Efficient reductions from a Gaussian source with applications to statistical-computational tradeoffs

Mengqi Lou[⋆], Guy Bresler[‡], Ashwin Pananjady[⋆,†]

Schools of [⋆]Industrial and Systems Engineering and [†]Electrical and Computer Engineering, Georgia Tech
[‡]Department of Electrical Engineering and Computer Science, MIT

October 9, 2025

Abstract

Given a single observation from a Gaussian distribution with unknown mean θ, we design computationally efficient procedures that can approximately generate an observation from a different target distribution $\mathcal{Q}_\theta$ uniformly for all θ in a parameter set. We leverage our technique to establish reduction-based computational lower bounds for several canonical high-dimensional statistical models under widely-believed conjectures in average-case complexity. In particular, we cover cases in which:

1. $\mathcal{Q}_\theta$ is a general location model with non-Gaussian distribution, including both light-tailed examples (e.g., generalized normal distributions) and heavy-tailed ones (e.g., Student's t-distributions). As a consequence, we show that computational lower bounds proved for spiked tensor PCA with Gaussian noise are universal, in that they extend to other non-Gaussian noise distributions within our class.
2. $\mathcal{Q}_\theta$ is a normal distribution with mean $f(\theta)$ for a general, smooth, and nonlinear link function $f : \mathbb{R} \to \mathbb{R}$. Using this reduction, we construct a reduction from symmetric mixtures of linear regressions to generalized linear models with link function f, and establish computational lower bounds for solving the k-sparse generalized linear model when f is an even function. This result constitutes the first reduction-based confirmation of a k-to-k^2 statistical-to-computational gap in k-sparse phase retrieval, resolving a conjecture posed by Cai et al. (2016). As a second application, we construct a reduction from the sparse rank-1 submatrix model to the planted submatrix model, establishing a pointwise correspondence between the phase diagrams of the two models that faithfully preserves regions of computational hardness and tractability.

Contents

1 Introduction and background

For some unknown real value θ from an uncountably large parameter set Θ, suppose we observe a random draw $X_\theta \sim \mathcal{N}(\theta, 1)$. Consider two simple-to-state puzzles, both of a flavor that dates back to classical questions in mathematical statistics (e.g., Blackwell, 1951, 1953; Karlin and Rubin, 1956; Lindley, 1956; Stone, 1961; Hansen and Torgersen, 1974):

(P1) Can we generate a sample that is close in total variation distance to a random variable drawn from a prespecified, non-Gaussian location model with location parameter θ?

(P2) For a prespecified nonlinear function f, can we generate a sample from a distribution that is close in total variation distance to $\mathcal{N}(f(\theta), \sigma^2)$?

The difficulty of these puzzles lies in the fact that θ is unknown a priori and impossible to estimate from a single observation—nevertheless, our procedure must work for infinitely many values of θ.

A naive approach in both cases is the plug-in estimator. For puzzle (P2) above, this would amount to outputting $f(X_\theta) + Z$ where $Z \sim \mathcal{N}(0, \sigma^2)$ is an independent random variable. Indeed, this is a reasonable strategy for large σ^2: We have

$$\|\mathcal{N}(f(X_\theta), \sigma^2) - \mathcal{N}(f(\theta), \sigma^2)\|_{\mathsf{TV}} \overset{\text{(i)}}{\leq} \sqrt{\frac{1}{2}\mathsf{KL}(\mathcal{N}(f(X_\theta), \sigma^2) \parallel \mathcal{N}(f(\theta), \sigma^2))} \leq \frac{\mathbb{E}|f(X_\theta) - f(\theta)|}{2\sigma} =: \frac{C_{f,\Theta}}{\sigma},$$

where step (i) holds by Pinsker's inequality, and $C_{f,\Theta}$ is uniformly bounded provided f is well-behaved on the parameter set Θ and Θ is bounded. Said a different way, for a natural class of functions f, if the variance σ^2 scales polynomially in $1/\epsilon$, then the total variation distance between $f(X_\theta) + Z$ and the desired $Y_\theta \sim \mathcal{N}(f(\theta), \sigma^2)$ is bounded above by ϵ. We will show later in the paper that even for simple functions f, such a variance blow-up is necessary for the plug-in approach.

One of the main results of this paper is a more sophisticated procedure that achieves this total variation guarantee—for a large class of functions f—with σ^2 scaling only polylogarithmically in $1/\epsilon$. This presents an exponential improvement over the plug-in approach, and together with our other main result, which is a procedure answering puzzle (P1) for large class of non-Gaussian location models, yields several consequences for establishing new statistical-computational tradeoffs in canonical high-dimensional statistical models.

To set up the problem formally in high dimensions, suppose we have two parametric statistical models. The first is the model $\mathcal{U} = (\mathbb{R}^D, \{\mathcal{P}_\xi\}_{\xi \in \Xi})$, also called the source, and the second is $\mathcal{V} = (\mathbb{R}^D, \{\mathcal{Q}_\xi\}_{\xi \in \Xi})$, also called the target. Suppose both the source model $\mathcal{U}$ and the target model $\mathcal{V}$ share the same sample space $\mathbb{R}^D$ and (possibly structured) parameter space $\Xi \subseteq \mathbb{R}^D$, but differ in their respective distributions $\mathcal{P}_\xi$ and $\mathcal{Q}_\xi$. For the moment, this is abstract notation for a high-dimensional model and accommodates isomorphisms of $\mathbb{R}^D$—for instance, the parameter set and observations could be matrices or tensors. Our focus in this paper is specifically on the setting where the source statistical model is the *Gaussian location model*, with $\mathcal{P}_\xi = \mathcal{N}(\xi, \sigma^2 I_D)$. Specifically, for a variance parameter σ^2 that should be thought of as fixed and known, we have

$$\mathcal{U} = (\mathbb{R}^D, \{\mathcal{N}(\xi, \sigma^2 I_D)\}_{\xi \in \Xi}). \tag{1}$$

Now suppose there is some *unknown* $\xi \in \Xi$ fixed by nature. We draw a *single* observation (or sample) from the source distribution, and denote this random variable by $X_\xi \sim \mathcal{N}(\xi, \sigma^2 I_D)$. We would like to transform this high-dimensional source random vector to a D-dimensional random vector over the target sample space that is close to $Y_\xi \sim \mathcal{Q}_\xi$. In particular, our goal is to design a randomized algorithm $\mathsf{K} : \mathbb{R}^D \to \mathbb{R}^D$ that additionally takes some $\epsilon \in (0, 1)$ as input and achieves the following twofold objective:

(A) **Statistical:** Uniformly over $\xi \in \Xi$, the random variable $\mathsf{K}(X_\xi)$ must be ϵ-close in total variation (TV) distance to an output random variable $Y_\xi \sim \mathcal{Q}_\xi$, i.e.,

$$\sup_{\xi \in \Xi} \mathsf{d}_{\mathsf{TV}}(\mathrm{K}(X_\xi), Y_\xi) \leq \epsilon, \qquad \text{and}$$

(B) **Computational:** The transformation algorithm K is computationally efficient, in that it must have (pointwise) runtime polynomial in the dimension D, and in $1/\epsilon$.

We refer to such a transformation algorithm as a polynomial-time *reduction*. It is worth noting that for some target models, there may not exist any reduction satisfying the statistical desideratum (A) above due to data processing considerations (see, e.g., Polyanskiy and Wu, 2025, Chapter 7), and also that it may be possible that there are reductions that satisfy the statistical desideratum (A), but that a computationally efficient one satisfying desideratum (B) is unlikely to exist under widely believed conjectures from average-case complexity (see Lou et al. (2025, Appendix A) for examples).

One of the main applications of such a polynomial-time reduction is that it can be used to transfer computational hardness from the source model to the target model. We briefly explain the underlying idea at a high level[1]. Suppose that—owing to evidence from average-case complexity—we believe that computationally efficient and accurate estimators do not exist for the source model. Now say we have a computationally efficient reduction as above, which transforms a sample from the source model into one that resembles a sample from the target model while keeping the underlying estimand the same. Then computationally efficient and accurate estimation for the target model must be impossible. To see this, suppose (for the sake of contradiction) that there exists a computationally efficient and accurate estimator for the target model. Then we can use this to produce an accurate estimator for the source model as follows: First draw a source sample, then use the reduction to transform it into a target sample and apply the estimator for the target model. Since the reduction and the estimator for the target model run in polynomial time, so does this entire procedure for the source model. But this is impossible by assumption and we have a contradiction. Thus, any accurate estimator for the target model cannot run in polynomial time.

1.1 Motivating examples

We now present concrete high-dimensional examples that motivate our paper. These examples serve as a preview; they are formally presented again in Section 3 along with explicit guarantees. In all these examples, reductions from the high-dimensional source model (1) are key. As we will see shortly, however, it suffices in these cases to design reductions from the *scalar* source

$$\mathcal{U} = (\mathbb{R}, \{\mathcal{N}(\theta, \sigma^2)\}_{\theta \in \Theta}), \tag{2}$$

where $\Theta \subseteq \mathbb{R}$ is some parameter set consisting of real values, and σ^2 is a known variance parameter. Indeed, at the core of these applications are puzzles of the form (P1) and (P2) mentioned above.

Motivation 1. In spiked tensor Principal Component Analysis (PCA) (Montanari and Richard, 2014), we observe a single order-s tensor $T \in \mathbb{R}^{n^{\otimes s}} = \mathbb{R}^{n \times \cdots \times n}$ given by

$$T = \beta(\sqrt{n}v)^{\otimes s} + \zeta, \tag{3}$$

[1]Rigorous formulations of these computational implications can be found in (e.g., Berthet and Rigollet, 2013b; Ma and Wu, 2015; Brennan et al., 2018).

where $\beta \in \mathbb{R}$ denotes the signal strength, $v \in \mathbb{S}^{n-1}$ is the unknown spike, and $\zeta \in \mathbb{R}^{n^{\otimes s}}$ is a symmetric Gaussian noise tensor (see Eq. (42)). Based on a reduction from the secret leakage planted clique conjecture (Brennan and Bresler, 2020) or hypergraph planted clique conjecture (Luo and Zhang, 2022), it is believed that no polynomial-time algorithm can succeed at either detection or recovery when $\beta = \widetilde{o}(n^{-s/4})$. In contrast, recovery is information-theoretically possible for a much weaker signal strength $\beta = \widetilde{\omega}(n^{(1-s)/2})$ via exhaustive search (Montanari and Richard, 2014; Lesieur et al., 2017b; Chen, 2019; Jagannath et al., 2020; Perry et al., 2020).

A key question concerns the universality of the computational barrier: Does the same computational lower bound persist when entry-wise Gaussian noise ζ is replaced by a zero-mean, non-Gaussian distribution, for instance one having lighter or heavier tails? As argued above, one way to affirmatively answer this question is to reduce the original tensor T to one with non-Gaussian noise. In addition, observe that for each index of the tensor $(i_1, \ldots, i_s)$, the entry satisfies $T_{i_1,\ldots,i_s} \sim \mathcal{N}(\theta, 1)$, where $\theta = \beta n^{s/2} v_{i_1} \cdots v_{i_s}$ is an unknown mean. Since the noise is independent across entries, the core problem is to design a one-dimensional reduction from the source model (2) (with $\mathcal{P}_\theta = \mathcal{N}(\theta, \sigma^2)$) to a general non-Gaussian target location model $\mathcal{Q}_\theta$, as posed in puzzle (P1). Executing these one-dimensional reductions entry-by-entry is one way to arrive at a polynomial-time, high-dimensional reduction.

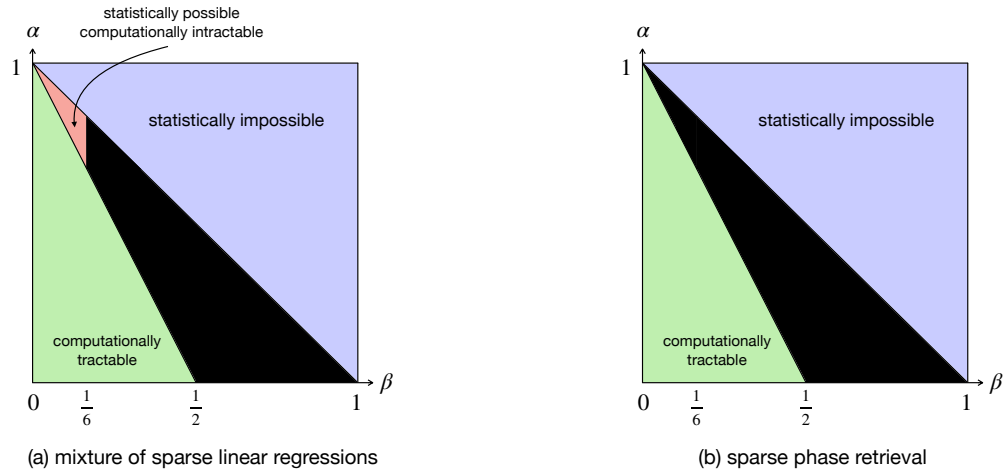


Figure 1. Phase diagrams for the sparse mixture of linear regressions (MixSLR; panel (a)) and sparse phase retrieval (SPR; panel (b)), under the parameterization $k = \widetilde{\Theta}(n^\beta)$ and $\eta^4 = \widetilde{\Theta}(n^{-\alpha})$, with constants $\alpha, \beta \in (0, 1)$. The purple region denotes the statistically impossible regime; the red region denotes the statistically possible but computationally intractable regime; the green region denotes the computationally tractable regime; and the black region denotes a statistically possible yet conjectured computationally hard regime, whose proof remains open. Our goal is to transfer all computationally hard regimes confirmed in panel (a) to panel (b).

Motivation 2. In symmetric mixtures of sparse linear regressions (MixSLR) (Quandt and Ramsey, 1978; Städler et al., 2010), we observe i.i.d. samples $\{(x_i, y_i)\}_{i=1}^n$ generated by

$$y_i = R_i \cdot \langle x_i, v \rangle + \zeta_i, \quad \text{where} \quad \zeta_i \sim \mathcal{N}(0, 1).$$

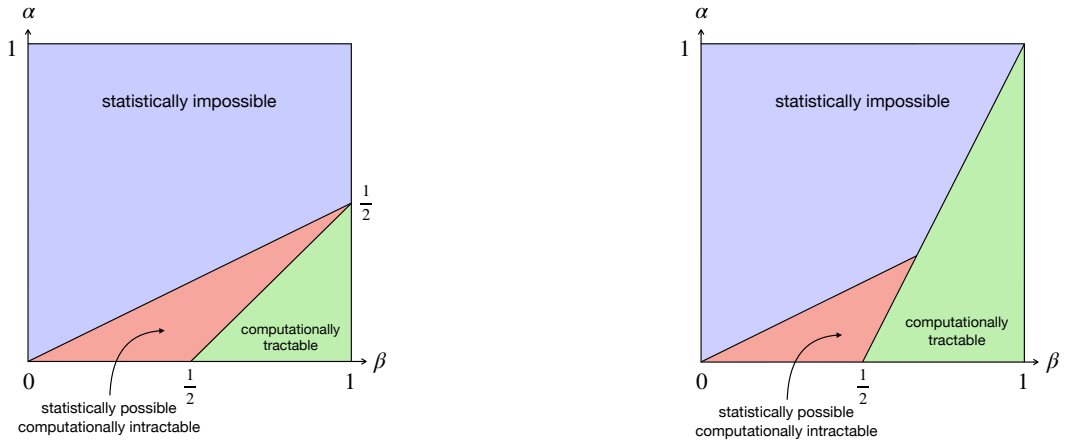
Here, $R_i \in \{-1, 1\}$ is a hidden label and $v \in \mathbb{R}^d$ is the unknown k-sparse parameter. There is a so-called k-to-k^2 statistical-to-computational gap in MixSLR. Information theoretically, we require $n = \widetilde{\Theta}\big(k \log(d)/\|v\|_2^4\big)$ for estimating v to within ℓ_2 distance $\|v\|_2/2$ (Fan et al., 2018), yet there is

reduction-based evidence from the secret leakage planted clique conjecture that no polynomial-time algorithm can succeed when $n = \widetilde{o}\big(k^2/\|v\|_2^4\big)$ (Brennan and Bresler, 2020).

A related but distinct model is the sparse generalized linear model (SpGLM), in which samples follow

$$\widetilde{y}_i = f\big(\langle x_i, v\rangle\big) + \zeta_i, \quad \text{where} \quad \zeta_i \sim \mathcal{N}(0,1),$$

where $f : \mathbb{R} \to \mathbb{R}$ is the link function. Note that the computational gap in MixSLR arises due to the latent signs—in light of this, it is natural to ask: Does the same k^2 computational barrier persist in SpGLM with even link functions, which also destroy sign information? In particular, when $f(t) = t^2$ we obtain the familiar sparse phase retrieval model (Li and Voroninski, 2013), and it was conjectured by Cai et al. (2016) that there is a k^2 computational barrier, i.e., that no polynomial time algorithm can succeed when $n = \widetilde{o}(k^2)$. However, no reduction-based evidence for this conjecture has been established. A natural way by which one might establish such computational hardness is by transforming samples $\{x_i, y_i\}_{i=1}^n$ from MixSLR to samples $\{x_i, \widetilde{y}_i\}_{i=1}^n$ from SpGLM, thereby allowing us to transfer any computationally hard regimes that have been confirmed in Figure 1(a) to Figure 1(b). Note that $y_i \sim \mathcal{N}(\theta, 1)$ and $\widetilde{y}_i \sim \mathcal{N}(f(\theta), 1)$ for some unknown mean $\theta = R_i \cdot \langle x_i, v\rangle$, and the noise is independent and Gaussian across samples. Thus, a one-dimensional reduction from the Gaussian source (2) suffices: We must design a reduction from $\mathcal{P}_\theta = \mathcal{N}(\theta, 1)$ to target $\mathcal{Q}_\theta = \mathcal{N}(f(\theta), 1)$, as alluded to in puzzle (P2).



(a) Sparse rank-1 submatrix model (b) Sparse planted submatrix model

Figure 2. Phase diagrams for the sparse rank-1 submatrix model (Rank1SubMat; panel (a)) and the planted submatrix model (PlantSubMat; panel (b)), under the parameterization $\mu/k = \widetilde{\Theta}(n^{-\alpha})$ and $k = \widetilde{\Theta}(n^{\beta})$ for $\alpha, \beta \in (0,1)$. The purple region indicates the statistically impossible regime. The green region corresponds to the computationally tractable regime where polynomial-time algorithms are known to succeed. The red region indicates the statistically possible but computationally intractable regime. In panel (a), the computational threshold is given by the line $\alpha = \beta - 1/2$, while in panel (b), the threshold is the line $\alpha = 2\beta - 1$.

Motivation 3. In the sparse rank-one submatrix model (Rank1SubMat), we observe

$$M = \mu\, rc^\top + G, \quad \text{where} \quad G_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1).$$

Here, $\mu \in \mathbb{R}$ is the signal strength and $r, c \in \mathbb{R}^n$ are k-sparse unit-norm vectors with both positive and negative entries. In the planted submatrix model (PlantSubMat), M has the same form but r, c are restricted to be k-sparse unit-norm vectors with only nonnegative entries. Existing

reduction-based computational hardness results for PlantSubMat and Rank1SubMat are based on the planted clique conjecture (Brennan and Bresler, 2020; Ma and Wu, 2015); see Figure 2 for their phase diagrams. However, computational hardness for each of these problems is established *individually* through reductions from the planted clique model. A natural question is whether, without relying on the planted clique conjecture, one can construct a *direct* reduction from Rank1SubMat to PlantSubMat. Note that once again, this would involve an entry-by-entry reduction from a Gaussian source. In principle, such a reduction would allow one to transfer computational hardness directly from instances of Rank1SubMat to instances of PlantSubMat, allowing us to move beyond computational hardness that can be established through reductions from planted clique. Concretely, while estimation in the planted clique model can be performed using a quasi-polynomial time algorithm, it is conjectured that Rank1SubMat requires exponential time for certain regimes of signal-to-noise ratio (Ding et al., 2024). Stronger computational hardness of this flavor cannot be transferred to PlantSubMat via a direct reduction from planted clique. Also note that unlike in the previous examples, Figure 2 suggests that the two models exhibit different computational thresholds. Can a reduction between them be designed to preserve both computational tractability and intractability (i.e., map tractable regimes to tractable regimes and computationally hard regimes to hard regimes)?

1.2 Setup and contributions

Having motivated the study of reductions from a Gaussian source, we now state our contributions. We begin with some setup. Given a real-valued parameter $\theta \in \Theta$, we let $u(\cdot;\theta) : \mathbb{R} \to \mathbb{R}$ denote the density of the source distribution $\mathcal{P}_\theta = \mathcal{N}(\theta, \sigma^2)$. Specifically, we let

$$u(x;\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right), \quad \text{for all } x, \theta \in \mathbb{R}. \tag{4}$$

Similarly, we let $v(\cdot;\theta) : \mathbb{R} \to \mathbb{R}$ denote the density of the target distribution $\mathcal{Q}_\theta$ with respect to Lebesgue measure. A mapping $\mathcal{T}(\cdot \mid \cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is called a Markov kernel if $\mathcal{T}(\cdot \mid x)$ is a density on $\mathbb{R}$ for all $x \in \mathbb{R}$. For a Markov kernel $\mathcal{T}$, we let

$$\delta(\mathcal{U}, \mathcal{V}; \mathcal{T}) := \sup_{\theta \in \Theta} \left\| \int_{\mathbb{R}} \mathcal{T}(\cdot \mid x) \cdot u(x;\theta) \mathrm{d}x - v(\cdot;\theta) \right\|_{\mathsf{TV}} \tag{5}$$

denote the TV-deficiency attained by $\mathcal{T}$ when mapping the source statistical model $\mathcal{U}$ to the target statistical model $\mathcal{V}$.

Main contributions. We now provide an overview of our main contributions.

(i) **Design and analysis of two reductions.** We construct a Markov kernel $\mathcal{T}$ from the Gaussian location model $\mathcal{U}$ in Eq. (2) to a general class of target models $\mathcal{V}$ and establish an upper bound on the TV-deficiency $\delta(\mathcal{U}, \mathcal{V}; \mathcal{T})$ in terms of properties of $\mathcal{V}$; see Section 2.1. We use this Markov kernel to construct our single-sample reduction algorithm K.

We then apply this framework to two concrete classes of target models and show that the reduction K is both statistically and computationally efficient. In particular:

– When $\mathcal{V}$ corresponds to a class of non-Gaussian location models with a suitably proxy for signal-to-noise ratio decaying polylogarithmically in $1/\epsilon$, Theorem 1 guarantees that our reduction achieves ϵ-TV deficiency and runs in time polynomial in $1/\epsilon$ for arbitrarily small $\epsilon \in (0, 1)$. We allow for the parameter to be any real value, and cover light-tailed

target models such as the generalized normal as well as heavy-tailed models such as Student's t distributions, among others. See Corollaries 1 and 2 for details.

- When $\mathcal{V}$ is a Gaussian model with mean $f(\theta)$ and a suitably proxy for signal-to-noise ratio decays polylogarithmically in $1/\epsilon$, Theorem 2 guarantees that our reduction achieves ϵ-TV deficiency and runs in time polynomial in $1/\epsilon$ for arbitrarily small $\epsilon \in (0, 1)$. Under only the assumption that the parameter set is suitably bounded, we cover functions that are constant-degree polynomials or globally analytic, among others. See Corollaries 3 and 4 for details.

(ii) **Applications to computational hardness.** Using these reductions, we obtain affirmative answers to several open questions on computational hardness in high-dimensional statistics:

- We prove a reduction-based universality result for Tensor PCA, showing that the same computational barrier as in the Gaussian case persists for several non-Gaussian noise distributions (including both lighter-tailed and heavier-tailed cases); see Corollary 5.
- We construct a reduction from MixSLR to SpGLM for a general class of even link function, thereby confirming the existence of a k^2 computational barrier for k-sparse generalized linear models with an even link; see Corollary 6. This provides the first reduction-based resolution of the conjectured k-to-k^2 statistical-to-computational gap in sparse phase retrieval posed by Cai et al. (2016). See Corollary 7 for details.
- We construct a reduction from Rank1SubMat to PlantSubMat, yielding a pointwise correspondence between their phase diagrams (Figure 4) and showing that the reduction faithfully preserves both computational tractability and intractability; see Theorem 5.

1.3 Related work

We review related work under two verticals. Given that our focus is on reductions, the first vertical will discuss other reduction-based approaches to proving computational hardness in high-dimensional statistical models. In our second vertical, we provide details on other approaches to proving hardness that are not reduction-based, wherein computational hardness is proved within some (restricted) classes of algorithms. Given that this latter literature is vast, we focus on papers that provide lower bounds for the high-dimensional problems outlined in Section 1.1.

Reduction-based computational hardness. The program of transferring computational hardness between problems in high-dimensional statistics was initiated roughly a decade ago (Berthet and Rigollet, 2013a,b; Wainwright, 2014; Ma and Wu, 2015; Hajek et al., 2015). Since then, the framework of reductions has been applied to an extensive list of high-dimensional statistical problems (Brennan et al., 2018; Brennan and Bresler, 2020). Specific examples include principal component analysis with structure (Brennan and Bresler, 2019; Wang et al., 2024; Bresler and Harbuzova, 2025), mixture models with adversarial corruptions (Brennan and Bresler, 2020), block models in matrix and tensor settings (Brennan and Bresler, 2020; Luo and Zhang, 2022; Han et al., 2022), tensor PCA (Brennan and Bresler, 2020), and adaptation lower bounds in shape-constrained regression problems (Shah et al., 2019; Pananjady and Samworth, 2022). Recently, reductions have been constructed from hard problems that have seen a long line of work in cryptography (Gupte et al., 2022; Bruna et al., 2021; Gupte et al., 2024; Bangachev et al., 2025).

This body of work introduces many powerful tools and techniques to transfer structure between problems, but it cannot handle continuous-valued inputs or parameter spaces like the Gaussian location model—indeed, all the distributional reductions in this literature are from source distributions

taking either two or three values (Hajek et al., 2015; Ma and Wu, 2015; Brennan et al., 2018; Brennan and Bresler, 2020). There are some classical reductions that handle Gaussian sources, but they are only effective when the target is also Gaussian (see the books by Shiryaev and Spokoiny (2000) and Torgersen (1991) for examples). To our knowledge, the only other single-sample reductions from continuous valued source models—apart from a few specialized source-target pairs (Hansen and Torgersen, 1974; Lehmann, 1988)—are from our own prior work (Lou et al., 2025), which addresses reductions from Laplace, Erlang, and uniform source models to general target models. However, the tools introduced in that paper do not apply to the Gaussian source model (2).

Other literature that is related in spirit to the reduction framework includes the problems of simultaneous optimal transport (Wang and Zhang, 2022) and sample amplification (Axelrod et al., 2024). However, both these problems differ in their focus.

Computational hardness for restricted algorithm classes. There are several techniques that have been developed to prove computational hardness over restricted classes of algorithms, including those based on low-degree polynomials (Kunisky et al., 2019; Wein, 2025), sum-of-squares hierarchies (Raghavendra et al., 2018), and the statistical query model (Kearns, 1998). As mentioned before, each is a vast field, so we focus on particular papers that address our problems of interest in Section 1.1.

Tensor PCA. For tensor PCA with Gaussian noise (3), a large body of work has demonstrated that many natural classes of algorithms fail when the signal strength falls below the computational threshold $\beta = \widetilde{o}(n^{-s/4})$. This includes low-degree polynomial algorithms (Kunisky et al., 2019), sum-of-squares relaxations (Hopkins et al., 2015, 2016, 2017; Hopkins, 2018; Kim et al., 2017), approximate message passing (AMP) algorithms (Montanari and Richard, 2014; Lesieur et al., 2017b), algorithms based on the Kikuchi hierarchy (Wein et al., 2019), statistical query algorithms (Dudeja and Hsu, 2021; Brennan et al., 2021), and Langevin dynamics applied to the maximum likelihood objective (Ben Arous et al., 2020). Additional evidence for computational hardness has been shown via landscape analysis (Ben Arous et al., 2019; Ros et al., 2019), predictions from statistical physics (Bandeira et al., 2018), and communication-complexity lower bounds (Dudeja and Hsu, 2024). Universality of hardness for other noise distributions was recently shown by Kunisky (2025) for the class of low-coordinate degree algorithms.

Sparse generalized linear models. For sparse generalized linear models (SpGLM), several works have established computational barriers. In particular, low-degree polynomial methods require at least k^2 samples for exact support recovery in k-sparse phase retrieval with Gaussian covariates in the noiseless setting (Arpino and Venkataramanan, 2023). Note that this problem is identical to the noiseless MixSLR problem with Gaussian covariates. Using the statistical query framework, it has also been shown that no computationally efficient algorithm can attain the information-theoretic minimax risk for SpGLM with even link functions (Wang et al., 2019). This work builds on a general framework based on showing statistical-query lower bounds for *non-Gaussian component analysis* problems (Diakonikolas et al., 2017; Diakonikolas and Kane, 2023). Beyond sparsity, statistical-computational gaps in generalized linear models arise can depending on the choice of link function, and lower bounds have been established in this setting via both statistical query and low-degree polynomial methods (see, e.g., Damian et al., 2024).

Sparse submatrix models. For sparse submatrix detection and recovery problems (Rank1SubMat and PlantSubMat), evidence of computational hardness comes from multiple fronts: low-degree polynomial lower bounds (Ding et al., 2024; Schramm and Wein, 2022; Wein, 2025; Kunisky et al., 2019), failure of AMP (Lesieur et al., 2015, 2017a; Hajek et al., 2018; Macris et al., 2020), sum-of-squares lower bounds (Ma and Wigderson, 2015; Hopkins et al., 2017), and the overlap gap

property (Gamarnik et al., 2021; Ben Arous et al., 2023). Statistical query lower bounds have also been connected to the low-degree approach in such problems (Brennan et al., 2021). Notably, in PlantSubMat, there exists a gap between detection and recovery thresholds (Chen and Xu, 2016; Schramm and Wein, 2022; Ben Arous et al., 2023).

Remark 1. *In spite of the many different methods now in use to prove computational hardness in statistical problems, reductions remain an ironclad route to establishing fundamental relationships between the average-case computational complexity of different problems. As discussed above, the reduction approach differs from proving computational hardness for restricted classes of algorithms—instead, it directly addresses the question of polynomial-time solvability* by any algorithm *assuming hardness of some base problem. While showing hardness for a restricted class of algorithms does not depend on assumed hardness for any other problem, the ramification of such a result depends crucially on the belief that that class of algorithms is actually optimal for the problem at hand. It is by now well-known that each of these restricted classes of algorithms can be suboptimal for simple problems, and it is still unclear what type of problem each algorithm class solves optimally. For instance, the Lenstra–Lenstra–Lovasz (LLL) lattice basis reduction algorithm can be used to solve some of these problems in polynomial time with state-of-the-art guarantees (e.g., Andoni et al., 2017; Zadik et al., 2022; Diakonikolas and Kane, 2023), but it is not a low-degree polynomial algorithm or a statistical query algorithm. For another example, it has recently been observed that algorithmically simple problems such as shortest path on random graphs can exhibit the overlap gap property (Li and Schramm, 2024). Reductions, however, are immune to such specialized counterexamples. More generally, a key property of reductions is that they can work synergistically with analyses for restricted algorithmic approaches by clarifying what transformations preserve polynomial-time solutions to problems.*

1.4 Notation and organization

For $a, b \in \mathbb{R}$, we let $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. We use $\mathcal{L}[X]$ to denote the law of a random variable X. For two (possibly signed) measures μ and ν, we use $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ to denote the Radon–Nikodym derivative of ν with respect to μ. We use $\mathcal{N}(\theta, \sigma^2)$ to denote a normal distribution with mean θ and variance σ^2. We use $\mathsf{GN}(\theta, \tau, \beta)$ to denote a Generalized normal distribution with mean θ, scale τ, and shape β. Let $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\mathrm{d}t$ for all $z > 0$ be the Gamma function. Let $\mathbb{N} = \{1, 2, \dots\}$ denote the set of natural numbers and let $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\}$ denote the set of non-negative integers. Let $[n]$ denote the set of natural numbers less than or equal to n. For $n \in \mathbb{N}$, we define double factorial $n!! = \prod_{k=1}^{n/2}(2k)$ if n is even and $n!! = \prod_{k=1}^{(n+1)/2}(2k-1)$ if n is odd. For a function of two variables $v(y;\theta)$ and $k \in \mathbb{Z}_{\geq 0}$, we write $\nabla_\theta^{(k)} v(y;\theta)\big|_{\theta=\widetilde{\theta}}$ to denote the k-th derivative of v with respect to its second argument θ, evaluated at $\theta = \widetilde{\theta}$. Similarly, for a function $f : \mathbb{R} \to \mathbb{R}$, we let $f^{(k)}(x)$ to denote the k-th derivative of f evaluated at x, i.e., $f^{(k)}(x) = \frac{\mathrm{d}^k f(t)}{\mathrm{d}t^k}\big|_{t=x}$. For a positive integer s, we let $\mathfrak{S}_s$ denote the set of all permutations of $[s]$, i.e., $\mathfrak{S}_s = \{\pi : [s] \to [s] \mid \pi \text{ is a bijection}\}$. Following the notation in Raginsky (2011), we use $\|\mu - \nu\|_{\mathsf{TV}}$ to denote the total variation distance between two measures (μ, ν) defined on the same space. If $X \sim \mu$ and $Y \sim \nu$, then we use the notation $\mathsf{d}_{\mathsf{TV}}(X, Y) := \|\mu - \nu\|_{\mathsf{TV}}$. We define the (physicist's) Hermite polynomials $H_n : \mathbb{R} \to \mathbb{R}$ as

$$H_n(x) = (-1)^n e^{x^2} \cdot \frac{\mathrm{d}^n\, e^{-x^2}}{\mathrm{d}x^n}, \quad \text{for all} \quad x \in \mathbb{R} \text{ and } n = 0, 1, 2, \dots. \tag{6}$$

It will be convenient later to work with the normalized Hermite polynomials

$$\overline{H}_n(x) = \frac{H_n(x)}{\left(\sqrt{\pi} 2^n n!\right)^{1/2}}, \quad \text{for all} \quad x \in \mathbb{R} \text{ and } n = 0, 1, 2, \ldots. \tag{7}$$

For two functions $f, g : \mathbb{R} \to \mathbb{R}$, define their inner product as

$$\langle f, g \rangle := \int_{\mathbb{R}} f(x) g(x) \exp(-x^2) \mathrm{d}x. \tag{8}$$

Note that the normalized Hermite polynomials are orthonormal in that $\left\langle \overline{H}_n(\cdot), \overline{H}_m(\cdot) \right\rangle = \delta_{nm}$, where δ_{nm} is the Kronecker delta with $\delta_{nm} = \mathbb{I}\{n = m\}$. We use (c, c', C, C') to denote universal positive constants that may take different values in each instantiation. For two sequences of non-negative reals $\{f_n\}_{n \geq 1}$ and $\{g_n\}_{n \geq 1}$, we use $f_n \lesssim g_n$ to indicate that there is a universal positive constant C such that $f_n \leq C g_n$ for all $n \geq 1$. The relation $f_n \gtrsim g_n$ indicates that $g_n \lesssim f_n$, and we say that $f_n \asymp g_n$ if both $f_n \lesssim g_n$ and $f_n \gtrsim g_n$ hold simultaneously. We also use standard order notation $f_n = \mathcal{O}(g_n)$ to indicate that $f_n \lesssim g_n$ and $f_n = \widetilde{\mathcal{O}}(g_n)$ to indicate that $f_n \lesssim g_n \log^c n$, for a universal constant $c > 0$. We say that $f_n = \Omega(g_n)$ (resp. $f_n = \widetilde{\Omega}(g_n)$) if $g_n = \mathcal{O}(f_n)$ (resp. $g_n = \widetilde{\mathcal{O}}(f_n)$). We say $f_n = \Theta(g_n)$ (resp. $f_n = \widetilde{\Theta}(g_n)$) if $f_n = \Omega(g_n)$ and $f_n = \mathcal{O}(g_n)$ hold simultaneously (resp. $f_n = \widetilde{\Omega}(g_n)$ and $f_n = \widetilde{\mathcal{O}}(g_n)$). The notation $f_n = o(g_n)$ is used when $\lim_{n \to \infty} f_n / g_n = 0$, and $f_n = \omega(g_n)$ when $g_n = o(f_n)$. We write $\Omega_n(1)$ to denote a sequence that is bounded below by some universal and positive constant for all n large enough. Similarly, $o_n(1)$ denotes a non-negative sequence that converges to 0 as n goes to infinity. For $x > 0$, we write $\mathsf{poly}(x)$ to denote a polynomial function of x and write $\mathsf{polylog}(x)$ to denote a polynomial function of $\log(x)$. All logarithms are to the natural base e unless otherwise stated. We write $\mathsf{sign}(z) = 1$ if $z \geq 0$ and -1 otherwise. We use $\mathbb{I}\{\cdot\}$ to denote the indicator function.

Organization. The remainder of this paper is organized as follows. In Section 2, we present concrete reductions from the Gaussian source model to general target models, focusing on the cases where the target is either a general location model or a Gaussian model with a nonlinear mean transformation. In Section 3, we apply these reductions to establish computational lower bounds for several high-dimensional models, thereby addressing the questions raised in Motivations 1–3. We conclude in Section 4 with a discussion of open problems. Proofs of all results are provided in Sections 5, 6, 7, and 8.

2 General families of reductions from the Gaussian source

In this section, we present a general family of reductions from the Gaussian location model. Specifically, in Section 2.1, we develop methodology for constructing a Markov kernel $\mathcal{T}$ that maps the Gaussian source $\mathcal{U}$ (2) to a general target $\mathcal{V}$, and supply a bound on its TV deficiency—in Lemma 1—as a function of various properties of the target model. In Section 2.2, we apply our Markov kernel to cases where the target $\mathcal{V}$ is a non-Gaussian location model and prove explicit bounds on the TV deficiency $\delta(\mathcal{U}, \mathcal{V}; \mathcal{T})$. In Section 2.3, we consider the case where the target is Gaussian with mean given by a nonlinear transformation $f : \mathbb{R} \to \mathbb{R}$, i.e., $\mathcal{V} = (\mathbb{R}, \{\mathcal{N}(f(\theta), 1)\}_{\theta \in \Theta})$, and again provide explicit bounds on $\delta(\mathcal{U}, \mathcal{V}; \mathcal{T})$.

2.1 Constructing a general Markov kernel

In this section, we describe the main idea behind constructing a reduction from the Gaussian location model (2) to a general target statistical model

$$\mathcal{V} = (\mathbb{R}, \{\mathcal{Q}_\theta\}_{\theta\in\Theta}).$$

Since our approach is based on taking higher-order derivatives, we require some mild regularity conditions on the target density $v(\cdot;\theta)$.

Assumption 1. *The following conditions hold:*

(a) The higher-order derivatives $\nabla_x^{(k)} v(y;x)$ exist for all $x, y \in \mathbb{R}$ and $k \in \mathbb{Z}_{\geq 0}$.

(b) The higher-order derivatives vanish at infinity. In particular, for all $k \in \mathbb{Z}_{\geq 0}$, $y \in \mathbb{R}$, and $\theta \in \Theta$,

$$\lim_{x\uparrow+\infty} \left[\nabla_x^{(k)} v(y;x)\right] \cdot e^{-\frac{(x-\theta)^2}{2\sigma^2}} = \lim_{x\downarrow-\infty} \left[\nabla_x^{(k)} v(y;x)\right] \cdot e^{-\frac{(x-\theta)^2}{2\sigma^2}} = 0.$$

(c) The function $v\big(y; \sqrt{2}\sigma x + \theta\big)$ satisfies, for all $y \in \mathbb{R}$ and $\theta \in \Theta$,

$$\int_{\mathbb{R}} \Big(v\big(y; \sqrt{2}\sigma x + \theta\big)\Big)^2 \exp(-x^2)\mathrm{d}x < +\infty.$$

In words, Assumption 1(a) requires $v(y;x)$ to be infinitely differentiable with respect to x; Assumption 1(b) requires that the partial derivatives $\nabla_x^{(k)} v(y;x)$ do not grow faster than exponentially, which is usually satisfied in most applications of interest; and Assumption 1(c) places a boundedness condition that typically holds since $v(y;x)$ is uniformly bounded, which in turn ensures that the integral is finite.

