(\left|\sum_{i=1}^n \boldsymbol{\omega}_i f(X_i) - f(\mathbf{x})\right|^r \geqslant t\right) \leqslant 2\exp\left(-\frac{1}{2}\frac{nh^d p(\mathbf{x})\left(\mathsf{c}_t t^{1/p} - \mathsf{c}_h\frac{h^2}{p(\mathbf{x})}\right)^2}{\mathsf{c}'_t t_{\max}^{2/r} + \mathsf{c}'_h h^2 + \frac{R_K}{3}\left(\frac{L_f h}{er_K} + t_{\max}^{1/r}\right)\mathsf{c}_t t_{\max}^{1/r}}\right).$$

From the above, we obtain

$$\begin{aligned}(25) &\leqslant \int_0^{t_0} dt + 2\int_{t_0}^{t_{\max}} \exp\left(-\frac{1}{2}\frac{nh^d p(\mathbf{x})\left(\mathsf{c}_t t^{1/p} - \mathsf{c}_h\frac{h^2}{p(\mathbf{x})}\right)^2}{\mathsf{c}'_t t_{\max}^{2/r} + \mathsf{c}'_h h^2 + \frac{R_K}{3}\left(\frac{L_f h}{er_K} + t_{\max}^{1/r}\right)\mathsf{c}_t t_{\max}^{1/r}}\right) dt \\ &= t_0 + 2r\int_0^{t_{\max}^{1/r}} y^{r-1}\exp\left(-\frac{1}{2}\frac{nh^d p(\mathbf{x})\left(\mathsf{c}_t y - \mathsf{c}_h\frac{h^2}{p(\mathbf{x})}\right)^2}{\mathsf{c}'_t t_{\max}^{2/r} + \mathsf{c}'_h h^2 + \frac{R_K}{3}\left(\frac{L_f h}{er_K} + t_{\max}^{1/r}\right)\mathsf{c}_t t_{\max}^{1/r}}\right) dy \\ &\leqslant t_0 + \frac{r2^{r/2+1}\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}}\left(\frac{\mathsf{c}'_t t_{\max}^{2/r} + \mathsf{c}'_h h^2 + \frac{R_K}{3}\left(\frac{L_f h}{er_K} + t_{\max}^{1/r}\right)\mathsf{c}_t t_{\max}^{1/r}}{nh^d p(\mathbf{x})\mathsf{c}_t^2}\right)^{p/2}.\end{aligned} \tag{25}$$

Applying the bounds (23), (24) to the sum (21), we obtain the statement. ■

Appendix C. Proof of Theorem 4

We require one additional proposition.

Proposition 19 *Suppose Assumptions 3-4 hold. Assume* $p(\mathbf{x}) \geqslant 2L_p bh$ *and* $d(\mathbf{x}, \partial\mathcal{S}) \geqslant$. *Then*

$$\mathbb{P}\left(\frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right) \geqslant 2p(\mathbf{x}) + L_p h \int_{\mathbb{R}^d} \|\mathbf{t}\| K(\mathbf{t}) d\mu(\mathbf{t})\right) \leqslant \exp\left(-\mathsf{z}_{prop.2}\right),$$

where

$$\mathsf{z}_{prop.2} = 2nh^d p(\mathbf{x}) \cdot \left\{R_K/3 + \frac{\pi^{d/2} R_K^2}{\Gamma(d/2) r_K} + \frac{\pi^{d/2} R_K^2}{b\Gamma(d/2) r_K^2}\right\}.$$

Proof First, we study the mathematical expectation of $\widehat{p}_n(\mathbf{x}) = \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{X_i - \mathbf{x}}{h}\right)$. We have

$$\begin{aligned}
\mathbb{E}K\left(\frac{X_i - \mathbf{x}}{h}\right) &= h^d \int_{\mathbb{R}^n} K(\mathbf{t}) p(\mathbf{x} + h\mathbf{t}) d\mu(\mathbf{t}) \\
&\leqslant h^d \left\{\int_{\mathbb{R}^d} K(\mathbf{t}) p(\mathbf{x}) d\mu(\mathbf{t}) + hL_p \int_{\mathbb{R}^d} K(\mathbf{t}) \|\mathbf{t}\| d\mu(\mathbf{t})\right\} \\
&= h^d p(\mathbf{x}) + L_p h^{d+1} \int_{\mathbb{R}^d} \|\mathbf{t}\| K(\mathbf{t}) d\mu(\mathbf{t}).
\end{aligned}$$

Next, we have

$$\begin{aligned}
\operatorname{Var} K\left(\frac{X_i - \mathbf{x}}{h}\right) \leqslant \mathbb{E}K^2\left(\frac{X_i - \mathbf{x}}{h}\right) &= h^d \int_{\mathbb{R}^d} K^2(\mathbf{t}) p(\mathbf{x} + h\mathbf{t}) d\mu(\mathbf{t}) \\
&\leqslant h^d \int_{\mathbb{R}^d} K^2(\mathbf{t}) p(\mathbf{x}) d\mu(\mathbf{t}) + L_p h^{d+1} \int K^2(\mathbf{t}) \|\mathbf{t}\| d\mu(\mathbf{t}).
\end{aligned}$$

Finally, we have $(nh^d)^{-1} K\left(\frac{\mathbf{x} - X_i}{h}\right) \leqslant (nh^d)^{-1} R_K$. Consequently, we may apply the Bernstein inequality to bound $\widehat{p}_n(\mathbf{x})$ in probability:

$$\begin{aligned}
&\mathbb{P}\left(\widehat{p}_n(\mathbf{x}) \geqslant p(\mathbf{x}) + L_p h \int_{\mathbb{R}^d} \|\mathbf{t}\| K(\mathbf{t}) d\mu(\mathbf{t}) + t\right) \\
&\quad \leqslant \exp\left(-\frac{t^2/2}{\left\{p(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mu(\mathbf{t}) + L_p h \int K^2(\mathbf{t}) \|\mathbf{t}\| d\mu(\mathbf{t})\right\} / (nh^d) + R_k (nh^d)^{-1} t/3}\right).
\end{aligned}$$

Set $t = p(\mathbf{x})$. Then,

$$\mathbb{P}\left(\widehat{p}_n(\mathbf{x}) \geqslant 2p(\mathbf{x}) + L_p h \int_{\mathbb{R}^d} \|\mathbf{t}\| K(\mathbf{t}) d\mu(\mathbf{t})\right)$$
$$\leqslant \exp\left(-\frac{nh^d p(\mathbf{x})/2}{\left\{\int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mu(\mathbf{t}) + L_p h/p(\mathbf{x}) \int K^2(\mathbf{t})\|\mathbf{t}\| d\mu(\mathbf{t})\right\} + R_K/3}\right).$$

The condition of the proposition ensures that $h/p(\mathbf{x}) \leqslant 1/(2L_p b)$. Combining it with bounds on integrals provided by Proposition 7, we simplify the denominator and finalize the proof. ∎

Now, we prove Theorem 4.

Proof Due to Proposition 3, we have

$$\mathbb{E}_{\mathcal{D}} \mathcal{R}_\lambda(\mathbf{x}) - \mathcal{R}^*_\lambda(\mathbf{x}) = \mathbb{E}\left[\left(f(\mathbf{x}) - \widehat{f}_n(\mathbf{x})\right)^2 \mathrm{I}\left\{\widehat{a}_n(\mathbf{x}) = 0\right\}\right] + \Delta(\mathbf{x}) \cdot \mathbb{P}\left(\widehat{a}_n(\mathbf{x}) \neq a^*(\mathbf{x})\right). \quad (26)$$

We consider two cases.