Let us now describe the core observations behind our construction. Clearly, any reduction K corresponds to a Markov kernel $\mathcal{T}(\cdot \mid \cdot) : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$—given a sample $X_\theta \sim \mathcal{N}(\theta, \sigma^2)$, the reduction generates the random variable $\mathrm{K}(X_\theta)$ by sampling from the conditional density $\mathcal{T}(\cdot \mid X_\theta)$. As a result, the marginal density of $\mathrm{K}(X_\theta)$ is given by $\int_{\mathbb{R}} \mathcal{T}(\cdot \mid x) \cdot u(x;\theta)\,\mathrm{d}x$. Minimizing the total variation (TV) distance between the distribution of $\mathrm{K}(X_\theta)$ and the target distribution $\mathcal{Q}_\theta$ thus amounts to solving the following optimization problem:

$$\begin{aligned} \delta(\mathcal{U},\mathcal{V}) := \inf_{\mathcal{T}(\cdot|\cdot):\mathbb{R}\times\mathbb{R}\to\mathbb{R}} \sup_{\theta\in\Theta} \ & \frac{1}{2}\left\| \int_{\mathbb{R}} \mathcal{T}(\cdot \mid x)\cdot u(x;\theta)\,\mathrm{d}x - v(\cdot;\theta)\right\|_1, \\ \text{subject to} \quad & \mathcal{T}(y \mid x) \geq 0 \quad \text{and} \quad \int_{\mathbb{R}} \mathcal{T}(y' \mid x)\,\mathrm{d}y' = 1 \quad \text{for all } x, y \in \mathbb{R}. \end{aligned} \tag{9}$$

The optimization problem (9), while linear, is infinite-dimensional. It can thus be computationally inefficient to solve, and analyzing its optimal value is challenging due to the complexity of the constraint set. Our approach is thus to pursue a feasible (instead of optimal) Markov kernel $\mathcal{T}$ and to bound its TV deficiency $\delta(\mathcal{U},\mathcal{V};\mathcal{T})$. We proceed in a few steps:

Step 1: Construct a good signed kernel.

Let $\mathcal{S}(\cdot \mid \cdot) : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be a bivariate function such that $S(\cdot \mid x)$ is a measurable function for all $x \in \mathbb{R}$. For any such *signed kernel* $\mathcal{S}$, define

$$\varsigma(\mathcal{U},\mathcal{V};\mathcal{S}) := \sup_{\theta\in\Theta} \frac{1}{2}\left\| \int_{\mathbb{R}} \mathcal{S}(\cdot \mid x)\cdot u(x;\theta)\mathrm{d}x - v(\cdot;\theta)\right\|_1, \tag{10}$$

which is analogous to $\delta(\mathcal{U}, \mathcal{V}; \mathcal{T})$ defined in Eq. (5). The analogous optimization problem to Eq. (9) is then given by the unconstrained problem

$$\varsigma(\mathcal{U}, \mathcal{V}) := \inf_{\mathcal{S}(\cdot|\cdot):\mathbb{R}\times\mathbb{R}\to\mathbb{R}} \sup_{\theta\in\Theta} \frac{1}{2} \left\| \int_{\mathbb{R}} \mathcal{S}(\cdot \mid x) \cdot u(x;\theta)\, \mathrm{d}x - v(\cdot;\theta) \right\|_1. \tag{11}$$

Our first observation is that the optimization problem (11) admits a closed-form solution. We state this result as the following proposition.

Proposition 1. *Consider the source model $\mathcal{U} = (\mathbb{R}, \{\mathcal{N}(\theta, \sigma^2)\}_{\theta\in\Theta})$ and general target model $\mathcal{V} = (\mathbb{R}, \{\mathcal{Q}_\theta\}_{\theta\in\Theta})$. Let $u(\cdot;\theta)$ be defined in Eq. (4) and let $v(\cdot;\theta) : \mathbb{R} \to \mathbb{R}$ be the density of distribution $\mathcal{Q}_\theta$. Suppose Assumption 1 holds for $\Theta \subseteq \mathbb{R}$ and for all $\theta \in \Theta$,*

$$\sum_{k=0}^{+\infty} \int_{\mathbb{R}} \frac{\sigma^{2k}}{(2k)!!} \left| \nabla_x^{(2k)} v(y;x) \right| \cdot u(x;\theta)\, \mathrm{d}x < \infty. \tag{12}$$

Consider the signed kernel defined as

$$\mathcal{S}^\star(y \mid x) := \sum_{k=0}^{+\infty} \frac{(-1)^k \sigma^{2k}}{(2k)!!} \nabla_\theta^{(2k)} v(y;\theta)\big|_{\theta=x} \quad \text{for all } x, y \in \mathbb{R}. \tag{13}$$

Then we have $\varsigma(\mathcal{U}, \mathcal{V}; \mathcal{S}^\star) = 0$.

We provide the proof of Proposition 1 in Section 5.1. Note that the regularity condition (12) is a technical assumption imposed to justify the interchange of the integral and the infinite sum via Fubini's theorem. In later sections, we will make use of a modified signed kernel that does not require this condition.

While Proposition 1 provides a closed-form expression for the optimal signed kernel, computing the signed kernel $\mathcal{S}^\star$ defined in Eq. (13) is computationally intractable since it requires evaluating $2k$-th order derivatives $\nabla_\theta^{(2k)} v(y;\theta)$ for infinitely many values of k. To address this, we consider a truncated version of the kernel. Specifically, for $N \in \mathbb{N}$, we define the truncated signed kernel as

$$\mathcal{S}_N^\star(y \mid x) = \sum_{k=0}^{N} \frac{(-1)^k \sigma^{2k}}{(2k)!!} \nabla_\theta^{(2k)} v(y;\theta)\big|_{\theta=x} \quad \text{for all } x, y \in \mathbb{R}. \tag{14}$$

In light of Proposition 1, we should expect that the signed deficiency $\varsigma(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star)$ is small for large N. We quantify the computational cost of evaluating the signed kernel $\mathcal{S}_N^\star$ (14) using the following definition.

Definition 1. *For each $N \in \mathbb{N}$, define $T_{\mathsf{eval}}(N)$ as the maximum time taken to evaluate $\mathcal{S}_N^\star(y \mid x)$ for any $x, y \in \mathbb{R}$.*

In the specific applications that follow, we will provide explicit bounds on $T_{\mathsf{eval}}(N)$.

Step 2: Jordan decomposition to find a valid Markov kernel.

Having motivated (through Proposition 1) why the signed kernel $\mathcal{S}_N^\star$ defined in Eq. (14) may serve as a good choice to solve the optimization problem (11), we now construct a valid Markov kernel

by retaining only the positive part of $\mathcal{S}_N^\star$ and normalizing. Specifically, since we have the Jordan decomposition of the signed measure $\mathcal{S}_N^\star(\cdot \mid x) = [\mathcal{S}_N^\star(y \mid x) \vee 0] + [\mathcal{S}_N^\star(\cdot \mid x) \wedge 0]$, we define

$$\mathcal{T}_N^\star(y \mid x) = \frac{\mathcal{S}_N^\star(y \mid x) \vee 0}{\int_{\mathbb{R}} \left[\mathcal{S}_N^\star(y' \mid x) \vee 0\right] \mathrm{d}y'}, \quad \text{for all } x, y \in \mathbb{R}. \tag{15}$$

Note that by construction, for each $x \in \mathbb{R}$, $\mathcal{T}_N^\star(\cdot \mid x)$ is a density.

Lemma 1 of Lou et al. (2025) yields the Markov kernel deficiency bound

$$\delta\big(\mathcal{U}, \mathcal{V}; \mathcal{T}_N^\star\big) \le \varsigma\big(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star\big) + \frac{1}{2} \sup_{\theta \in \Theta} \int_{\mathbb{R}} \big(|p(x) - 1| + q(x)\big)\, u(x; \theta)\, \mathrm{d}x, \tag{16}$$

where the functions $p, q : \mathbb{R} \to [0, +\infty)$ are given by

$$p(x) := \int_{\mathbb{R}} \left[\mathcal{S}_N^\star(y' \mid x) \vee 0\right] \mathrm{d}y' \quad \text{and} \quad q(x) := -\int_{\mathbb{R}} \left[\mathcal{S}_N^\star(y' \mid x) \wedge 0\right] \mathrm{d}y'. \tag{17}$$

Specifically, the second term on the right-hand side of Eq. (16) quantifies the discrepancy between the signed kernel $\mathcal{S}_N^\star$ and the Markov kernel $\mathcal{T}_N^\star$ introduced by the truncation and normalization in Eq. (15). The following lemma provides an upper bound on the first term $\varsigma\big(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star\big)$.

Lemma 1. *Consider the source model $\mathcal{U} = (\mathbb{R}, \{\mathcal{N}(\theta, \sigma^2)\}_{\theta \in \Theta})$ and general target model $\mathcal{V} = (\mathbb{R}, \{\mathcal{Q}_\theta\}_{\theta \in \Theta})$. Let $u(\cdot; \theta)$ be defined in Eq. (4) and let $v(\cdot; \theta) : \mathbb{R} \to \mathbb{R}$ denote the density of distribution $\mathcal{Q}_\theta$. For each $N \in \mathbb{N}$, consider the signed kernel $\mathcal{S}_N^\star$ defined in Eq. (14). Suppose Assumption 1 holds. Then*

$$\varsigma\big(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star\big) \le \frac{1}{2} \sup_{\theta \in \Theta} \int_{\mathbb{R}} \sum_{k=N+1}^{+\infty} \left| \int_{\mathbb{R}} v\big(y; \theta + \sqrt{2}\sigma x\big) \cdot \overline{H}_{2k}(x) e^{-x^2} \mathrm{d}x \right| \mathrm{d}y, \tag{18}$$

where $\overline{H}_{2k}(x)$ is the normalized Hermite polynomial defined in Eq. (7).

We provide the proof of Lemma 1 in Section 5.2. Recall that the Hermite polynomials $\big\{\overline{H}_n(x)\big\}_{n \in \mathbb{Z}_{\ge 0}}$ form an orthonormal basis for the function space

$$\left\{ f : \mathbb{R} \to \mathbb{R} \,\middle|\, \int_{\mathbb{R}} |f(x)|^2 e^{-x^2}\, \mathrm{d}x < \infty \right\}.$$

According to Ineq. (18), the signed deficiency $\varsigma\big(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star\big)$ can be bounded by the Hermite tail coefficients of the function $x \mapsto v\big(y; \theta + \sqrt{2}\sigma x\big)$, onto the subspace spanned by $\big\{\overline{H}_{2n}(x)\big\}_{n=N+1}^{\infty}$. These coefficients typically decay exponentially in $\sqrt{N}$ when $v\big(y; \theta + \sqrt{2}\sigma x\big)$ is sufficiently smooth in x; see Theorem 1 and Theorem 2 for details.

Step 3: Approximately sample from the Markov kernel.

Our final step is to use a rejection sampling algorithm to approximately sample from the family of distributions $\{\mathcal{T}_N^\star(\cdot \mid x)\}_{x \in \mathbb{R}}$. We use the particular algorithm from Lou et al. (2025, Section 3), and denote it by RK. We do not reproduce a description of the entire algorithm here, but note that the algorithm RK requires:

1. A family of base measures $\{\mathcal{D}(\cdot \mid x)\}_{x \in \mathbb{R}}$ such that the ratio between $\mathcal{S}_N^\star(y \mid x)$ and $\mathcal{D}(y \mid x)$ is uniformly bounded. That is, for some constant $M > 0$,

$$\frac{\mathcal{S}_N^\star(y \mid x) \vee 0}{\mathcal{D}(y \mid x)} \le M, \quad \text{for all } x, y \in \mathbb{R}.$$

2. Moreover, the base measures $\{\mathcal{D}(\cdot \mid x)\}_{x\in\mathbb{R}}$ should be chosen so that it is computationally efficient to sample from them. We let T_{samp} denote the maximum time taken to sample from any base measure $\mathcal{D}(\cdot \mid x)$.

Given these, the output of the algorithm is a sample that is close in total variation to one having density proportional to $\mathcal{S}_N^\star(y \mid x) \vee 0$. Concretely, denote the output of the algorithm is by $\mathrm{RK}(x, T_{\mathsf{iter}}, M, y_0)$, where $T_{\mathsf{iter}} \in \mathbb{N}$ is the number of iterations and $y_0 \in \mathbb{R}$ is an arbitrary initialization. As shown in Lou et al. (2025), the law of this random variable satisfies

$$\big\|\mathcal{L}\big[\mathrm{RK}(x, T_{\mathsf{iter}}, M, y_0)\big] - \mathcal{T}_N(\cdot \mid x)\big\|_{\mathsf{TV}} \le 2\exp\left(-\frac{T_{\mathsf{iter}}}{M}\, p(x)\right), \quad \text{for all } x \in \mathbb{R}, \tag{19}$$

where $p(x)$ is defined in Eq. (17). Thus, provided that $p(x) \ge c$ for some universal constant $c > 0$, the error from sampling procedure decreases geometrically as the number of iterations T_{iter} increases. Computationally, the implementation of $\mathrm{RK}(x, T_{\mathsf{iter}}, M, y_0)$ requires, in each iteration, generating a sample from $\mathcal{D}(\cdot \mid x)$ and evaluating the signed kernel $\mathcal{S}_N^\star$ once. Hence, RK can be implemented in time $\mathcal{O}\big(T_{\mathsf{iter}} \cdot T_{\mathsf{samp}} \cdot T_{\mathsf{eval}}(N)\big)$.

Putting together the steps.

Our reduction algorithm is defined as

$$\mathsf{K}(X_\theta) := \mathrm{RK}(X_\theta, T_{\mathsf{iter}}, M, y_0),$$

where $X_\theta \sim \mathcal{N}(\theta, \sigma^2)$ is the input sample from the source. The initialization y_0 can be arbitrary[2]. Recall that our goal was to control the total variation distance between $\mathsf{K}(X_\theta)$ and the target sample $Y_\theta \sim v(\cdot;\theta)$, which we denoted by $\mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big)$. Error arises from two sources: (i) the sampling error of the rejection sampling procedure RK, which can be bounded as in Eq. (19); and (ii) the deficiency of the Markov kernel $\mathcal{T}_N$, which is quantified by $\delta(\mathcal{U}, \mathcal{V}; \mathcal{T}_N)$ and can be bounded by using Ineq. (16) and Ineq. (18).

Note that while the reduction is general, its TV deficiency may not always be small. We next turn to two specific class of target statistical models and evaluate both the sampling error and the deficiency $\delta(\mathcal{U}, \mathcal{V}; \mathcal{T}_N)$ in each case.

2.2 General location target models

In this section, we consider non-Gaussian target models that share the same location parameter as the Gaussian source model (2). Specifically, we consider a target model $\mathcal{V} = (\mathbb{R}, \{\mathcal{Q}_\theta\}_{\theta\in\mathbb{R}})$, where each distribution $\mathcal{Q}_\theta$ has a density of the form

$$v(y;\theta) = \frac{1}{\tau}\exp\left(-\psi\left(\frac{y-\theta}{\tau}\right)\right), \quad \text{for all } y, \theta \in \mathbb{R}. \tag{20}$$

Here, $\psi : \mathbb{R} \to \mathbb{R}$ denotes the negative log-density function, and $\tau > 0$ is a scale parameter that controls (among other things) the variance of the distribution. For our reduction to be effective, we require some mild regularity conditions on the negative log-density function ψ.

Assumption 2. *Suppose $\psi : \mathbb{R} \to \mathbb{R}$ satisfies the following conditions:*

[2] As discussed in Lou et al. (2025), an arbitrary initialization suffices for a TV guarantee. Specific initializations may be desired if we want to prove bounds in other metrics besides TV.

(a) $\int_{\mathbb{R}} \exp\big(-\psi(t)\big)\mathrm{d}t = 1$ *and* $\int_{\mathbb{R}} \exp\big(-2\psi(t)\big)\mathrm{d}t < +\infty$.

(b) ψ *is infinitely differentiable on* $\mathbb{R}$ *and there exist constants* $C_\psi \in \mathbb{Z}_{\geq 0}$ *and* $\widetilde{C}_\psi > 0$ *depending only on* ψ *such that for all* $n \in \mathbb{N}$ *and* $k \in [n]$*, we have*

$$\int_{\mathbb{R}} \exp\big(-\psi(t)\big)\left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{(\widetilde{C}_\psi)^j j!}\right)^k \mathrm{d}t \leq (C_\psi n)!.$$

Note that Assumption 2 is very mild and encompasses a broad class of smooth densities. In particular, Assumption 2(a) ensures that the function $v(y;\theta)$ defined in Eq. (20) is indeed a valid density function. To interpret Assumption 2(b), recall the definition of analytic functions: a function $\psi : \mathbb{R} \to \mathbb{R}$ is called *analytic* if for every compact set $D \subset \mathbb{R}$, there exists a constant $C > 0$ (depending only on D) such that

$$|\psi^{(j)}(x)| \leq C^{j+1} j! \quad \text{for all } x \in D \text{ and } j \in \mathbb{Z}_{\geq 0}. \tag{21}$$

Assumption 2(b) requires that ψ is analytic with respect to the measure $e^{-\psi(x)}\,\mathrm{d}x$, which includes a rich class of functions. In particular, it contains all *globally analytic* functions, which are functions that satisfy this condition uniformly over $\mathbb{R}$:

Definition 2. *For a parameter* $R > 0$*, we define a class of globally analytic functions:*

$$\mathcal{F}(R) := \left\{ \psi : \mathbb{R} \to \mathbb{R} \;\middle|\; \sum_{j=1}^{\infty} \frac{|\psi^{(j)}(x)|}{R^j j!} \leq 1 \quad \textit{for all } x \in \mathbb{R} \right\}. \tag{22}$$

See Corollary 2 and Example 1 to follow, which illustrate that many canonical distributions satisfy Assumption 2 and belong to the class $\mathcal{F}(R)$.

Before stating our theorem showcasing the reduction guarantee, it is useful to define a few quantities that will appear in the theorem statement. For the constants $(C_\psi, \widetilde{C}_\psi)$ appearing in Assumption 2(b), we define the constant $\overline{C}_\psi$ for convenience as

$$\overline{C}_\psi := 16\pi^2 (2e)^{4(1+C_\psi)} e^{4(1+C_\psi)^2}. \tag{23}$$

For $\lambda > 0$ and $n \in \mathbb{N}$, we define a set $\mathcal{A}(\psi, n, \lambda) \subseteq \mathbb{R}$ and also its complement as

$$\mathcal{A}(\psi, n, \lambda) := \left\{ t \in \mathbb{R} \;\middle|\; \sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{\lambda^j j!} \leq 1 \right\} \quad \text{and} \quad \mathcal{A}^{\complement}(\psi, n, \lambda) := \mathbb{R} \setminus \mathcal{A}(\psi, n, \lambda). \tag{24}$$

Using these objects, we now define two quantities that will be used to measure the deficiency and computational complexity of the reduction, respectively. First, for $\lambda > 0$ and $n \in \mathbb{N}$, define

$$\mathsf{Err}(\psi, n, \lambda) := \int_{\mathcal{A}^{\complement}(\psi,n,\lambda)} e^{-\psi(t)}\mathrm{d}t + \max_{k \in [n]} \int_{\mathcal{A}^{\complement}(\psi,n,\lambda)} e^{-\psi(t)} \left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{\lambda^j j!}\right)^k \mathrm{d}t. \tag{25}$$

Second, for $\lambda > 0$ and $n \in \mathbb{N}$, define

$$M(\psi, n, \lambda) := \max_{k \in [n]} \sup_{t \in \mathbb{R}} \left\{ e^{-\psi(t)+\psi(t/2)} \cdot \left[1 + \left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{\lambda^j j!}\right)^k\right] \right\}. \tag{26}$$

Note that both $\mathsf{Err}(\psi, n, \lambda)$ and $M(\psi, n, \lambda)$ decrease as λ increases. In applications, one can choose a sufficiently large λ to ensure that these two quantities are small. We are now ready to state the main result.

Theorem 1. *Consider source model* $\mathcal{U} = (\mathbb{R}, \{\mathcal{N}(\theta, \sigma^2)\}_{\theta\in\mathbb{R}})$ *and target model* $\mathcal{V} = (\mathbb{R}, \{\mathcal{Q}_\theta\}_{\theta\in\mathbb{R}})$, *where* $\mathcal{Q}_\theta$ *has density function* $v(\cdot;\theta)$ *defined in Eq. (20). For each* $N \in \mathbb{N}$, *consider the signed kernel* $\mathcal{S}_N^\star$ *defined in Eq. (14) and the corresponding Markov kernel* $\mathcal{T}_N^\star$ *defined in Eq. (15). For* $\lambda > 0$ *and* $n \in \mathbb{N}$, *let* $\mathsf{Err}(\psi, n, \lambda)$ *and* $M(\psi, n, \lambda)$ *be defined as in Eq. (25) and Eq. (26), respectively. Suppose the negative log-density function* $\psi : \mathbb{R} \to \mathbb{R}$ *satisfies Assumption 2 with a pair of constants* $(C_\psi, \widetilde{C}_\psi)$, *and that the density* $e^{-\psi}$ *can be sampled in time* T_{samp}. *Suppose the tuple of parameters* $(N, \tau, \lambda, \sigma)$ *satisfies*

$$2N + 1 \geq \overline{C}_\psi, \quad \tau \geq 8e\sigma \max\left\{\widetilde{C}_\psi, \lambda N\right\}, \quad \text{and} \quad \sigma, \lambda > 0, \tag{27}$$

where $\overline{C}_\psi$ *is defined in Eq. (23). Then the following statements are true:*

(a) We have

$$\delta\big(\mathcal{U}, \mathcal{V}; \mathcal{T}_N^\star\big) \leq \overline{C}_\psi \exp\left(-(2e)^{-1}(2N+1)^{\frac{1}{2(1+C_\psi)}}\right) + 4\mathsf{Err}(\psi, 2N, \lambda). \tag{28}$$

(b) Consequently, for each $\epsilon \in (0,1)$, *if the tuple* (N, τ, λ) *satisfies Ineq. (27) and also*

$$2N + 1 \geq \Big(2e \log\big(3\overline{C}_\psi/\epsilon\big)\Big)^{2(1+C_\psi)}, \tag{29}$$

then there is a reduction algorithm $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ *that runs in time*

$$\mathcal{O}\Big(T_{\mathsf{iter}} \cdot T_{\mathsf{samp}} \cdot T_{\mathsf{eval}}(N)\Big) \quad \text{for any} \quad T_{\mathsf{iter}} \in \mathbb{N},$$

and satisfies that, for $X_\theta \sim \mathcal{N}(\theta, \sigma^2)$ *and* $Y_\theta \sim Q_\theta$,

$$\sup_{\theta\in\mathbb{R}} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \leq \exp\left(-\frac{T_{\mathsf{iter}}}{4M(\psi, 2N, \lambda)}\right) + \frac{\epsilon}{3} + 4\mathsf{Err}(\psi, 2N, \lambda). \tag{30}$$

We provide the proof of Theorem 1 in Section 6. A few remarks are in order. To achieve ϵ-TV deficiency, we must choose N, λ, and T_{iter} such that the right-hand sides of Ineqs. (28) and (30) are both smaller than ϵ. As a first step, observe that since the constants $(C_\psi, \widetilde{C}_\psi)$ are independent of ϵ, it suffices to set

$$N = \mathcal{O}\big(\mathsf{polylog}(1/\epsilon)\big).$$

Given such a choice of N specified by Ineq. (29), we next select a sufficiently large λ so that the quantities $\mathsf{Err}(\psi, 2N, \lambda)$ and $M(\psi, 2N, \lambda)$ become small. Ideally, if we are able to choose

$$\lambda = \mathcal{O}\big(\mathsf{polylog}(1/\epsilon)\big) \quad \text{such that} \quad \mathsf{Err}(\psi, 2N, \lambda) \leq \frac{\epsilon}{12} \quad \text{and} \quad M(\psi, 2N, \lambda) = \mathcal{O}(1),$$

then, by taking $T_{\mathsf{iter}} = \mathcal{O}(\log(3/\epsilon))$, Ineq. (30) ensures

$$\sup_{\theta\in\mathbb{R}} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \leq \epsilon.$$

Ultimately, the choices of N and λ necessitate an increase in the scale parameter τ^2 of the target distribution, as stipulated by condition (27). In other words, the reduction incurs a cost in terms of the scale parameter (or equivalently, a decrease in signal-to-noise ratio). However, this cost is mild: we only require $\tau \gtrsim \sigma\,\mathsf{polylog}(1/\epsilon)$, which is merely *polylogarithmic* in $1/\epsilon$. This is a significant

improvement over the plug-in approach, which typically requires τ to grow *polynomially* in $1/\epsilon$; see Lou et al. (2025, Proposition 1) for a discussion of the inefficiency of the plug-in approach in a related context.

Furthermore, provided that the evaluation time $T_{\mathsf{eval}}(N)$ of the signed kernel (see Definition 1) is polynomial in $1/\epsilon$, the overall reduction can be carried out in time $\mathcal{O}(\mathsf{poly}(1/\epsilon))$, thereby ensuring that the reduction is computationally efficient.

Operationally, Theorem 1 can be viewed as a universality result in the setting of a *single* observation, and affirmatively answers puzzle (P1) from the introduction. Specifically, given only *one* observation $X_\theta \sim \mathcal{N}(\theta, \sigma^2)$——without knowledge of the true mean θ, which can take any real value——one can generate a new observation with the same mean, but whose distribution is transformed into a general non-Gaussian form. Such a reduction technique is particularly valuable for establishing the universality of computational hardness: lower bounds derived under Gaussian observations can be transferred to settings with general non-Gaussian observations. We demonstrate the use of this reduction in Section 3.1 through the example of tensor PCA with non-Gaussian noise.

While Theorem 1 is abstractly stated, it applies to many concrete families of target densities. We prove two broad examples of such families, beginning with log-densities that are globally analytic according to Definition 2.

Corollary 1. *For $R > 0$, let the class of functions $\mathcal{F}(R)$ be defined as in Definition 2. Suppose $\psi \in \mathcal{F}(R)$ and define $M := \sup_{t\in\mathbb{R}} \{e^{-\psi(t)+\psi(t/2)}\}$. Then Assumption 2 holds with $C_\psi = 0$ and $\widetilde{C}_\psi = R$, and*

$$\mathsf{Err}(\psi, n, R) = 0 \quad \textit{and} \quad M(\psi, n, R) = 2M \quad \textit{for all } n \in \mathbb{N}.$$

Moreover, if $\psi^{(n)}(x)$ can be evaluated in computational time $\mathcal{O}\big(e^{\mathcal{O}(\sqrt{n})}\big)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$, then $T_{\mathsf{eval}}(n) = \mathcal{O}\big(e^{\mathcal{O}(\sqrt{n})}\big)$. Consequently, for each $\epsilon \in (0, 1)$, there exists a universal and positive constant $\widetilde{C}$ such that for

$$\tau \geq \widetilde{C} R\sigma \log^2\big(\widetilde{C}/\epsilon\big),$$

there is a reduction algorithm $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ that runs in time

$$\mathcal{O}\Big(M \cdot T_{\mathsf{samp}} \cdot \mathsf{poly}(1/\epsilon)\Big) \quad \textit{and satisfies} \quad \sup_{\theta\in\mathbb{R}} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \leq \epsilon,$$

where $X_\theta \sim \mathcal{N}(\theta, \sigma^2)$ and the random variable Y_θ has density $v(\cdot; \theta)$ defined in Eq. (20).

We provide the proof of Corollary in Section 6.3. Note that it is common for the higher-order derivatives $\psi^{(n)}(x)$ to be computable in time $\mathcal{O}(e^{\mathcal{O}(\sqrt{n})})$. For a canonical example, consider the case where $\psi(x) = g \circ h(x)$ is the composition of two basic functions. By Faà di Bruno's formula, we have

$$\psi^{(n)}(x) = \sum_{m_1,\ldots,m_n} \frac{n!}{m_1!\, m_2! \cdots m_n!} \cdot g^{(m_1+\cdots+m_n)}(h(x)) \cdot \prod_{j=1}^{n} \left(\frac{h^{(j)}(x)}{j!}\right)^{m_j}, \tag{31}$$

where the summation is over all n-tuples of nonnegative integers $(m_1, \ldots, m_n)$ satisfying $\sum_{j=1}^n j m_j = n$. The total number of such tuples is at most $\mathcal{O}(e^{\pi\sqrt{n}})$. If each term in the summation can be computed in time $\mathcal{O}(n)$—as g and h are assumed to be simple functions—then the total time to evaluate $\psi^{(n)}(x)$ is $\mathcal{O}(n \cdot e^{\pi\sqrt{n}})$, which satisfies our computational assumption.

In the following example, we further specialize Corollary 1 to illustrate that it covers natural target distributions.

Example 1. *The following distributions have globally analytic log-densities and so Corollary 1 applies. In particular, we obtain the following results.*

1. *Student't t-distribution: let $\nu \geq 1$ be the degree of freedom and consider*

$$\psi(t) = \frac{\nu+1}{2}\log\left(1+\frac{t^2}{\nu}\right) - \log\left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}\right) \quad \text{and} \quad v(y;\theta) = \frac{1}{\tau}\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}\left(1+\frac{(\frac{y-\theta}{\tau})^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

   Then $\psi \in \mathcal{F}\big(6e(\nu+1)\big)$ and $M \leq 1$. Computing $\psi^{(n)}(x)$ takes time $\mathcal{O}(\mathsf{poly}(n))$ for all $x \in \mathbb{R}$, $n \in \mathbb{N}$.

2. *Hyperbolic secant distribution: consider*

$$\psi(t) = \log\left(e^{\frac{\pi t}{2}} + e^{-\frac{\pi t}{2}}\right) \quad \text{and} \quad v(y;\theta) = \frac{1}{\tau}\cdot\frac{1}{\exp\left(\frac{\pi}{2}\frac{y-\theta}{\tau}\right)+\exp\left(-\frac{\pi}{2}\frac{y-\theta}{\tau}\right)}.$$

   Then $\psi \in \mathcal{F}\big(2\pi e\big)$ and $M \leq 2$. Computing $\psi^{(n)}(x)$ takes time $\mathcal{O}\big(e^{5\sqrt{n}}\big)$ for all $x \in \mathbb{R}$, $n \in \mathbb{N}$.

3. *Logistic distribution: consider*

$$\psi(t) = 2\log\left(e^{t/2} + e^{-t/2}\right) \quad \text{and} \quad v(y;\theta) = \frac{1}{\tau}\cdot\frac{1}{\left[\exp\left(\frac{y-\theta}{2\tau}\right)+\exp\left(-\frac{y-\theta}{2\tau}\right)\right]^2}.$$

   Then $\psi \in \mathcal{F}\big(2\pi e\big)$ and $M \leq 4$. Computing $\psi^{(n)}(x)$ takes time $\mathcal{O}\big(e^{5\sqrt{n}}\big)$ for all $x \in \mathbb{R}$, $n \in \mathbb{N}$.

Remark 2. *Note that evaluating the signed kernel $\mathcal{S}_N^\star(y \mid x)$ (14) for the target density $v(y;\theta)$ (20) requires computing the higher-order derivative $\psi^{(2N)}(\cdot)$. Condition (29) dictates that $N \gtrsim \big(\log(1/\epsilon)\big)^{2(1+C_\psi)}$. In the setting of Corollary 1, we have $C_\psi = 0$, so we take $N \asymp \log^2(1/\epsilon)$. In this case, evaluating $\psi^{(2N)}(\cdot)$ using formula (31) requires computational time*

$$\mathcal{O}\big(e^{\mathcal{O}(\sqrt{N})}\big) \;=\; \mathsf{poly}(1/\epsilon).$$

However, in the more general case $C_\psi \geq 1$, a naive application of formula (31) leads to computational time of order

$$e^{\sqrt{N}} \;=\; \exp\Big(\big(\log(1/\epsilon)\big)^{1+C_\psi}\Big),$$

which grows quasi-polynomially in $1/\epsilon$. Therefore, when $C_\psi \geq 1$, a more computationally efficient algorithm for evaluating higher-order derivatives is required. The following Corollary 2 serves as an illustration.

Note that Theorem 1 also applies when the negative log density ψ is not globally analytic. We next illustrate this with the case where ψ is a monomial of fixed degree, which corresponds to the family of generalized normal distributions.

Corollary 2 (Generalized normal target)**.** *Let $\beta \in \mathbb{N}$ be a positive and even integer. Consider*

$$\psi(t) = t^\beta - \log\left(\frac{\beta}{2\Gamma(1/\beta)}\right) \quad \text{and} \quad v(y;\theta) = \frac{\beta}{2\tau\Gamma(1/\beta)}\exp\left(-\frac{|y-\theta|^\beta}{\tau^\beta}\right),$$

where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\mathrm{d}t$ for all $z > 0$. Assumption 2 holds with $C_\psi = 1$ and $\widetilde{C}_\psi = 2\beta^2$, and

$$\mathsf{Err}(\psi, n, \lambda) \leq 3e^{-\sqrt{\lambda}}, \quad M(\psi, n, \lambda) \leq 3, \quad T_{\mathsf{eval}}(n) = \mathcal{O}\big(n^{\beta+2}\big) \quad \text{for all } n \in \mathbb{N}, \lambda \geq 4n^2\beta^2.$$

Consequently, for each $\epsilon \in (0,1)$, there exists a pair of universal and positive constants (C_1, C_2) such that if

$$\tau \geq C_1 \sigma \beta^2 \log^{12}\left(\frac{C_2}{\epsilon}\right),$$

then there is a reduction algorithm $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ that runs in time

$$\mathcal{O}\Big(\log^{4\beta+5}\left(C_2/\epsilon\right)\Big) \quad \text{and satisfies} \quad \sup_{\theta \in \mathbb{R}} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \leq \epsilon,$$

where $X_\theta \sim \mathcal{N}(\theta, \sigma^2)$ and $Y_\theta \sim \mathsf{GN}(\theta, \tau, \beta)$.

We provide the proof of Corollary 2 in Section 6.5. Note that in Corollary 2, the target distribution has much lighter tails than the source distribution when $\beta \geq 4$. Also note that fast computation is still possible by using specific formulas for the derivatives (cf. Remark 2).

Remark 3. *First, Corollary 2 does not cover the case when β is odd, e.g., the Laplace distribution with $\beta = 1$. When β is odd, the target density is non-differentiable at one point, which violates Assumption 2(b), and our technique relies crucially on the differentiability of the target density. One could apply a mollification technique to smooth the target density; however, calculations show that to achieve ϵ-TV deficiency, this would require inflating the scale parameter τ* polynomially *in $1/\epsilon$. Second, when β is chosen on the order of $1/\epsilon$, the generalized normal distribution is ϵ-close to the uniform distribution in TV distance. In this case, Corollary 1 also requires inflating τ polynomially in $1/\epsilon$. Thus, Corollary 1 does not cover targets with extremely light tails such as the uniform distribution. Interestingly, a reduction from the uniform to the Gaussian location model is known (Lou et al., 2025).*

2.3 Gaussian target models with mean nonlinearly transformed

In this section, we consider a Gaussian target model where the mean parameter θ is transformed nonlinearly. Specifically, we study

$$\mathcal{V} = \left(\Theta, \{\mathcal{N}\left(f\left(\theta/\tau\right), 1\right)\}_{\theta \in \Theta}\right), \tag{32}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a link function that nonlinearly transforms the mean parameter θ, and $\tau > 0$ is a fixed parameter that adjusts the signal-to-noise ratio. Since τ already serves this purpose, we fix the variance of the target distribution to be 1 for convenience.[3] Note that the target distribution $\mathcal{Q}_\theta$ has the density

$$v(y;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\left(y - f\left(\frac{\theta}{\tau}\right)\right)^2}{2}\right), \quad \text{for all } y \in \mathbb{R},\ \theta \in \Theta. \tag{33}$$

We next state some mild regularity conditions on the link function $f : \mathbb{R} \to \mathbb{R}$.

Assumption 3. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is infinitely differentiable on $\mathbb{R}$ and there exists a pair of universal constants $(C_f, \widetilde{C}_f)$, where $C_f \in \mathbb{Z}_{\geq 0}$ and $\widetilde{C}_f > 0$, that only depends on the link function $f : \mathbb{R} \to \mathbb{R}$ such that*

$$\int_{\mathbb{R}} \left(\sum_{j=1}^{n} \frac{\left|f^{(j)}\left(\left(\theta + \sqrt{2}x\right)/\tau\right)\right|}{\left(\widetilde{C}_f\right)^j j!}\right)^n \frac{\exp(-x^2/2)}{\sqrt{2\pi}} \mathrm{d}x \leq \left(C_f n\right)!, \tag{34}$$

for all $n \in \mathbb{N}$ and $\tau \geq 2\sqrt{2}(1 \vee |\theta|)$.

[3] Both the reduction and the analysis can be readily extended to the general case where the target model (32) has arbitrary variance. However, fixing the variance to 1 simplifies exposition and suffices for our applications.

Note that Assumption 3 imposes the analyticity of the function f with respect to the standard Gaussian measure, which is analogous to Assumption 2(b). This condition encompasses a broad class of functions, including polynomials of constant degree (see Corollary 3) as well as any globally analytic function (see Corollary 4).

To state our theorem, we require formal definitions of several quantities derived from the link function f. Let $(C_f, \widetilde{C}_f)$ be the same pair of constants as in Assumption 3 and for all $n \in \mathbb{N}$ and $\tau > 0$, define

$$\begin{aligned} \overline{C}_f &:= 32\pi^2 (2e)^{4(1+C_f)} e^{4(1+C_f)^2}, \qquad L_f(n) := \sup_{|x|\le 1} \sum_{j=1}^{n} \frac{|f^{(j)}(x)|}{(\widetilde{C}_f)^j j!}, \\ \text{and} \quad \widetilde{L}_f(n,\tau) &:= \sup_{|x|\ge 1} \sum_{j=1}^{n} \frac{|f^{(j)}(x)|}{(\widetilde{C}_f)^j j!} \cdot \exp\Big(-\frac{\tau^2 x^2}{8n}\Big). \end{aligned} \tag{35}$$

Note that the quantity $L_f(n)$ ought to be bounded for any reasonable function f, given that the supremum is calculated over the range $|x| \le 1$. Moreover, by increasing τ, we can effectively ensure that $\widetilde{L}_f(n,\tau)$ also remains bounded. We are now ready to state the theorem.

Theorem 2. *Consider Gaussian source model $\mathcal{U} = (\mathbb{R}, \{\mathcal{N}(\theta,1)\}_{\theta\in\Theta})$ and Gaussian target model $\mathcal{V} = \big(\mathbb{R}, \{\mathcal{N}\big(f(\theta/\tau),1\big)\}_{\theta\in\Theta}\big)$. For each $N \in \mathbb{N}$, consider the signed kernel $\mathcal{S}_N^\star$ defined in Eq. (14) and the corresponding Markov kernel $\mathcal{T}_N^\star$ defined in Eq. (15). Suppose the link function $f : \mathbb{R} \to \mathbb{R}$ satisfies Assumption 3 with a pair of constants $(C_f, \widetilde{C}_f)$. For $\tau > 0$ and $n \in \mathbb{N}$, let $\overline{C}_f$, $L_f(n)$ and $\widetilde{L}_f(n,\tau)$ be defined in Eq. (35). There exists a universal and positive constant C such that if*

$$2N+1 \ge \overline{C}_f \quad \text{and} \quad \tau^2 \ge C\bigg[\big(\widetilde{C}_f \vee 1\big)^4 \Big(L_f(2N) \vee \widetilde{L}_f(2N,\tau) \vee 1\Big)^2 N^2 \ \vee \ \sup_{\theta\in\Theta} \theta^2\bigg], \tag{36}$$

then the following statements hold.