Case 1. If $\sigma^2(\mathbf{x}) \geqslant \lambda$, then

$$\mathrm{I}\left\{\widehat{a}_n(\mathbf{x}) = 0\right\} = \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{4a}{nh^d}\right\} \cdot \mathrm{I}\left\{\widehat{a}_n(\mathbf{x}) \neq a^*(\mathbf{x})\right\},$$

consequently,

$$\mathbb{E}\left[\left(f(\mathbf{x}) - \widehat{f}_n(\mathbf{x})\right)^2 \mathrm{I}\left\{\widehat{a}_n(\mathbf{x}) = 0\right\}\right] \leqslant \sqrt{\mathbb{E}\left[\left(f(\mathbf{x}) - \widehat{f}_n(\mathbf{x})\right)^4 \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{4a}{nh^d}\right\}\right]} \cdot \mathbb{P}^{1/2}\left(\widehat{a}_n(\mathbf{x}) \neq a^*(\mathbf{x})\right).$$

Hence,

$$\mathbb{E}_{\mathcal{D}} \mathcal{R}_\lambda(\mathbf{x}) - \mathcal{R}^*_\lambda(\mathbf{x}) \leqslant \left\{\sqrt{\mathbb{E}\left[\left(f(\mathbf{x}) - \widehat{f}_n(\mathbf{x})\right)^4 \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{4a}{nh^d}\right\}\right]} + \Delta(\mathbf{x})\right\} \cdot \mathbb{P}^{1/2}\left(\widehat{\alpha}_n(\mathbf{x}) \neq \alpha(\mathbf{x})\right).$$

Using $\sqrt{\sum_{i=1}^n a_i} \leqslant \sum_{i=1}^n \sqrt{a_i}$ for any sequence of positive numbers $a_1, \ldots, a_n$, we bound the square root of the expectation with Lemma 18. We may consider $\mathsf{c}_t, \mathsf{c}_h, \mathsf{c}'_t, \mathsf{c}'_h$ as constants since the conditions of the theorem bound $h/p(\mathbf{x})$ and h. At the same time, $\widehat{\alpha}_n(\mathbf{x}) \neq \alpha(\mathbf{x})$ implies

$$\widehat{\sigma}^2_n(\mathbf{x}) \leqslant \lambda\left[1 - \frac{\sqrt{2}\|K\|_2 z_{1-\beta}}{\sqrt{nh^d \widehat{p}_n(\mathbf{x})}}\right] \quad \text{or} \quad \widehat{p}_n(\mathbf{x}) < \frac{4a}{nh^d}.$$

Since $p(\mathbf{x}) \geqslant 8/(\omega_d b^d n h^d)$, we have $\widehat{p}_n(\mathbf{x}) < \frac{4a}{nh^d}$ with probability $e^{-\mathbf{z}_{prop.1}/26}$ due to Proposition 5. At the same time, $\widehat{p}_n(\mathbf{x}) < 2p(\mathbf{x}) + L_p h \int_{\mathbb{R}^d} \|\mathbf{t}\| K(\mathbf{t}) d\mu(\mathbf{t})$ with probability $1 - \exp\left(-\Omega(nh^d p(\mathbf{x}))\right)$. Consequently,

$$\begin{aligned}\mathbb{P}\left(\widehat{\alpha}_n(\mathbf{x}) \neq \alpha(\mathbf{x})\right) &\leqslant \mathbb{P}\left(\widehat{\sigma}_n^2(\mathbf{x}) \leqslant \lambda\left[1 - \frac{\sqrt{2}\|K\|_2 z_{1-\beta}}{\sqrt{nh^d\left(2p(\mathbf{x}) + L_p h\int_{\mathbb{R}^d}\|\mathbf{t}\|K(\mathbf{t})d\mu(\mathbf{t})\right)}}\right]\right) \\ &\quad + 2\exp\left(-\Omega(nh^d p(\mathbf{x}))\right) \\ &= \mathbb{P}\left(\widehat{\sigma}_n^2(\mathbf{x}) - \sigma^2(\mathbf{x}) \leqslant -\Delta(\mathbf{x}) - \frac{\sqrt{2}\lambda\|K\|_2 z_{1-\beta}}{\sqrt{nh^d\left(2p(\mathbf{x}) + L_p h\int_{\mathbb{R}^d}\|\mathbf{t}\|K(\mathbf{t})d\mu(\mathbf{t})\right)}}\right) \\ &\quad + 2\exp\left(-\Omega(nh^d p(\mathbf{x}))\right).\end{aligned}$$

Thus, we may apply Corollary 15 with $\delta(\mathbf{x}) = \Delta(\mathbf{x}) + \frac{\sqrt{2}\lambda\|K\|_2 z_{1-\beta}}{\sqrt{nh^d\left(2p(\mathbf{x}) + L_p h\int_{\mathbb{R}^d}\|\mathbf{t}\|K(\mathbf{t})d\mu(\mathbf{t})\right)}}$ and obtain the first part of the statement.