(a) We have

$$\delta\big(\mathcal{U},\mathcal{V};\mathcal{T}_N^\star\big) \le \overline{C}_f \exp\bigg(-\frac{(2N+1)^{\frac{1}{2(1+C_f)}}}{2e}\bigg) + 12e^{-\tau/8} + 15 \sup_{\theta\in\Theta} \exp\bigg(\frac{\theta^2}{2} - \frac{\tau^2}{8}\bigg). \tag{37}$$

(b) Consequently, for each $\epsilon \in (0,1)$, if N and τ are setting to satisfy Eq. (36) and also

$$2N+1 \ge \Big(2e\log\big(4\overline{C}_f/\epsilon\big)\Big)^{2(1+C_f)}, \quad \tau^2 \ge 64\log^2\big(96/\epsilon\big) \vee \Big[8\log\big(120/\epsilon\big) + 4\sup_{\theta\in\Theta}\theta^2\Big], \tag{38}$$

then there exists a reduction algorithm $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ that runs in time

$$\mathcal{O}\big(\log(4/\epsilon) T_{\mathsf{eval}}(N)\big) \quad \text{and satisfies} \quad \sup_{\theta\in\Theta} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \le \epsilon, \tag{39}$$

where $X_\theta \sim \mathcal{N}(\theta,1)$ and $Y_\theta \sim \mathcal{N}\big(f(\theta/\tau),1\big)$.

We provide the proof of Theorem 2 in Section 7. A few remarks about Theorem 2 are in order. To achieve ϵ-TV deficiency, it suffices to set $N = \Theta(\mathsf{polylog}(1/\epsilon))$, since the constants $(C_f, \widetilde{C}_f)$ are independent of ϵ. Consequently, if the link function f satisfies

$$L_f(2N) = \mathcal{O}(\mathsf{polylog}(1/\epsilon)), \quad \widetilde{L}_f(2N,\tau) = \mathcal{O}(\mathsf{polylog}(1/\epsilon)) \quad \text{for some } \tau = \mathcal{O}(\mathsf{polylog}(1/\epsilon)), \tag{40}$$

and if the parameter set $\Theta \subset \mathbb{R}$ is bounded, then the reduction can attain ϵ-TV deficiency provided that

$$\tau^2 \gtrsim \mathsf{polylog}(1/\epsilon).$$

We will see shortly that condition (40) is satisfied for many canonical link functions that obey natural regularity conditions.

As in Theorem 1, this reduction comes at the cost of requiring a lower signal-to-noise ratio (SNR) in the target model (32)—via inflating τ—compared to the source model (2). However, the degradation remains mild: the SNR is reduced only by a $\mathsf{polylog}(1/\epsilon)$ multiplicative factor. This represents a significant improvement over the plug-in approach, which typically requires reducing the SNR by a polynomial factor in $1/\epsilon$. For concreteness, we show in Appendix A.2 that the plug-in approach must incur this blow-up even when $f(\theta) = \theta^2$.

Operationally, Theorem 2 provides an answer to puzzle (P2) posed in the introduction for a large class of functions. In Section 3.2, we demonstrate a concrete application by establishing the computational hardness of the k-sparse generalized linear model with even link functions via reduction from sparse mixtures of linear regressions. In Section 3.3, we employ Theorem 2 to construct a reduction between the two high-dimensional models Rank1SubMat and PlantSubMat.

Let us now instantiate the abstract Theorem 2 for several canonical function families.

Corollary 3 (Monomial)**.** *Consider link function $f(\theta) = \theta^B$ for some $B \in \mathbb{N}$. Then Assumption 3 holds with $C_f = 2B + 1$ and $\widetilde{C}_f = 4eB$. Moreover, we have for all $n \in \mathbb{N}$ and $\tau^2 \geq 8nB$ that*

$$L_f(n) \leq 1, \quad \widetilde{L}_f(n, \tau) \leq 1, \quad \text{and} \quad T_{\mathsf{eval}}(n) = \mathcal{O}(n^{B+2}).$$

Consequently, for each $\epsilon \in (0, 1)$, there exists a universal constant $C_B \geq 1$ that only depends on B such that if

$$\tau^2 \geq C_B \cdot \log^{8(B+1)}(C_B/\epsilon) + 4 \sup_{\theta \in \Theta} \theta^2,$$

then there is a reduction algorithm $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ that runs in time

$$\mathcal{O}\Big(C_B \log^{4(B+2)^2}(C_B/\epsilon)\Big) \quad \text{and satisfies} \quad \sup_{\theta \in \Theta} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \leq \epsilon,$$

where $X_\theta \sim \mathcal{N}(\theta, 1)$ and $Y_\theta \sim \mathcal{N}\big(f(\theta/\tau), 1\big)$.

We provide the proof of Corollary 3 in Section 7.3.

To produce other examples, we use the following lemma, which can be used to verify Assumption 3 and bound the quantities defined in Eq. (35) when the link function is a linear combination of a set of basic functions.

Lemma 2. *Let $f_k : \mathbb{R} \to \mathbb{R}$ be a function satisfying Assumption 3 with a pair of constants $\big(\widetilde{C}_{f_k}, C_{f_k}\big)$ for all $k \in [K]$. Consider $f(x) = \sum_{k=1}^K c_k f_k(x)$ for all $x \in \mathbb{R}$ with linear coefficients $c_k \in \mathbb{R}$ for all $k \in [K]$. Then f satisfies Assumption 3 with the pair of constants $(C_f, \widetilde{C}_f)$ defined as*

$$C_f = \max_{k \in [K]} \{C_{f_k}\} \quad \text{and} \quad \widetilde{C}_f = \sum_{\ell=1}^K |c_\ell| \cdot \max_{k \in [K]} \{\widetilde{C}_{f_k}\}.$$

Moreover, we have for all $n \in \mathbb{N}$ and $\tau > 0$,

$$L_f(n) \leq \frac{\sum_{k=1}^K |c_k| \cdot L_{f_k}(n)}{\sum_{\ell=1}^K |c_\ell|} \quad \text{and} \quad \widetilde{L}_f(n, \tau) \leq \frac{\sum_{k=1}^K |c_k| \cdot \widetilde{L}_{f_k}(n, \tau)}{\sum_{\ell=1}^K |c_\ell|}.$$

The proof of Lemma 2 is provided in Appendix B.1.

Combining Corollary 3 with Lemma 2, we obtain a reduction when the function f is a polynomial of constant degree. This already encompasses several natural nonlinear transformations. To cover others that do not have a natural polynomial representation, we consider the class of link functions that are globally analytic in the sense of Definition 2.

Corollary 4. *For $R > 0$, let the class of functions $\mathcal{F}(R)$ be defined as in Definition 2. Let $f : \mathbb{R} \to \mathbb{R}$ be any element in $\mathcal{F}(R)$ and suppose $f^{(n)}(x)$ can be evaluated in computational time $\mathcal{O}\big(e^{\mathcal{O}(\sqrt{n})}\big)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$. Then Assumption 3 holds with $C_f = 0$ and $\widetilde{C}_f = 2R$. We have for all $n \in \mathbb{N}$ and $\tau > 0$,*

$$L_f(n) \le 1, \quad \widetilde{L}_f(n, \tau) \le 1, \quad T_{\mathsf{eval}}(n) = \mathcal{O}\big(e^{\mathcal{O}(\sqrt{n})}\big).$$

Consequently, there exists a universal and positive constant C such that if

$$\tau^2 \ge CR^2 \log^2(C/\epsilon) + 4 \sup_{\theta \in \Theta} \theta^2,$$

there is a reduction algorithm $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ that runs in time

$$\mathcal{O}\Big(\mathsf{poly}(1/\epsilon)\Big) \quad \text{and satisfies} \quad \sup_{\theta \in \Theta} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \le \epsilon,$$

where $X_\theta \sim \mathcal{N}(\theta, 1)$ and $Y_\theta \sim \mathcal{N}\big(f(\theta/\tau), 1\big)$.

There are many interesting link functions satisfying the conditions in Corollary 4. We specify a few concrete examples below.

Example 2. *The following functions are all globally analytic in the sense of Definition 2.*

1. *Trigonometric function: $f(\theta) = \cos(a\theta + b)$ or $f(\theta) = \sin(a\theta + b)$ for any $a, b \in \mathbb{R}$. We have $f \in \mathcal{F}(2|a|)$, and one can evaluate $f^{(n)}(x)$ in time $\mathcal{O}(1)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$.*
2. *Hyperbolic function: $f(\theta) = \mathsf{sech}(\theta) = \frac{2}{e^\theta + e^{-\theta}}$. We have $f \in \mathcal{F}(8e)$, and one can evaluate $f^{(n)}(x)$ in time $\mathcal{O}(e^{\pi\sqrt{n}})$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$.*
3. *Other examples: $f(\theta) = (1 + \theta^2)^{-1}$ or $f(\theta) = \exp(-\theta^2)$ or $f(\theta) = (1 + e^{-\theta})^{-1}$. We have $f \in \mathcal{F}(8e)$, and one can evaluate $f^{(n)}(x)$ in time $\mathcal{O}\big(e^{\pi\sqrt{n}}\big)$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$.*

We provide proof of the above claims in Section 7.5.

3 Applications to statistical-computational tradeoffs

We now present three applications of our reductions in Section 2 to the study of statistical-computational tradeoffs in some canonical high-dimensional statistics and machine learning models. Our computational hardness results are based on the k-partite hypergraph planted clique (k-HPC^s) conjecture and the k-part bipartite planted clique conjecture (k-BPC); see Appendix A.1 for a detailed description. The two conjectures are also used in Brennan and Bresler (2020). We will take these conjectures as given and state consequences of these hardness results as appropriate in the sequel. Our goal in this section is to cover the motivating examples mentioned in Section 1 through the lens of reductions.

3.1 Tensor principal component analysis with non-Gaussian noise

In Tensor PCA (Montanari and Richard, 2014), we observe a single order s tensor $T(\beta, v)$ with dimensions $n^{\otimes s} = \underbrace{n \times \cdots \times n}_{s \text{ times}}$ given by

$$T(\beta, v) = \beta(\sqrt{n}v)^{\otimes s} + \zeta. \tag{41}$$

The positive scalar β denotes the signal-to-noise ratio, the unknown spike $v \in \mathbb{R}^n$ is typically sampled from a prior distribution, and the observations are corrupted by a symmetric Gaussian noise tensor ζ. Specifically, let $G \sim \mathcal{N}(0,1)^{\otimes n^{\otimes s}}$ be an asymmetric tensor with entries that are i.i.d. $\mathcal{N}(0,1)$, and define the symmetrized noise tensor ζ by

$$\zeta_{i_1,\ldots,i_s} = \frac{1}{s!} \sum_{\pi \in \mathfrak{S}_s} G_{i_{\pi(1)},\ldots,i_{\pi(s)}} \quad \text{for all } (i_1, \ldots, i_s) \in [n]^s. \tag{42}$$

For notational convenience, let $\sigma^2_{i_1,\ldots,i_s}$ denote the variance of $\zeta_{i_1,\ldots,i_s}$. Since s is a constant, all entries of ζ have constant-order variance. In particular, the minimum variance is $\sigma^2_{i_1,\ldots,i_s} = 1/s!$ when all indices $(i_1, \ldots, i_s)$ are distinct, while the maximum variance is $\sigma^2_{i_1,\ldots,i_s} = 1$ when all indices are identical.

The task of recovering v to within nontrivial ℓ_2-error $o(1)$ exhibits a statistical-computational gap. On the one hand, it is information-theoretically possible to recover v when $\beta = \widetilde{\omega}(n^{(1-s)/2})$ via exhaustive search (Montanari and Richard, 2014; Lesieur et al., 2017b; Chen, 2019; Jagannath et al., 2020; Perry et al., 2020). On the other hand, all known polynomial-time algorithms require a stronger signal, namely $\beta = \widetilde{\Omega}(n^{-s/4})$, including spectral methods (Montanari and Richard, 2014), the sum-of-squares hierarchy (Hopkins et al., 2015, 2017), and spectral algorithms based on the Kikuchi hierarchy (Wein et al., 2019). This statistical-computational gap in tensor PCA is believed to be fundamental, and the following reduction-based computational hardness result from Brennan and Bresler (2020, Theorem 10 and Lemma 102) makes this precise.

Proposition 2. *Let n be a parameter and $s \geq 3$ be a constant. Let $\zeta \in \mathbb{R}^{n^{\otimes s}}$ be the symmetric Gaussian tensor defined in Eq. (42) and define*

$$T = \beta(\sqrt{n}v)^{\otimes s} + \zeta \quad \textit{where} \quad v \sim \mathsf{Unif}(\mathbb{S}^{n-1}). \tag{43}$$

Suppose the k-HPC^s conjecture holds. If $\beta = \widetilde{o}\big(n^{-s/4}\big)$ and $\beta = \omega\big(n^{-s/2}\sqrt{s\log(n)}\big)$, then there cannot *exist an algorithm $\mathcal{A} : \mathbb{R}^{n^{\otimes s}} \to \mathbb{S}^{d-1}$ that runs in time $\mathsf{poly}(n)$ and satisfies*

$$\big\langle \mathcal{A}(T), v \big\rangle = \Omega(1) \quad \textit{with probability at least} \quad \Omega_n(1). \tag{44}$$

A few comments on Proposition 2 are in order. First, the condition $\beta = \omega\big(n^{-s/2}\sqrt{s\log(n)}\big)$ is not an essential feature of the result and can be ignored, since it is weaker than the information-theoretic threshold $\beta = \widetilde{\omega}(n^{(1-s)/2})$ necessary for recovering v with constant probability (Montanari and Richard, 2014; Chen, 2019). Second, we remark that the requirement

$$\big\langle \mathcal{A}(T), v \big\rangle = \Omega(1)$$

is weaker than the condition

$$\big\| \mathcal{A}(T) - v \big\|_2 = o(1).$$

Thus, any estimation algorithm $\mathcal{A}$ that achieves ℓ_2 error $o(1)$ automatically satisfies requirement (44). Third, Proposition 2 as stated differs from the results of Brennan and Bresler (2020,

Theorem 10 and Lemma 102) in two ways: In their setting, (i) the spike $\sqrt{n}v$ is a discrete vector in $\{-1,1\}^n$, and (ii) the Gaussian noise ζ is asymmetric. Nevertheless, the model we consider exhibits the same gap as the model in Brennan and Bresler (2020) because there is an exact reduction from the Tensor PCA model in Brennan and Bresler (2020, Theorem 10) to our model (41). This reduction proceeds by multiplying a uniformly random rotation matrix to each mode of the tensor, and achieves zero total variation deficiency. For completeness, we provide the proof of this claim in Appendix B.5 and a formal proof of Proposition 2 in Section 8.1.

The key takeaway from Proposition 2 is that computational hardness result holds under the Gaussian noise model. This restriction is somewhat unsatisfactory, as real-world observations can be affected by various types of non-Gaussian noise—indeed, this has also been noticed in recent work (Kunisky, 2025). We now use our results from Section 2 to extend reduction-based hardness results for tensor PCA to a larger class of non-Gaussian distributions.

To be concrete, consider the same tensor PCA model as in Eq. (41), but with observations corrupted by general non-Gaussian noise. Specifically, we study the following model:

$$\widetilde{T}(\beta, v) = \beta(\sqrt{n}v)^{\otimes s} + \widetilde{\zeta}, \tag{45}$$

Here the noise tensor $\widetilde{\zeta} \in \mathbb{R}^{n^{\otimes s}}$ is still symmetric—specifically, let $\mathcal{Q}_0$ denote the distribution of the noise, then the entries $\widetilde{\zeta}_{i_1,\ldots,i_s} \sim \mathcal{Q}_0$ and satisfy $\widetilde{\zeta}_{i_1,\ldots,i_s} = \widetilde{\zeta}_{i_{\pi(1)},\ldots,i_{\pi(s)}}$ for any permutation π on $[s]$. We write the density of the noise distribution $\mathcal{Q}_0$ as

$$f_{\mathcal{Q}_0}(y) = \frac{1}{\tau} \exp\big(-\psi\big(t/\tau\big)\big) \quad \text{for all } y \in \mathbb{R}, \tag{46}$$

where $\tau > 0$ is the scale parameter of distribution $\mathcal{Q}_0$ and $\psi : \mathbb{R} \to \mathbb{R}$ is a general function. Recall the definitions of $\mathsf{Err}(\psi, 2N, \lambda)$ in Eq. (25), $M(\psi, 2N, \lambda)$ in Eq. (26), as well as the evaluation time $T_{\mathsf{eval}}(N)$ from Definition 1. We will make use of the signed kernel $\mathcal{S}_N^\star$ (14), with target density defined in Eq. (20). We make the following assumption on the function ψ.

Assumption 4. *The function ψ satisfies Assumption 2 with a pair of constants $(C_\psi, \widetilde{C}_\psi)$. For each $\epsilon \in (0,1)$, if the pair of parameters (N, λ) are set to*

$$N = \Theta\Big(\overline{C}_\psi \log^{(2(1+C_\psi))}\big(3\overline{C}_\psi/\epsilon\big)\Big) \quad \textit{and} \quad \lambda = \Theta\Big(\mathsf{polylog}\big(1/\epsilon\big)\Big),$$

$$\textit{then} \quad \mathsf{Err}(\psi, 2N, \lambda) \le \epsilon/6, \qquad M(\psi, 2N, \lambda) = \mathcal{O}(1), \quad \textit{and} \quad T_{\mathsf{eval}}(N) = \mathcal{O}\big(\mathsf{poly}(1/\epsilon)\big).$$

Note that a broad class of noise distributions satisfies Assumption 4. As illustrated in Corollary 2, Corollary 1, and Example 1, this class includes the generalized normal distribution $\mathsf{GN}(0, \tau, \beta)$, which has lighter tails than the Gaussian distribution, as well as the Student's t-distribution (including the Cauchy distribution), which has heavier tails. It also includes any distribution whose negative log-density ψ is a globally analytic function—such as the logistic distribution, the hyperbolic secant distribution, etc. We are now ready to state the result that relates the Gaussian noise model (41) to the general noise model (45).

Theorem 3. *Let n be a parameter and $s \ge 3$ be a constant. Let $\beta \in \mathbb{R}$ denote the signal strength and $v \in \mathbb{R}^n$ denote the spike vector. Let tensor $T(\beta, v) \in \mathbb{R}^{n^{\otimes s}}$ be distributed according to model (41). Let tensor $\widetilde{T}(\beta, v) \in \mathbb{R}^{n^{\otimes s}}$ be distributed according to model (45) with noise distribution $\mathcal{Q}_0$. Suppose Assumption 4 holds. Then there is a reduction algorithm $\mathcal{R} : \mathbb{R}^{n^{\otimes s}} \to \mathbb{R}^{n^{\otimes s}}$ satisfying the following: For each $\delta \in (0,1)$, if the scale of $\mathcal{Q}_0$ is set to $\tau^2 = \Theta\big(\mathsf{polylog}(n^s/\delta)\big)$, then $\mathcal{R}$ runs in time*

$$\mathcal{O}\Big(\mathsf{poly}\big(n^s/\delta\big)\Big) \quad \textit{and} \quad \sup_{\beta \in \mathbb{R}, v \in \mathbb{R}^n} \mathsf{d}_{\mathsf{TV}}\Big(\mathcal{R}\big(T(\beta, v)\big), \widetilde{T}(\beta, v)\Big) \le \delta.$$

We provide the proof of Theorem 3 in Section 8.2. A few comments on Theorem 3 are in order. First, the reduction guarantee does not impose any assumptions on the signal strength β or the spike vector v. In particular, there is no need to assume any prior distribution on β or v. Second, the reduction algorithm $\mathcal{R}$ is constructed by applying the entrywise transformation $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ from Theorem 1 to each entry of the tensor $T(\beta, v)$. Specifically, recall that the distribution $\mathcal{Q}_\theta$ defined in Eq. (20) is equivalent to a location model centered at θ, but with noise distribution $\mathcal{Q}_0$. Then, conditional on β and v, for each index tuple $(i_1, \dots, i_s)$, we have

$$T(\beta, v)_{i_1,\dots,i_s} \sim \mathcal{N}(\theta, \sigma^2_{i_1,\dots,i_s}) \quad \text{and} \quad \widetilde{T}(\beta, v)_{i_1,\dots,i_s} \sim \mathcal{Q}_\theta, \quad \text{where } \theta = \beta n^{s/2} v_{i_1} \times v_{i_2} \times \cdots \times v_{i_s}.$$

This setup exactly matches the source and target distributions considered in Theorem 1.

The following hardness result immediately follows by combing the hardness of the source problem (Proposition 2) with the reduction (Theorem 3).

Corollary 5. *Suppose $s \geq 3$ is a constant. Consider the tensor $\widetilde{T} = \beta(\sqrt{n}v)^{\otimes n} + \widetilde{\zeta}$, where $v \sim \mathsf{Unif}(\mathbb{S}^{n-1})$ and $\widetilde{\zeta} \in \mathbb{R}^{n^{\otimes s}}$ is the symmetric noise tensor whose entries have density $f_{\mathcal{Q}_0}$ defined in Eq. (46). Suppose the noise distribution $\mathcal{Q}_0$ satisfies Assumption 4 and the scale parameter satisfies $\tau^2 = \Theta\big(\mathsf{polylog}(n)\big)$. Suppose the k-HPC^s conjecture holds. If $\beta = \widetilde{o}\big(n^{-s/4}\big)$ and $\beta = \omega\big(n^{-s/2}\sqrt{s\log(n)}\big)$, then there* cannot *exist an algorithm $\mathcal{A} : \mathbb{R}^{n^{\otimes s}} \to \{0, 1\}$ that runs in time $\mathsf{poly}(n)$ and satisfies*

$$\big\langle \mathcal{A}\big(\widetilde{T}\big), v \big\rangle = \Omega(1) \quad \textit{with probability at least} \quad \Omega_n(1).$$

We prove Corollary 5 in Section 8.3. A few comments on Corollary 5 are in order. First, note that the scale parameter (which roughly corresponds to variance inflation) $\tau^2 = \Theta\big(\mathsf{polylog}(n)\big)$ only affects the statistical-computational gap up to a polylogarithmic factor. This is because the signal-to-noise ratio in the non-Gaussian tensor PCA model (45) is given by β/τ, and we only require τ to be *polylogarithmic* in n. Note that the gap itself is polynomial, so this polylogarithmic change should not be thought of as substantial. Second, to the best of our knowledge, Corollary 5 is the first reduction-based hardness result for tensor PCA with non-Gaussian noise. It demonstrates a striking universality phenomenon: the same computational hardness persists for a wide class of noise distributions, including both light-tailed (e.g., generalized normal) and heavy-tailed (e.g., Student's t or Cauchy) distributions.

Remark 4. *While Kunisky (2025) also establish a form of universality, their result applies only to a restricted class of algorithms, namely low coordinate degree algorithms. In contrast, our result applies to all polynomial-time algorithms, a strictly broader class. Similarly, Brennan et al. (2019) demonstrate universality of computational lower bounds for the submatrix model; however, their techniques apply only when the matrix entries are drawn from at most two distributions. In contrast, our results hold under a continuous prior on the rank-one tensor, so that the entries of the tensor may be drawn from infinitely many possible distributions.*

3.2 Symmetric mixtures of linear regressions and generalized linear models

In the symmetric version of the mixture of linear regressions model (Quandt and Ramsey, 1978; Städler et al., 2010), there is some unknown vector $v \in \mathbb{R}^d$ and we observe i.i.d. data $\{(x_i, y_i)\}_{i=1}^n$ generated according to

$$y_i = R_i \cdot \langle x_i, v \rangle + \zeta_i, \quad \text{where} \quad \zeta_i \sim \mathcal{N}(0, 1). \tag{47}$$

Here, the random variable $R_i \in \{-1, 1\}$ and the noise ζ_i are unobserved and independent of each other. For convenience, we denote $\eta = \|v\|_2$. In the setting that $v \in \mathbb{R}^d$ is k-sparse, the task is of estimating v in ℓ_2-error $o(\eta)$ exhibits a statistical-computational gap. In particular, the information-theoretic limit occurs at sample complexity $n = \widetilde{\Theta}(k \log(d)/\eta^4)$; see Fan et al. (2018, Section E.3) for details[4]. On the other hand, there is a computational lower bound of $n = \widetilde{\Theta}(k^2/\eta^4)$, below which no polynomial-time algorithm can succeed at the estimation task, under the k-BPC conjecture; see Brennan and Bresler (2020, Theorem 8) for the precise statement. A refined computational lower bound—along with runtime tradeoffs—was obtained by Arpino and Venkataramanan (2023) for the class of low-degree polynomial algorithms. We restate the reduction-based hardness result of Brennan and Bresler (2020, Theorem 8) as the following proposition.

Proposition 3. *Let $k, d, n \in \mathbb{N}$ and $\eta > 0$ be polynomial in each other, $k = o(\sqrt{d})$, and $k = o(n^{1/6})$. Consider the setting $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$, $\{R_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathsf{Rad}$, $v = \eta \mathbf{1}_S/\sqrt{k}$ for some k-subset $S \subseteq [d]$, and y_i follows model (47) for all $i \in [n]$. Suppose the k-BPC conjecture holds. If $n = \widetilde{o}\big(k^2/\eta^4\big)$, then there cannot exist an algorithm $\mathcal{A} : (\mathbb{R}^d \times \mathbb{R})^n \to \mathbb{R}^d$ that runs in time $\mathsf{poly}(n)$ and satisfies*

$$\big\|\mathcal{A}\big(\{x_i, y_i\}_{i=1}^n\big) - v\big\|_2 = o(\eta) \quad \textit{with probability at least} \quad 1 - o_n(1).$$

While Proposition 3 imposes the constraint $k = o(\sqrt{d})$ and $k = o(n^{1/6})$ on the sparsity level for technical reasons, it still establishes computational hardness in a nontrivial regime: By reparametrizing with

$$k = \widetilde{\Theta}\big(n^{\beta}\big) \quad \text{and} \quad \eta^4 = \widetilde{\Theta}\big(n^{-\alpha}\big), \tag{48}$$

for constants $\beta \in (0, 1)$ and $\alpha > 0$, we obtain that every instance in the regime

$$\big\{(\alpha, \beta) : 0 < \beta \le 1/6,\ 1 - 2\beta \le \alpha \le 1 - \beta\big\} \tag{49}$$

is statistically possible but remains computationally intractable for estimation; see Figure 3 for an illustration.

At a heuristic level, the k^2 computational lower bound arises because in the model (47), the Rademacher random variable R_i destroys information about the sign of the regression term $\langle x_i, v\rangle$. Contrast this with the case of sparse linear regression, where no such statistical-computational gap appears. This observation naturally raises the following question: Does the $\Omega(k^2)$ computational lower bound arise for any signless observation model in variants of sparse linear regression?

Toward answering this question, we consider the class of generalized linear models (Barbier et al., 2019) with even link functions. Specifically, suppose there is an unknown vector $v \in \mathbb{R}^d$ and we observe i.i.d. data $\{(x_i, \widetilde{y}_i)\}_{i=1}^n$ generated according to

$$\widetilde{y}_i = f\big(\langle x_i, v\rangle/\tau\big) + \widetilde{\zeta}_i, \quad \text{where} \quad \widetilde{\zeta}_i \sim \mathcal{N}(0, 1). \tag{50}$$

In the above, $f : \mathbb{R} \to \mathbb{R}$ is a general even link function and the parameter $\tau > 0$ is a proxy for the signal-to-noise ratio, which will be specified in concrete applications. Also suppose that the signal v is k-sparse. We now use our results from Section 2 to extend reduction-based hardness results for mixture of sparse linear regressions to a class of generalized sparse linear models.

We make the following assumption on the link function f. Recall the definitions of $L_f(2N)$ and $\widetilde{L}_f(2N, \tau)$ in Eq. (35) as well as the evaluation time $T_{\mathsf{eval}}(N)$ from Definition 1. We will make use of the signed kernel $\mathcal{S}_N^\star$ (14), with target density defined in Eq. (33).

[4] Although the information-theoretic lower bound in Fan et al. (2018) is stated for a hypothesis testing problem, it also applies to the estimation setting, since any estimator achieving ℓ_2-error $o(\eta)$ can be used to solve the corresponding testing problem; see Brennan and Bresler (2020, Appendix P.2) for a rigorous argument.

Assumption 5. *The link function $f : \mathbb{R} \to \mathbb{R}$ is even, i.e., $f(x) = f(-x)$ for all $x \in \mathbb{R}$, and it satisfies Assumption 3 with a pair of constants $(C_f, \widetilde{C}_f)$. For each $\epsilon \in (0,1)$, if $N = \Theta\Big(\overline{C}_f \log^{2(1+C_f)}\big(4\overline{C}_f/\epsilon\big)\Big)$ where $\overline{C}_f$ is defined in Eq. (35), then we have for some $\tau = \mathcal{O}\big(\mathsf{polylog}(1/\epsilon)\big)$,*

$$L_f(2N) = \mathcal{O}\Big(\mathsf{polylog}(1/\epsilon)\Big), \quad \widetilde{L}_f(2N,\tau) = \mathcal{O}\Big(\mathsf{polylog}(1/\epsilon)\Big), \quad \text{and} \quad T_{\mathsf{eval}}(N) = \mathcal{O}\Big(\mathsf{poly}(1/\epsilon)\Big).$$

Note that a broad class of functions satisfies Assumption 5. As illustrated in Corollary 3, Corollary 4, and Example 2, this includes monomials of constant degree—for instance, the function $f(\theta) = \theta^2$ is covered, which corresponds to the phase retrieval model in the context of generalized linear models. Moreover, Assumption 5 also encompasses the class of globally analytic functions (see Definition 2), which includes many canonical and practically relevant link functions such as $f(\theta) = \cos(\theta)$, $f(\theta) = \mathsf{sech}(\theta)$, and $f(\theta) = (1+\theta^2)^{-1}$, among others. Additionally, by Lemma 2, any linear combination of these basic functions is also covered. We are now ready to relate mixtures of linear regressions (47) to the generalized linear model (50) with a link function satisfying Assumption 5.

Theorem 4. *Let $v \in \mathbb{R}^d$ denote the unknown vector and let $\{R_i\}_{i=1}^n \subseteq \{-1,1\}^n$ be any vector of signs. Conditioned on $\{R_i\}_{i=1}^n$ and v, suppose samples $\{(x_i, y_i)\}_{i=1}^n$ are drawn* i.i.d. *according to model (47) and samples $\{(x_i, \widetilde{y}_i)\}_{i=1}^n$ are drawn* i.i.d. *according to model (50). Suppose the link function $f : \mathbb{R} \to \mathbb{R}$ satisfies Assumption 5. Then there is a reduction algorithm $\mathcal{R} : (\mathbb{R}^d \times \mathbb{R})^n \to (\mathbb{R}^d \times \mathbb{R})^n$ satisfying the following: For each $\delta \in (0,1)$, if we set*

$$\tau = \mathsf{polylog}(n/\delta) \quad \text{and large enough so that} \quad \tau \geq 2 \max_{i \in [n]} \big|\langle x_i, v\rangle\big|, \tag{51}$$

then the reduction $\mathcal{R}$ runs in time

$$\mathcal{O}\Big(\mathsf{poly}(n/\delta)\Big) \quad \text{and satisfies} \quad \mathsf{d_{TV}}\Big(\mathcal{R}\big(\{(x_i,y_i)\}_{i=1}^n\big), \{(x_i,\widetilde{y}_i)\}_{i=1}^n\Big) \leq \delta.$$

We provide the proof in Section 8.4. Note that the reduction guarantee does not impose any assumptions on the sign variables $\{R_i\}_{i=1}^n$ and the covariates $\{x_i\}_{i=1}^n$. It also does not place any structural assumptions on v such as sparsity. In particular, there is no need to assume any prior distribution on $\{R_i\}_{i=1}^n$, $\{x_i\}_{i=1}^n$, and v. This is because the reduction algorithm $\mathcal{R}$ is constructed by applying the entrywise transformation $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ from Theorem 2 to each sample y_i. Specifically, conditional on the tuple (R_i, x_i, v), we have

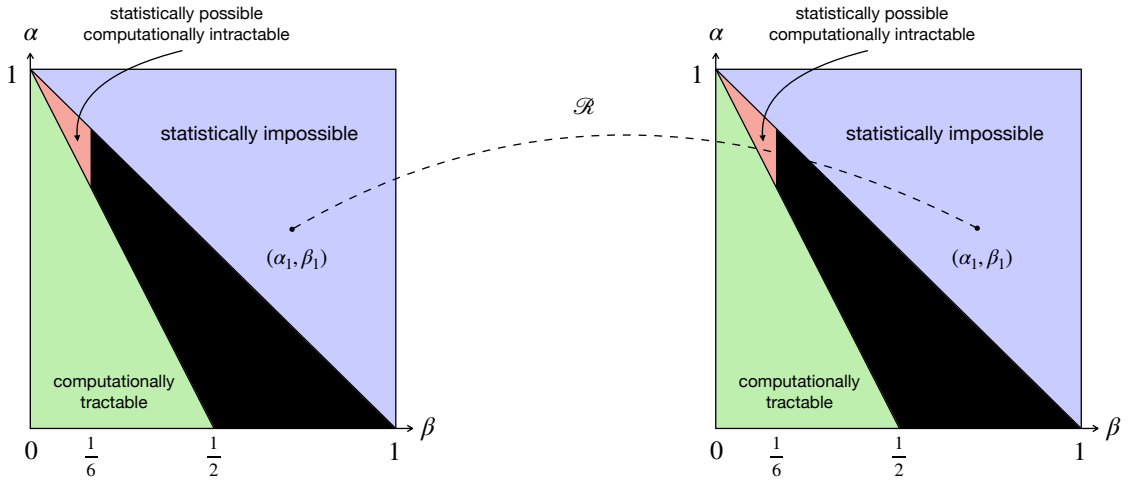
$$y_i \sim \mathcal{N}(\theta, 1) \quad \text{and} \quad \widetilde{y}_i \sim \mathcal{N}\big(f(\theta/\tau), 1\big), \quad \text{where } \theta = R_i \cdot \langle x_i, v\rangle.$$

This setup exactly matches the source and target distributions considered in Theorem 2.

Remark 5. *Note that any instance of the mixture of sparse linear regressions (*MixSLR*) (47) with signal v is mapped under the reduction $\mathcal{R}$ to an instance of the sparse generalized linear model (*SpGLM*) (50) with signal v/τ. In most applications of Theorem 4, it suffices to take δ to be an inverse polynomial of n and in the Gaussian covariance setting $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$, standard Gaussian tail bounds and a union bound yield*

$$\mathbb{P}\Big\{\max_{i\in[n]} |\langle x_i, v\rangle| \leq 2\sqrt{\log(n)}\Big\} \geq 1 - n^{-1} \quad \text{for} \quad \|v\|_2 \leq 1.$$

Thus, condition (51) is satisfied by letting $\tau = \Theta(\mathsf{polylog}(n))$. Therefore, the reduction $\mathcal{R}$ preserves the sparsity k and only changes the signal strength η by a polylogarithmic factor.



(a) mixture of sparse linear regressions (b) sparse phase retrieval

Figure 3. Phase diagrams for the mixture of sparse linear regressions (MixSLR; panel (a)) and sparse phase retrieval (SPR; panel (b)), under the parameterization $k = \widetilde{\Theta}(n^{\beta})$ and $\eta^4 = \widetilde{\Theta}(n^{-\alpha})$, with constants $\alpha, \beta \in (0, 1)$. The purple region denotes the statistically impossible regime; the red region denotes the statistically possible but computationally intractable regime; the green region denotes the computationally tractable regime; and the black region denotes a statistically possible yet conjectured computationally hard regime, whose proof remains open. The reduction $\mathcal{R}$ maps MixSLR instances with parameters (α, β) to SPR instances with the same (α, β) for all (α, β) in the phase diagram, so any computational hardness regime established in panel (a) transfers directly to panel (b).

The hardness result of MixSLR (Proposition 3) and the reduction (Theorem 4) imply the following computational hardness of SpGLM.

Corollary 6. *Let $k, d, n \in \mathbb{N}$ and $\eta > 0$ be polynomial in each other, $k = o(\sqrt{d})$, and $k = o(n^{1/6})$. Consider the setting $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$, $v = \eta \mathbf{1}_S/\sqrt{k}$ for some k-subset $S \subseteq [d]$, and $\widetilde{y}_i$ follows model (50) for all $i \in [n]$. Suppose in model (50), the link function f satisfies Assumption 5 and the scale parameter satisfies $\tau = \Theta\big(\mathsf{polylog}(n)\big)$. Suppose the k-BPC conjecture holds. If $n = \widetilde{o}\big(k^2/\eta^4\big)$, then there cannot exist an algorithm $\mathcal{A} : (\mathbb{R}^d \times \mathbb{R})^n \to \mathbb{R}^d$ that runs in time $\mathsf{poly}(n)$ and satisfies*

$$\big\|\mathcal{A}\big(\{x_i, \widetilde{y}_i\}\big) - v\big\|_2 = o(\eta) \quad \textit{with probability at least} \quad 1 - o_n(1).$$

We provide the proof of Corollary 6 in Section 8.5, and record a few remarks here. First, the constraints $k = o(\sqrt{d})$ and $k = o(n^{1/6})$ are inherited from Proposition 3, which is the source of the computational hardness. Nevertheless, this still establishes computational hardness in the nontrivial regime defined in Eq. (49) under the parameterization (48); see the illustration of this computationally intractable regime in Figure 3. Second, our reduction itself imposes no constraints on k, so any hardness shown for MixSLR automatically transfers to SpGLM. In particular, to understand the computational complexity of SpGLM in the black region of Figure 3(b), it suffices to understand the corresponding black region of Figure 3(a). Third, the scaling factor τ does not affect the signal strength beyond a polylogarithmic factor, since in model (50) the effective signal strength is $\|v\|_2/\tau = \eta/\tau$, and we set $\tau = \Theta(\mathsf{polylog}(n))$. We have thus demonstrated that the $\widetilde{\Omega}(k^2)$ computational barrier is universal for models with signless observations arising from even link functions. This resolves an open question posed in Brennan and Bresler (2020, Appendix D). We remark that a k-to-k^2 gap has been established under the statistical query model (Wang et al.,

2019) for a class of even link functions, but this only applies to a subset of algorithms and link functions.

We now highlight the computational hardness result for sparse phase retrieval (SPR), where we have the model

$$\widetilde{y}_i = |\langle x_i, v\rangle|^2 + \widetilde{\zeta}_i, \quad \text{where} \quad \widetilde{\zeta}_i \sim \mathcal{N}(0,1). \tag{52}$$

The following computational hardness result immediately follows from Corollary 6.

Corollary 7 (Sparse Phase Retrieval). *Let $k, d, n \in \mathbb{N}$ and $\eta > 0$ be polynomial in each other, $k = o(\sqrt{d})$, and $k = o(n^{1/6})$. Consider the setting $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$, $v = \eta \mathbf{1}_S/\sqrt{k}$ for some k-subset $S \subseteq [d]$, and $\widetilde{y}_i$ follows the phase retrieval model (52) for all $i \in [n]$. Suppose the k-BPC conjecture holds. If $n = \widetilde{o}\big(k^2/\eta^4\big)$, then there cannot exist an algorithm $\mathcal{A} : (\mathbb{R}^d \times \mathbb{R})^n \to \mathbb{R}^d$ that runs in time $\mathsf{poly}(n)$ and satisfies*

$$\big\|\mathcal{A}\big(\{x_i, \widetilde{y}_i\}\big) - v\big\|_2 = o(\eta) \quad \textit{with probability at least} \quad 1 - o_n(1).$$

Note that the square link function $f(t) = t^2$ satisfies Assumption 5 (see Corollary 3). Hence, Corollary 7 follows directly from Corollary 6 by setting $f(\theta) = \theta^2$ and absorbing τ into η. The information-theoretic limit for k-sparse phase retrieval is known to occur at sample complexity $n = \Theta\big(k \log(d)/\eta^4\big)$; see Cai et al. (2016); Lecué and Mendelson (2015). In contrast, all polynomial-time algorithms require sample complexity in the order of k^2; see Li and Voroninski (2013); Cai et al. (2016); Wang et al. (2017); Celentano et al. (2020). Motiveted by this gap, Cai et al. (2016) conjectured that any computationally efficient method must require $n = \widetilde{\Omega}(k^2/\eta^4)$ samples. Corollary 4 provides the first reduction-based confirmation of this conjecture by establishing a computational lower bound: no polynomial-time algorithm can succeed when $n = \widetilde{o}(k^2/\eta^4)$. This demonstrates a statistical–computational gap between the information-theoretic rate k and the computational barrier k^2. In particular, it establishes computational hardness in a nontrivial regime; see Figure 3(b) for an illustration.