Case 2. If $\sigma^2(\mathbf{x}) \leqslant \lambda$, we bound

$$\mathcal{R}(\mathbf{x}) - \mathcal{R}^*(\mathbf{x}) \leqslant \mathbb{E}\left[\left(f(\mathbf{x}) - \widehat{f}_n(\mathbf{x})\right)^2 \mathrm{I}\left\{\widehat{p}_n(\mathbf{x}) \geqslant \frac{4a}{nh^d}\right\}\right] + \Delta(\mathbf{x}) \cdot \mathbb{P}\left(\widehat{\alpha}_n(\mathbf{x}) \neq \alpha(\mathbf{x})\right).$$

The expectation can be bounded via Lemma 18. The event $\widehat{\alpha}_n(\mathbf{x}) \neq \alpha(\mathbf{x})$ means that

$$\widehat{\sigma}_n^2(\mathbf{x}) \geqslant \lambda\left[1 - \frac{\sqrt{2}(4\pi)^{-d/4} z_{1-\beta}}{\sqrt{nh^d \widehat{p}_n(\mathbf{x})}}\right].$$

Consider the event $\widehat{p}_n(\mathbf{x}) \geqslant a b^d \omega_d \cdot p(\mathbf{x})/2$. Its complement has probability $\exp\left(-\Omega(nh^d p(\mathbf{x}))\right)$ due to Proposition 5. Thus, we just consider the event

$$\widehat{\sigma}_n^2(\mathbf{x}) \geqslant \lambda\left[1 - \frac{2(4\pi)^{-d/4} z_{1-\beta}}{\sqrt{a b^d \omega_d \cdot p(\mathbf{x}) \cdot nh^d}}\right].$$

It implies

$$\widehat{\sigma}_n^2(\mathbf{x}) - \sigma^2(\mathbf{x}) \geqslant \Delta(\mathbf{x}) - \frac{2\lambda(4\pi)^{-d/4} z_{1-\beta}}{\sqrt{a b^d \omega_d \cdot p(\mathbf{x}) \cdot nh^d}},$$

and bounding its probability with Corollary 15, we finalize case 2.