3.3 Sparse rank-1 submatrix model and planted submatrix model

In the sparse rank-1 submatrix model (Rank1SubMat), we observe an $n \times n$ matrix

$$M = \mu\, r c^\top + \zeta, \quad \text{where } r, c \in S_{n,k},\ \zeta \sim \mathcal{N}(0,1)^{\otimes n \times n}. \tag{53}$$

Here, $\mu \in \mathbb{R}$ is the signal strength, $\mathcal{N}(0,1)^{\otimes n \times n}$ denotes the distribution of an $n \times n$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries, and $S_{n,k}$ denotes set of k-sparse unit vectors defined as

$$S_{n,k} := \left\{ x \in \mathbb{S}^{n-1} :\ \|x\|_0 = k,\ x_i \in \left\{0, \pm\tfrac{1}{\sqrt{k}}\right\} \text{ for all } i \in [n] \right\}. \tag{54}$$

The corresponding detection problem can be formalized as distinguishing between the following two hypotheses upon observing the matrix M:

$$\mathcal{H}_0 : M \text{ follows model (53) with } \mu = 0, \quad \text{and} \tag{55a}$$
$$\mathcal{H}_1 : M \text{ follows model (53) for some } \mu > 0, r, c \in S_{n,k}. \tag{55b}$$

The detection problem in Eq. (55) exhibits a computational barrier, which we briefly describe. Define the computational threshold

$$\mu^{\text{comp}}_{\mathsf{Rank1}} := \min\left\{k, \sqrt{n}\right\}. \tag{56}$$

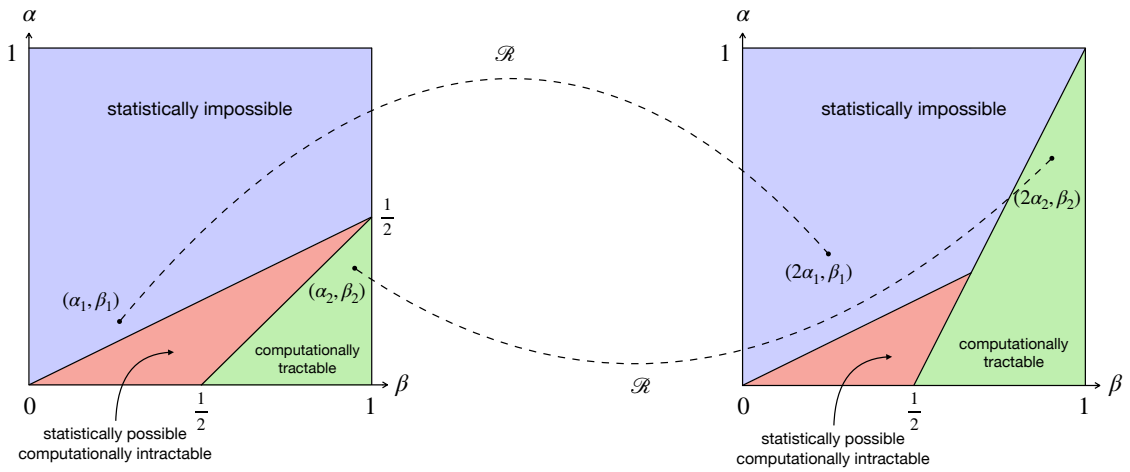


Figure 4. Phase diagrams for the sparse rank-1 submatrix model (Rank1SubMat; panel (a)) and the planted submatrix model (PlantSubMat; panel (b)), under the parameterization $\mu/k = \widetilde{\Theta}(n^{-\alpha})$ and $k = \widetilde{\Theta}(n^{\beta})$ for $\alpha, \beta \in (0,1)$. The purple region indicates the statistically impossible regime. The green region corresponds to the computationally tractable regime where polynomial-time algorithms are known to succeed. The red region indicates the statistically possible but computationally intractable regime. In panel (a), the computational threshold is given by the line $\alpha = \beta - 1/2$, while in panel (b), the threshold is the line $\alpha = 2\beta - 1$. The reduction $\mathcal{R}$ established in Theorem 5 maps instances with parameters (α, β) in Rank1SubMat to instances with parameters $(2\alpha, \beta)$ in PlantSubMat. This transformation demonstrates a pointwise correspondence between the phase diagrams, preserving the computational tractability and intractability of problem instances.

When the signal strength satisfies $\mu = \widetilde{\omega}\big(\mu^{\mathrm{comp}}_{\mathsf{Rank1}}\big)$, polynomial-time algorithms exist for detection, and these also succeed for recovering r, c under H_1. Specifically, if $\mu = \widetilde{\omega}(k)$, then each entry of the signal matrix $\mu\, rc^\top$ dominates the noise ζ. If $\mu = \widetilde{\omega}(\sqrt{n})$, then spectral methods succeed (Montanari et al., 2015), since the spectral norm of the noise matrix is only $\mathcal{O}(\sqrt{n})$ and the top singular vectors of M therefore reveal information about the signal $\mu\, rc^\top$. Conversely, it has been shown via reductions from the planted clique conjecture that no polynomial-time algorithm can succeed when $\mu = \widetilde{o}\big(\mu^{\mathrm{comp}}_{\mathsf{Rank1}}\big)$; see Brennan et al. (2018).

In the planted submatrix model (PlantSubMat), we observe an $n \times n$ matrix

$$M = \mu\, rc^\top + \zeta, \quad \text{where } r, c \in S^+_{n,k},\ \zeta \sim \mathcal{N}(0,1)^{\otimes n \times n}. \tag{57}$$

Here, the spike vectors r and c are restricted to have nonnegative entries. Specifically, $S^+_{n,k}$ denotes the set of k-sparse unit vectors with nonnegative coordinates:

$$S^+_{n,k} := \left\{ x \in \mathbb{S}^{n-1} : \|x\|_0 = k,\ x_i \in \left\{0, \tfrac{1}{\sqrt{k}}\right\}\ \forall i \in [n] \right\}. \tag{58}$$

The corresponding detection problem involves distinguishing between the following hypotheses upon observing M:

$$\mathcal{H}_0 : M \text{ follows model (57) with } \mu = 0, \quad \text{and} \tag{59a}$$

$$\mathcal{H}_1 : M \text{ follows model (57) for some } \mu > 0, r, c \in S^+_{n,k}. \tag{59b}$$

Compared to the sparse rank-1 submatrix model, the planted submatrix model exhibits a different computational threshold, defined as

$$\mu^{\mathrm{comp}}_{\mathsf{Plant}} := \min\left\{k, \tfrac{n}{k}\right\}. \tag{60}$$

When $\mu = \widetilde{\omega}\big(\mu_{\mathsf{Plant}}^{\mathrm{comp}}\big)$, polynomial-time algorithms exist. For instance, when $\mu = \widetilde{\omega}(k)$, each entry of the signal matrix $\mu\, rc^\top$ again dominates the noise. Alternatively, when $\mu = \widetilde{\omega}(n/k)$, summing all entries of $\mu\, rc^\top$ yields a signal $\mu k = \widetilde{\omega}(n)$, while the total sum of entries in the noise matrix is only $\widetilde{O}(n)$, making the two hypotheses distinguishable. Conversely, when $\mu = \widetilde{o}\big(\mu_{\mathsf{Plant}}^{\mathrm{comp}}\big)$, then no polynomial-time algorithm is known, and evidence of computational hardness has been provided through reductions from the planted clique conjecture (Ma and Wu, 2015; Brennan et al., 2018).

As observed, the computational threshold $\mu_{\mathsf{Plant}}^{\mathrm{comp}}$ is significantly smaller than $\mu_{\mathsf{Rank1}}^{\mathrm{comp}}$ in the regime $k \geq \sqrt{n}$, indicating that PlantSubMat is computationally easier than Rank1SubMat. To rigorously establish this relation, we construct a reduction from Rank1SubMat to PlantSubMat, thereby showing that the computational threshold of Rank1SubMat directly implies the corresponding threshold of PlantSubMat, without relying on the planted clique conjecture. We now formally state the result.

Theorem 5. *Let $S_{n,k} \subseteq \mathbb{R}^n$ be defined as in Eq. (54). There exists a universal and positive constant C and a reduction algorithm $\mathcal{R} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ such that the following holds: For each $\delta \in (0,1)$, if the tuple of parameters (μ, k, n, τ) satisfies*

$$\tau^2 \geq C \log^{24}(n^2/\delta) + \frac{4\mu^2}{k^2}, \tag{61}$$

then the reduction $\mathcal{R}$ runs in time $\mathcal{O}\Big(n^2 \log^{64}\big(Cn^2/\delta\big)\Big)$ and

$$\sup_{\mu\in\mathbb{R},\, r,c\in S_{n,k}} \mathsf{d}_{\mathsf{TV}}\bigg(\mathcal{R}\Big(\mu rc^\top + \zeta\Big),\ \frac{\mu^2}{\tau^2 k}|r||c|^\top + \zeta\bigg) \leq \delta,$$

where $\zeta \sim \mathcal{N}(0,1)^{\otimes n\times n}$ and $|r|, |c| \in \mathbb{R}^n$ denote the vectors obtained by taking the entrywise absolute value of r, c, respectively.

We provide the proof of Theorem 5 in Section 8.6. Note that it is typical to set $\delta = \mathcal{O}(n^{-C'})$ for some constant $C' > 0$ to achieve meaningful reductions, so that condition (61) is satisfied by choosing $\tau = \Theta(\mathrm{polylog}(n))$.

Let us now discuss some consequences of the theorem, focusing on the regime where $\mu/k = \widetilde{\Theta}(n^{-\alpha})$ for some $\alpha \in (0,1)$. The reduction maps an instance of Rank1SubMat with parameters (μ, k, n) to an instance of PlantSubMat with parameters $\big(\mu^2/(\tau^2 k),\ k,\ n\big)$. Since $\tau = \Theta(\mathrm{polylog}(n))$, its effect on the signal strength can be neglected. This ensures a pointwise correspondence between the phase diagrams of the two models: if the Rank1SubMat instance with parameters (μ, k, n) is computationally tractable, then its image under the reduction, an instance of PlantSubMat with parameters $(\mu^2/(\tau^2 k), k, n)$, is also tractable, and the same holds for intractability. This correspondence can be formalized by parameterizing $\mu/k = \widetilde{\Theta}(n^{-\alpha})$ and $k = \widetilde{\Theta}(n^{\beta})$ for $\alpha, \beta \in (0,1)$. Under this parameterization, a point (α, β) in the Rank1SubMat phase diagram is mapped to $(2\alpha, \beta)$ in the PlantSubMat phase diagram, confirming that the reduction preserves computational feasibility pointwise; see Figure 4 for an illustration.

Finally, we emphasize that Theorem 5 establishes computational lower bounds without relying on the planted clique conjecture. Indeed, even if the planted clique conjecture were false, a computational lower bound for Rank1SubMat would still imply the corresponding lower bound for PlantSubMat. Moreover, reductions from Rank1SubMat provide evidence of hardness beyond polynomial time. Since existing hardness results for PlantSubMat are conditioned on the planted clique conjecture (Brennan et al., 2018; Ma and Wu, 2015), they rule out only polynomial-time algorithms. However, these reductions lose strength once we go beyond polynomial time, since the planted clique problem is known to be solvable in quasi-polynomial time, specifically $\mathcal{O}\big(n^{\mathcal{O}(\log n)}\big)$, whenever the

clique size exceeds the information-theoretic threshold. In contrast, reductions from Rank1SubMat remain meaningful in this regime, since Rank1SubMat is believed to require subexponential time in the hard regime. More concretely, if we fix $\mu = \widetilde{\Theta}(\sqrt{n})$ and let the sparsity ratio be $\rho = k/n$, then in the statistically feasible but computationally intractable regime $n^{-1/2} \ll \rho \ll 1$, Ding et al. (2024) showed that recovery of the spike requires time $\mathcal{O}\big(\exp(\rho^2 n)\big)$, and evidence for the optimality of this recovery algorithm was also provided. Consequently, hardness based on Rank1SubMat can exclude a strictly larger class of algorithms than hardness based on planted clique.

4 Discussion

Motivated by the goal of characterizing fundamental statistical-computational tradeoffs in canonical high-dimensional models, we constructed a Markov kernel that maps the Gaussian location model (the source) to general target models within small total variation deficiency. The target models we covered include (i) general non-Gaussian location models and (ii) Gaussian models with a nonlinearly transformed location parameter. Building on this kernel, we designed a computationally efficient reduction that transforms a single source observation into a single target observation without knowledge of the underlying location parameter. This single-sample reduction allows us to establish new computational lower bounds for several high-dimensional problems including tensor PCA and sparse generalized linear models, as well as novel connections between the sparse rank-one submatrix model and the planted submatrix model.

Our work leaves several open questions, which we briefly outline here. First, our construction of the optimal signed kernel relies on the fact that, under the Gaussian measure, the Hermite polynomials form an orthonormal basis for a suitable function space. Many other distributions also admit orthonormal polynomial bases—for instance, the Laguerre polynomials with respect to the Gamma distribution, and the Gegenbauer polynomials with respect to the Beta distribution; see Goldstein and Reinert (2005) for more examples. Can our design of the optimal signed kernel be generalized to these source distributions as well? Furthermore, for Laplace, Uniform, and Erlang sources, the optimal signed kernels have been derived using different techniques, such as Fourier inversion; see Lou et al. (2025). Can we unify these approaches under a single framework based on orthonormal polynomials?

Second, our technique relies critically on the differentiability of the target density with respect to the location parameter. For instance, when the target is the Laplace location model, the reduction fails due to the density being non-differentiable at a single point. Nevertheless, Gaussian-to-Laplace reductions are of particular interest, for example in differential privacy. Can the optimal Markov kernel defined in Eq. (9) overcome this limitation, or does a lower bound—perhaps based on the data processing inequality—rule out such reductions entirely? Furthermore, our reduction requires the target density to satisfy smoothness assumptions such as Assumption 2 and Assumption 3. While these are met by a broad class of distributions, an important open question is whether these conditions can be relaxed or even removed, or conversely, whether one can establish lower bounds showing that they are indeed necessary.

Finally, we believe the Gaussian-source reduction framework has applications well beyond the settings explored in this paper. For instance, we chose to focus our discussion of universality of computational hardness on tensor PCA. However, the reduction in Theorem 1 is quite general, and could be adapted to regression and denoising models where the observations are corrupted by Gaussian noise. By transferring these to analogous models with non-Gaussian noise satisfying Assumption 4, one could potentially extend universality results across a broad class of problems, complementing the directions studied in Brennan et al. (2018); Brennan and Bresler (2020).

Acknowledgments

ML was supported in part by the ARC-ACO fellowship through the Algorithms and Randomness Center at Georgia Tech. ML and AP were supported in part by the NSF under grants CCF-2107455 and DMS-2210734, a Google Research Scholar award, and by research awards/gifts from Adobe, Amazon, and Mathworks. GB gratefully acknowledges support from NSF award CCF-2428619.

5 Proofs of results from Section 2.1

We provide proofs of Proposition 1 and Lemma 1 below. We begin by reviewing some basic properties of the Hermite polynomials defined in Eqs. (6) and (7), which will be used frequently in our proofs.

First, note that the normalized Hermite polynomials are orthonormal in that $\langle \overline{H}_n(\cdot), \overline{H}_m(\cdot) \rangle = \delta_{nm}$, where δ_{nm} is the Kronecker delta with $\delta_{nm} = \mathbb{I}\{n = m\}$. The other properties we use are:

$$\left|H_n(x)\exp(-x^2/2)\right| \le \left(\sqrt{\pi}2^n n!\right)^{1/2}, \qquad \left|\overline{H}_n(x)\exp(-x^2/2)\right| \le 1, \tag{62a}$$

$$\frac{\mathrm{d}H_{n+1}(x)}{\mathrm{d}x} = 2(n+1)H_n(x), \qquad \frac{\mathrm{d}\overline{H}_{n+1}(x)}{\mathrm{d}x} = \sqrt{2(n+1)}\cdot\overline{H}_n(x), \tag{62b}$$

$$\text{and} \qquad H_n(x) = \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{n!}{m!(n-2m)!}\cdot(-1)^m(2x)^{n-2m}. \tag{62c}$$

Inequality (62a) can be found in Szego (1939, page 190), and Eqs. (62b) and (62c) correspond to Szego (1939, Eq. (5.5.10) and Eq. (5.5.4)), respectively.

5.1 Proof of Proposition 1

Recall the signed $\mathcal{S}^\star$ from Eq. (13) and the source density $u(x;\theta)$ from Eq. (4). It suffices to verify

$$\int_{\mathbb{R}} \mathcal{S}^\star(y \mid x)\cdot u(x;\theta)\mathrm{d}x = v(y;\theta) \quad \text{for all} \quad y,\theta \in \mathbb{R}. \tag{63}$$

To reduce the notational burden, we let $c_k = (-1)^k\sigma^{2k}/(2k)!!$ for $k \in \mathbb{Z}_{\ge 0}$ and $\phi(x) = e^{-x^2}/\sqrt{\pi}$ for $x \in \mathbb{R}$. In our proof, we separately evaluate the LHS and RHS of Eq. (63) in terms of Hermite projections, and then show that these are equal.

LHS of Eq. (63). By applying integration by parts and using Assumption 1(b), we obtain

$$\begin{aligned}
\int_{\mathbb{R}} \nabla_x^{(2k)} v(y;x)\cdot\phi\left(\frac{x-\theta}{\sqrt{2}\sigma}\right)\mathrm{d}x &= \nabla_x^{(2k-1)} v(y;x)\cdot\phi\left(\frac{x-\theta}{\sqrt{2}\sigma}\right)\Big|_{-\infty}^{+\infty} - \int_{\mathbb{R}} \nabla_x^{(2k-1)} v(y;x)\cdot\phi'\left(\frac{x-\theta}{\sqrt{2}\sigma}\right)\cdot\frac{1}{\sqrt{2}\sigma}\mathrm{d}x \\
&= -\frac{1}{\sqrt{2}\sigma}\int_{\mathbb{R}} \nabla_x^{(2k-1)} v(y;x)\cdot\phi'\left(\frac{x-\theta}{\sqrt{2}\sigma}\right)\mathrm{d}x \\
&\overset{\text{(i)}}{=} \left(\frac{1}{\sqrt{2}\sigma}\right)^{2k}\cdot\int_{\mathbb{R}} v(y;x)\cdot\phi^{(2k)}\left(\frac{x-\theta}{\sqrt{2}\sigma}\right)\mathrm{d}x \\
&\overset{\text{(ii)}}{=} \frac{\sqrt{2}\sigma}{2^k\sigma^{2k}}\int_{\mathbb{R}} v\left(y;\theta+\sqrt{2}\sigma x\right)\cdot\phi^{(2k)}(x)\mathrm{d}x \\
&\overset{\text{(iii)}}{=} \frac{\sqrt{2}\sigma}{2^k\sigma^{2k}}\int_{\mathbb{R}} v\left(y;\theta+\sqrt{2}\sigma x\right)\cdot H_{2k}(x)\frac{e^{-x^2}}{\sqrt{\pi}}\mathrm{d}x,
\end{aligned}$$

where in step (i) we perform the integration by parts from the previous step $2k$ times. In step (ii) we use a change of variables, and in step (iii) we use the definition of Hermite polynomials (6). Consequently, we obtain

$$\begin{aligned}\int_{\mathbb{R}} c_k \nabla_x^{(2k)} v(y;x) \cdot u(x;\theta) \mathrm{d}x &= \int_{\mathbb{R}} c_k \nabla_x^{(2k)} v(y;x) \cdot \frac{1}{\sqrt{2}\sigma} \phi\left(\frac{x-\theta}{\sqrt{2}\sigma}\right) \mathrm{d}x \\ &= \frac{c_k}{2^k \sigma^{2k}} \int_{\mathbb{R}} v\big(y; \theta + \sqrt{2}\sigma x\big) \cdot H_{2k}(x) \frac{e^{-x^2}}{\sqrt{\pi}} \mathrm{d}x.\end{aligned}$$

Now define $f(x;\theta,y) = v\big(y;\theta+\sqrt{2}\sigma x\big)$ for convenience, and recall our definition of the inner product operation $\langle \cdot, \cdot \rangle$ from Eq. (8). Then

$$\begin{aligned}\int_{\mathbb{R}} c_k \nabla_x^{(2k)} v(y;x) \cdot u(x;\theta) \mathrm{d}x &= \frac{c_k}{\sqrt{\pi} 2^k \sigma^{2k}} \big\langle f(\cdot;\theta,y), H_{2k}(\cdot) \big\rangle \\ &= \big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot) \big\rangle \cdot \frac{c_k}{\sqrt{\pi} 2^k \sigma^{2k}} \cdot (\sqrt{\pi} 2^{2k} (2k)!)^{1/2} \\ &= \big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot) \big\rangle \cdot \overline{H}_{2k}(0). \end{aligned} \tag{64}$$

To justify Eq. (64), recall that $c_k = (-1)^k \sigma^{2k}/(2k)!!$. Then by definition of $\overline{H}_{2k}(x)$ in Eq. (6), we have

$$\overline{H}_{2k}(0) = \frac{H_{2k}(0)}{\big(\sqrt{\pi} 2^{2k} (2k)!\big)^{1/2}} \overset{\text{(i)}}{=} \frac{(-1)^k (2k)!/k!}{\big(\sqrt{\pi} 2^{2k} (2k)!\big)^{1/2}} = \frac{c_k}{\sqrt{\pi} 2^k \sigma^{2k}} \cdot (\sqrt{\pi} 2^{2k} (2k)!)^{1/2},$$

where in step (i) we use Szego (1939, Eq. (5.5.5)) so that $H_{2k}(0) = (-1)^k (2k)!/k!$.

Putting together the pieces, we obtain

$$\begin{aligned}\int_{\mathbb{R}} \mathcal{S}^\star(y \mid x) \cdot u(x;\theta) \mathrm{d}x &= \int_{\mathbb{R}} \sum_{k=0}^{+\infty} c_k \nabla_x^{(2k)} v(y;x) \cdot u(x;\theta) \mathrm{d}x \\ &\overset{\text{(i)}}{=} \sum_{k=0}^{+\infty} \int_{\mathbb{R}} c_k \nabla_x^{(2k)} v(y;x) \cdot u(x;\theta) \mathrm{d}x \\ &\overset{\text{(ii)}}{=} \sum_{k=0}^{+\infty} \big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot) \big\rangle \cdot \overline{H}_{2k}(0),\end{aligned}$$

where in step (i) we use Assumption 1(d) and Fubini's theorem so that switching the summation and integral is valid, and in step (ii) we use Eq. (64).

RHS of Eq. (63). By Assumption 1(c), $f(x;\theta,y)$ admits a Hermite expansion, that is,

$$f(x;\theta,y) = \sum_{k=0}^{+\infty} \big\langle f(\cdot;\theta,y), \overline{H}_k(\cdot) \big\rangle \cdot \overline{H}_k(x), \quad \text{for all} \quad x, y \in \mathbb{R}.$$

Setting $x = 0$ yields for all $y \in \mathbb{R}$,

$$v(y;\theta) = f(0;\theta,y) = \sum_{k=0}^{+\infty} \big\langle f(\cdot;\theta,y), \overline{H}_k(\cdot) \big\rangle \cdot \overline{H}_k(0) = \sum_{k=0}^{+\infty} \big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot) \big\rangle \cdot \overline{H}_{2k}(0), \tag{65}$$

where in the last step we use the fact that $\overline{H}_k(0) = 0$ when k is odd; see Szego (1939, Eq. (5.5.5)).

Combing the pieces. Our two steps yield

$$\int_{\mathbb{R}} \mathcal{S}^\star(y \mid x) \cdot u(x;\theta)\mathrm{d}x = f(0;\theta,y) = v(y;\theta) \quad \text{for all } y \in \mathbb{R}, \theta \in \Theta.$$

This concludes the proof. □

5.2 Proof of Lemma 1

Before proving the lemma, let us establish the claimed inequality (16) for completeness. Applying the triangle inequality, we have

$$\begin{aligned}\left\| \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) \cdot u(x;\theta)\mathrm{d}x - v(\cdot;\theta) \right\|_1 \le & \left\| \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) \cdot u(x;\theta)\mathrm{d}x - \int_{\mathbb{R}} \mathcal{S}_N^\star(\cdot \mid x) \cdot u(x;\theta)\mathrm{d}x \right\|_1 \\ & + \left\| \int_{\mathbb{R}} \mathcal{S}_N^\star(\cdot \mid x) \cdot u(x;\theta)\mathrm{d}x - v(\cdot;\theta) \right\|_1. \end{aligned} \tag{66}$$

By following the same steps as in Lou et al. (2025, Section 7.2), we obtain

$$\left\| \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) \cdot u(x;\theta)\mathrm{d}x - \int_{\mathbb{R}} \mathcal{S}_N^\star(\cdot \mid x) \cdot u(x;\theta)\mathrm{d}x \right\|_1 \le \int_{\mathbb{R}} \big(|p(x)-1| + q(x)\big) \cdot u(x;\theta)\mathrm{d}x. \tag{67}$$

Consequently, substituting Ineq. (67) into the RHS of Ineq. (66) and taking a supremum over $\theta \in \Theta$ yields inequality (16).

We now turn to proving Ineq. (18); the reader is advised to go through the proof of Proposition 1 for context. In that proof, we defined $f(x;\theta,y) = v\big(y;\theta + \sqrt{2}\sigma x\big)$ and $c_k = (-1)^k \sigma^{2k}/(2k)!!$ for convenience. By Assumption 1(c) and Eq. (65), we have

$$v(y;\theta) = \sum_{k=0}^{+\infty} \big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot) \big\rangle \cdot \overline{H}_{2k}(0) \quad \text{for all } y \in \mathbb{R}, \theta \in \Theta.$$

Next,

$$\begin{aligned}\int_{\mathbb{R}} \mathcal{S}_N^\star(y \mid x) \cdot u(x;\theta)\mathrm{d}x &= \int_{\mathbb{R}} \sum_{k=0}^{N} c_k \nabla_x^{(2k)} v(y;x) \cdot u(x;\theta)\mathrm{d}x \\ &= \sum_{k=0}^{N} \int_{\mathbb{R}} c_k \nabla_x^{(2k)} v(y;x) \cdot u(x;\theta)\mathrm{d}x \\ &\overset{\text{(i)}}{=} \sum_{k=0}^{N} \big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot) \big\rangle \cdot \overline{H}_{2k}(0), \end{aligned}$$

where in step (i) we use Eq. (64). Combining the two equations above yields

$$\begin{aligned}
\Big\| \int_{\mathbb{R}} \mathcal{S}_N^\star(\cdot \mid x) \cdot u(x;\theta)\mathrm{d}x - v(\cdot;\theta)\Big\|_1 &= \int_{\mathbb{R}} \Big| \int_{\mathbb{R}} \mathcal{S}_N^\star(y \mid x) u(x;\theta)\mathrm{d}x - v(y;\theta)\Big|\mathrm{d}y \\
&= \int_{\mathbb{R}} \Big| \sum_{k=0}^{N} \big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot)\big\rangle \cdot \overline{H}_{2k}(0) - \sum_{k=0}^{+\infty} \big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot)\big\rangle \cdot \overline{H}_{2k}(0)\Big|\mathrm{d}y \\
&= \int_{\mathbb{R}} \Big| \sum_{k=N+1}^{+\infty} \big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot)\big\rangle \cdot \overline{H}_{2k}(0)\Big|\mathrm{d}y \\
&\overset{\text{(i)}}{\leq} \int_{\mathbb{R}} \sum_{k=N+1}^{+\infty} \Big|\big\langle f(\cdot;\theta,y), \overline{H}_{2k}(\cdot)\big\rangle\Big|\mathrm{d}y \\
&= \int_{\mathbb{R}} \sum_{k=N+1}^{+\infty} \Big| \int_{\mathbb{R}} v(y;\theta+\sqrt{2}\sigma x) \cdot \overline{H}_{2k}(x)\exp(-x^2)\mathrm{d}x\Big|\mathrm{d}y,
\end{aligned}$$

where in step (i) we apply the triangle inequality and use the sequence of bounds

$$|\overline{H}_{2k}(0)| = \frac{|H_{2k}(0)|}{\big(\sqrt{\pi}2^{2k}(2k)!\big)^{1/2}} = \frac{(2k)!/k!}{\big(\sqrt{\pi}2^{2k}(2k)!\big)^{1/2}} = \frac{(2k-1)!!}{\big(\sqrt{\pi}(2k)!\big)^{1/2}} \leq 1.$$

Taking supremum over $\theta \in \Theta$ on both sides of the inequality in the display above yields the claim. □

6 Proofs of results from Section 2.2

We prove Theorem 1(a) in Section 6.1 and Theorem 1(b) in Section 6.2. We provide the proof of the corollaries and examples following Theorem 1 in Sections 6.3–6.5.

6.1 Proof of Theorem 1(a)

We will use Lemma 1 to bound $\delta(\mathcal{U},\mathcal{V};\mathcal{T}_N^\star)$ by the signed deficiency and other terms, and then proceed to bounding those quantities. The first part of the proof is dedicated to verifying Assumption 1 for the target density $v(y;\theta)$ defined in Eq. (20). This will then allow us to apply Lemma 1.

Assumption 1(a) immediately holds by Assumption 2(b). To verify Assumption 1(b), note that by definition, $v(y;x) = g(h(x))$ where $g(x) = \frac{1}{\tau}\exp(-x)$ and $h(x) = \psi((y-x)/\tau)$ for all $x, y \in \mathbb{R}$. Applying Faà di Bruno's formula yields

$$\begin{aligned}
\nabla_x^{(n)} v(y;x) &= \sum_{m_1,\ldots,m_n} \frac{n!}{m_1!\,m_2!\cdots m_n!} \cdot g^{(m_1+\cdots+m_n)}(h(x)) \cdot \prod_{j=1}^{n} \left(\frac{h^{(j)}(x)}{j!}\right)^{m_j} \\
&= \frac{\exp(-h(x))}{\tau} \cdot (n!) \cdot \sum_{m_1,\ldots,m_n} \frac{(-1)^{m_1+\cdots+m_n}}{m_1!\,m_2!\cdots m_n!} \prod_{j=1}^{n} \left(\psi^{(j)}\left(\frac{y-x}{\tau}\right)\frac{(-1)^j}{\tau^j j!}\right)^{m_j}, \qquad (68)
\end{aligned}$$

where the summation is over all n-tuples of nonnegative integers $(m_1,\cdots,m_n)$ satisfying $\sum_{j=1}^n jm_j = n$. Note that for all $n \in \mathbb{N}$, the following bound holds and will be used repeatedly:

$$\left|\left\{(m_1,\ldots,m_n) : \sum_{j=1}^{n} jm_j = n, m_j \geq 0, m_j \in \mathbb{Z} \text{ for all } j \in [n]\right\}\right| \leq \frac{n}{1}\cdot\frac{n}{2}\cdots\frac{n}{n} \leq \frac{n^n}{n!} \leq e^n. \qquad (69)$$

By applying the triangle inequality, we obtain

$$\begin{aligned}\left|\nabla_x^{(n)} v(y;x)\right| &\le \frac{\exp(-h(x))}{\tau}\cdot(n!)\cdot\sum_{m_1,\dots,m_n}\frac{\prod_{j=1}^n\left|\psi^{(j)}\left(\frac{y-x}{\tau}\right)\frac{1}{\tau^j j!}\right|^{m_j}}{m_1!\,m_2!\cdots m_n!}\\ &\le \frac{\exp(-h(x))}{\tau}\cdot(n!)\cdot\sum_{m_1,\dots,m_n}\left(\sum_{j=1}^n\left|\psi^{(j)}\left(\frac{y-x}{\tau}\right)\frac{1}{\tau^j j!}\right|\right)^{m_1+\cdots+m_n},\end{aligned}$$

where in the last step we use the multinomial theorem so that

$$\frac{\prod_{j=1}^n\left|\psi^{(j)}\left(\frac{y-x}{\tau}\right)\frac{1}{\tau^j j!}\right|^{m_j}}{m_1!\,m_2!\cdots m_n!}\le\left(\sum_{j=1}^n\left|\psi^{(j)}\left(\frac{y-x}{\tau}\right)\frac{1}{\tau^j j!}\right|\right)^{m_1+\cdots+m_n}.$$

Integrating over $x\in\mathbb{R}$ and using the change of variables $t=(y-x)/\tau$ yields

$$\begin{aligned}\int_{\mathbb{R}}\left|\nabla_x^{(n)} v(y;x)\right|\mathrm{d}x &\le n!\cdot\sum_{m_1,\dots,m_n}\int_{\mathbb{R}}\exp(-\psi(t))\left(\sum_{j=1}^n\frac{|\psi^{(j)}(t)|}{\tau^j j!}\right)^{m_1+\cdots+m_n}\mathrm{d}t\\ &\overset{\text{(i)}}{\le} n!\cdot\sum_{m_1,\dots,m_n}(C_\psi n)!\overset{\text{(ii)}}{\le} n!\cdot e^n\cdot(C_\psi n)!<+\infty,\end{aligned}$$

where in step (i) we use Assumption 2(b), $\tau\ge\widetilde{C}_\psi$, and $1\le m_1+\cdots+m_n\le n$, and in step (ii) we use the fact that there are at most e^n terms in the summation; see Ineq. (69). Note that $\left|\nabla_x^{(n)} v(y;x)\right|$ is a continuous function in x by definition (20) and Assumption 2. Using the bounded integral in the display above, we obtain

$$\lim_{x\uparrow+\infty}\left|\nabla_x^{(n)} v(y;x)\right|=\lim_{x\downarrow-\infty}\left|\nabla_x^{(n)} v(y;x)\right|=0.$$

This verifies Assumption 1(b). We next turn to prove Assumption 1(c). By definition (20),

$$v\big(y;\sqrt{2}\sigma x+\theta\big)=\frac{1}{\tau}\exp\bigg(-\psi\Big(\frac{y-\sqrt{2}\sigma x-\theta}{\tau}\Big)\bigg).$$

Thus,

$$\begin{aligned}\int_{\mathbb{R}}\big(v\big(y;\sqrt{2}\sigma x+\theta\big)\big)^2 e^{-x^2}\mathrm{d}x &\le\int_{\mathbb{R}}\big(v\big(y;\sqrt{2}\sigma x+\theta\big)\big)^2\mathrm{d}x\\ &=\int_{\mathbb{R}}\frac{1}{\sqrt{2}\sigma\tau}\exp\big(-2\psi(t)\big)\mathrm{d}t<+\infty,\end{aligned}$$

where in the second step we use the change of variables $t=\frac{y-\sqrt{2}\sigma x-\theta}{\tau}$, and in the last step we use Assumption 2(a). This proves Assumption 1(c).

Having verified Assumption 2, we can now apply Lemma 1 to obtain

$$\delta\big(\mathcal{U},\mathcal{V};\mathcal{T}_N^\star\big)\le\varsigma\big(\mathcal{U},\mathcal{V};\mathcal{S}_N^\star\big)+\frac{1}{2}\sup_{\theta\in\mathbb{R}}\int_{\mathbb{R}}\big(|p(x)-1|+q(x)\big)\cdot u(x;\theta)\mathrm{d}x.\tag{70}$$

To bound the terms on the RHS of the above display, we use the following lemma.

Lemma 3. *Let the source and target statistical models $\mathcal{U}$ and $\mathcal{V}$ be given by Eq. (2) and Eq. (20), respectively. Let the signed kernel $\mathcal{S}_N^\star$ be defined in Eq. (14) and let $\varsigma\big(\mathcal{U},\mathcal{V};\mathcal{S}_N^\star\big)$ be defined in Eq. (10). Suppose the tuple of parameters (N,τ,λ,σ) satisfies condition (27) and Assumption 2 holds. Let $\overline{C}_\psi$ be defined in Eq. (23). Then the following two parts hold.*

(a) We have

$$\varsigma\big(\mathcal{U},\mathcal{V};\mathcal{S}_N^\star\big) \le \overline{C}_\psi \exp\bigg(-(2e)^{-1}(2N+1)^{\frac{1}{2(1+C_\psi)}}\bigg), \tag{71}$$

(b) Let $p(x)$ and $q(x)$ be defined in Eq. (17). Let $\mathsf{Err}(\psi,n,\lambda)$ be defined in Eq. (25). Then

$$\frac{1}{2}\sup_{\theta\in\mathbb{R}}\int_{\mathbb{R}}\big(|p(x)-1|+q(x)\big)\cdot u(x;\theta)\mathrm{d}x \le 4\mathsf{Err}(\psi,2N,\lambda) \tag{72}$$

Indeed, substituting Ineq. (71) and Ineq. (72) into the RHS of Ineq. (70) proves Theorem 1(a). □

It remains to prove Lemma 3; we do so in the following subsections.

6.1.1 Proof of Lemma 3(a)

By Lemma 1 (Ineq. (18)), we have

$$\varsigma\big(\mathcal{U},\mathcal{V};\mathcal{S}_N^\star\big) \le \frac{1}{2}\sup_{\theta\in\mathbb{R}}\int_{\mathbb{R}}\sum_{k=N+1}^{+\infty}\bigg|\int_{\mathbb{R}} v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot\overline{H}_{2k}(x)e^{-x^2}\mathrm{d}x\bigg|\mathrm{d}y. \tag{73}$$

For each $n\in\mathbb{N}$, let $a_n=\sqrt{2(n+1)}$ for convenience. En route to bounding the RHS of the inequality in the display above, let us apply integration by parts and use Eq. (62b). We obtain that for all $n\in\mathbb{N}$,

$$\begin{aligned}
&\int_{\mathbb{R}} v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot\overline{H}_n(x)e^{-x^2}\mathrm{d}x\\
&=\frac{1}{a_n}\int_{\mathbb{R}} v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot e^{-x^2}\mathrm{d}\overline{H}_{n+1}(x)\\
&=\frac{1}{a_n}\Big[v\big(y;\theta+\sqrt{2}\sigma x\big)e^{-x^2}\overline{H}_{n+1}(x)\Big]\Big|_{-\infty}^{+\infty}-\frac{1}{a_n}\int_{\mathbb{R}}\frac{\mathrm{d}\big[v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot e^{-x^2}\big]}{\mathrm{d}x}\overline{H}_{n+1}(x)\mathrm{d}x.
\end{aligned}$$

Performing the same integration by parts r times yields

$$\begin{aligned}
\int_{\mathbb{R}} v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot\overline{H}_n(x)e^{-x^2}\mathrm{d}x &= \sum_{\ell=0}^{r-1}\frac{(-1)^\ell}{\prod_{j=0}^{\ell}a_{n+j}}\bigg[\frac{\mathrm{d}^\ell\big[v\big(y;\theta+\sqrt{2}\sigma x\big)e^{-x^2}\big]}{\mathrm{d}x^\ell}\overline{H}_{n+1+\ell}(x)\bigg]\bigg|_{-\infty}^{+\infty}\\
&\quad+\frac{(-1)^r}{\prod_{j=0}^{r-1}a_{n+j}}\int_{\mathbb{R}}\frac{\mathrm{d}^r\big[v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot e^{-x^2}\big]}{\mathrm{d}x^r}\overline{H}_{n+r}(x)\mathrm{d}x.
\end{aligned} \tag{74}$$

We claim the following equation holds for all $n\in\mathbb{N}$, $\ell\in\mathbb{Z}_{\ge 0}$, $y\in\mathbb{R}$, and $\tau\ge\sqrt{2}\sigma\widetilde{C}_\psi,\sigma>0$:

$$\bigg[\frac{\mathrm{d}^\ell\big[v\big(y;\theta+\sqrt{2}\sigma x\big)e^{-x^2}\big]}{\mathrm{d}x^\ell}\overline{H}_{n+1+\ell}(x)\bigg]\bigg|_{-\infty}^{+\infty}=0, \tag{75}$$

where the evaluation at infinity is with respect to variable x. We defer the proof of claim (75) to the end of this section. Applying the product rule for higher-order derivative yields for all $r \in \mathbb{Z}_{\geq 0}$,

$$\begin{aligned}\frac{\mathrm{d}^r\left[v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot e^{-x^2}\right]}{\mathrm{d}x^r} &= \sum_{k=0}^{r}\binom{r}{k}\nabla_x^{(r-k)}v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot\frac{\mathrm{d}^k e^{-x^2}}{\mathrm{d}x^k}\\ &= \sum_{k=0}^{r}\binom{r}{k}\nabla_x^{(r-k)}v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot(-1)^k H_k(x)e^{-x^2}. \end{aligned} \tag{76}$$

Substituting Eq. (76) and Eq. (75) into Eq. (74) and applying the triangle inequality yields

$$\begin{aligned}&\int_{\mathbb{R}}\left|\int_{\mathbb{R}} v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot\overline{H}_n(x)e^{-x^2}\mathrm{d}x\right|\mathrm{d}y\\ &\leq \frac{1}{\prod_{j=0}^{r-1}a_{n+j}}\sum_{k=0}^{r}\binom{r}{k}\int_{\mathbb{R}}\int_{\mathbb{R}}\left|\nabla_x^{(r-k)}v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot H_k(x)e^{-x^2}\cdot\overline{H}_{n+r}(x)\right|\mathrm{d}x\mathrm{d}y\\ &\leq \frac{1}{\prod_{j=0}^{r-1}a_{n+j}}\sum_{k=0}^{r}\binom{r}{k}\cdot\sqrt{\gamma_k}\int_{\mathbb{R}}\int_{\mathbb{R}}\left|\nabla_x^{(r-k)}v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot e^{-x^2/2}\right|\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{77}$$

where in the last step we let $\gamma_k = \sqrt{\pi}2^k k!$ and use $\left|H_k(x)e^{-x^2/2}\right| \leq \sqrt{\gamma_k}$ and also $\left|\overline{H}_{n+r}(x)\right| \leq 1$ for all $x \in \mathbb{R}$, by Ineq. (62a). We further claim that for all $r, k \in \mathbb{Z}_{\geq 0}$ and $k \leq r$, and for all $\tau \geq 4\sqrt{2}e\sigma\widetilde{C}_\psi$, we have

$$\int_{\mathbb{R}}\int_{\mathbb{R}}\left|\nabla_x^{(r-k)}v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot e^{-x^2/2}\right|\mathrm{d}x\mathrm{d}y \leq 4\pi(r-k)!\cdot\frac{1}{4^{r-k}}\cdot\big(C_\psi(r-k)\big)! \tag{78}$$

We defer the proof of Claim (78) to the end of this section. Substituting Ineq. (78) into Ineq. (77) yields

$$\begin{aligned}\int_{\mathbb{R}}\left|\int_{\mathbb{R}} v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot\overline{H}_n(x)e^{-x^2}\mathrm{d}x\right|\mathrm{d}y &\leq \frac{4\pi}{\prod_{j=0}^{r-1}a_{n+j}}\sum_{k=0}^{r}\binom{r}{k}\cdot\sqrt{\gamma_k}\cdot\frac{(r-k)!}{4^{r-k}}\cdot\big(C_\psi(r-k)\big)!\\ &\overset{\text{(i)}}{\leq} \frac{4\pi^{5/4}\cdot r!\cdot\big(C_\psi\cdot r\big)!\cdot 2^{r/2}}{\prod_{j=0}^{r-1}a_{n+j}}\sum_{k=0}^{r}\frac{1}{4^{r-k}}\\ &\overset{\text{(ii)}}{\leq} \frac{8\pi^{5/4}\cdot r!\cdot\big(C_\psi\cdot r\big)!\cdot 2^{r/2}}{2^{r/2}\cdot n^{r/2}} = \frac{8\pi^{5/4}\cdot r!\cdot\big(C_\psi\cdot r\big)!}{n^{r/2}}, \end{aligned} \tag{79}$$

where in the step (i) we use $\binom{r}{k} = r!/\big((r-k)!\cdot k!\big)$ and $\gamma_k = \sqrt{\pi}2^k\cdot k!$, and in step (ii) we recall that $a_{n+j} = \sqrt{2(n+j+1)} \geq \sqrt{2n}$ for all j and $\sum_{k=0}^{r}4^{-(r-k)} \leq 2$. Note that the inequality in the display above holds for all $r \in \mathbb{Z}_{\geq 0}$. We now set

$$r = \left\lfloor n^{\frac{1}{2(1+C_\psi)}}/\big(e\big(C_\psi\vee 1\big)\big)\right\rfloor \quad \text{so that} \quad \frac{\big(\big(C_\psi\vee 1\big)r\big)^{1+C_\psi}}{\sqrt{n}} \leq e^{-(1+C_\psi)}.$$

Consequently, we obtain

$$\begin{aligned}\frac{r!\cdot\big(C_\psi\cdot r\big)!}{n^{r/2}} \leq \frac{r^r\cdot\big(\big(C_\psi\vee 1\big)r\big)^{C_\psi r}}{n^{r/2}} &= \left(\frac{\big(\big(C_\psi\vee 1\big)r\big)^{1+C_\psi}}{\sqrt{n}}\right)^r\\ &\leq e^{-(1+C_\psi)r} \leq e^{1+C_\psi}\exp\Big(-e^{-1}n^{\frac{1}{2(1+C_\psi)}}\Big), \end{aligned} \tag{80}$$

where in the last step we use that by definition, we have $r \geq \frac{n^{\frac{1}{2(1+C_\psi)}}}{e(C_\psi \vee 1)} - 1$. Substituting Ineq. (80) into the RHS of Ineq. (79) yields that for all $n \in \mathbb{Z}_{\geq 0}$,

$$\int_{\mathbb{R}} \left| \int_{\mathbb{R}} v\big(y; \theta + \sqrt{2}\sigma x\big) \cdot \overline{H}_n(x) e^{-x^2} \mathrm{d}x \right| \mathrm{d}y \leq 8\pi^{5/4} e^{1+C_\psi} \exp\Big(- e^{-1} n^{\frac{1}{2(1+C_\psi)}} \Big).$$

Substituting the inequality in the display above into the RHS of Ineq. (73) yields

$$\begin{aligned}
\varsigma\big(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star\big) &\leq 4\pi^{5/4} e^{1+C_\psi} \sum_{n=2N+2}^{+\infty} \exp\Big(- e^{-1} n^{\frac{1}{2(1+C_\psi)}} \Big) \\
&\leq 4\pi^{5/4} e^{1+C_\psi} \int_{2N+1}^{+\infty} \exp\Big(- e^{-1} t^{\frac{1}{2(1+C_\psi)}} \Big) \mathrm{d}t.
\end{aligned}$$

To conclude, we must bound the integral in the display above. To reduce the notational burden, let $a = 2N+1$, $b = 2(1+C_\psi)$, and $C = e^{e^{-1}} > 1$. Note that $\frac{b a^{-1/b}}{\log(C)} \leq 1$ by assumption. We obtain

$$\int_{2N+1}^{+\infty} \exp\Big(- e^{-1} t^{\frac{1}{2(1+C_\psi)}} \Big) \mathrm{d}t = \int_a^{+\infty} C^{-t^{1/b}} \mathrm{d}t = b \int_{a^{1/b}}^{+\infty} z^{b-1} e^{-\log(C) z} \mathrm{d}z,$$

where in the last step we use the change of variables $z = t^{1/b}$. Integrating by parts then yields

$$\begin{aligned}
\int_{a^{1/b}}^{+\infty} z^{b-1} e^{-\log(C) z} \mathrm{d}z &= e^{-\log(C) a^{1/b}} \cdot \left(\frac{a^{\frac{b-1}{b}}}{\log(C)} + \frac{(b-1) a^{\frac{b-2}{b}}}{\log(C)^2} + \frac{(b-1)(b-2) a^{\frac{b-3}{b}}}{\log(C)^3} + \cdots + \frac{b!}{\log(C)^b} \right) \\
&\leq e^{-\log(C) a^{1/b}} \cdot \frac{a^{\frac{b-1}{b}} b}{\log(C)}.
\end{aligned}$$

Here the second step uses the fact that since $\frac{b a^{-1/b}}{\log(C)} \leq 1$, the series is decreasing, and thus the total sum can be bounded by the first term multiplied by the total number of terms. Putting all the pieces together, we have obtained for $2N+1 \geq (2e(1+C_\psi))^{2(1+C_\psi)}$,

$$\begin{aligned}
\varsigma\big(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star\big) &\leq 16\pi^{\frac{5}{4}} e^{2+C_\psi} (1+C_\psi)^2 (2N+1) \exp\Big(- e^{-1} (2N+1)^{\frac{1}{2(1+C_\psi)}} \Big) \\
&= 16\pi^{\frac{5}{4}} e^{2+C_\psi} (1+C_\psi)^2 \exp\Big(- (2e)^{-1} (2N+1)^{\frac{1}{2(1+C_\psi)}} \Big) \\
&\quad \times (2e)^{2(1+C_\psi)} \cdot \frac{2N+1}{(2e)^{2(1+C_\psi)}} \cdot \exp\Big(- (2e)^{-1} (2N+1)^{\frac{1}{2(1+C_\psi)}} \Big) \\
&\overset{\text{(i)}}{\leq} 16\pi^{\frac{5}{4}} e^{2+C_\psi} (1+C_\psi)^2 (2e)^{2(1+C_\psi)} \exp\Big(- (2e)^{-1} (2N+1)^{\frac{1}{2(1+C_\psi)}} \Big),
\end{aligned}$$

where in step (i) use the numerical inequality

$$\frac{t}{(2e)^{2(1+C_\psi)}} \leq \exp\Big((2e)^{-1} t^{\frac{1}{2(1+C_\psi)}} \Big) \quad \text{for all } t \geq (2e)^{2(1+C_\psi)} e^{4(1+C_\psi)^2}. \tag{81}$$

In particular, we apply this inequality for $t = 2N+1$, which is in the desired range by assumption. To conclude, note that Ineq. (71) holds by recalling the definition $\overline{C}_\psi$ in Eq. (23).

It remains to verify the claims (75), (78), and (81).

Proof of Claim (75). By definition (20), we obtain $v\big(y;\theta+\sqrt{2}\sigma x\big)=g(h(x))$, where $g(x)=\exp(-x)/\tau$ and $h(x)=\psi\big(\big(y-\theta-\sqrt{2}\sigma x\big)/\tau\big)$. Applying the Faà di Bruno's formula yields

$$\begin{aligned}\nabla_x^{(n)}v\big(y;\theta+\sqrt{2}\sigma x\big)&=\sum_{m_1,\ldots,m_n}\frac{n!}{m_1!\,m_2!\cdots m_n!}\cdot g^{(m_1+\cdots+m_n)}(h(x))\cdot\prod_{j=1}^{n}\left(\frac{h^{(j)}(x)}{j!}\right)^{m_j}\\&=\frac{\exp(-h(x))}{\tau}\cdot(n!)\cdot\sum_{m_1,\ldots,m_n}\frac{(-1)^{m_1+\cdots+m_n}}{m_1!\,m_2!\cdots m_n!}\prod_{j=1}^{n}\left(\frac{\psi^{(j)}\big(\frac{y-\theta-\sqrt{2}\sigma x}{\tau}\big)\cdot\big(-\sqrt{2}\sigma\big)^j}{\tau^j j!}\right)^{m_j},\end{aligned}\tag{82}$$

where the summation is over all n-tuples of nonnegative integers $(m_1,\cdots,m_n)$ satisfying $\sum_{j=1}^n jm_j=n$. Applying the triangle inequality and using the multinomial theorem yield

$$\begin{aligned}\big|\nabla_x^{(n)}v\big(y;\theta+\sqrt{2}\sigma x\big)\big|&\le\frac{\exp(-h(x))}{\tau}\cdot(n!)\cdot\sum_{m_1,\ldots,m_n}\left(\sum_{j=1}^{n}\frac{\big|\psi^{(j)}\big(\frac{y-\theta-\sqrt{2}\sigma x}{\tau}\big)\big|\cdot\big(\sqrt{2}\sigma\big)^j}{\tau^j j!}\right)^{m_1+\cdots+m_n}\\&\le\frac{\exp(-h(x))}{\tau}\cdot(n!)\cdot\sum_{m_1,\ldots,m_n}\left(\sum_{j=1}^{n}\frac{\big|\psi^{(j)}\big(\frac{y-\theta-\sqrt{2}\sigma x}{\tau}\big)\big|}{\big(\widetilde{C}_\psi\big)^j j!}\right)^{m_1+\cdots+m_n},\end{aligned}$$

where in the last step we use $\tau\ge\sqrt{2}\sigma\widetilde{C}_\psi$. Integrating over $x\in\mathbb{R}$ and using change of variables $t=\frac{y-\theta-\sqrt{2}\sigma x}{\tau}$ yield that for all $n\in\mathbb{Z}_{\ge0}$ and $y\in\mathbb{R}$,

$$\begin{aligned}\int_{\mathbb{R}}\big|\nabla_x^{(n)}v\big(y;\theta+\sqrt{2}\sigma x\big)\big|\mathrm{d}x&\le\frac{n!}{\sqrt{2}\sigma}\sum_{m_1,\ldots,m_n}\int_{\mathbb{R}}\exp(-\psi(t))\left(\sum_{j=1}^{n}\frac{|\psi^{(j)}(t)|}{\big(\widetilde{C}_\psi\big)^j j!}\right)^{m_1+\cdots+m_n}\mathrm{d}t\\&\le\frac{n!}{\sqrt{2}\sigma}\cdot e^n\cdot(C_\psi n)!<+\infty,\end{aligned}$$

where we have used Assumption 2 and the fact that there are at most e^n terms in the summation. Combining the inequality in the display with Eq. (76) and applying the triangle inequality yield

$$\begin{aligned}&\int_{\mathbb{R}}\left|\frac{\mathrm{d}^\ell\big[v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot e^{-x^2}\big]}{\mathrm{d}x^\ell}\cdot\overline{H}_{n+1+\ell}(x)\right|\mathrm{d}x\\&\le\sum_{k=0}^{\ell}\binom{\ell}{k}\int_{\mathbb{R}}\big|\nabla_x^{(\ell-k)}v\big(y;\theta+\sqrt{2}\sigma x\big)\big|\cdot\big|\overline{H}_{n+1+\ell}(x)\cdot H_k(x)e^{-x^2}\big|\mathrm{d}x\\&\overset{\text{(i)}}{\le}\sum_{k=0}^{\ell}\binom{\ell}{k}\big(\sqrt{\pi}2^k k!\big)^{1/2}\int_{\mathbb{R}}\big|\nabla_x^{(\ell-k)}v\big(y;\theta+\sqrt{2}\sigma x\big)\big|\mathrm{d}x<+\infty\end{aligned}$$

where in step (i) we use the boundness property of Hermite polynomial (62a). Consequently, for a positive and continuous function, if the integral over $\mathbb{R}$ is finite, then this function must converge to zero at infinity. Thus,

$$\lim_{|x|\uparrow+\infty}\frac{\mathrm{d}^\ell\big[v\big(y;\theta+\sqrt{2}\sigma x\big)\cdot e^{-x^2}\big]}{\mathrm{d}x^\ell}\cdot\overline{H}_{n+1+\ell}(x)=0.$$

This proves Claim (75).

Proof of Claim (78). Using Eq. (82) with the triangle inequality yields for all $n \in \mathbb{N}$,

$$\begin{aligned}\left|\nabla_x^{(n)} v\big(y;\theta+\sqrt{2}\sigma x\big)\right| &\le \frac{\exp(-h(x))}{\tau}\sum_{m_1,\ldots,m_n}\frac{n!}{m_1!\,m_2!\cdots m_n!}\prod_{j=1}^{n}\left(\frac{\left|\psi^{(j)}\left(\frac{y-\theta-\sqrt{2}\sigma x}{\tau}\right)\right|\cdot\left(\sqrt{2}\sigma\right)^j}{\tau^j j!}\right)^{m_j}\\ &= \left(\frac{\sqrt{2}\widetilde{C}_\psi\sigma}{\tau}\right)^n\cdot\frac{\exp(-h(x))}{\tau}\sum_{m_1,\ldots,m_n}\frac{n!}{m_1!\,m_2!\cdots m_n!}\prod_{j=1}^{n}\left(\frac{\left|\psi^{(j)}\left(\frac{y-\theta-\sqrt{2}\sigma x}{\tau}\right)\right|}{\left(\widetilde{C}_\psi\right)^j j!}\right)^{m_j},\end{aligned}$$

where in the second step we use $\sum_{j=1}^n jm_j = n$ so that

$$\left(\frac{\sqrt{2}\widetilde{C}_\psi\sigma}{\tau}\right)^n\prod_{j=1}^{n}\left(\frac{1}{\left(\widetilde{C}_\psi\right)^j}\right)^{m_j} = \left(\frac{\sqrt{2}\sigma}{\tau}\right)^n = \prod_{j=1}^{n}\left(\frac{\left(\sqrt{2}\sigma\right)^j}{\tau^j}\right)^{m_j}.$$

Continuing and applying the multinomial theorem yields

$$\left|\nabla_x^{(n)} v\big(y;\theta+\sqrt{2}\sigma x\big)\right| \le \left(\frac{\sqrt{2}\widetilde{C}_\psi\sigma}{\tau}\right)^n\cdot(n!)\cdot\frac{\exp(-h(x))}{\tau}\cdot\sum_{m_1,\ldots,m_n}\left(\sum_{j=1}^{n}\frac{\left|\psi^{(j)}\left(\frac{y-\theta-\sqrt{2}\sigma x}{\tau}\right)\right|}{\left(\widetilde{C}_\psi\right)^j j!}\right)^{m_1+\cdots+m_n}.$$

Recall that we let $h(x) = \psi\big((y-\theta-\sqrt{2}\sigma x)/\tau\big)$. Integrating the above expression over $x, y \in \mathbb{R}$ then yields

$$\begin{aligned}&\int_{\mathbb{R}}\int_{\mathbb{R}}\left|\nabla_x^{(n)} v\big(y;\theta+\sqrt{2}\sigma x\big)\right|e^{-x^2/2}\mathrm{d}x\mathrm{d}y\\ &\le \left(\frac{\sqrt{2}\widetilde{C}_\psi\sigma}{\tau}\right)^n\cdot(n!)\cdot\sum_{m_1,\ldots,m_n}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{1}{\tau}e^{-\psi\left(\frac{y-\theta-\sqrt{2}\sigma x}{\tau}\right)}\left(\sum_{j=1}^{n}\frac{\left|\psi^{(j)}\left(\frac{y-\theta-\sqrt{2}\sigma x}{\tau}\right)\right|}{\left(\widetilde{C}_\psi\right)^j j!}\right)^{m_1+\cdots+m_n}\mathrm{d}y\cdot e^{-x^2/2}\mathrm{d}x\\ &\overset{\text{(i)}}{=} \left(\frac{\sqrt{2}\widetilde{C}_\psi\sigma}{\tau}\right)^n\cdot(n!)\cdot\sum_{m_1,\ldots,m_n}\int_{\mathbb{R}}\int_{\mathbb{R}}\exp(-\psi(t))\left(\sum_{j=1}^{n}\frac{\left|\psi^{(j)}(t)\right|}{\left(\widetilde{C}_\psi\right)^j j!}\right)^{m_1+\cdots+m_n}\mathrm{d}t\cdot e^{-x^2/2}\mathrm{d}x\\ &\overset{\text{(ii)}}{\le} \left(\frac{\sqrt{2}\widetilde{C}_\psi\sigma}{\tau}\right)^n\cdot(n!)\cdot\sum_{m_1,\ldots,m_n}(C_\psi n)!\cdot\int_{\mathbb{R}}e^{-x^2/2}\mathrm{d}x\\ &\le \left(\frac{\sqrt{2}\widetilde{C}_\psi\sigma}{\tau}\right)^n\cdot(n!)\cdot e^n\cdot(C_\psi n)!\cdot\sqrt{2\pi},\end{aligned}$$

where in step (i) we first integrate over $y \in \mathbb{R}$ and use the change of variables $t = (y-\theta-\sqrt{2}\sigma x)/\tau$, in step (ii) we use Assumption 2, and in the last step we use there are at most e^n terms in the summation (see Ineq. (69)) and also that $\int_{\mathbb{R}} e^{-x^2/2}\mathrm{d}x = \sqrt{2\pi}$. Thus, by letting $\tau \ge 4\sqrt{2}e\sigma\widetilde{C}_\psi$, we obtain

$$\int_{\mathbb{R}}\int_{\mathbb{R}}\left|\nabla_x^{(n)} v\big(y;\theta+\sqrt{2}\sigma x\big)\right|e^{-x^2/2}\mathrm{d}x\mathrm{d}y \le \sqrt{2\pi}(n!)\cdot(C_\psi n)!\cdot\frac{1}{4^n} \quad \text{for all } n \in \mathbb{Z}_{\ge 0}.$$

This proves Claim (78).

Proof of Claim (81). By letting $x = (2e)^{-1}t^{\frac{1}{2(1+C_\psi)}}$, it is equivalent to show

$$x^{2(1+C_\psi)} \le e^x \quad \text{for all } x \ge e^{2(1+C_\psi)},$$

which is equivalent to, by taking the log on both sides and re-arranging the terms,

$$h(x) := 2(1+C_\psi)\frac{\log(x)}{x} \le 1 \quad \text{for all } x \ge e^{2(1+C_\psi)}.$$

To show this, simply note that the function $h(x)$ is monotone decreasing for all $x \ge 1$. Thus,

$$h(x) \le h\big(e^{2(1+C_\psi)}\big) = \frac{(2(1+C_\psi))^2}{e^{2(1+C_\psi)}} \le \sup_{a>0}\frac{a^2}{e^a} = \frac{2^2}{e^2} \le 1 \quad \text{for all } x \ge e^{2(1+C_\psi)},$$

as claimed. □

6.1.2 Proof of Lemma 3(b)

We first state an auxiliary lemma and defer its proof to Appendix B.2.

Lemma 4. *Consider the signed kernel $\mathcal{S}_N^\star(y \mid x)$ defined in Eq. (14) for the target density $v(y;\theta)$ defined in Eq. (20). Let the set $\mathcal{A}(\psi,n,\lambda)$, its complement $\mathcal{A}^\complement(\psi,n,\lambda)$, and the scalar $\mathsf{Err}(\psi,n,\lambda)$ be defined as in Eq. (24) and Eq. (25), respectively. Suppose Assumption 2 holds with the pair of constants $(C_\psi, \widetilde{C}_\psi)$. Then the following holds for all $N \in \mathbb{N}$ and $\lambda > 0$.*

(i) If $\tau \ge 8e\lambda\sigma N$, then $\mathcal{S}_N^\star(y \mid x) \ge 0$ as long as $(y-x)/\tau \in \mathcal{A}(\psi, 2N, \lambda)$.

(ii) Let $\rho(t) = \exp(-\psi(t))$ for all $t \in \mathbb{R}$. Then for all $1 \le n \le 2N$ and $x \in \mathbb{R}$, we have

$$\int_{\mathbb{R}} \rho^{(n)}(t)\mathrm{d}t = 0 \quad \text{and} \quad \int_{\frac{y-x}{\tau}\in\mathcal{A}^\complement(\psi,2N,\lambda)} \big|\nabla_x^{(n)} v(y;x)\big|\mathrm{d}y \le \frac{n!\cdot\lambda^n\cdot e^n}{\tau^n}\cdot\mathsf{Err}(\psi,2N,\lambda). \tag{83}$$

We now use the lemma above to prove inequality (72). Note that $v(y;x) = \frac{1}{\tau}\rho(\frac{y-x}{\tau})$. Consequently, we obtain for all $n \in \mathbb{Z}_{\ge 0}$ and $x, y \in \mathbb{R}$ that

$$\nabla_x^{(n)} v(y;x) = \frac{1}{\tau}\rho^{(n)}\left(\frac{y-x}{\tau}\right)\left(\frac{-1}{\tau}\right)^n. \tag{84}$$

Let us now upper bound $p(x)$ (17). Using part (i) of Lemma 4 and definition (17), we obtain

$$\begin{aligned} p(x) &\le \int_{\frac{y-x}{\tau}\in\mathcal{A}(\psi,2N,\lambda)} \mathcal{S}_N^\star(y \mid x)\mathrm{d}y + \int_{\frac{y-x}{\tau}\in\mathcal{A}^\complement(\psi,2N,\lambda)} \big|\mathcal{S}_N^\star(y \mid x)\big|\mathrm{d}y \\ &\le \int_{\mathbb{R}} \mathcal{S}_N^\star(y \mid x)\mathrm{d}y + 2\int_{\frac{y-x}{\tau}\in\mathcal{A}^\complement(\psi,2N,\lambda)} \big|\mathcal{S}_N^\star(y \mid x)\big|\mathrm{d}y. \end{aligned} \tag{85}$$

Next, by definition, we have

$$\int_{\mathbb{R}} \mathcal{S}_N^\star(y \mid x)\mathrm{d}y = \sum_{k=0}^{N}\frac{(-1)^k\sigma^{2k}}{(2k)!!}\int_{\mathbb{R}}\nabla_x^{(2k)}v(y;x)\mathrm{d}y \overset{\text{(i)}}{=} \sum_{k=0}^{N}\frac{(-1)^k\sigma^{2k}}{(2k)!!}\int_{\mathbb{R}}\frac{\rho^{(2k)}(t)}{\tau^{2k}}\mathrm{d}t = 1, \tag{86}$$

where in step (i) we use Eq. (84) and use the change of variables $t = (y-x)/\tau$, and in the last step we use Ineq. (83) and note that $\int_{\mathbb{R}}\rho(t)\mathrm{d}t = 1$. Towards bounding the second term on the RHS of Ineq. (85), we obtain by using the triangle inequality that

$$\begin{aligned} \int_{\frac{y-x}{\tau}\in\mathcal{A}^\complement(\psi,2N,\lambda)} \big|\mathcal{S}_N^\star(y \mid x)\big|\mathrm{d}y &\le \sum_{k=0}^{N}\frac{\sigma^{2k}}{(2k)!!}\int_{\frac{y-x}{\tau}\in\mathcal{A}^\complement(\psi,2N,\lambda)}\big|\nabla_x^{(2k)}v(y;x)\big|\mathrm{d}y \\ &\overset{\text{(i)}}{\le} \int_{\frac{y-x}{\tau}\in\mathcal{A}^\complement(\psi,2N,\lambda)} v(y;x)\mathrm{d}y + \sum_{k=1}^{N}\frac{\sigma^{2k}}{(2k)!!}\frac{(2k)!\cdot\lambda^{2k}\cdot e^{2k}}{\tau^{2k}}\cdot\mathsf{Err}(\psi,2N,\lambda) \\ &\overset{\text{(ii)}}{\le} 2\mathsf{Err}(\psi,2N,\lambda), \end{aligned} \tag{87}$$

where in step (i) we use Ineq. (83). In step (ii) we use

$$\int_{\frac{y-x}{\tau}\in\mathcal{A}^{\complement}(\psi,2N,\lambda)} v(y;x)\mathrm{d}y = \int_{\frac{y-x}{\tau}\in\mathcal{A}^{\complement}(\psi,2N,\lambda)} \frac{\rho(\frac{y-x}{\tau})}{\tau}\mathrm{d}y = \int_{\mathcal{A}^{\complement}(\psi,2N,\lambda)} e^{-\psi(t)}\mathrm{d}t \le \mathsf{Err}(\psi,2N,\lambda),$$

and also use $\tau \ge 8e\lambda\sigma N$ so that we have the chain of bounds

$$\sum_{k=1}^{N} \frac{\sigma^{2k}\cdot(2k)!\cdot(e\lambda)^{2k}}{(2k)!!\cdot\tau^{2k}} \le \sum_{k=1}^{N}\left(\frac{2e\lambda\sigma k}{\tau}\right)^{2k} \le \sum_{k=1}^{N}\frac{1}{4^k} \le 1.$$

Substituting Eq. (86) and Ineq. (87) into the RHS of Ineq. (85) yields

$$p(x) \le 1 + 4\mathsf{Err}(\psi,2N,\lambda) \quad \text{for all } x\in\mathbb{R}.$$

We now bound $q(x)$ (17). We obtain by definition that

$$p(x) - q(x) = \int_{\mathbb{R}}\big[\mathcal{S}_N^{\star}(y\mid x)\vee 0\big]\mathrm{d}y + \int_{\mathbb{R}}\big[\mathcal{S}_N^{\star}(y\mid x)\wedge 0\big]\mathrm{d}y = \int_{\mathbb{R}}\mathcal{S}_N^{\star}(y\mid x)\mathrm{d}y = 1,$$

where in the last step we use Eq. (86). Note that $q(x)\ge 0$ by its definition (17) and thus,

$$p(x) = q(x) + 1 \ge 1 \quad \text{for all} \quad x\in\mathbb{R}. \tag{88}$$

Putting the pieces together yields

$$\big|p(x)-1\big| + q(x) \le 2\big|p(x)-1\big| \le 8\mathsf{Err}(\psi,2N,\lambda) \quad \text{for all } x\in\mathbb{R}.$$

Note that the RHS of the inequality in the display is independent of x. Integrating it over the density $u(x;\theta)$ and taking a supremum over θ, we complete the proof of Ineq. (72). □

6.2 Proof of Theorem 1(b)

We split the proof into several parts.

Constructing the reduction based on a general rejection kernel. We define a class of base measures $\{\mathcal{D}(\cdot\mid x)\}_{x\in\mathbb{R}}$ as

$$\mathcal{D}(y\mid x) = \frac{1}{2\tau}\exp\left(-\psi\left(\frac{y-x}{2\tau}\right)\right), \quad \text{for all } x,y\in\mathbb{R}. \tag{89}$$

We recall the general rejection kernel $x\mapsto \textsc{rk}(x,T,M,y_0)$ from Lou et al. (2025, page 10), and employ it using the base measures $\mathcal{D}(y\mid x)$ defined in (89) and the signed kernel $\mathcal{S}_N^{\star}(y\mid x)$ defined in (14). The rejection kernel RK requires as input a source sample x, the total number of iterations T, and a scalar $M>0$ that is required to satisfy, for input x, the bound

$$\frac{\mathcal{S}_N^{\star}(y\mid x)\vee 0}{\mathcal{D}(y\mid x)} \le M, \quad \text{for all } y\in\mathbb{R}.$$

An output sample point y_0 is also required, and serves as an initialization for the algorithm.

Staying faithful to the theorem statement, we consider T_{iter} to be any nonnegative integer, set $y_0 = 0$, and recall $M(\psi,2N,\lambda)$ in Eq. (26). We will now construct our reduction algorithm

from the rejection kernel and show that it satisfies the conditions required for RK to succeed. For $X_\theta \sim \mathcal{N}(\theta, \sigma^2)$, define

$$\mathsf{K}(X_\theta) := \textsc{rk}\big(X_\theta, T_{\mathsf{iter}}, 4M(\psi, 2N, \lambda), y_0\big). \tag{90}$$

Note that for all $x, y \in \mathbb{R}$, we have by definition of $\mathcal{S}_N^\star(y \mid x)$ (14),

$$\begin{aligned}
\frac{\mathcal{S}_N^\star(y \mid x) \vee 0}{\mathcal{D}(y \mid x)} &\le \sum_{k=0}^{N} \frac{\sigma^{2k}}{(2k)!!} \frac{\big|\nabla_x^{(2k)} v(y; x)\big|}{\mathcal{D}(y \mid x)} \\
&\overset{\text{(i)}}{\le} 2e^{-\psi(\frac{y-x}{\tau})+\psi(\frac{y-x}{2\tau})} \left(1 + \sum_{k=1}^{N} \frac{\sigma^{2k}(2k)!\lambda^{2k}}{(2k)!!\tau^{2k}} \sum_{m_1,\ldots,m_{2k}} \left(\sum_{j=1}^{2k} \frac{|\psi^{(j)}(\frac{y-x}{\tau})|}{\lambda^j j!}\right)^{m_1+\cdots+m_{2k}}\right) \\
&\overset{\text{(ii)}}{\le} 2M(\psi, 2N, \lambda) + 2\sum_{k=1}^{N} \frac{\sigma^{2k}(2k)^{2k}\lambda^{2k}}{\tau^{2k}} \sum_{m_1,\ldots,m_{2k}} M(\psi, 2N, \lambda) \le 4M(\psi, 2N, \lambda),
\end{aligned}$$

where in step (i) we use Ineq. (142) and Eq. (89), in step (ii) we use definition (26) so that

$$\begin{aligned}
&e^{-\psi(\frac{y-x}{\tau})+\psi(\frac{y-x}{2\tau})} \le M(\psi, 2N, \lambda), \quad \text{and} \\
&e^{-\psi(\frac{y-x}{\tau})+\psi(\frac{y-x}{2\tau})} \left(\sum_{j=1}^{2k} \frac{|\psi^{(j)}(\frac{y-x}{\tau})|}{\lambda^j j!}\right)^{m_1+\cdots+m_{2k}} \le M(\psi, 2k, \lambda) \le M(\psi, 2N, \lambda).
\end{aligned}$$

In the last step, we use the fact that $\tau \ge 8e\sigma\lambda N$, and also that there are at most e^{2k} many $2k$-tuples of nonnegative integers $(m_1, \cdots, m_{2k})$ satisfying $\sum_{j=1}^{2k} j m_j = 2k$. Therefore, the reduction $\mathsf{K}(X_\theta)$ in Eq. (90) satisfies the conditions required by RK.

Computational complexity of the reduction. Note that from Lou et al. (2025, page 10), the reduction procedure defined in (90) requires, at each iteration, generating a sample form $\mathsf{Unif}([0,1])$ and a sample from base measures (89), which takes time $\mathcal{O}(T_{\mathsf{samp}})$ by Assumption 2(c). It also requires evaluating $\mathcal{S}_N^\star$ once per iteration, which takes time $\mathcal{O}(T_{\mathsf{eval}}(N))$ by Assumption 1. Moreover, the total number of iterations is T_{iter}. Consequently, the reduction K (90) has computational complexity

$$\mathcal{O}\Big(T_{\mathsf{iter}} \cdot T_{\mathsf{samp}} \cdot T_{\mathsf{eval}}(N)\Big).$$

Proof of TV deficiency guarantee of the reduction. We now turn to prove Eq. (30). Applying the triangle inequality as in Ineq. (125) yields

$$\sup_{\theta\in\mathbb{R}} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \le \frac{1}{2} \sup_{\theta\in\mathbb{R}} \left\| \mathsf{K}(X_\theta) - \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) \cdot u(x; \theta) \mathrm{d}x \right\|_1 + \delta\big(\mathcal{U}, \mathcal{V}; \mathcal{T}_N^\star\big). \tag{91}$$

We next turn to bound the two terms in the RHS of the inequality in the display. Applying Ineq. (28) and using $2N + 1 \ge \big(2e \log\big(3\overline{C}_\psi/\epsilon\big)\big)^{2(1+C_\psi)}$ yield

$$\delta\big(\mathcal{U}, \mathcal{V}; \mathcal{T}_N^\star\big) \le \frac{\epsilon}{3} + 4\mathsf{Err}(\psi, 2N, \lambda). \tag{92}$$

Towards bounding the first term of the RHS of Ineq. (91), we let $Y = \textsc{rk}\big(x, T_{\mathsf{iter}}, 4M(\psi, 2N, \lambda), y_0\big)$ for each $x \in \mathbb{R}$. Then from Lou et al. (2025, Lemma 2), we obtain that the random variable Y is drawn from the mixture distribution

$$f_Y(y \mid x) = \frac{\mathcal{S}_N^\star(y \mid x) \vee 0}{p(x)} \cdot \big(1 - \ell(x)\big) + \delta_{y_0}(y) \cdot \ell(x), \quad \text{for all } y \in \mathbb{R}, \tag{93}$$

where $p(x)$ is defined in Eq. (17), $\delta_{y_0}(y)$ is the Dirac delta function centered at y_0, and

$$\ell(x) = \left(1 - \frac{p(x)}{4M(\psi, 2N, \lambda)}\right)^{T_{\mathsf{iter}}}. \tag{94}$$

By the law of total probability and definition of $\mathsf{K}(X_\theta)$ in Eq. (90), we obtain that the marginal distribution of $\mathsf{K}(X_\theta)$ is given by

$$\begin{aligned} f_{\mathsf{K}(X_\theta)}(y) &= \int_{\mathbb{R}} f_Y(y \mid x) \cdot u(x;\theta)\mathrm{d}x \\ &= \int_{\mathbb{R}} \Big(\mathcal{T}_N^\star(y \mid x)\big(1 - \ell(x)\big) + \delta_{y_0}(y) \cdot \ell(x)\Big) \cdot u(x;\theta)\mathrm{d}x, \end{aligned} \tag{95}$$

where we use Eq. (93) and also note by definition (15), $\mathcal{T}_N^\star(y \mid x) = \frac{\mathcal{S}_N^\star(y|x) \vee 0}{p(x)}$. Continuing, we obtain

$$\begin{aligned} \left\| \mathsf{K}(X_\theta) - \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) \cdot u(x;\theta)\mathrm{d}x \right\|_1 &= \int_{\mathbb{R}} \left| f_{\mathsf{K}(X_\theta)}(y) - \int_{\mathbb{R}} \mathcal{T}_N^\star(y \mid x) \cdot u(x;\theta)\mathrm{d}x \right| \mathrm{d}y \\ &\overset{\text{(i)}}{=} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} \Big(- \ell(x) \cdot \mathcal{T}_N^\star(y \mid x) + \delta_{y_0}(y) \cdot \ell(x)\Big) u(x;\theta)\mathrm{d}x \right| \mathrm{d}y \\ &\overset{\text{(ii)}}{\leq} 2 \int_{\mathbb{R}} |\ell(x)| u(x;\theta)\mathrm{d}x \\ &\overset{\text{(iii)}}{\leq} 2 \exp\left(- \frac{T_{\mathsf{iter}}}{4M(\psi, 2N, \lambda)}\right), \end{aligned} \tag{96}$$

where in step (i) we use Eq. (95), and in step (ii) we apply triangle inequality and use

$$\int_{\mathbb{R}} \mathcal{T}_N^\star(y \mid x)\mathrm{d}y = 1 \quad \text{and} \quad \int_{\mathbb{R}} \delta_{y_0}(y)\mathrm{d}y = 1.$$

In step (iii), we use Ineq. (88): $p(x) \geq 1$ for all $x \in \mathbb{R}$, so that from Eq. (94),

$$\ell(x) \leq \exp\left(- \frac{T_{\mathsf{iter}} \cdot p(x)}{4M(\psi, 2N, \lambda)}\right) \leq \exp\left(- \frac{T_{\mathsf{iter}}}{4M(\psi, 2N, \lambda)}\right)$$

Substituting Ineq. (96) and Ineq. (92) into the RHS of Ineq. (91) completes the proof of guarantee (30). □

6.3 Proof of Corollary 1

Note that since $\psi \in \mathcal{F}(R)$ (2), we immediately obtain that Assumption 2(b) holds with $\widetilde{C}_\psi = R$ and $C_\psi = 0$. Similarly, we obtain $\mathcal{A}(\psi, n, R) = \mathbb{R}$ and $\mathcal{A}(\psi, n, R) = \emptyset$ for all $n \in \mathbb{N}$. Consequently, we have

$$\mathsf{Err}(\psi, n, R) = 0 \quad \text{and} \quad M(\psi, n, R) \leq 2 \sup_{t \in \mathbb{R}} \left\{ e^{-\psi(t) + \psi(t/2)} \right\} = \mathcal{O}(1) \quad \text{for all } n \in \mathbb{N}.$$

Applying Ineq. (30), by setting $T_{\mathsf{iter}} = \mathcal{O}\big(\log(3/\epsilon)\big)$, we obtain $\sup_{\theta \in \Theta} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \leq \epsilon$. Note that eventually we take

$$\lambda = R, \quad N = \mathcal{O}\big(\log^2(C/\epsilon)\big), \quad \text{and} \quad \tau = \mathcal{O}\big(\sigma R \log^2(C/\epsilon)\big),$$

so that conditions (27) and (29) are satisfied.