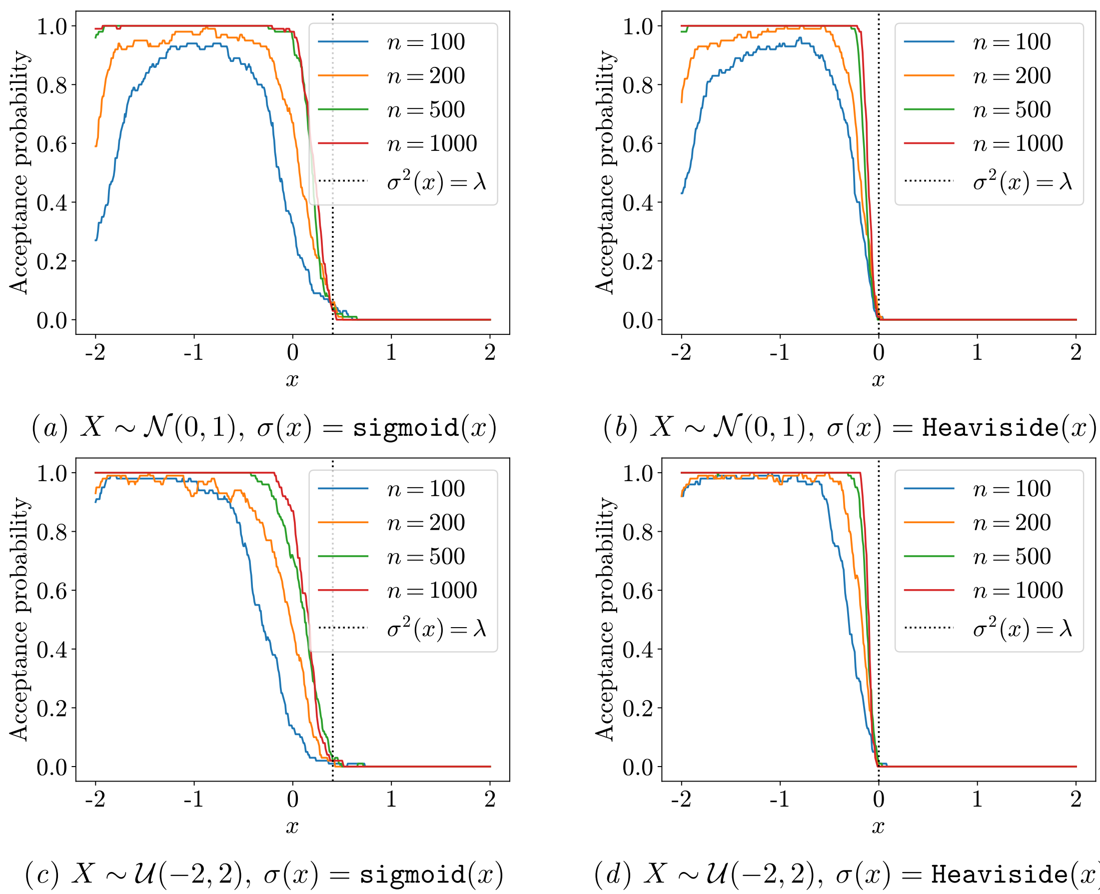
For any $\sigma^2(\mathbf{x})$, we may bound the probability $\mathbb{P}\big(\widehat{\alpha}_n(\mathbf{x}) \neq \alpha(\mathbf{x})\big)$ by one. The remaining term can be bounded via Lemma 18. That finalizes our proof. ■

Appendix D. Additional experiments

D.1. Synthetic data

D.1.1. Acceptance probability

On Figure 6 we present additional experiments for acceptance experiments with different combinations of data distribution and standard deviation. For normal data we see the expected decline on the left side due to lower $\widehat{p}_n(x)$.



(a) $X \sim \mathcal{N}(0,1)$, $\sigma(x) = \texttt{sigmoid}(x)$

(b) $X \sim \mathcal{N}(0,1)$, $\sigma(x) = \texttt{Heaviside}(x)$

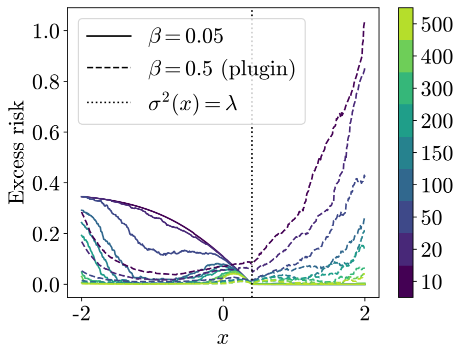
(c) $X \sim \mathcal{U}(-2,2)$, $\sigma(x) = \texttt{sigmoid}(x)$

(d) $X \sim \mathcal{U}(-2,2)$, $\sigma(x) = \texttt{Heaviside}(x)$

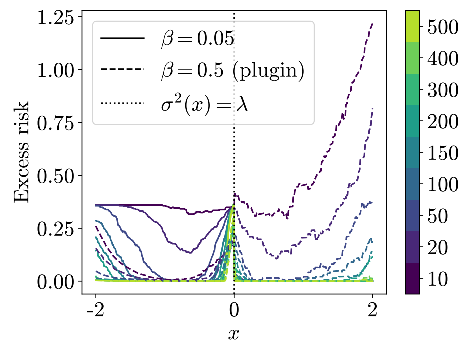
Figure 6: Acceptance probability. We sample 100 datasets of sizes $[100, 200, 500, 1000]$ and for points in $[-2, 2]$ calculate ratio of points where our method accepts the regression result. Optimal bandwidth is selected using leave-one-out cross-validation with mean squared error.