Finally, we bound the computational complexity of the reduction. Note that we can use Eq. (68) to evaluate the signed kernel $\mathcal{S}_N^\star$ (14). There are at most $\mathcal{O}\big(e^{\pi\sqrt{n}}\big)$ many terms in the summation of Eq. (68) and each term in the summation can be computed in time $\mathcal{O}\big(e^{C\sqrt{n}}\big)$. Also, evaluating $\mathcal{S}_N^\star$ (14) only requires computing $\nabla_x^{(n)} v(y;x)$ for $1 \le n \le \mathcal{O}\big(\log^2(C/\epsilon)\big)$. Thus, we obtain

$$T_{\mathsf{eval}}(N) = \mathcal{O}\Big(e^{C'\sqrt{N}}\Big) = \mathsf{poly}(1/\epsilon).$$

Combining with $T_{\mathsf{iter}} = \mathcal{O}\big(\log(3/\epsilon)\big)$ yields that the overall computational complexity is of the order $T_{\mathsf{samp}} \cdot \mathsf{poly}(1/\epsilon)$. □

6.4 Proof of claims in Example 1

We provide proofs for the different families of distributions separately.

Student's t-distribution. Note that $\psi(x) = \frac{\nu+1}{2} h(g(x))$ for $h(x) = \log(x)$ and $g(x) = 1 + x^2/\nu$. Applying Faà di Bruno's formula yields

$$\begin{aligned}
\psi^{(n)}(x) &= \frac{\nu+1}{2} \sum_{m_1+2m_2=n} \frac{n!}{m_1! m_2!} h^{(m_1+m_2)}\big(g(x)\big) \prod_{j=1}^{2} \left(\frac{g^{(j)}(x)}{j!}\right)^{m_j} \\
&= \frac{\nu+1}{2} \cdot n! \cdot \sum_{m_1+2m_2=n} \frac{(-1)^{m_1+m_2-1}(m_1+m_2-1)!}{m_1! m_2!} \frac{(2x/\nu)^{m_1} (1/\nu)^{m_2}}{(1+x^2/\nu)^{m_1+m_2-1}},
\end{aligned}$$

where the summation is over all nonnegative integers m_1 and m_2 such that $m_1 + 2m_2 = n$. Using the above closed form equation, we can evaluate $\psi^{(n)}(x)$ in time $\mathcal{O}\big(\mathsf{poly}(n)\big)$. Applying the triangle inequality yields

$$\begin{aligned}
\big|\psi^{(n)}(x)\big| &\le \frac{\nu+1}{2} \cdot n! \cdot \sum_{m_1+2m_2=n} \frac{(m_1+m_2)!}{m_1! m_2!} \frac{(2|x|/\nu)^{m_1} (1/\nu)^{m_2}}{(1+x^2/\nu)^{m_1+m_2}} \\
&\overset{\text{(i)}}{\le} \frac{\nu+1}{2} \cdot n! \cdot \sum_{m_1+2m_2=n} \left(\frac{\frac{2|x|+1}{\nu}}{1+x^2/\nu}\right)^{m_1+m_2} \\
&\le \frac{\nu+1}{2} \cdot n! \cdot (3e)^n,
\end{aligned}$$

where in step (i) we use the multinomial theorem. In step (ii) we use

$$\frac{\frac{2|x|+1}{\nu}}{1+x^2/\nu} \le \frac{x^2+2}{\nu+x^2} \le 1 + \frac{2}{\nu} \le 3,$$

the fact that $m_1 + m_2 \le n$ and that there are at most e^n terms in the summation. Consequently, by letting $R = 6e(\nu+1)$, we obtain

$$\sum_{j=1}^{n} \frac{|\psi^{(j)}(x)|}{R^j j!} \le \sum_{j=1}^{n} \frac{1}{2^j} \le 1 \quad \text{for all} \quad x \in \mathbb{R}, n \in \mathbb{N}.$$

Thus, $\psi \in \mathcal{F}(6e(\nu+1))$. Note that for all $t \in \mathbb{R}$, we have

$$\psi(t) - \psi(t/2) = \frac{\nu+1}{2} \log\left(\frac{1+t^2/v}{1+t^2/4v}\right) \ge \frac{1}{2}\log(5/4).$$

Consequently, $\sup_{t\in\mathbb{R}} \big\{e^{-\psi(t)+\psi(t/2)}\big\} \le \sqrt{4/5} \le 1$. This completes the proof for Student's t-distribution. ♣

Hyperbolic secant distribution. By letting $h(t) = \log(t)$ and $g(t) = e^{\frac{\pi t}{2}} + e^{-\frac{\pi t}{2}}$, we have $\psi(t) = h(g(t))$. Now, applying Faà di Bruno's formula yields

$$\begin{aligned}
\left|\psi^{(n)}(t)\right| &\le \sum_{m_1,\ldots,m_n} \frac{n!}{(m_1!)\cdots(m_n!)} \cdot \left|h^{(m_1+\cdots+m_n)}(g(t))\right| \cdot \prod_{j=1}^{n} \left(\frac{\left|g^{(j)}(t)\right|}{j!}\right)^{m_j} \\
&\le n! \sum_{m_1,\ldots,m_n} \frac{(m_1+\cdots+m_n)!}{(m_1!)\cdots(m_n!)} \frac{1}{(g(t))^{m_1+\cdots+m_n}} \prod_{j=1}^{n} \left(\frac{(\frac{\pi}{2})^j g(t)}{j!}\right)^{m_j} \\
&\le n! \cdot (\pi/2)^n \sum_{m_1,\ldots,m_n} \frac{(m_1+\cdots+m_n)!}{(m_1!)\cdots(m_n!)} \prod_{j=1}^{n} \left(\frac{1}{j!}\right)^{m_j} \\
&\le n! \cdot (\pi/2)^n \sum_{m_1,\ldots,m_n} \left(\sum_{j=1}^{n} \frac{1}{j!}\right)^{m_1+\cdots+m_n} \le n! \cdot (\pi/2)^n \cdot (2e)^n.
\end{aligned}$$

Consequently, for $R = 2\pi e$ and uniformly for all $n \in \mathbb{N}$ and $t \in \mathbb{R}$, we have

$$\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{(R)^j j!} \le \sum_{j=1}^{n} \frac{1}{2^j} \le 1.$$

Thus, $\psi \in \mathcal{F}(2\pi e)$. Finally, note that for all $t \in \mathbb{R}$, we have

$$\psi(t) - \psi(t/2) = \log\left(\frac{e^{\frac{\pi t}{2}} + e^{-\frac{\pi t}{2}}}{e^{\frac{\pi t}{4}} + e^{-\frac{\pi t}{4}}}\right) \ge \log(1/2).$$

Consequently, we obtain $\sup_{t\in\mathbb{R}} \left\{e^{-\psi(t)+\psi(t/2)}\right\} \le 2$. One can use Faà di Bruno's formula to evaluate $\psi^{(n)}(t)$, since there are at must $\mathcal{O}(e^{\pi\sqrt{n}})$ terms in the summation and each term can be computed in time $\mathcal{O}(\mathsf{poly}(n))$. Thus, the total computational time is $\mathcal{O}(e^{5\sqrt{n}})$. ♣

The proof for the logistic distribution is identical so we omit it.

6.5 Proof of Corollary 2

We must verify the assumption, bound the error quantities, and prove the consequence; we do so separately.

Verifying Assumption 2. By definition, we have

$$\psi^{(j)}(t) = \begin{cases} \beta! \cdot t^{\beta-j}/(\beta-j)! & \text{if } 1 \le j \le \beta \\ 0 & \text{if } j \ge \beta+1. \end{cases}$$

We thus obtain for all $n \in \mathbb{N}$ and $k \in [n]$ that

$$\left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{(\widetilde{C}_\psi)^j j!}\right)^k \le \left(\sum_{j=1}^{\beta} \frac{\binom{\beta}{j}|t|^{\beta-j}}{(\widetilde{C}_\psi)^j}\right)^k \le \beta^{k-1} \sum_{j=1}^{\beta} \left(\frac{\beta}{\widetilde{C}_\psi}\right)^{jk} |t|^{(\beta-j)k} \le \frac{1}{2\beta} \sum_{j=1}^{\beta} |t|^{(\beta-j)k},$$

where in the last step we set $\widetilde{C}_\psi = 2\beta^2$. Integrating both sides over $t \in \mathbb{R}$ yields

$$\int_{\mathbb{R}} e^{-\psi(t)} \cdot \left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{(\widetilde{C}_\psi)^j j!}\right)^k \mathrm{d}t \le \frac{1}{2\beta} \sum_{j=1}^{\beta} \int_{\mathbb{R}} e^{-\psi(t)} |t|^{(\beta-j)k} \mathrm{d}t \overset{\text{(i)}}{\le} k! \le n!.$$

In step (i) we use that for any $j \in [\beta]$,

$$\begin{aligned}\int_{\mathbb{R}} e^{-\psi(t)}|t|^{(\beta-j)k}\mathrm{d}t &= \frac{\beta}{\Gamma(1/\beta)}\int_0^{+\infty} e^{-t^\beta}t^{(\beta-j)k}\mathrm{d}t \\ &\le \frac{\beta}{\Gamma(1/\beta)}\int_0^{+\infty} e^{-t^\beta}\mathrm{d}t + \frac{\beta}{\Gamma(1/\beta)}\int_0^{+\infty} e^{-t^\beta}t^{\beta k}\mathrm{d}t \\ &= \frac{1}{\Gamma(1/\beta)}\int_0^{+\infty} e^{-x}x^{1/\beta-1}\mathrm{d}x + \frac{1}{\Gamma(1/\beta)}\int_0^{+\infty} e^{-x}x^{k+1/\beta-1}\mathrm{d}x \\ &= 1 + \frac{\Gamma(k+1/\beta)}{\Gamma(1/\beta)} = 1 + k! \le 2\cdot(k!)\end{aligned}$$

Thus, Assumption 2 holds with $\widetilde{C}_\psi = 2\beta^2$ and $C_\psi = 1$.

Bounding $\mathsf{Err}(\psi,n,\lambda)$ and $M(\psi,n,\lambda)$. Note that for $|t| \le \lambda^{\frac{1}{2\beta}}$ and $\lambda \ge 4\beta^2$, we have

$$\sum_{j=1}^{n}\frac{|\psi^{(j)}(t)|}{\lambda^j j!} = \sum_{j=1}^{\beta}\frac{\binom{\beta}{j}|t|^{\beta-j}}{\lambda^j} \le \sum_{j=1}^{\beta}\frac{\beta^j\lambda^{1/2}}{\lambda^j} \le \sum_{j=1}^{\beta}\frac{1}{2^j} \le 1.$$

Consequently, we obtain that for $\lambda \ge 4\beta^2$ and for all $n \in \mathbb{N}$,

$$\left\{t \in \mathbb{R} : |t| \le \lambda^{\frac{1}{2\beta}}\right\} \subseteq \mathcal{A}(\psi,n,\lambda) \quad \text{and} \quad \mathcal{A}^{\complement}(\psi,n,\lambda) \subseteq \left\{t \in \mathbb{R} : |t| \ge \lambda^{\frac{1}{2\beta}}\right\}.$$

We now turn to bound $\mathsf{Err}(\psi,n,\lambda)$ (25). Using the relation in the display above, we obtain

$$\begin{aligned}\int_{\mathcal{A}^{\complement}(\psi,n,\lambda)} e^{-\psi(t)}\mathrm{d}t &\le \frac{\beta}{\Gamma(1/\beta)}\int_{\lambda^{\frac{1}{2\beta}}}^{+\infty} e^{-t^\beta}\mathrm{d}t \\ &= \frac{\beta}{\Gamma(1/\beta)}\int_0^{+\infty} e^{-\left(t+\lambda^{\frac{1}{2\beta}}\right)^\beta}\mathrm{d}t \le \frac{\beta}{\Gamma(1/\beta)}\int_0^{+\infty} e^{-t^\beta}\mathrm{d}t\cdot e^{-\sqrt{\lambda}} = e^{-\sqrt{\lambda}}. \qquad (97)\end{aligned}$$

Continuing, we have that for all $n \in \mathbb{N}$ and $k \in [n]$,

$$\begin{aligned}\int_{\mathcal{A}^{\complement}(\psi,n,\lambda)} e^{-\psi(t)}\left(\sum_{j=1}^{\beta}\frac{|\psi^{(j)}(t)|}{\lambda^j j!}\right)^k\mathrm{d}t &\le \int_{\mathcal{A}^{\complement}(\psi,n,\lambda)} e^{-\psi(t)}\left(\sum_{j=1}^{\beta}\frac{\beta^j|t|^{\beta-j}}{\lambda^j}\right)^k\mathrm{d}t \\ &\le \frac{\beta}{\Gamma(1/\beta)}\int_{\lambda^{\frac{1}{2\beta}}}^{+\infty} e^{-t^\beta}\left(\sum_{j=1}^{\beta}\frac{\beta^j|t|^{\beta-j}}{\lambda^j}\right)^k\mathrm{d}t \\ &\overset{\text{(i)}}{\le} \frac{\beta}{\Gamma(1/\beta)}\int_{\lambda^{\frac{1}{2\beta}}}^{+\infty} e^{-t^\beta}\frac{t^{\beta k}}{\lambda^{k/2}}\mathrm{d}t \\ &\overset{\text{(ii)}}{\le} e^{-\sqrt{\lambda}}\int_0^{+\infty}\frac{\beta}{\Gamma(1/\beta)}e^{-\frac{t^\beta}{2}}\mathrm{d}t \le 2e^{-\sqrt{\lambda}}. \qquad (98)\end{aligned}$$

Step (i) above holds for all $\lambda \ge 4\beta^2$. In this case, for $|t| \ge \lambda^{\frac{1}{2\beta}} \ge 1$, we have $\sum_{j=1}^{\beta}\frac{\beta^j|t|^{-j}}{\lambda^{j/2}} \le \sum_{j=1}^{\beta}\frac{1}{2^j} \le 1$. Step (ii) above holds for all $\lambda \ge 4k^2$; in this case

$$\begin{aligned}\int_{\lambda^{\frac{1}{2\beta}}}^{+\infty} e^{-t^\beta}t^{\beta k}\mathrm{d}t &\le \sup_{t \ge \lambda^{\frac{1}{2\beta}}}\left\{t^{\beta k}e^{-t^\beta/2}\right\}\cdot\int_{\lambda^{\frac{1}{2\beta}}}^{+\infty} e^{-t^\beta/2}\mathrm{d}t \\ &\le \lambda^{k/2}e^{-\sqrt{\lambda}/2}\cdot\int_0^{+\infty} e^{-t^\beta/2}\mathrm{d}t\cdot e^{-\sqrt{\lambda}/2} = \lambda^{k/2}e^{-\sqrt{\lambda}}\cdot\int_0^{+\infty} e^{-t^\beta/2}\mathrm{d}t.\end{aligned}$$

Combining Ineq. (97) and Ineq. (98) yields that

$$\mathsf{Err}(\psi, n, \lambda) \le 3e^{-\sqrt{\lambda}} \quad \text{for all } \lambda \ge 4n^2 \vee 4\beta^2 \text{ and } n \in \mathbb{N}.$$

We next bound $M(\psi, n, \lambda)$ (26). Since $-\psi(t) + \psi(t/2) \le -t^\beta/2$, we obtain

$$\begin{aligned}\sup_{t\in\mathbb{R}} \left\{ e^{-\psi(t)+\psi(t/2)} \left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{\lambda^j j!} \right)^k \right\} &\le \beta^k \sum_{j=1}^{\beta} \frac{\beta^{jk}}{\lambda^{jk}} \cdot \sup_{t\in\mathbb{R}} \left\{ e^{-\frac{t^\beta}{2}} \cdot |t|^{(\beta-j)k} \right\} \\ &\overset{\text{(i)}}{\le} \beta^k k^k \sum_{j=1}^{\beta} \left(\frac{\beta}{\lambda}\right)^{jk} \overset{\text{(ii)}}{\le} \frac{\beta^{2k}k^k}{\lambda^k} \sum_{j=1}^{\beta} \frac{1}{2^{j-1}} \overset{\text{(iii)}}{\le} 2.\end{aligned}$$

In step (i) above, we use the chain of bounds

$$\sup_{t\in\mathbb{R}} \left\{ e^{-\frac{t^\beta}{2}} \cdot |t|^{(\beta-j)k} \right\} \le 1 \vee \sup_{t>1} \left\{ e^{-\frac{t^\beta}{2}} \cdot |t|^{\beta k} \right\} \le 1 \vee k^k \le k^k.$$

Step (ii) above holds for all $\lambda \ge \beta^2$, since $\beta \ge 2$. Finally, step (iii) holds for all $\lambda \ge \beta^2 n$. Putting together the pieces, we obtain by definition (26) that

$$M(\psi, n, \lambda) \le 3, \quad \text{for all } n \in \mathbb{N} \text{ and } \lambda \ge \beta^2 n.$$

Proof of the consequence. Note that since $C_\psi = 1$, the requirement on N from conditions (27) and (29) is $N \ge C \log^4(C/\epsilon)$ for a universal and positive constant C. It suffices to set $\lambda \ge 4\beta^2 N$ to satisfy the conditions from the previous parts. We thus have

$$\mathsf{Err}(\psi, 2N, \lambda) \le \frac{\epsilon}{6} \quad \text{and} \quad M(\psi, 2N, \lambda) \le 3, \quad \text{if} \quad \lambda \gtrsim \beta^2 \log^8(C/\epsilon).$$

Consequently, by letting $T_{\mathsf{iter}} \gtrsim \log(3/\epsilon)$, we obtain from guarantee (30) that

$$\sup_{\theta\in\mathbb{R}} \mathsf{d}_{\mathsf{TV}}\left(\mathsf{K}(X_\theta), Y_\theta\right) \le \epsilon.$$

Note that the eventual requirement on τ from condition (27) reads as $\tau \gtrsim \sigma\beta^2 \log^{12}(C/\epsilon)$.

We now turn to the computational complexity of the reduction. Note that we can use Eq. (68) to evaluate the signed kernel $\mathcal{S}_N^\star$ (14). Since $\psi^{(j)}(t) = 0$ for $j \ge \beta + 1$, then the summation in Eq. (68) is over all tuples of nonnegative integers $(m_1, \cdots, m_\beta)$ satisfying $\sum_{j=1}^{\beta} j m_j = n$. There are at most n^β such tuples. Thus, evaluating $\nabla_x^{(n)} v(y; x)$ takes time $\mathcal{O}(n^{\beta+1})$. Since in addition we have $N = \mathcal{O}(\log^4(C/\epsilon))$, we can bound the evaluation time $T_{\mathsf{eval}}(N) = \mathcal{O}(\log^{4\beta+4}(C/\epsilon))$. We view the sampling time as a constant ($T_{\mathsf{samp}} = \mathcal{O}(1)$), since the generalized normal distribution is straightforward to sample from. Putting this together with number of iterations $T_{\mathsf{iter}} \asymp \log(3/\epsilon)$ yields a total computational complexity of $\mathcal{O}\Big(\log^{4\beta+5}(C/\epsilon)\Big)$. □

7 Proofs of results from Section 2.3

We prove Theorem 2(a) in Section 7.1 and Theorem 2(b) in Section 7.2. We provide the proof of the corollaries and concrete examples of Theorem 2 in Sections 7.3–7.5.

7.1 Proof of Theorem 2(a)

Throughout this section, to reduce the notational burden, we use the shorthand:

$$h_j(x) = \frac{f^{(j)}(x/\tau)}{(\widetilde{C}_f)^j j!}, \quad \phi(x) = \frac{e^{-x^2}}{\sqrt{2\pi}}, \quad g(x) = \frac{f(x/\tau) - y}{2}, \quad \text{for all } x, y \in \mathbb{R}, j \in \mathbb{Z}_{\geq 0}. \tag{99}$$

We first state an auxiliary lemma, whose proof we defer to Appendix B.3.

Lemma 5. *Let $v(y;\theta)$ be defined as in Eq. (33). Let $h_j(x)$, $\phi(x)$, and $g(x)$ be defined in Eq. (99). Then we have for all $n \in \mathbb{Z}_{\geq 0}$ and $x, y \in \mathbb{R}$,*

$$\nabla_\theta^{(n)} v(y;\theta)\big|_{\theta=x} = \frac{(\widetilde{C}_f)^n}{\tau^n} \cdot n! \cdot \sum_{m_1,\ldots,m_n} \frac{(-1)^{m_1+\cdots+m_n} \cdot \phi(g(x)) \cdot H_{m_1+\cdots+m_n}(g(x))}{m_1!\, m_2! \cdots m_n!} \cdot \prod_{j=1}^{n} \left(\frac{h_j(x)}{2}\right)^{m_j}, \tag{100}$$

where the summation is over all n-tuples of nonnegative integers $(m_1, \cdots, m_n)$ satisfying $\sum_{j=1}^n j m_j = n$. Moreover, we have

$$\left|\nabla_\theta^{(n)} v(y;\theta)\big|_{\theta=x}\right| \leq e^{-\frac{g(x)^2}{2}} \cdot \frac{(2\widetilde{C}_f)^n}{\tau^n} \cdot n! \cdot e^n \cdot \left[\left(\sum_{j=1}^{n} |h_j(x)|\right)^n \vee 1\right]. \tag{101}$$

We now use the Lemma 5 to verify that the target density $v(y;\theta)$ defined in Eq. (33) satisfies Assumption 1, ensuring that Lemma 1 is applicable.

Part (a) of the assumption immediately holds by definition of $v(y;\theta)$ (33). We now verify Assumption 1(b). Ineq. (101) implies that

$$\begin{aligned}
\int_{\mathbb{R}} \left|\nabla_x^{(n)} v(y;x)\right| e^{-\frac{(x-\theta)^2}{2}} \mathrm{d}x &\leq \frac{(2e\widetilde{C}_f)^n}{\tau^n} \cdot n! \cdot \left[\int_{\mathbb{R}} \left(\sum_{j=1}^{n} |h_j(x)|\right)^n e^{-\frac{(x-\theta)^2}{2}} \mathrm{d}x + \int_{\mathbb{R}} e^{-\frac{(x-\theta)^2}{2}} \mathrm{d}x\right] \\
&\overset{\text{(i)}}{=} \frac{(2e\widetilde{C}_f)^n}{\tau^n} \cdot n! \cdot \left[\sqrt{2}\int_{\mathbb{R}} \left(\sum_{j=1}^{n} \frac{|f^{(j)}((\theta+\sqrt{2}t)/\tau)|}{(\widetilde{C}_f)^j j!}\right)^n e^{-t^2} \mathrm{d}t + \sqrt{2\pi}\right] \\
&\overset{\text{(ii)}}{\leq} \frac{(2e\widetilde{C}_f)^n}{\tau^n} \cdot n! \cdot \left(2\sqrt{\pi} \cdot (C_f n)! + \sqrt{2\pi}\right) < +\infty,
\end{aligned}$$

where in step (i) we use the change of variable $t = (x-\theta)/\sqrt{2}$, and in step (ii) we use Assumption 3. Note that $\left|\nabla_x^{(n)} v(y;x)\right| e^{-\frac{(x-\theta)^2}{2}}$ is a continuous and nonnegative function on $\mathbb{R}$, and its integral is bounded. This means it converges to zero at infinity, that is,

$$\lim_{x\uparrow+\infty} \left[\nabla_x^{(n)} v(y;x)\right] e^{-\frac{(x-\theta)^2}{2}} = \lim_{x\downarrow-\infty} \left[\nabla_x^{(n)} v(y;x)\right] e^{-\frac{(x-\theta)^2}{2}} = 0.$$

This proves Assumption 1(b). We now verify Assumption 1(c). Note that by definition (33), $v(y;\sqrt{2}x+\theta) \leq 1/\sqrt{\pi}$ for all $x, y, \theta \in \mathbb{R}$, and consequently

$$\int_{\mathbb{R}} \left(v\big(y;\sqrt{2}x+\theta\big)\right)^2 \exp(-x^2)\mathrm{d}x \leq \int_{\mathbb{R}} e^{-x^2}/\sqrt{\pi}\mathrm{d}x = 1 < +\infty.$$

This proves that Assumption 1(c) holds.

Consequently, Lemma 1 holds and we obtain

$$\delta(\mathcal{U}, \mathcal{V}; \mathcal{T}_N^\star) \leq \varsigma(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star) + \frac{1}{2} \sup_{\theta\in\Theta} \int_{\mathbb{R}} \big(|p(x) - 1| + q(x)\big) \cdot u(x;\theta)\mathrm{d}x. \tag{102}$$

To bound the terms appearing on the RHS above, we state an auxiliary lemma, whose proof we defer to the following subsections.

Lemma 6. *Let the source and target statistical models $\mathcal{U}$ and $\mathcal{V}$ be given by Eq. (2) (with $\sigma = 1$) and Eq. (32), respectively. Let the signed kernel $\mathcal{S}_N^\star$ be defined in Eq. (14) and let $\varsigma(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star)$ be defined in Eq. (10). Suppose the pair of parameters (N, τ) satisfies condition (36) and Assumption 3 holds. Let $\overline{C}_f$ be defined in Eq. (35). Then the following statements hold:*

(a) We have

$$\varsigma(\mathcal{U}, \mathcal{V}; \mathcal{S}_N^\star) \le \overline{C}_f \exp\bigg(- (2e)^{-1}(2N+1)^{\frac{1}{2(1+C_f)}}\bigg). \tag{103}$$

(b) Let $p(x)$ and $q(x)$ be defined in Eq. (17). Then

$$\frac{1}{2}\sup_{\theta\in\mathbb{R}} \int_{\mathbb{R}} \big(|p(x)-1| + q(x)\big)\cdot u(x;\theta)\mathrm{d}x \le 12\exp(-\tau/8) + 15 \sup_{\theta\in\Theta} \exp\bigg(\frac{\theta^2}{2} - \frac{\tau^2}{8}\bigg). \tag{104}$$

Substituting Ineq. (103) and Ineq. (104) into the RHS of Ineq. (102) proves part (a) of the theorem. □

We next prove Ineq. (103) in Section 7.1.1, and prove Ineq. (104) in Section 7.1.2.

7.1.1 Proof of Lemma 6(a)

The proof is identical to that of Lemma 3(a) in Section 6.1.1. We state two key claims. First, for all $\tau \ge 2 \vee 2|\theta| \vee 2\widetilde{C}_f$ and $n, \ell \in \mathbb{Z}_{\ge 0}$ and $y \in \mathbb{R}$, we have

$$\bigg[\frac{\mathrm{d}^\ell\big[v\big(y;\theta+\sqrt{2}x\big)e^{-x^2}\big]}{\mathrm{d}x^\ell}\overline{H}_{n+1+\ell}(x)\bigg]\bigg|_{-\infty}^{+\infty} = 0, \tag{105}$$

where the evaluation at infinity is with respect to variable x. Second, for all $\tau \ge 8\sqrt{2}e\widetilde{C}_f$ and for all $r, k \in \mathbb{Z}_{\ge 0}$ and $k \le r$, we have

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \Big|\nabla_x^{(r-k)} v\big(y;\theta+\sqrt{2}x\big)\cdot e^{-x^2/2}\Big|\mathrm{d}x\mathrm{d}y \le 4\sqrt{2}\pi(r-k)!\cdot\frac{1}{4^{r-k}}\cdot\big(C_f(r-k)\big)!. \tag{106}$$

We prove the two claims above, which are analogous to Claims (75) and (78) in Section 6.1.1, and we omit the remaining steps to prove the lemma, since these are exactly parallel to before.

Proof of Claim (105). By definition, we have for all $n \in \mathbb{Z}_{\ge 0}$ that

$$\nabla_x^{(n)} v\big(y;\theta+\sqrt{2}x\big) = \nabla_z^{(n)} v\big(y;z\big)\big|_{z=\theta+\sqrt{2}x}\cdot(\sqrt{2})^n.$$

Applying Ineq. (101), we obtain

$$\begin{aligned}
\big|\nabla_x^{(n)} v\big(y;\theta+\sqrt{2}x\big)\big| &\le 2^{n/2}\cdot\Big|\nabla_z^{(n)} v\big(y;z\big)\big|_{z=\theta+\sqrt{2}x}\Big| \\
&\le e^{-\frac{g(\theta+\sqrt{2}x)^2}{2}}\cdot\frac{(2\sqrt{2}e\widetilde{C}_f)^n}{\tau^n}\cdot n!\cdot\bigg[\bigg(\sum_{j=1}^n\big|h_j\big(\theta+\sqrt{2}x\big)\big|\bigg)^n \vee 1\bigg] \qquad (107)\\
&\le \frac{(2\sqrt{2}e\widetilde{C}_f)^n}{\tau^n}\cdot n!\cdot\bigg[\bigg(\sum_{j=1}^n\big|h_j\big(\theta+\sqrt{2}x\big)\big|\bigg)^n \vee 1\bigg]. \qquad (108)
\end{aligned}$$

Using the product rule (76) and Ineq. (62a), we obtain

$$\left|\frac{\mathrm{d}^\ell\left[v\big(y;\theta+\sqrt{2}x\big)e^{-x^2}\right]}{\mathrm{d}x^\ell}\overline{H}_{n+1+\ell}(x)\right| \le \sum_{k=0}^{\ell}\binom{\ell}{k}\left|\nabla_x^{(\ell-k)}v\big(y;\theta+\sqrt{2}x\big)\cdot H_k(x)e^{-x^2}\right|$$
$$\le \sum_{k=0}^{\ell}\binom{\ell}{k}\frac{(2\sqrt{2}e\widetilde{C}_f)^{\ell-k}}{\tau^{\ell-k}}\cdot(\ell-k)!\cdot\left[\left(\sum_{j=1}^{\ell-k}\left|h_j\big(\theta+\sqrt{2}x\big)\right|\right)^{\ell-k}\vee 1\right]\cdot\sqrt{\gamma_k}\cdot e^{-x^2/2},$$

where in the last step we have used Ineq. (108) and $|H_k(x)e^{-x^2/2}| \le \sqrt{\gamma_k}$ with $\gamma_k = \sqrt{\pi}2^k k!$ from Ineq. (62a). Note that from Assumption 3 and Eq. (99), we obtain

$$\int_{\mathbb{R}}\left(\sum_{j=1}^{\ell-k}\left|h_j\big(\theta+\sqrt{2}x\big)\right|\right)^{\ell-k}e^{-x^2/2}\mathrm{d}x \le \sqrt{2\pi}(C_f(\ell-k))! < +\infty.$$

If the integral of a continuous and nonnegative function over $\mathbb{R}$ is finite, then it must converge to zero at infinity. Thus, for all $\ell-k\ge 0$

$$\lim_{x\to\infty}\left(\sum_{j=1}^{\ell-k}\left|h_j\big(\theta+\sqrt{2}x\big)\right|\right)^{\ell-k}e^{-x^2/2} = 0.$$

Combining the pieces yields

$$\lim_{x\to\infty}\frac{\mathrm{d}^\ell\left[v\big(y;\theta+\sqrt{2}x\big)e^{-x^2}\right]}{\mathrm{d}x^\ell}\cdot\overline{H}_{n+1+\ell}(x) = 0,$$

which proves Claim (105).

Proof of Claim (106)**.** Applying Ineq. (107) yields

$$\int_{\mathbb{R}}\int_{\mathbb{R}}\left|\nabla_x^{(r-k)}v\big(y;\theta+\sqrt{2}x\big)\cdot e^{-x^2/2}\right|\mathrm{d}y\mathrm{d}x$$
$$\le \frac{\big(2\sqrt{2}e\widetilde{C}_f\big)^{r-k}}{\tau^{r-k}}\cdot(r-k)!\cdot\int_{\mathbb{R}}\int_{\mathbb{R}}e^{-\frac{g(\theta+\sqrt{2}x)^2}{2}}\mathrm{d}y\cdot\left[\left(\sum_{j=1}^{r-k}\left|h_j\big(\theta+\sqrt{2}x\big)\right|\right)^{r-k}+1\right]e^{-x^2/2}\mathrm{d}x.$$

Recall the function $g(x) = (f(x/\tau)-y)/2$ defined in Eq. (99). We obtain

$$\int_{\mathbb{R}}e^{-\frac{g(\theta+\sqrt{2}x)^2}{2}}\mathrm{d}y = \int_{\mathbb{R}}e^{-y^2/4}\mathrm{d}y = 2\sqrt{\pi}.$$

By Assumption 3 and by definition of $h_j(x)$ in Eq. (99), we obtain

$$\int_{\mathbb{R}}\left[\left(\sum_{j=1}^{r-k}\left|h_j\big(\theta+\sqrt{2}x\big)\right|\right)^{r-k}+1\right]e^{-x^2/2}\mathrm{d}x \le \sqrt{2\pi}\big(1+(C_f(r-k))!\big) \le 2\sqrt{2\pi}\cdot(C_f(r-k))!.$$

Combining all the pieces and letting $\tau \ge 8\sqrt{2}e\widetilde{C}_f$, we have

$$\int_{\mathbb{R}}\int_{\mathbb{R}}\left|\nabla_x^{(r-k)}v\big(y;\theta+\sqrt{2}x\big)\cdot e^{-x^2/2}\right|\mathrm{d}y\mathrm{d}x \le \frac{4\sqrt{2}\pi}{4^{r-k}}\cdot(C_f(r-k))!\cdot(r-k)!.$$

This concludes the proof of Claim. (106). □

7.1.2 Proof of Lemma 6(b)

We first state an auxiliary lemma and defer its proof to Appendix B.4.

Lemma 7. *Consider the signed kernel $\mathcal{S}_N^\star(y \mid x)$ defined in Eq. (14) for target density $v(y;\theta)$ defined in Eq. (33). For all $n \in \mathbb{N}$, let $L_f(n)$ be defined as in Eq. (35). Recall the shorthand in Eq. (99). Suppose Assumption 3 holds with the pair of constants $(C_f, \widetilde{C}_f)$. Then we have:*

(a) If $\tau \geq 4e^3\big(\widetilde{C}_f\big)^2\big(L_f(2N) \vee 1\big)N$, then

$$\mathcal{S}_N^\star(y \mid x) \geq 0 \quad \text{for all} \quad |x| \leq \tau, \quad |y - f(x/\tau)| \leq \sqrt{\tau}, \quad N \in \mathbb{Z}_{\geq 0}. \tag{109}$$

(b) For all $|x| \leq \tau$ and $k \geq 1$ and $k \in \mathbb{N}$,

$$\begin{aligned}
&\int_{f(\frac{x}{\tau})+\sqrt{\tau}}^{+\infty} \big|\nabla_x^{(2k)} v(y;x)\big| \mathrm{d}y + \int_{-\infty}^{f(\frac{x}{\tau})-\sqrt{\tau}} \big|\nabla_x^{(2k)} v(y;x)\big| \mathrm{d}y \\
&\leq 2\sqrt{2\pi}\exp(-\tau/8) \cdot (2k)! \cdot \left(\frac{2e\widetilde{C}_f}{\tau}\right)^{2k} \cdot \big(L_f(2k) \vee 1\big)^{2k}.
\end{aligned} \tag{110}$$

(c) For all $x \in \mathbb{R}$ and $n \geq 1$ and $n \in \mathbb{N}$,

$$\int_{\mathbb{R}} \nabla_x^{(n)} v(y;x) \mathrm{d}y = 0. \tag{111}$$

Taking the above lemma as given, let us now prove Lemma 6(b). We split the proof into two cases.

Case 1: bound $|p(x) - 1| + q(x)$ for $|x| \leq \tau$. By definition (17), we obtain for all $x \in \mathbb{R}$

$$\begin{aligned}
p(x) - q(x) &= \int_{\mathbb{R}} \big[\mathcal{S}_N^\star(y \mid x) \vee 0\big] + \big[\mathcal{S}_N^\star(y \mid x) \wedge 0\big] \mathrm{d}y \\
&= \int_{\mathbb{R}} \mathcal{S}_N^\star(y \mid x) \mathrm{d}y \\
&= \int_{\mathbb{R}} v(y;x) \mathrm{d}y + \sum_{k=1}^{N} \frac{(-1)^k}{(2k)!!} \int_{\mathbb{R}} \nabla_x^{(2k)} v(y;x) \mathrm{d}y \overset{\text{(i)}}{=} 1 + 0 = 1,
\end{aligned}$$

where in step (i) we use Eq. (111) to control the second term and definition (33) to control the first, whereby $\int_{\mathbb{R}} v(y;x)\mathrm{d}y = \int_{\mathbb{R}} e^{-y^2/2}/\sqrt{2\pi}\mathrm{d}y = 1$. We consequently obtain $q(x) = p(x) - 1$ for all $x \in \mathbb{R}$. Moreover, note that since $q(x) \geq 0$ by definition (17), we obtain $p(x) \geq 1$ for all $x \in \mathbb{R}$. We thus obtain

$$|p(x) - 1| + q(x) \leq 2(p(x) - 1) \quad \text{for all } x \in \mathbb{R}. \tag{112}$$

We next turn to upper bound $p(x)$ when $|x| \le \tau$. Using Ineq. (109) and definition of $p(x)$ (17), we have

$$\begin{aligned}
p(x) &\le \int_{f(\frac{x}{\tau})-\sqrt{\tau}}^{f(\frac{x}{\tau})+\sqrt{\tau}} \mathcal{S}_N^\star(y \mid x)\mathrm{d}y + \int_{f(\frac{x}{\tau})+\sqrt{\tau}}^{+\infty} \big|\mathcal{S}_N^\star(y \mid x)\big|\mathrm{d}y + \int_{-\infty}^{f(\frac{x}{\tau})-\sqrt{\tau}} \big|\mathcal{S}_N^\star(y \mid x)\big|\mathrm{d}y \\
&\le \int_{\mathbb{R}} v(y;x)\mathrm{d}y + \sum_{k=1}^{N} \frac{1}{(2k)!!} \int_{f(\frac{x}{\tau})-\sqrt{\tau}}^{f(\frac{x}{\tau})+\sqrt{\tau}} \nabla_x^{(2k)} v(y;x)\mathrm{d}y \\
&\quad + \sum_{k=1}^{N} \frac{1}{(2k)!!} \bigg(\int_{f(\frac{x}{\tau})+\sqrt{\tau}}^{+\infty} \big|\nabla_x^{(2k)} v(y;x)\big|\mathrm{d}y + \int_{-\infty}^{f(\frac{x}{\tau})-\sqrt{\tau}} \big|\nabla_x^{(2k)} v(y;x)\big|\mathrm{d}y \bigg) \\
&\le 1 + 2\sum_{k=1}^{N} \frac{1}{(2k)!!} \bigg(\int_{f(\frac{x}{\tau})+\sqrt{\tau}}^{+\infty} \big|\nabla_x^{(2k)} v(y;x)\big|\mathrm{d}y + \int_{-\infty}^{f(\frac{x}{\tau})-\sqrt{\tau}} \big|\nabla_x^{(2k)} v(y;x)\big|\mathrm{d}y \bigg) / \qquad (113)
\end{aligned}$$

Here in the last step we use $\int_{\mathbb{R}} v(y;x)\mathrm{d}y = 1$ by definition (33). We also and use Eq. (111) so that for $k \ge 1$, we have

$$\begin{aligned}
\int_{f(\frac{x}{\tau})-\sqrt{\tau}}^{f(\frac{x}{\tau})+\sqrt{\tau}} \nabla_x^{(2k)} v(y;x)\mathrm{d}y &= -\int_{f(\frac{x}{\tau})+\sqrt{\tau}}^{+\infty} \nabla_x^{(2k)} v(y;x)\mathrm{d}y - \int_{-\infty}^{f(\frac{x}{\tau})-\sqrt{\tau}} \nabla_x^{(2k)} v(y;x)\mathrm{d}y \\
&\le \int_{f(\frac{x}{\tau})+\sqrt{\tau}}^{+\infty} \big|\nabla_x^{(2k)} v(y;x)\big|\mathrm{d}y + \int_{-\infty}^{f(\frac{x}{\tau})-\sqrt{\tau}} \big|\nabla_x^{(2k)} v(y;x)\big|\mathrm{d}y.
\end{aligned}$$

Substituting Ineq. (110) into the RHS of Ineq. (113) yields

$$\begin{aligned}
p(x) &\le 1 + 4\sqrt{2\pi} e^{-\tau/8} \sum_{k=1}^{N} \frac{(2k)!}{(2k)!!} \cdot (4e^2)^k \cdot \left(\frac{\widetilde{C}_f}{\tau}\right)^{2k} \cdot \big(L_f(2k) \vee 1\big)^{2k} \\
&\overset{\text{(i)}}{\le} 1 + 4\sqrt{2\pi} e^{-\tau/8} \sum_{k=1}^{N} \left(\frac{8ke^2 \big(\widetilde{C}_f\big)^2 \big(L_f(2k) \vee 1\big)^2}{\tau^2}\right)^k \\
&\overset{\text{(ii)}}{\le} 1 + 4\sqrt{2\pi} e^{-\tau/8} \sum_{k=1}^{N} 2^{-k} \le 1 + 12 e^{-\tau/8}. \qquad (114)
\end{aligned}$$

Above, in step (i) we use $(2k)!/(2k)!! = (2k-1)!! \le (2k)^k$ and in step (ii) we use

$$\tau^2 \ge 16Ne^2 \big(\widetilde{C}_f\big)^2 \big(L_f(2N) \vee 1\big)^2 \ge 16ke^2 \big(\widetilde{C}_f\big)^2 \big(L_f(2k) \vee 1\big)^2, \quad \text{for all } k \in [N].$$

Combining Ineq. (114) with Ineq. (112) yields

$$\big|p(x) - 1\big| + q(x) \le 2\big(p(x) - 1\big) \le 24 \exp(-\tau/8), \quad \text{for all } |x| \le \tau. \qquad (115)$$

This concludes the proof for the first case. Now we consider the second case.

Case 2: bound $|p(x)-1|+q(x)$ for $|x|\geq\tau$. By definition, we obtain

$$\begin{aligned}
p(x)\vee q(x) &\leq \int_{\mathbb{R}} \left|\mathcal{S}_N^\star(y\mid x)\right|\mathrm{d}y \\
&\leq 1+\sum_{k=1}^{N}\frac{1}{(2k)!!}\int_{\mathbb{R}}\left|\nabla_x^{(2k)}v(y;x)\right|\mathrm{d}y \\
&\overset{\text{(i)}}{\leq} 1+\sum_{k=1}^{N}\frac{(2k)!}{(2k)!!}\cdot(4e^2)^k\cdot\left(\frac{\widetilde{C}_f}{\tau}\right)^{2k}\cdot\left[\left(\sum_{j=1}^{2k}\left|h_j(x)\right|\right)\vee 1\right]^{2k}\int_{\mathbb{R}}e^{-t^2/8}\mathrm{d}t \\
&\overset{\text{(ii)}}{\leq} 1+2\sqrt{2\pi}\sum_{k=1}^{N}\left(\frac{8e^2k\big(\widetilde{C}_f\big)^2}{\tau^2}\right)^k\cdot\left[\left(\sum_{j=1}^{2k}\left|h_j(x)\right|\right)\vee 1\right]^{2k}.
\end{aligned}$$

In step (i) we use Ineq. (101) and note that since $g(x)=\frac{f(x/\tau)-y}{2}$, we also use

$$\int_{\mathbb{R}}e^{-\frac{g(x)^2}{2}}\mathrm{d}y=\int_{\mathbb{R}}e^{-\frac{(y-f(x/\tau))^2}{8}}\mathrm{d}y=\int_{\mathbb{R}}e^{-t^2/8}\mathrm{d}t,$$

where in the last step we use change of variables $t=y-f(x/\tau)$. In step (ii), we use the inequalities $(2k)!/(2k)!!=(2k-1)!!\leq(2k)^k$ and $\int_{\mathbb{R}}e^{-t^2/8}\mathrm{d}t\leq 2\sqrt{2\pi}$. Note that $p(x),q(x)\geq 0$ for all $x\in\mathbb{R}$ by definition (17). Consequently, we obtain

$$\begin{aligned}
|p(x)-1|+q(x) &\leq 1+p(x)+q(x) \\
&\leq 3+4\sqrt{2\pi}\sum_{k=1}^{N}\left(\frac{8e^2k\big(\widetilde{C}_f\big)^2}{\tau^2}\right)^k\cdot\left[\left(\sum_{j=1}^{2k}\left|h_j(x)\right|\right)\vee 1\right]^{2k}, \quad \text{for all } x\in\mathbb{R}. \quad (116)
\end{aligned}$$

Putting the two cases together. We have

$$\begin{aligned}
\int_{\mathbb{R}}\big(|p(x)-1|+q(x)\big)\cdot u(x;\theta)\mathrm{d}x &\leq \int_{|x|\leq\tau}\big(|p(x)-1|+q(x)\big)\cdot u(x;\theta)\mathrm{d}x \\
&\quad+\int_{|x|\geq\tau}\big(|p(x)-1|+q(x)\big)\cdot u(x;\theta)\mathrm{d}x.
\end{aligned}$$

We then bound the two integrals in the display above. Applying Ineq. (115) yields

$$\int_{|x|\leq\tau}\big(|p(x)-1|+q(x)\big)\cdot u(x;\theta)\mathrm{d}x\leq 24\exp(-\tau/8)\cdot\int_{|x|\leq\tau}u(x;\theta)\mathrm{d}x\leq 24\exp(-\tau/8). \quad (117)$$

Applying Ineq. (116) yields

$$\begin{aligned}
&\int_{|x|\geq\tau}\big(|p(x)-1|+q(x)\big)\cdot u(x;\theta)\mathrm{d}x \\
&\leq 3\int_{|x|\geq\tau}u(x;\theta)\mathrm{d}x+4\sqrt{2\pi}\sum_{k=1}^{N}\left(\frac{8e^2k\big(\widetilde{C}_f\big)^2}{\tau^2}\right)^k\cdot\int_{|x|\geq\tau}\left[\left(\sum_{j=1}^{2k}\left|h_j(x)\right|\right)\vee 1\right]^{2k}\cdot u(x;\theta)\mathrm{d}x. \quad (118)
\end{aligned}$$

By definition of $u(x;\theta)$ (4) for $\sigma=1$, we obtain

$$u(x;\theta)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\theta)^2}{2}}\leq e^{\frac{\theta^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{4}},$$

where in the last step we use $(x-\theta)^2 \geq x^2+\theta^2-2|\theta x| \geq x^2+\theta^2-(x^2/2+2\theta^2) \geq x^2/2-\theta^2$. Consequently, we obtain

$$\int_{|x|\geq\tau} u(x;\theta)\mathrm{d}x \leq 2e^{\frac{\theta^2}{2}} \int_{\tau}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{4}}\mathrm{d}x \leq \sqrt{2}e^{\frac{\theta^2}{2}}e^{-\frac{\tau^2}{4}}, \tag{119}$$

and similarly for $k \geq 1$, we have

$$\begin{aligned}
\int_{|x|\geq\tau} \left[\left(\sum_{j=1}^{2k}\left|h_j(x)\right|\right)\vee 1\right]^{2k} \cdot u(x;\theta)\mathrm{d}x &\leq e^{\frac{\theta^2}{2}} \int_{|x|\geq\tau} \left[\left(\sum_{j=1}^{2k}\left|h_j(x)\right|\right)\vee 1\right]^{2k} \cdot \frac{e^{-\frac{x^2}{4}}}{\sqrt{2\pi}}\mathrm{d}x \\
&\overset{\text{(i)}}{\leq} e^{\frac{\theta^2}{2}} \int_{|x|\geq\tau} \frac{e^{-\frac{x^2}{8}}}{\sqrt{2\pi}}\mathrm{d}x \cdot \left(\widetilde{L}_f(2k,\tau)\vee 1\right)^{2k} \\
&\overset{\text{(ii)}}{\leq} 2e^{\frac{\theta^2}{2}}e^{-\frac{\tau^2}{8}} \cdot \left(\widetilde{L}_f(2k,\tau)\vee 1\right)^{2k}.
\end{aligned} \tag{120}$$

Above, in step (i) we use definition of $h_j(x)$ (99) and $\widetilde{L}_f(n,\tau,\sigma)$ (35) so that for all $|x| \geq \tau$,

$$\begin{aligned}
\left[\left(\sum_{j=1}^{2k}\left|h_j(x)\right|\right)\vee 1\right]e^{-\frac{x^2}{8(2k)}} &= \left[\left(\sum_{j=1}^{2k}\frac{|f(x/\tau)|}{(\widetilde{C}_f)^j j!}\right)\vee 1\right]e^{-\frac{x^2}{8(2k)}} \\
&\leq \sup_{|t|\geq 1}\left[\left(\sum_{j=1}^{2k}\frac{|f(t)|}{(\widetilde{C}_f)^j j!}\right)\vee 1\right]e^{-\frac{\tau^2 t^2}{8(2k)}} \\
&\leq \widetilde{L}_f(2k,\tau)\vee 1.
\end{aligned}$$

In step (ii), we use

$$\int_{|x|\geq\tau}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{8}}\mathrm{d}x \leq 2\exp\big(-\tau^2/8\big).$$

Substituting Ineq. (119) and Ineq. (120) into Ineq. (118) yields

$$\begin{aligned}
&\int_{|x|\geq\tau}\big(|p(x)-1|+q(x)\big)\cdot u(x;\theta)\mathrm{d}x \\
&\leq 2e^{\frac{\theta^2}{2}}e^{-\frac{\tau^2}{8}} \cdot \left(3+4\sqrt{2\pi}\sum_{k=1}^{N}\left(\frac{8e^2k\big(\widetilde{C}_f\big)^2}{\tau^2}\right)^k \cdot \left(\widetilde{L}_f(2k,\tau)\vee 1\right)^{2k}\right) \\
&\overset{\text{(i)}}{\leq} 2e^{\frac{\theta^2}{2}}e^{-\frac{\tau^2}{8}} \cdot \left(3+4\sqrt{2\pi}\sum_{k=1}^{N}\frac{1}{2^k}\right) \leq 30e^{\frac{\theta^2}{2}}e^{-\frac{\tau^2}{8}},
\end{aligned} \tag{121}$$

where in step (i) we use the condition

$$\tau^2 \geq 16e^2N\big(\widetilde{C}_f\big)^2\big(\widetilde{L}_f(2N,\tau)\vee 1\big)^2 \geq 8e^2k\big(\widetilde{C}_f\big)^2\big(\widetilde{L}_f(2k,\tau)\vee 1\big)^2 \quad \text{for all } k\in[N].$$

Finally, combining Ineq. (121) with Ineq. (117) yields

$$\int_{\mathbb{R}}\big(|p(x)-1|+q(x)\big)\cdot u(x;\theta)\mathrm{d}x \leq 24\exp(-\tau/8)+30e^{\frac{\theta^2}{2}}e^{-\frac{\tau^2}{8}}.$$

Taking a supremum over $\theta\in\Theta$ proves the desired result. □

7.2 Proof of Theorem 2(b)

We split the proof into three parts.

Construction of the reduction based on rejection kernel. We define a class of base measures $\{\mathcal{D}(\cdot \mid x)\}_{x\in\mathbb{R}}$ as

$$\mathcal{D}(y \mid x) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{(y - f(x/\tau))^2}{8}\right), \quad \text{for all } x, y \in \mathbb{R}. \tag{122}$$

As before, we use the general rejection kernel $x \mapsto \mathrm{RK}(x, T, M, y_0)$ from Lou et al. (2025, page 10) with the base measures $\mathcal{D}(y \mid x)$ defined in (122) and the signed kernel $\mathcal{S}_N^\star(y \mid x)$ defined in (14). As before, $M > 0$ is a constant that is required to satisfy: for input x,

$$\frac{\mathcal{S}_N^\star(y \mid x) \vee 0}{\mathcal{D}(y \mid x)} \le M, \quad \text{for all } y \in \mathbb{R},$$

and y_0 is the initialization. Throughout this section, we set

$$T = \left\lceil 4\sqrt{2\pi}\log(4/\epsilon) \right\rceil, \quad M = 2\sqrt{2\pi}, \quad \text{and} \quad y_0 = 0. \tag{123}$$

We are now ready to describe the algorithm K. For $X_\theta \sim \mathcal{N}(\theta, \sigma^2)$, define

$$\mathsf{K}(X_\theta) := \begin{cases} \mathrm{RK}(X_\theta, T, M, y_0), & \text{if } |X_\theta| \le \tau, \\ y_0, & \text{otherwise.} \end{cases} \tag{124}$$

Note that for all $|x| \le \tau$ and $y \in \mathbb{R}$, we have by definition of $\mathcal{S}_N^\star(y \mid x)$ (14),

$$\begin{aligned} \frac{\mathcal{S}_N^\star(y \mid x) \vee 0}{\mathcal{D}(y \mid x)} &\le \sum_{k=0}^{N} \frac{1}{(2k)!!} \frac{\left|\nabla_x^{(2k)} v(y; x)\right|}{\mathcal{D}(y \mid x)} \\ &\overset{\text{(i)}}{\le} 2\sqrt{2\pi} \sum_{k=0}^{N} \frac{(2k)!}{(2k)!!} \cdot \frac{(4e^2)^k \big(\widetilde{C}_f\big)^{2k}}{\tau^{2k}} \cdot \left[\left(\sum_{j=1}^{2k} \left|h_j(x)\right|\right) \vee 1\right]^{2k} \\ &\overset{\text{(ii)}}{\le} 2\sqrt{2\pi} \sum_{k=0}^{N} \frac{(2k)^k \cdot (4e^2)^k \big(\widetilde{C}_f\big)^{2k}}{\tau^{2k}} \cdot \big(L_f(2k) \vee 1\big)^{2k} \\ &\overset{\text{(iii)}}{\le} 2\sqrt{2\pi} \sum_{k=0}^{N} \frac{1}{2^k} = M. \end{aligned}$$

In step (i) we use Ineq. (143) and definition of $\mathcal{D}(y \mid x)$ (122). In step (ii) we use $(2k)!/(2k)!! = (2k-1)!! \le (2k)^k$ and recall $h_j(x)$ (99) and $L_f(2k)$ (35), so that

$$\sum_{j=1}^{2k} \left|h_j(x)\right| \le L_f(2k) \quad \text{for all } |x| \le \tau.$$

In step (iii) we use condition (36) that τ is large enough to guarantee that the summation is bounded. Therefore, the reduction $\mathsf{K}(X_\theta) = \mathrm{RK}(X_\theta, T, M, y_0)$ when $|X_\theta| \le \tau$ (124) is well defined.

Computational complexity of the reduction. The rejection kernel $x \mapsto \textsc{rk}(x, T, M, y_0)$ requires, at each iteration, generating a sample from the uniform distribution $\mathsf{Unif}([0,1])$ and a sample from the Gaussian base measure (122), both of which take $\mathcal{O}(1)$ time. Additionally, it requires evaluating the signed kernel $\mathcal{S}_N^\star$ once per iteration, which takes time $T_{\mathsf{eval}}(N)$. The procedure is repeated for at most $T = \mathcal{O}(\log(4/\epsilon))$ iterations. Therefore, the total computational complexity of the reduction $\mathsf{K}(X_\theta)$ is

$$\mathcal{O}\big(\log(4/\epsilon) \cdot T_{\mathsf{eval}}(N)\big).$$

Proof of TV deficiency guarantee of the reduction. Having defined the reduction, we turn to prove the claimed guarantee (39). Applying the triangle inequality, we obtain

$$\begin{aligned}\mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) &= \frac{1}{2}\big\|\mathsf{K}(X_\theta) - v(\cdot;\theta)\big\|_1 \\ &\le \frac{1}{2}\left\|\mathsf{K}(X_\theta) - \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) u(x;\theta)\mathrm{d}x\right\|_1 + \frac{1}{2}\left\|\int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) u(x;\theta)\mathrm{d}x - v(\cdot;\theta)\right\|_1.\end{aligned}$$

Taking the supremum over $\theta \in \Theta$ on both sides yields

$$\sup_{\theta\in\Theta} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}(X_\theta), Y_\theta\big) \le \frac{1}{2}\sup_{\theta\in\Theta}\left\|\mathsf{K}(X_\theta) - \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) u(x;\theta)\mathrm{d}x\right\|_1 + \delta\big(\mathcal{U}, \mathcal{V}; \mathcal{T}_N^\star\big). \tag{125}$$

We claim that under our setting of parameters (38) (123), the following holds.

$$\frac{1}{2}\sup_{\theta\in\Theta}\left\|\mathsf{K}(X_\theta) - \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) u(x;\theta)\mathrm{d}x\right\|_1 \le \frac{\epsilon}{2}, \tag{126a}$$

$$\text{and} \qquad \delta\big(\mathcal{U}, \mathcal{V}; \mathcal{T}_N^\star\big) \le \frac{\epsilon}{2}. \tag{126b}$$

The desired guarantee Ineq. (39) follows immediately. We now prove the above two inequalities separately.

Proof of Ineq. (126a). For each x satisfying $|x| \le \tau$, we let $Y = \textsc{rk}(x, T, M, y_0)$. Then from Lou et al. (2025, Lemma 2), we obtain that the random variable Y has the conditional distribution

$$f_Y(y \mid x) = \frac{\mathcal{S}_N^\star(y \mid x) \vee 0}{p(x)} \cdot \big(1 - (1 - p(x)/M)^T\big) + \delta_{y_0}(y) \cdot (1 - p(x)/M)^T, \quad \text{for all } y \in \mathbb{R}, \tag{127}$$

where $p(x)$ is defined in Eq. (17) and $\delta_{y_0}(y)$ is the Dirac delta function centered at y_0. By the law of total probability and definition of $\mathsf{K}(X_\theta)$ in Eq. (124), we obtain that $\mathsf{K}(X_\theta)$ has the marginal distribution

$$\begin{aligned} f_{\mathsf{K}(X_\theta)}(y) &= \int_{|x|\le\tau} f_Y(y \mid x) \cdot u(x;\theta)\mathrm{d}x + \delta_{y_0}(y) \cdot \int_{|x|\ge\tau} u(x;\theta)\mathrm{d}x \\ &= \int_{|x|\le\tau} \mathcal{T}_N^\star(y \mid x) \cdot (1 - \ell(x)) \cdot u(x;\theta)\mathrm{d}x \\ &\qquad + \delta_{y_0}(y) \cdot \left(\int_{|x|\ge\tau} u(x;\theta)\mathrm{d}x + \int_{|x|\le\tau} \ell(x) u(x;\theta)\mathrm{d}x\right), \end{aligned} \tag{128}$$

where in the last step we have used Eq. (127) along with the shorthand notation $\ell(x) = (1 - p(x)/M)^T$ and the previously defined notation for the Markov kernel (15), $\mathcal{T}_N^\star(y \mid x) = \frac{\mathcal{S}_N^\star(y|x)\vee 0}{p(x)}$.

Further recall from Ineq. (114) that $|p(x) - 1| \leq 12\exp(-\tau/8) \leq \epsilon/2 \leq 1/2$ for all $|x| \leq \tau$. Consequently, we obtain that for all $|x| \leq \tau$,

$$1/2 \leq p(x) \leq 3/2 \quad \text{and} \quad 0 \leq \ell(x) \leq e^{-p(x)T/M} \leq e^{-T/(2M)}. \tag{129}$$

Using Eq. (128), we obtain

$$\begin{aligned}
&\left\| \mathsf{K}(X_\theta) - \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) u(x;\theta) \mathrm{d}x \right\|_1 \\
&= \int_{\mathbb{R}} \left| \int_{|x|\leq\tau} \mathcal{T}_N^\star(y \mid x) \cdot (1 - \ell(x)) \cdot u(x;\theta) \mathrm{d}x - \int_{\mathbb{R}} \mathcal{T}_N^\star(y \mid x) u(x;\theta) \mathrm{d}x \right| \mathrm{d}y \\
&\qquad + \int_{|x|\geq\tau} u(x;\theta)\mathrm{d}x + \int_{|x|\leq\tau} \ell(x) u(x;\theta)\mathrm{d}x \\
&\leq \int_{\mathbb{R}} \int_{|x|\leq\tau} \mathcal{T}_N^\star(y \mid x) \cdot \ell(x) u(x;\theta)\mathrm{d}x\mathrm{d}y + \int_{\mathbb{R}} \int_{|x|\geq\tau} \mathcal{T}_N^\star(y \mid x) \cdot u(x;\theta)\mathrm{d}x\mathrm{d}y \\
&\qquad + \int_{|x|\geq\tau} u(x;\theta)\mathrm{d}x + \int_{|x|\leq\tau} \ell(x) u(x;\theta)\mathrm{d}x,
\end{aligned}$$

where in the last step we apply the triangle inequality. Continuing, we bound each of the terms in the display above

$$\begin{aligned}
\int_{\mathbb{R}} \int_{|x|\leq\tau} \mathcal{T}_N^\star(y \mid x) \cdot \ell(x) u(x;\theta)\mathrm{d}x\mathrm{d}y &= \int_{|x|\leq\tau} \int_{\mathbb{R}} \mathcal{T}_N^\star(y \mid x)\mathrm{d}y \cdot \ell(x) u(x;\theta)\mathrm{d}x\mathrm{d}y \\
&= \int_{|x|\leq\tau} \ell(x) u(x;\theta)\mathrm{d}x \leq e^{-\frac{T}{2M}},
\end{aligned}$$

where in the last step we use Ineq. (129). Also, we have

$$\int_{\mathbb{R}} \int_{|x|\geq\tau} \mathcal{T}_N^\star(y \mid x) \cdot u(x;\theta)\mathrm{d}x\mathrm{d}y = \int_{|x|\geq\tau} u(x;\theta)\mathrm{d}x \leq e^{\frac{\theta^2}{2}} \int_{|x|\geq\tau} \frac{e^{-x^2/(4)}}{\sqrt{2\pi}}\mathrm{d}x \leq \sqrt{2} e^{\frac{\theta^2}{2} - \frac{\tau^2}{4}}.$$

Combining the three pieces together yields

$$\left\| \mathsf{K}(X_\theta) - \int_{\mathbb{R}} \mathcal{T}_N^\star(\cdot \mid x) u(x;\theta)\mathrm{d}x \right\|_1 \leq 2e^{-\frac{T}{2M}} + 2\sqrt{2} e^{\frac{\theta^2}{2} - \frac{\tau^2}{4}} \leq \epsilon,$$

where in the last step we use $T \geq 2M\log(4/\epsilon)$ and $\tau^2 \geq 4\log(4\sqrt{2}/\epsilon) + 2\sup_{\theta\in\Theta}\theta^2$. This concludes the proof of Ineq. (126a).

Proof of Ineq. (126b). Since we set

$$2N + 1 \geq \left(2e \log\left(4\overline{C}_f/\epsilon\right)\right)^{2(1+C_f)}, \quad \tau^2 \geq 64\log^2\left(96/\epsilon\right) \vee \left[8\log\left(120/\epsilon\right) + 4\sup_{\theta\in\Theta}\theta^2\right],$$

we obtain

$$\begin{aligned}
&\overline{C}_f \exp\left(-(2e)^{-1}(2N+1)^{\frac{1}{2(1+C_f)}} \right) \leq \frac{\epsilon}{4}, \quad \text{and} \\
&12\exp(-\tau/8) + 15\sup_{\theta\in\Theta} \exp\left(\frac{\theta^2}{2} - \frac{\tau^2}{8}\right) \leq \frac{\epsilon}{4}.
\end{aligned}$$

Substituting the two inequalities in the display above into Ineq. (37) completes the proof. □

7.3 Proof of Corollary 3

We verify each part separately.

Verifying Assumption 3. By definition, we obtain $f^{(j)}(x) = B! \cdot x^{B-j}/(B-j)!$ for all $x \in \mathbb{R}$ and $0 \le j \le B$, and $f^{(j)}(x) = 0$ for all $x \in \mathbb{R}$ and $j \ge B+1$. We now verify Assumption 3 in two cases.

Case 1: $n \le B$. For $n \le B$ and $x \in \mathbb{R}$, we have

$$\begin{aligned}\sum_{j=1}^{n} \frac{\left|f^{(j)}\big((\theta+\sqrt{2}x)/\tau\big)\right|}{(\widetilde{C}_f)^j j!} &= \sum_{j=1}^{n} \frac{B!}{(\widetilde{C}_f)^j j! \cdot (B-j)!} \cdot \frac{\left|\theta+\sqrt{2}x\right|^{B-j}}{\tau^{B-j}} \\ &\le \sum_{j=1}^{n} \frac{\binom{B}{j}}{(\widetilde{C}_f)^j} \cdot \frac{(2|\theta|)^{B-j} + (2\sqrt{2}|x|)^{B-j}}{\tau^{B-j}} \\ &\overset{\text{(i)}}{\le} \sum_{j=1}^{n} \frac{1+|x|^{B-j}}{(4e)^j},\end{aligned}$$

where in step (i) we use $\binom{B}{j} \le B^j$ and let $\widetilde{C}_f = 4eB$, and we use $\tau \ge 2\sup_{\theta\in\Theta}|\theta|$ and $\tau \ge 2\sqrt{2}$. Consequently, we obtain for all $x \in \mathbb{R}$,

$$\begin{aligned}\left(\sum_{j=1}^{n} \frac{\left|f^{(j)}\big((\theta+\sqrt{2}x)/\tau\big)\right|}{(\widetilde{C}_f)^j j!}\right)^n \le \left(\sum_{j=1}^{n} \frac{1+|x|^{B-j}}{(4e)^j}\right)^n &\le n^n \sum_{j=1}^{n} \left(\frac{1+|x|^{B-j}}{(4e)^j}\right)^n \\ &\le 2^n n^n \sum_{j=1}^{n} \frac{1+|x|^{(B-j)n}}{(4e)^{jn}}. \qquad (130)\end{aligned}$$

Integrating over x yields

$$\begin{aligned}\int_{\mathbb{R}} \left(\sum_{j=1}^{n} \frac{\left|f^{(j)}\big((\theta+\sqrt{2}x)/\tau\big)\right|}{(\widetilde{C}_f)^j j!}\right)^n \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathrm{d}x &\le \frac{2^n n^n}{(2e)^n} \sum_{j=1}^{n} \left[2^{-j} + 2^{-j}\big((B-j)n!!\big)\right] \\ &\overset{\text{(i)}}{\le} n! \cdot (2Bn)! \le \big((2B+1)n\big)!, \qquad (131)\end{aligned}$$

where in step (i) we use $n^n \le e^n n!$. Thus, Assumption 3 holds with $C_f = 2B+1$ and $C_f = 4eB$ for all $n \le B$.

Case 2: $n \ge B+1$. Note that $f^{(j)}(x) = 0$ for all $x \in \mathbb{R}$ when $j \ge B+1$. Consequently, following the same steps as in Ineq. (130) yields

$$\begin{aligned}\left(\sum_{j=1}^{n} \frac{\left|f^{(j)}\big((\theta+\sqrt{2}x)/\tau\big)\right|}{(\widetilde{C}_f)^j j!}\right)^n &= \left(\sum_{j=1}^{B} \frac{\left|f^{(j)}\big((\theta+\sqrt{2}x)/\tau\big)\right|}{(\widetilde{C}_f)^j j!}\right)^n \\ &\le 2^n B^n \sum_{j=1}^{B} \frac{1+|x|^{(B-j)n}}{(4e)^{jn}}.\end{aligned}$$

Integrating over x and using $B \le n$ yields the same bound as Ineq. (131).

Bounding the quantities $L_f(n)$ and $\widetilde{L}_f(n,\tau)$. We next turn to bound $L_f(n)$ and $\widetilde{L}_f(n,\tau)$ defined in Eq. (35). We obtain by definition that for all $n \in \mathbb{N}$ and $n \geq 1$,

$$L_f(n) \leq \sup_{|x|\leq 1} \sum_{j=1}^{B} \frac{B! \cdot |x|^{B-j}}{(B-j)! \cdot (\widetilde{C}_f)^j j!} \leq \sum_{j=1}^{B} (2e)^{-j} \leq 1,$$

where we use $\frac{B!}{(B-j)!\cdot j!} = \binom{B}{j} \leq B^j$ and $\widetilde{C}_f = 2eB$. Reasoning similarly, we obtain

$$\widetilde{L}_f(n,\tau) \leq \sup_{|x|\geq 1} \sum_{j=1}^{B} \frac{|x|^B}{(2e)^j} \cdot \exp\left(-\frac{\tau^2 x^2}{8n}\right) \leq \sup_{|x|\geq 1} |x|^B \exp\left(-\frac{\tau^2 x^2}{8n}\right) \leq 1,$$

where in the last step we let $\tau^2 \geq 8nB$, whereby the supremum is achieved at $|x| = 1$.

Bounding the evaluation time $T_{\mathsf{eval}}(n)$. We next bound the evaluation time $T_{\mathsf{eval}}(n)$ of the signed kernel $\mathsf{S}_n^\star(y \mid x)$ (14), noting the closed-form expression (100). Note that $h_j(x) = 0$ for all $j \geq B+1$ and $x \in \mathbb{R}$. Thus, we obtain

$$\nabla_x^{(n)} v(y;x) = n! \sum_{m_1,\ldots,m_B} \frac{(-1)^{m_1+\cdots+m_B}}{m_1!\, m_2! \cdots m_B!} \cdot H_{m_1+\cdots+m_B}(g(x))\phi(g(x)) \cdot \prod_{j=1}^{B} \left(\frac{h_j(x)}{2}\right)^{m_j}.$$

Here the summation is over all B-tuples of nonnegative integers $(m_1, \cdots, m_B)$ satisfying $\sum_{j=1}^{B} j m_j = n$; thus there is at most n^B such B-tuples. Also, each term in the summation can be computed in time $\mathcal{O}(n^2)$ using Eq. (62c). Consequently, evaluating $\nabla_x^{(n)} v(y;x)$ for any pair of (x,y) takes time $\mathcal{O}(n^{B+2})$. Using the expression (14), we conclude that $T_{\mathsf{eval}}(n) = \mathcal{O}(n^{B+2})$.

Proof of the consequence. Finally, note that conditions (36) and (38) are equivalent to

$$N \geq C_B \log^{4(B+1)}(C_B/\epsilon) \quad \text{and} \quad \tau^2 \geq C_B \cdot \log^{8(B+1)}(C_B/\epsilon) + 4 \sup_{\theta\in\Theta} \theta^2,$$

where C_B is a constant that depends only on B. Applying Theorem 2 with the guarantees above yields the desired result. □

7.4 Proof of Corollary 4

We verify each part separately.

Verifying Assumption 3. Let $\widetilde{C}_f = 2R$. Since $f \in \mathcal{F}(R)$, then we obtain for all $x \in \mathbb{R}$,

$$\sum_{j=1}^{n} \frac{\left|f^{(j)}\left((\theta + \sqrt{2}x)/\tau\right)\right|}{(\widetilde{C}_f)^j j!} \leq \sum_{j=1}^{n} \frac{1}{2^j} \leq 1.$$

Consequently, Assumption 3 holds with $\widetilde{C}_f = 2R$ and $C_f = 0$.

Bounding the quantities $L_f(n)$ and $\widetilde{L}_f(n,\tau)$. We obtain by definition (35),

$$L_f(n) \leq \sum_{j=1}^{n} \frac{1}{2^j} \leq 1 \quad \text{and} \quad \widetilde{L}_f(n,\tau) \leq 1 \quad \text{for all } n \in \mathbb{N}, \tau > 0.$$

Bounding the evaluation time $T_{\mathsf{eval}}(N)$. We first derive the computational time of evaluating $\nabla_x^{(n)} v(y;x)$. Using Faà di Bruno's formula yields

$$\nabla_x^{(n)} v(y;x) = n! \sum_{m_1,\ldots,m_n} \frac{(-1)^{m_1+\cdots+m_n}}{m_1!\, m_2! \cdots m_n!} \cdot H_{m_1+\cdots+m_n}(g(x))\phi(g(x)) \cdot \prod_{j=1}^{n} \left(\frac{f^{(j)}(x/\tau)}{\tau^j j!} \right)^{m_j},$$

where the summation is over all nonnegative integers $(m_1, \ldots, m_n)$ satisfying $\sum_{j=1}^n j m_j = n$. Let $P(n)$ denote the number of n-tuples of nonnegative integers $(m_1, \cdots, m_n)$ satisfying $\sum_{j=1}^n j m_j = n$. Note that $P(n)$ corresponds to the partition function and asymptotically satisfies

$$P(n) \sim \frac{1}{4n\sqrt{3}} \exp\left(\pi\sqrt{2n/3}\right) \quad \text{as} \quad n \uparrow +\infty.$$

Thus, there are at most $\mathcal{O}\big(\exp\left(\pi\sqrt{n}\right)\big)$ terms in the above summation, and each term in the summation can be computed in time $\mathcal{O}(\mathsf{poly}(n)\exp\left(C\sqrt{n}\right))$ provided we can evaluate $f^{(n)}(x)$ in time $\mathcal{O}\big(\exp\left(C\sqrt{n}\right)\big)$, which is in turn true by assumption. Thus we can evaluate $\mathcal{S}_N^\star(y \mid x)$ (14) in time

$$\mathcal{O}\Big(\mathsf{poly}(N)\exp\left((C+\pi)\sqrt{N}\right)\Big) = \mathcal{O}\big(\mathsf{poly}(1/\epsilon)\big),$$

where in the last step we use the following logic: Since $C_f = 0$, it suffices to take $N = \mathcal{O}(\log^2(1/\epsilon))$. Thus N satisfies conditions (36) and (38).

Proof of the consequence. Finally, note that conditions (36) and (38) are equivalent to

$$N \geq \widetilde{C}\log^2(\widetilde{C}/\epsilon) \quad \text{and} \quad \tau^2 \geq \widetilde{C}R^2 \cdot \log^2(\widetilde{C}/\epsilon) + 4\sup_{\theta\in\Theta}\theta^2,$$

where $\widetilde{C}$ is a universal constant. Applying Theorem 2 with the guarantees above yields the desired result. □

7.5 Proofs of claims in Example 2

We perform derivations for the different functions separately.

Trigonometric links. Consider $f(\theta) = \cos(a\theta + b)$. Clearly, we have $|f^{(n)}(x)| \leq |a|^n$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$. Thus by letting $R = 2|a|$, we have

$$\sum_{j=1}^{n} \frac{|f^{(j)}(x)|}{R^j j!} \leq \sum_{j=1}^{n} \frac{1}{2^j} \leq 1.$$

Consequently, $f \in \mathcal{F}(2|a|)$. The same holds when $f(\theta) = \sin(a\theta + b)$. This concludes the proof.

Function $f(\theta) = \mathsf{sech}(\theta) = \frac{2}{e^\theta + e^{-\theta}}$. We let $g(x) = \frac{2}{x}$ and $h(x) = e^x + e^{-x}$. Note that $f(x) = g(h(x))$. Applying Faà di Bruno's formula yields

$$\begin{aligned} f^{(n)}(x) &= \sum_{m_1,\ldots,m_n} \frac{n!}{m_1!\, m_2! \cdots m_n!} g^{(m_1+\cdots+m_n)}(h(x)) \prod_{j=1}^{n} \left(\frac{h^{(j)}(x)}{j!} \right)^{m_j} \\ &= n! \sum_{m_1,\ldots,m_n} \frac{(-1)^{m_1+\cdots+m_n}}{m_1!\, m_2! \cdots m_n!} \cdot \frac{2(m_1+\cdots+m_n)!}{h(x)^{m_1+\cdots+m_n}} \prod_{j=1}^{n} \left(\frac{h^{(j)}(x)}{j!} \right)^{m_j}. \end{aligned} \tag{132}$$

Note that by definition $|h^{(j)}(x)| \le h(x)$ for all $x \in \mathbb{R}$. Consequently, we obtain

$$\frac{(m_1+\cdots+m_n)!}{m_1!\,m_2!\cdots m_n!}\cdot\frac{1}{h(x)^{m_1+\cdots+m_n}}\prod_{j=1}^{n}\left(\frac{|h^{(j)}(x)|}{j!}\right)^{m_j}$$
$$\le \frac{(m_1+\cdots+m_n)!}{m_1!\,m_2!\cdots m_n!}\prod_{j=1}^{n}\frac{1}{j!} \overset{\text{(i)}}{\le} \left(\sum_{j=1}^{n}\frac{1}{j!}\right)^{m_1+\cdots+m_n} \le 2^n,$$

where in step (i) we use the multinomial theorem and in the last step we use $m_1+\cdots+m_n \le n$. Putting the pieces together and applying the triangle inequality yields for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$,

$$|f^{(n)}(x)| \le 2n! \sum_{m_1,\ldots,m_n} 2^n \le 2\cdot n!\cdot(2e)^n,$$

where in the last step we use that there are at most e^n terms in the summation. Consequently, by letting $R = 8e$, we obtain

$$\sum_{j=1}^{n}\frac{|f^{(j)}(x)|}{R^j j!} \le \sum_{j=1}^{n} 2^{-j} \le 1.$$

Note that one can evaluate $f^{(n)}(x)$ using Eq. (132), where one need to sum over at most $\mathcal{O}\Big(\exp\Big(\pi\sqrt{2n/3}\Big)\Big)$ terms and each term in the summation can be computed in time $\mathcal{O}(\mathsf{poly}(n))$. Thus, the evaluation time of $f^{(n)}(x)$ can be bounded by

$$\mathcal{O}\Big(\mathsf{poly}(n)\cdot\exp\Big(\pi\sqrt{2n/3}\Big)\Big) = \mathcal{O}\Big(\exp\Big(\pi\sqrt{n}\Big)\Big).$$

Other functions. For $f(\theta) = (1+\theta^2)^{-1}$ and $f(\theta) = (1+e^{-\theta})^{-1}$. we can follow the identical steps as above to show $f \in \mathcal{F}(8e)$. For $f(\theta) = \exp(-\theta^2)$, we obtain

$$f^{(n)}(x) = (-1)^n H_n(x)\exp(-x^2).$$

By using the properties of Hermite polynomials (62a), it is straightforward to show $f \in \mathcal{F}(8e)$. We omit the details for brevity. □

8 Proofs of results from Section 3

In this section, we prove all the results related to the applications presented in this paper.

8.1 Proof of Proposition 2

We first state a result which directly follows by combining Theorem 10 and Lemma 102 of Brennan and Bresler (2020).

Proposition 4. *Let $s \ge 3$ be a constant. Let*

$$\widetilde{T} = \beta(\sqrt{n}\widetilde{v})^{\otimes s} + \widetilde{\zeta}, \quad \text{where } \widetilde{v} \sim \mathsf{Unif}\big(\{\tfrac{-1}{\sqrt{n}},\tfrac{1}{\sqrt{n}}\}^n\big) \quad \text{and} \quad \widetilde{\zeta} \sim \mathcal{N}(0,1)^{\otimes n^{\otimes s}}. \tag{133}$$

Suppose the k-HPC^s conjecture holds. If $\beta = \tilde{o}\big(n^{-s/4}\big)$ and $\beta = \omega\big(n^{-s/2}\sqrt{s\log(n)}\big)$, then there exists no *algorithm $\mathcal{A}' : \mathbb{R}^{n^{\otimes s}} \to \mathbb{S}^{d-1}$ that runs in time $\mathsf{poly}(n)$ and satisfies*

$$\big\langle \mathcal{A}'\big(\widetilde{T}\big), \widetilde{v}\big\rangle = \Omega(1) \quad \textit{with probability at least} \quad \Omega_n(1)$$

We next present an auxiliary lemma establishing that there exists a polynomial-time reduction that maps $\widetilde{T}$ defined in Eq. (133) to the tensor T defined in Eq. (43) and achieves zero TV deficiency.