D.1.2. Expected excess risk

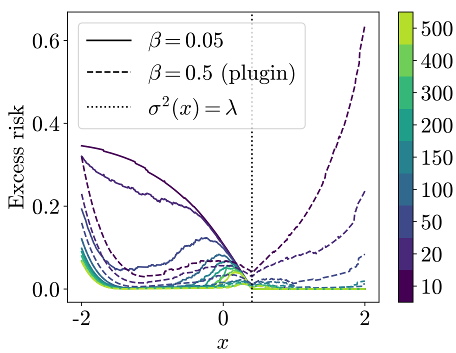
Point-wise expected excess risk with varying sample size is presented on Figure 7. With low sample size our method struggles in the low variance regions, while in high variance the performance is much better than the baseline method.



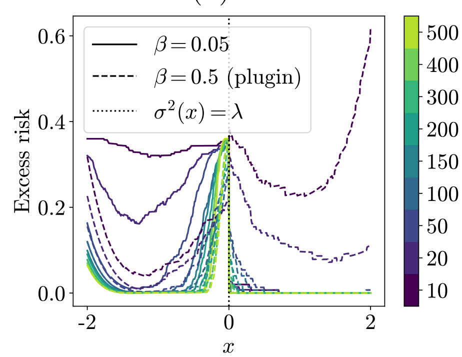
(a) $X \sim \mathcal{N}(0, 1)$, $\sigma(x) = \texttt{sigmoid}(x)$



(b) $X \sim \mathcal{N}(0, 1)$, $\sigma(x) = \texttt{Heaviside}(x)$



(c) $X \sim \mathcal{U}(-2, 2)$, $\sigma(x) = \texttt{sigmoid}(x)$



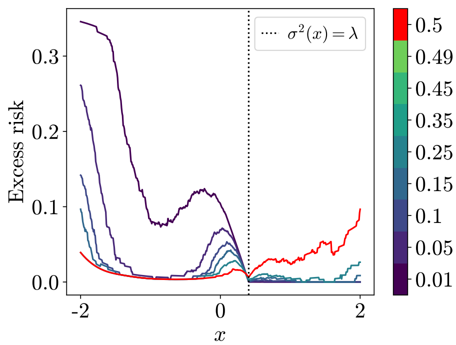
(d) $X \sim \mathcal{U}(-2, 2)$, $\sigma(x) = \texttt{Heaviside}(x)$

Figure 7: Expected excess risk. We sample 100 datasets of each size and for points in $[-2, 2]$.

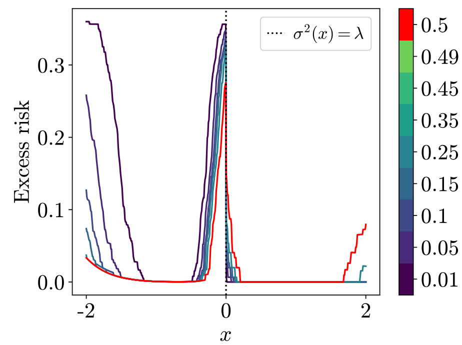
On Figure 8 we present the same quantity for a fixed sample size. For simple data with step-like variance all settings perform nearly identical. For sigmoid variance, we observe the characteristic bump left of zero.

D.2. Airfoil Self-Noise Data Set

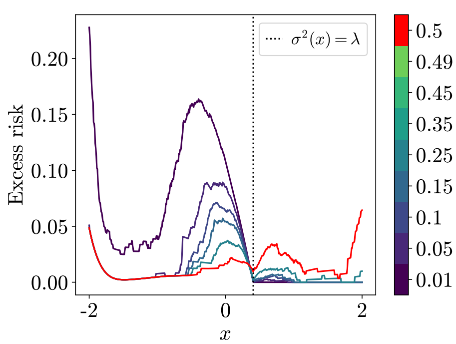
For the Airfoil dataset, we present additional charts for a different feature split. We can see that while acceptance has a similar dependence on λ, the actual mean squared error of accepted results highly depends on the data split. We plan to expand our study to higher-dimensional datasets to further investigate the behavior of the method.



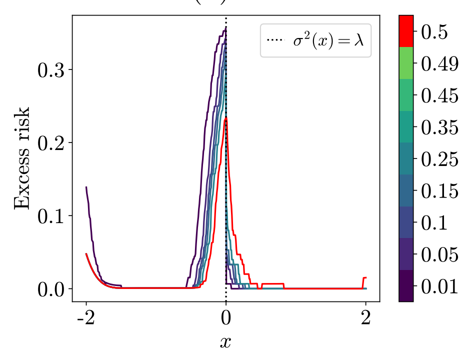
(a) $X \sim \mathcal{N}(0, 1)$, $\sigma(x) = \texttt{sigmoid}(x)$



(b) $X \sim \mathcal{N}(0, 1)$, $\sigma(x) = \texttt{Heaviside}(x)$

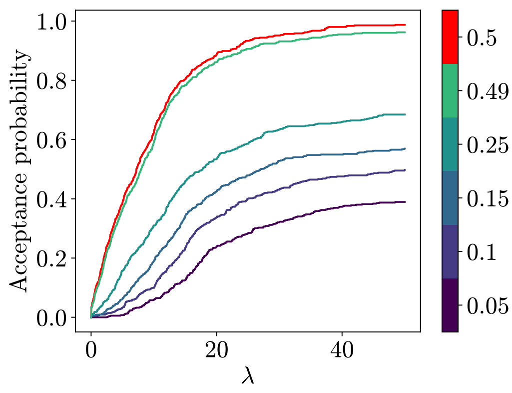


(c) $X \sim \mathcal{U}(-2, 2)$, $\sigma(x) = \texttt{sigmoid}(x)$

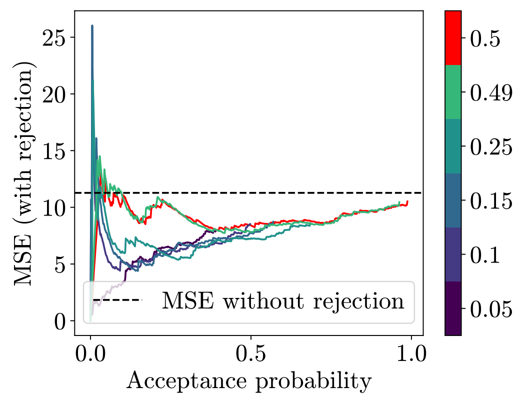


(d) $X \sim \mathcal{U}(-2, 2)$, $\sigma(x) = \texttt{Heaviside}(x)$

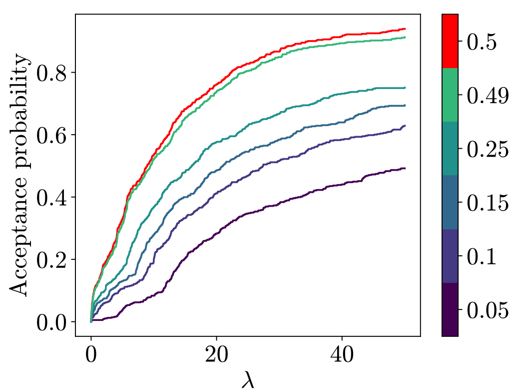
Figure 8: Expected excess risk for a fixed sample size of 100 and different values of β. Baseline method or "plugin" ($\beta = 0.5$) is shown in red.



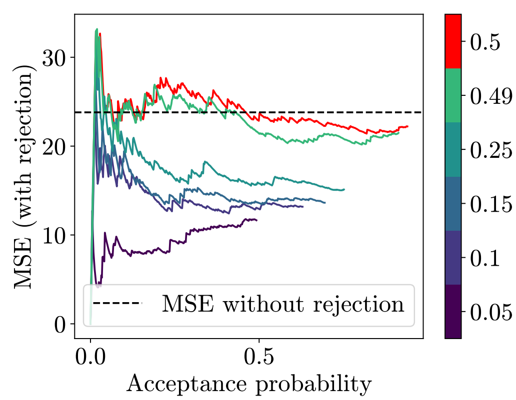
(a) Split by feature 0, acceptance



(b) Split by feature 0, MSE



(c) Split by feature 1, acceptance



(d) Split by feature 1, MSE

Figure 9: Experiments on Airfoil data, split 70/30 by different features.