Lemma 8. *Let $\widetilde{T}$ denote the tensor defined in Eq. (133) with signal strength $\beta \in \mathbb{R}$ and spike $\sqrt{n}\widetilde{v}$. Then there exists a reduction $\mathcal{R} : \mathbb{R}^{n^{\otimes s}} \to \mathbb{R}^{n^{\otimes s}}$ that runs in time $\mathsf{poly}(n)$ and satisfies*

$$\mathcal{R}(\widetilde{T}) = \beta\big(\sqrt{n}U\widetilde{v}\big)^{\otimes s} + \zeta,$$

where U is an $n \times n$ unitary matrix chosen uniformly at random and ζ is as defined in Eq. (42).

We provide the proof Lemma 8 in Appendix B.5. Note that $U\widetilde{v} \sim \mathsf{Unif}\big(\mathbb{S}^{n-1}\big)$ and thus $\mathcal{R}\big(\widetilde{T}\big)$ is equal in distribution to the tensor T defined in Eq. (43).

We now prove Proposition 2 by combining Proposition 4 with Lemma 8. We prove it by contradiction. Let T be defined as in Eq. (43) and assume that there is an algorithm $\mathcal{A} : \mathbb{R}^{n^{\otimes s}} \to \mathbb{S}^{d-1}$ that runs in time $\mathsf{poly}(n)$ and satisfies

$$\big\langle \mathcal{A}\big(T\big), v\big\rangle = \Omega(1) \quad \text{with probability at least} \quad \Omega_n(1).$$

Then the algorithm $\mathcal{A} \circ \mathcal{R} : \mathbb{R}^{n^{\otimes s}} \to \mathbb{S}^{d-1}$ runs in time $\mathsf{poly}(n)$ and satisfies

$$\big\langle \mathcal{A} \circ \mathcal{R}\big(\widetilde{T}\big), U\widetilde{v}\big\rangle = \Omega(1) \quad \text{with probability at least} \quad \Omega_n(1).$$

This implies that

$$\big\langle U^\top \mathcal{A} \circ \mathcal{R}\big(\widetilde{T}\big), \widetilde{v}\big\rangle = \Omega(1) \quad \text{with probability at least} \quad \Omega_n(1).$$

In words, the algorithm $U^\top \mathcal{A} \circ \mathcal{R} : \mathbb{R}^{n^{\otimes s}} \to \mathbb{S}^{d-1}$ runs in time $\mathsf{poly}(n)$ and produces accurate recovery. However, this violates Proposition 4. Consequently, our assumption that algorithm $\mathcal{A}$ runs in polynomial time must be false, thereby proving Proposition 2. □

8.2 Proof of Theorem 3

We partition the set of all index tuples

$$\big\{(i_1, \dots, i_s) \;\big|\; i_k \in [n],\ \forall\, k \in [s]\big\}$$

into disjoint subsets $S_1, S_2, \dots, S_L$, where each subset S_ℓ consists of index tuples that are equivalent up to permutation—that is, any two tuples in S_k are permutations of each other. Let $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ be the reduction in Theorem 1. We design the tensor-to-tensor reduction algorithm $\mathcal{R} : \mathbb{R}^{n^{\otimes s}} \to \mathbb{R}^{n^{\otimes s}}$ as follows.

Input: $T \in \mathbb{R}^{n^{\otimes s}}$. For each $\ell = 1, 2, \dots, L$:

1. Pick one index tuple $\big(i_1^\ell, \dots, i_s^\ell\big) \in S_\ell$. Let $\mathcal{R}(T)_{i_1^\ell, \dots, i_s^\ell} = \mathsf{K}\big(T_{i_1^\ell, \dots, i_s^\ell}\big)$.
2. For all other index tuples $(i_1', \dots, i_s') \in S_\ell \setminus \big\{\big(i_1^\ell, \dots, i_s^\ell\big)\big\}$, set $\mathcal{R}(T)_{(i_1', \dots, i_s')} = \mathcal{R}(T)_{i_1^\ell, \dots, i_s^\ell}$ to ensure the output is a symmetric tensor.

We next bound the TV distance between $\mathcal{R}(T(\beta, v))$ and $\widetilde{T}(\beta, v)$, where we use $T(\beta, v)$ to denote the tensor from model (41) when β and v are the fixed parameters, and likewise $\widetilde{T}(\beta, v)$ is the tensor from model (45). Recall the density of $\mathcal{Q}_0$ in Eq. (46) and the density of $\mathcal{Q}_\theta$ in Eq. (20). Conditional on β and v, we have by definition for all $\ell \in [L]$ that

$$T(\beta, v)_{i_1^\ell, \dots, i_s^\ell} \sim \mathcal{N}(\theta, \sigma^2_{i_1, \dots, i_s}) \quad \text{and} \quad \widetilde{T}(\beta, v)_{i_1^\ell, \dots, i_s^\ell} \sim \mathcal{Q}_\theta, \quad \text{where } \theta = \beta n^{s/2} \times v_{i_1^\ell} \times v_{i_2^\ell} \times \cdots \times v_{i_s^\ell}.$$

Note that $\sigma_{i_1,\dots,i_s}$ is a constant, since s is a fixed parameter and $1/s! \le \sigma^2_{i_1,\dots,i_s} \le 1$. Using Assumption 4 and Theorem 1, we then obtain our reduction. In particular, for all $\epsilon \in (0,1)$, if $\tau = \Theta\big(\mathsf{polylog}(1/\epsilon)\big)$ then setting $T_{\mathsf{iter}} = \Theta\big(\log(1/\epsilon)\big)$ yields, for all $\ell \in [L]$, that

$$\sup_{\beta\in\mathbb{R}, v\in\mathbb{R}^n} \mathsf{d}_{\mathsf{TV}}\Big(\mathsf{K}\big(T(\beta,v)_{i_1^\ell,\dots,i_s^\ell}\big), \widetilde{T}(\beta,v)_{i_1^\ell,\dots,i_s^\ell}\Big) \le \epsilon.$$

Applying the triangle inequality for the TV distance (or equivalently, a union bound) (Brennan et al., 2018, Lemma 6), we obtain

$$\begin{aligned}
\mathsf{d}_{\mathsf{TV}}\Big(\mathcal{R}\big(T(\beta,v)\big), \widetilde{T}(\beta,v)\Big) &\le \sum_{\ell=1}^{L} \mathsf{d}_{\mathsf{TV}}\Big(\mathcal{R}\big(T(\beta,v)\big)_{i_1^\ell,\dots,i_s^\ell}, \widetilde{T}(\beta,v)_{i_1^\ell,\dots,i_s^\ell}\Big) \\
&= \sum_{\ell=1}^{L} \mathsf{d}_{\mathsf{TV}}\Big(\mathsf{K}\big(T(\beta,v)_{i_1^\ell,\dots,i_s^\ell}\big), \widetilde{T}(\beta,v)_{i_1^\ell,\dots,i_s^\ell}\Big) \\
&\le L \cdot \epsilon \le n^s \cdot \epsilon.
\end{aligned}$$

For each $\delta \in (0,1)$, let $\epsilon = \delta/n^s$. Then taking the supremum over $\beta \in \mathbb{R}$ and $v \in \mathbb{R}^n$ of the inequality in the display above completes the proof of Theorem 3. Note that from Assumption 4 and Theorem 1, the reduction algorithm $\mathcal{R}$ runs in time

$$\mathcal{O}\Big(n^s \cdot T_{\mathsf{samp}} \cdot T_{\mathsf{iter}} \cdot T_{\mathsf{eval}}(N)\Big) = \mathcal{O}\Big(\mathsf{poly}\big(n^s/\delta\big)\Big),$$

where T_{samp} is independent of δ and n. □

8.3 Proof of Corollary 5

Let T be a tensor defined as in Eq. (43) and let $\widetilde{T}$ be the tensor defined in Theorem 5. Note that T and $\widetilde{T}$ both have the same signal strength $\beta \in \mathbb{R}$ and spike $v \in \mathbb{S}^{n-1}$. Let $\mathcal{R}$ be the reduction defined in Proposition 3 with TV deficiency $\delta = n^{-1}$. By Proposition 3, we have that $\mathcal{R}$ runs in time $\mathsf{poly}(n)$ and satisfies $\mathsf{d}_{\mathsf{TV}}\big(\mathcal{R}(T), \widetilde{T}\big) \le n^{-1}$. We now prove Corollary 5 by contradiction. Assume that there exists an algorithm $\mathcal{A} : \mathbb{R}^{n^{\otimes s}} \to \mathbb{S}^{n-1}$ that runs in time $\mathsf{poly}(n)$ and satisfies

$$\big\langle \mathcal{A}\big(\widetilde{T}\big), v\big\rangle = \Omega(1) \quad \text{with probability at least} \quad \Omega_n(1).$$

Then $\mathcal{A} \circ \mathcal{R} : \mathbb{R}^{n^{\otimes s}} \to \mathbb{S}^{n-1}$ also runs in time $\mathsf{poly}(n)$ and satisfies

$$\big\langle \mathcal{A} \circ \mathcal{R}\big(T\big), v\big\rangle = \Omega(1) \quad \text{with probability at least} \quad \Omega_n(1) - n^{-1}.$$

However, this contradicts Proposition 2. Thus, our assumption that the algorithm $\mathcal{A}$ runs in polynomial time cannot hold, thereby proving Corollary 5. □

8.4 Proof of Theorem 4

Let $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ be the reduction in Theorem 2. We design the reduction algorithm $\mathcal{R} : (\mathbb{R}^d \times \mathbb{R})^n \to (\mathbb{R}^d \times \mathbb{R})^n$ as follows: For inputs $\{(x_i, y_i)\}_{i=1}^n$ from mixtures of linear regressions, return $\mathcal{R}\big(\{(x_i,y_i)\}_{i=1}^n\big) := \{(x_i, \mathsf{K}(y_i))\}_{i=1}^n$. Note that by definition, conditioned on x_i and R_i,

$$y_i \sim \mathcal{N}(\theta, 1) \quad \text{and} \quad \widetilde{y}_i \sim \mathcal{N}(f(\theta/\tau), 1) \quad \text{where } \theta = R_i \cdot \langle x_i, v\rangle.$$

Consequently, applying Assumption 5 and Theorem 2, we obtain that for each $\epsilon \in (0,1)$,

$$\text{if} \quad \tau = \mathsf{polylog}(1/\epsilon) + 2\max_{i\in[n]} |\langle x_i, v\rangle|, \quad \text{then} \quad \mathsf{d_{TV}}\big(\mathsf{K}(y_i), \widetilde{y}_i\big) \le \epsilon.$$

Applying the triangle inequality for the TV distance (Brennan et al., 2018, Lemma 6), we obtain

$$\begin{aligned}\mathsf{d_{TV}}\Big(\mathcal{R}\big(\{(x_i,y_i)\}_{i=1}^n\big), \{(x_i,\widetilde{y}_i)\}_{i=1}^n\Big) &= \mathsf{d_{TV}}\Big(\big\{\big(x_i, \mathsf{K}(y_i)\big)\big\}_{i=1}^n, \{(x_i,\widetilde{y}_i)\}_{i=1}^n\Big) \\ &\le \sum_{i=1}^n \mathsf{d_{TV}}\big(\mathsf{K}(y_i), \widetilde{y}_i\big) \le n\epsilon.\end{aligned}$$

The desired TV guarantee follows by letting $\epsilon = \delta/n$. From Assumption 5 and Theorem 2, the total time taken to compute $\mathcal{R}$ is

$$\mathcal{O}\Big(n \cdot \log(4n/\delta) \cdot T_{\mathsf{eval}}(N)\Big) = \mathcal{O}\Big(\mathsf{poly}(n/\delta)\Big),$$

as claimed. □

8.5 Proof of Corollary 6

We define the following event

$$\mathcal{E} = \Big\{ \max_{i\in[n]} \big|\langle x_i, v\rangle\big| \le 10\log^2(n)\Big\}.$$

Note that $\langle x_i, v\rangle \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\eta^2)$ since $\|v\|_2 = \eta$ and $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d)$. Moreover, by the condition that $k = o(n^{1/6})$ and $n = \tilde{o}(k^2/\eta^4)$, we obtain $\eta = \tilde{o}(k^{-1}) \le 1$. Consequently, applying a standard bound on Gaussian maxima yields

$$\mathbb{P}\big[\mathcal{E}\big] \ge 1 - n^{-3}. \tag{134}$$

By setting $\tau = \mathsf{polylog}(n/\delta)$ for any $\delta \in (0,1)$, the reduction $\mathcal{R}$ in Proposition 4 runs in time $\mathsf{poly}(n/\delta)$. The TV-deficiency guarantee in Proposition 4 requires condition (51), which holds with high probability according to Ineq. (134). We now combine the high probability guarantee (134) with Proposition 4 to derive the desired TV-deficiency guarantee.

Let $\mathsf{K} : \mathbb{R} \to \mathbb{R}$ be the reduction in Theorem 2 and note that

$$y_i \mid x_i \sim \mathcal{N}\big(R_i \cdot \langle x_i, v\rangle, 1\big) \quad \text{and} \quad \widetilde{y}_i \mid x_i \sim \mathcal{N}\big(f\big(\langle x_i, v\rangle/\tau\big), 1\big).$$

To reduce the notational burden, we let $\rho : \mathbb{R}^d \to \mathbb{R}$ denote the density function of $x_i \sim \mathcal{N}(0, I_d)$, let $\gamma(\cdot; x), \widetilde{\gamma}(\cdot; x) : \mathbb{R} \to \mathbb{R}$ denote the conditional densities of $\mathsf{K}(y_i) \mid x_i = x$ and $\widetilde{y}_i \mid x_i = x$, respectively. Then we have

$$\begin{aligned}&\mathsf{d_{TV}}\left(\mathcal{R}\big(\{(x_i,y_i)\}_{i=1}^n\big), \{(x_i,\widetilde{y}_i)\}_{i=1}^n\right) \\ &\overset{\text{(i)}}{\le} n \cdot \mathsf{d_{TV}}\big((x_i, \mathsf{K}(y_i)), (x_i, \widetilde{y}_i)\big) \\ &= \frac{n}{2}\int_{\mathbb{R}^d} \rho(x) \cdot \int_{\mathbb{R}} \Big|\gamma(t;x) - \widetilde{\gamma}(t;x)\Big| \mathrm{d}t\,\mathrm{d}x \\ &= \frac{n}{2}\int_{\mathcal{E}} \rho(x) \cdot \int_{\mathbb{R}} \Big|\gamma(t;x) - \widetilde{\gamma}(t;x)\Big| \mathrm{d}t\,\mathrm{d}x + \frac{n}{2}\int_{\mathcal{E}^{\complement}} \rho(x) \cdot \int_{\mathbb{R}} \Big|\gamma(t;x) - \widetilde{\gamma}(t;x)\Big| \mathrm{d}t\,\mathrm{d}x,\end{aligned}$$

where step (i) follows since the samples are i.i.d. Towards bounding the two terms in the display above, we obtain

$$\frac{n}{2}\int_{\mathcal{E}}\rho(x)\cdot\int_{\mathbb{R}}\Big|\gamma(t;x)-\widetilde{\gamma}(t;x)\Big|\mathrm{d}t\,\mathrm{d}x\le n\cdot\int_{\mathcal{E}}\rho(x)\mathrm{d}x\cdot\frac{\delta}{n}\le\delta,$$

where we use the guarantee that on event $\mathcal{E}$, $\mathsf{d_{TV}}(\mathsf{K}(y_i),\widetilde{y}_i)\le\delta/n$. Moreover, we obtain

$$\frac{n}{2}\int_{\mathcal{E}^{\complement}}\rho(x)\cdot\int_{\mathbb{R}}\Big|\gamma(t;x)-\widetilde{\gamma}(t;x)\Big|\mathrm{d}t\,\mathrm{d}x\le n\cdot\int_{\mathcal{E}^{\complement}}\rho(x)\mathrm{d}x\le n\cdot\mathbb{P}\Big[\mathcal{E}^{\complement}\Big]\le n^{-2}.$$

Putting the three pieces together and setting $\delta=n^{-1}$ yields

$$\mathsf{d_{TV}}\Big(\mathcal{R}\big(\{(x_i,y_i)\}_{i=1}^n\big),\{(x_i,\widetilde{y}_i)\}_{i=1}^n\Big)\le n^{-1}+n^{-2}\le 2n^{-1}. \tag{135}$$

For the sake of contradiction, we assume that there exists an algorithm $\mathcal{A}:(\mathbb{R}^d\times\mathbb{R})^n\to\mathbb{R}^d$ that runs in time $\mathsf{poly}(n)$ and for sample size $n=\tilde{o}\big(k^2/\eta^4\big)$, satisfies

$$\big\|\mathcal{A}\big(\{x_i,\widetilde{y}_i\}_{i=1}^n\big)-v\big\|_2=o(\eta)\quad\text{with probability at least}\quad 1-o_n(1). \tag{136}$$

Then we note that $\mathcal{A}\circ\mathcal{R}:(\mathbb{R}^d\times\mathbb{R})^n\to\mathbb{R}^d$ also runs in polynomial time and it is an estimation algorithm for data $\{x_i,y_i\}_{i=1}^n$ from MixSLR (47). Moreover, the TV-deficiency guarantee (135) and the guarantee of $\mathcal{A}$ (136) imply

$$\big\|\mathcal{A}\circ\mathcal{R}\big(\{x_i,y_i\}_{i=1}^n\big)-v\big\|_2=o(\eta)\quad\text{with probability at least}\quad 1-o_n(1)-2n^{-1}=1-o_n(1).$$

This contradicts the computational hardness claimed in Proposition 3. Thus, our assumption that algorithm $\mathcal{A}$ runs in polynomial time must be false, thereby proving Corollary 6. □

8.6 Proof of Theorem 5

We construct the reduction $\mathcal{R}:\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n}$ using the scalar reduction $\mathsf{K}:\mathbb{R}\to\mathbb{R}$ developed in Corollary 3 with the parameter setting $B=2$.

Input: A matrix $M=\mu rc^\top+\mathcal{N}(0,1)^{\otimes n\times n}$ for $r,c\in S_{n,k}$.
Procedure: Construct a new matrix $\widetilde{M}\in\mathbb{R}^{n\times n}$: For all $i,j\in[n]$, set $\widetilde{M}_{ij}=\mathsf{K}(M_{ij})$.
Output: The reduced matrix $\mathcal{R}(M):=\widetilde{M}$.

We next bound the TV deficiency of the reduction $\mathcal{R}$. For notational convenience, we let $f:\mathbb{R}\to\mathbb{R}$ be the square function, i.e., $f(t)=t^2$ for all $t\in\mathbb{R}$ and let

$$\overline{M}:=\frac{\mu^2}{\tau^2k}|r||c|^\top+\mathcal{N}(0,1)^{\otimes n\times n},$$

where $|r|,|c|\in\mathbb{R}^n$ denotes the vectors obtained by taking the entrywise absolute value of r and c, respectively. Note that for all $r,c\in S_{n,k}$, we have

$$r_i^2c_j^2=\frac{|r_i|\cdot|c_j|}{k}\quad\text{for all}\quad i,j\in[n].$$

Then observe that for all $i,j\in[n]$,

$$M_{ij}\sim\mathcal{N}(\mu r_ic_j,1)\quad\text{and}\quad\overline{M}_{ij}\sim\mathcal{N}\Big(f\Big(\frac{\mu r_ic_j}{\tau}\Big),1\Big).$$

Consequently, since $\left(\mu r_i c_j\right)^2 \leq \mu^2/k^2$, applying Corollary 3 with $\epsilon = \delta/n^2$ for any $\delta \in (0,1)$ and using the condition on τ (61) yield the TV guarantee

$$\mathsf{d}_{\mathsf{TV}}\Big(\mathsf{K}(M_{ij}), \overline{M}_{ij}\Big) \leq \frac{\delta}{n^2} \quad \text{for all } i,j \in [n].$$

Putting the pieces together yields

$$\mathsf{d}_{\mathsf{TV}}\Big(\mathcal{R}(M), \overline{M}\Big) \leq \sum_{i=1}^{n}\sum_{j=1}^{n} \mathsf{d}_{\mathsf{TV}}\Big(\mathsf{K}(M_{ij}), \overline{M}_{ij}\Big) \leq \delta.$$

Note that the bound in the display is independent of $r, c \in S_{n,k}$ and $\mu \in \mathbb{R}$. Taking the supremum over $r, c \in S_{n,k}$ and $\mu \in \mathbb{R}$ on both sides yields the desired TV deficiency guarantee. The computational complexity of the reduction $\mathcal{R}$ follows from the computational complexity of the scalar reduction K in Corollary 3, and we use this reduction n^2 times in total. □

References

A. Andoni, D. Hsu, K. Shi, and X. Sun. Correspondence retrieval. In *Conference on Learning Theory*, pages 105–126. PMLR, 2017.

G. Arpino and R. Venkataramanan. Statistical-computational tradeoffs in mixed sparse linear regression. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 921–986. PMLR, 2023.

B. Axelrod, S. Garg, Y. Han, V. Sharan, and G. Valiant. On the statistical complexity of sample amplification. *The Annals of Statistics*, 52(6):2767–2790, 2024.

A. Bandeira, A. Perry, and A. S. Wein. Notes on computational-to-statistical gaps: predictions using statistical physics. *Portugaliae mathematica*, 75(2):159–186, 2018.

K. Bangachev, G. Bresler, S. Tiegel, and V. Vaikuntanathan. Near-optimal time-sparsity trade-offs for solving noisy linear equations. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1910–1920, 2025.

J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12): 5451–5460, 2019.

G. Ben Arous, S. Mei, A. Montanari, and M. Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330, 2019.

G. Ben Arous, R. Gheissari, and A. Jagannath. Algorithmic thresholds for tensor PCA. *The Annals of Probability*, 48(4):2052–2087, 2020.

G. Ben Arous, A. S. Wein, and I. Zadik. Free energy wells and overlap gap property in sparse PCA. *Communications on Pure and Applied Mathematics*, 76(10):2410–2473, 2023.

Q. Berthet and P. Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on learning theory*, pages 1046–1066. PMLR, 2013a.

Q. Berthet and P. Rigollet. Optimal detection of sparse principal components in high dimension. *The Annals of Statistics*, 41(4):1780 – 1815, 2013b. doi: 10.1214/13-AOS1127. URL https://doi.org/10.1214/13-AOS1127.

D. Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 93–103. University of California Press, 1951.

D. Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, pages 265–272, 1953.

M. Brennan and G. Bresler. Optimal average-case reductions to sparse PCA: From weak assumptions to strong hardness. In *Conference on Learning Theory*, pages 469–470. PMLR, 2019.

M. Brennan and G. Bresler. Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory*, pages 648–847. PMLR, 2020.

M. Brennan, G. Bresler, and W. Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166. PMLR, 2018.

M. Brennan, G. Bresler, and W. Huleihel. Universality of computational lower bounds for submatrix detection. In *Conference on Learning Theory*, pages 417–468. PMLR, 2019.

M. Brennan, G. Bresler, S. Hopkins, J. Li, and T. Schramm. Statistical query algorithms and low degree tests are almost equivalent. In *Conference on Learning Theory*, pages 774–774. PMLR, 2021.

G. Bresler and A. Harbuzova. Computational equivalence of spiked covariance and spiked wigner models via gram-schmidt perturbation. *arXiv preprint arXiv:2503.02802*, 2025.

J. Bruna, O. Regev, M. J. Song, and Y. Tang. Continuous LWE. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 694–707, 2021.

T. Cai, X. Li, and Z. Ma. Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow. *The Annals of Statistics*, 44(82):2221–2251, 2016.

M. Celentano, A. Montanari, and Y. Wu. The estimation error of general first order methods. In *Conference on Learning Theory*, pages 1078–1141. PMLR, 2020.

W.-K. Chen. Phase transition in the spiked random tensor with Rademacher prior. *The Annals of Statistics*, 47(5):2734–2756, 2019.

Y. Chen and J. Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *Journal of Machine Learning Research*, 17(27):1–57, 2016.

A. Damian, L. Pillaud-Vivien, J. D. Lee, and J. Bruna. Computational-statistical gaps in Gaussian single-index models. *arXiv preprint arXiv:2403.05529*, 2024.

I. Diakonikolas and D. M. Kane. *Algorithmic high-dimensional robust statistics*. Cambridge university press, 2023.

I. Diakonikolas, D. M. Kane, and A. Stewart. Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–84. IEEE, 2017.

Y. Ding, D. Kunisky, A. S. Wein, and A. S. Bandeira. Subexponential-time algorithms for sparse PCA. *Foundations of Computational Mathematics*, 24(3):865–914, 2024.

R. Dudeja and D. Hsu. Statistical query lower bounds for tensor pca. *Journal of Machine Learning Research*, 22(83):1–51, 2021.

R. Dudeja and D. Hsu. Statistical-computational trade-offs in tensor pca and related problems via communication complexity. *The Annals of Statistics*, 52(1):131–156, 2024.

J. Fan, H. Liu, Z. Wang, and Z. Yang. Curse of heterogeneity: Computational barriers in sparse mixture models and phase retrieval. *arXiv preprint arXiv:1808.06996*, 2018.

D. Gamarnik, A. Jagannath, and S. Sen. The overlap gap property in principal submatrix recovery. *Probability Theory and Related Fields*, 181(4):757–814, 2021.

L. Goldstein and G. Reinert. Distributional transformations, orthogonal polynomials, and Stein characterizations. *Journal of Theoretical Probability*, 18(1):237–260, 2005.

A. Gupte, N. Vafa, and V. Vaikuntanathan. Continuous LWE is as hard as LWE and applications to learning Gaussian mixtures. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1162–1173. IEEE, 2022.

A. Gupte, N. Vafa, and V. Vaikuntanathan. Sparse linear regression and lattice problems. In *Theory of Cryptography Conference*, pages 276–307. Springer, 2024.

B. Hajek, Y. Wu, and J. Xu. Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory*, pages 899–928. PMLR, 2015.

B. Hajek, Y. Wu, and J. Xu. Submatrix localization via message passing. *Journal of Machine Learning Research*, 18(186):1–52, 2018.

R. Han, Y. Luo, M. Wang, and A. R. Zhang. Exact clustering in tensor block model: Statistical optimality and computational limit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5): 1666–1698, 2022.

O. H. Hansen and E. N. Torgersen. Comparison of linear normal experiments. *The Annals of Statistics*, pages 367–373, 1974.

S. Hopkins. *Statistical inference and the sum of squares method*. PhD thesis, Cornell University, 2018.

S. B. Hopkins, J. Shi, and D. Steurer. Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory*, pages 956–1006. PMLR, 2015.

S. B. Hopkins, T. Schramm, J. Shi, and D. Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 178–191, 2016.

S. B. Hopkins, P. K. Kothari, A. Potechin, P. Raghavendra, T. Schramm, and D. Steurer. The power of sum-of-squares for detecting hidden structures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2017.

A. Jagannath, P. Lopatto, and L. Miolane. Statistical thresholds for tensor PCA. *The Annals of Applied Probability*, 30(4):1910–1933, 2020.

S. Karlin and H. Rubin. The theory of decision procedures for distributions with monotone likelihood ratio. *The Annals of Mathematical Statistics*, pages 272–299, 1956.

M. Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6): 983–1006, 1998.

C. Kim, A. S. Bandeira, and M. X. Goemans. Community detection in hypergraphs, spiked tensor models, and sum-of-squares. In *2017 international conference on sampling theory and applications (sampta)*, pages 124–128. IEEE, 2017.

D. Kunisky. Low coordinate degree algorithms I: Universality of computational thresholds for hypothesis testing. *The Annals of Statistics*, 53(2):774–801, 2025.

D. Kunisky, A. S. Wein, and A. S. Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019.

G. Lecué and S. Mendelson. Minimax rate of convergence and the performance of empirical risk minimization in phase retrieval. *Electron. J. Probab*, 20(57):1–29, 2015.

E. L. Lehmann. Comparing Location Experiments. *The Annals of Statistics*, 16(2):521 – 533, 1988.

T. Lesieur, F. Krzakala, and L. Zdeborová. Phase transitions in sparse PCA. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1635–1639. IEEE, 2015.

T. Lesieur, F. Krzakala, and L. Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017a.

T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 ieee international symposium on information theory (isit)*, pages 511–515. IEEE, 2017b.

S. Li and T. Schramm. Some easy optimization problems have the overlap-gap property. *arXiv preprint arXiv:2411.01836*, 2024.

X. Li and V. Voroninski. Sparse signal recovery from quadratic measurements via convex programming. *SIAM Journal on Mathematical Analysis*, 45(5):3019–3033, 2013.

D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.

M. Lou, G. Bresler, and A. Pananjady. Computationally efficient reductions between some statistical models. *IEEE Transactions on Information Theory*, 71(9):7097–7133, 2025. doi: 10.1109/TIT.2025.3587354.

Y. Luo and A. R. Zhang. Tensor clustering with planted structures: Statistical optimality and computational limits. *The Annals of Statistics*, 50(1):584–613, 2022.

T. Ma and A. Wigderson. Sum-of-squares lower bounds for sparse PCA. *Advances in Neural Information Processing Systems*, 28, 2015.

Z. Ma and Y. Wu. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, 43 (3):1089 – 1116, 2015. doi: 10.1214/14-AOS1300.

N. Macris, C. Rush, et al. All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation. *Advances in Neural Information Processing Systems*, 33:14915–14926, 2020.

A. Montanari and E. Richard. A statistical model for tensor PCA. *Advances in neural information processing systems*, 27, 2014.

A. Montanari, D. Reichman, and O. Zeitouni. On the limitation of spectral methods: From the Gaussian hidden clique problem to rank-one perturbations of Gaussian tensors. *Advances in Neural Information Processing Systems*, 28, 2015.

A. Pananjady and R. J. Samworth. Isotonic regression with unknown permutations: Statistics, computation and adaptation. *The Annals of Statistics*, 50(1):324–350, 2022.

A. Perry, A. S. Wein, and A. S. Bandeira. Statistical limits of spiked tensor models. 2020.

Y. Polyanskiy and Y. Wu. *Information theory: From coding to learning*. Cambridge university press, 2025.

R. E. Quandt and J. B. Ramsey. Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73(364):730–738, 1978.

P. Raghavendra, T. Schramm, and D. Steurer. High dimensional estimation via sum-of-squares proofs. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3389–3423. World Scientific, 2018.

M. Raginsky. Shannon meets Blackwell and Le Cam: Channels, codes, and statistical experiments. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 1220–1224. IEEE, 2011.

V. Ros, G. Ben Arous, G. Biroli, and C. Cammarota. Complex energy landscapes in spiked-tensor and simple glassy models: Ruggedness, arrangements of local minima, and phase transitions. *Physical Review X*, 9(1):011003, 2019.

T. Schramm and A. S. Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.

N. B. Shah, S. Balakrishnan, and M. J. Wainwright. Feeling the Bern: Adaptive estimators for Bernoulli probabilities of pairwise comparisons. *IEEE Transactions on Information Theory*, 65(8):4854–4874, 2019.

A. N. Shiryaev and V. G. Spokoiny. *Statistical Experiments And Decisions, Asymptotic Theory*, volume 8. World Scientific, 2000.

N. Städler, P. Bühlmann, and S. van de Geer. ℓ_1-penalization for mixture regression models. *Test*, 19: 209–256, 2010.

M. Stone. Non-equivalent comparisons of experiments and their use for experiments involving location parameters. *The Annals of Mathematical Statistics*, pages 326–332, 1961.

G. Szego. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939.

E. Torgersen. *Comparison of statistical experiments*, volume 36. Cambridge University Press, 1991.

M. J. Wainwright. Constrained forms of statistical minimax: Computation, communication and privacy. In *Proceedings of the International Congress of Mathematicians*, pages 13–21, 2014.

G. Wang, L. Zhang, G. B. Giannakis, M. Akçakaya, and J. Chen. Sparse phase retrieval via truncated amplitude flow. *IEEE Transactions on Signal Processing*, 66(2):479–491, 2017.

G. Wang, M. Lou, and A. Pananjady. Do algorithms and barriers for sparse principal component analysis extend to other structured settings? *IEEE Transactions on Signal Processing*, 72:3187–3200, 2024.

L. Wang, Z. Yang, and Z. Wang. Statistical-computational tradeoff in single index models. *Advances in neural information processing systems*, 32, 2019.

R. Wang and Z. Zhang. Simultaneous optimal transport. *arXiv preprint arXiv:2201.03483*, 2022.

A. S. Wein. Computational Complexity of Statistics: New Insights from Low-Degree Polynomials. *arXiv preprint arXiv:2506.10748*, 2025.

A. S. Wein, A. El Alaoui, and C. Moore. The Kikuchi hierarchy and tensor PCA. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1446–1468. IEEE, 2019.

I. Zadik, M. J. Song, A. S. Wein, and J. Bruna. Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory*, pages 1247–1248. PMLR, 2022.

Appendix

A Supplementary results

We state the computational hardness conjectures and prove some auxiliary lemmas and claims in this section.

A.1 Explicit statements of hardness conjectures

In this section, we explicitly state the conjectures that are used in Section 3.

Given two positive integers m and n, let $\mathcal{G}(m, n, 1/2)$ denote the distribution on bipartite graphs G with parts of size m and n wherein each edge between the two parts is included independently with probability $1/2$. Let $k_m \in [m]$ and $k_n \in [n]$ such that k_n divides n. Let $E = \{E_1, \ldots, E_{k_n}\}$ be a partition of $[n]$ into k_n parts, each of size n/k_n. We define the distribution

$$k\text{-BPC}_E(m, n, k_m, k_n, 1/2)$$

on bipartite graphs with left side $[m]$ and right side $[n]$ as follows:

1. Sample $G \sim \mathcal{G}(m, n, 1/2)$.
2. Choose a subset $S \subseteq [m]$ of size k_m uniformly at random.
3. Choose a subset $T \subseteq [n]$ of size k_n uniformly at random subject to the constraint
$$|T \cap E_j| = 1 \quad \text{for all } j \in [k_n].$$
4. Plant the complete bipartite graph between S and T in the graph G, i.e. set all edges $\{(u, v) : u \in S, v \in T\}$ to be present.

The resulting distribution on G is denoted $k\text{-BPC}_E(m, n, k_m, k_n, 1/2)$. We define the following two hypotheses:

$$\mathcal{H}_0 : G \sim \mathcal{G}(m, n, 1/2) \qquad \text{and} \qquad \mathcal{H}_1 : G \sim k\text{-BPC}_E(m, n, k_m, k_n, 1/2). \tag{137}$$

We now state the k-BPC conjecture.

Conjecture 1 (k-BPC)**.** *Let m and n be two positive integers and be polynomial in one another. Let $k_m \in [m]$ and $k_n \in [n]$ such that k_n divides n. Let E be any partition of $[n]$ into k_n parts, each of size n/k_n, and E is released. Let $\mathcal{H}_0$ and $\mathcal{H}_1$ be defined in Eq. (137). If*

$$k_m = o(\sqrt{m}) \qquad \text{and} \qquad k_n = o(\sqrt{n}),$$

then there cannot exist an algorithm $\mathcal{A} : \{0,1\}^{m \times n} \to \{0,1\}$ that runs in time $\mathsf{poly}(n)$ and satisfies

$$\mathbb{P}_{\mathcal{H}_0}\Big\{\mathcal{A}(G) = 1\Big\} + \mathbb{P}_{\mathcal{H}_1}\Big\{\mathcal{A}(G) = 0\Big\} = 1 - \Omega_n(1).$$

Let n and s be non-negative integers with $s \geq 3$. Let $\mathcal{G}^s(n, 1/2)$ denote the distribution on s-uniform hypergraphs G with n vertices and each hyperedge is included independently with probability $1/2$. Let $k \in [n]$ and k divides n. Let $E = \{E_1, \ldots, E_k\}$ be a partition of $[n]$ into k parts, each of size n/k. We define the distribution $k\text{-HPC}^s_E(n, k, 1/2)$ as follows:

1. Sample $G \sim \mathcal{G}^s(n, 1/2)$.
2. Choose a subset $S \subseteq [n]$ of size k uniformly at random subject to the constraint
$$|S \cap E_j| = 1 \quad \text{for all } j \in [k].$$
3. Plant the complete s-uniform hypergraph with vertex set S in the graph G, i.e., add all $\binom{k}{s}$ hyperedges within S to G.

The resulting distribution on G is denoted $k\text{-HPC}^s_E(n, k, 1/2)$. We define two hypotheses:

$$\mathcal{H}_0 : G \sim \mathcal{G}^s(n, 1/2) \qquad \text{and} \qquad \mathcal{H}_1 : G \sim k\text{-HPC}^s_E(n, k, 1/2). \tag{138}$$

We now state the k-HPC^s conjecture.

Conjecture 2 (k-HPC)**.** *Let n be a positive integer and let $k \in [n]$ divides n. Let E be any partition of $[n]$ into k parts, each of size n/k, and E is released. Let $\mathcal{H}_0$ and $\mathcal{H}_1$ be defined in Eq. (138). If $k = o(\sqrt{n})$, then there cannot exist an algorithm $\mathcal{A} : \{0,1\}^{n^s} \to \{0,1\}$ that runs in time $\mathsf{poly}(n)$ and satisfies*

$$\mathbb{P}_{\mathcal{H}_0}\Big\{\mathcal{A}(G) = 1\Big\} + \mathbb{P}_{\mathcal{H}_1}\Big\{\mathcal{A}(G) = 0\Big\} = 1 - \Omega_n(1).$$

A.2 Inefficiency of the plug-in approach

We next show that a contraction of the signal-to-noise ratio (equivalently, an inflation of the variance) by a factor of order polynomially in $1/\epsilon$ is necessary for the plug-in approach. To compare with the guarantee of Theorem 2, we consider the source model $\mathcal{U}$ and target model $\mathcal{V}$ defined as

$$\mathcal{U} = (\mathbb{R}, \{\mathcal{N}(\theta, 1)\}_{\theta\in\Theta}) \quad \text{and} \quad \mathcal{V} = \big(\mathbb{R}, \big\{\mathcal{N}\big(\theta^2/\tau^2, 1\big)\big\}_{\theta\in\Theta}\big) \quad \text{with} \quad \Theta = [-1, 1].$$

Given $X_\theta \sim \mathcal{N}(\theta, 1)$, the plug-in approach returns the random variable

$$\widetilde{\mathsf{K}}_{\mathsf{plugin}}(X_\theta) := \frac{X_\theta^2}{\tau^2} + G, \quad G \sim \mathcal{N}(0, 1). \tag{139}$$

In the above, the random variable G is independent of X_θ. Operationally, the plug-in approach uses the single sample X_θ as an estimator of the true mean θ and attempts to approximate the random variable $Y_\theta \sim \mathcal{N}(\theta^2/\tau^2, 1)$.

Lemma 9. *For all positive signal-to-noise ratio parameters τ^2, we have*

$$\sup_{\theta\in\Theta} \mathsf{d}_{\mathsf{TV}}\big(\widetilde{\mathsf{K}}_{\mathsf{plugin}}(X_\theta), Y_\theta\big) \geq \frac{(\sqrt{2}-1)e^{-3/2}}{2\pi} \cdot \min\{1, \tau^{-2}\}.$$

Consequently, it is necessary to take $\tau^2 \gtrsim \epsilon^{-1}$ to guarantee ϵ-TV deficiency.

Note that Lemma 9 illustrates the inefficiency of the plug-in approach. In contrast, Theorem 2 and Corollary 3 show that it suffices to take $\tau^2 = \mathcal{O}(\mathsf{polylog}(1/\epsilon))$ for the approach we proposed.

Proof. By definition of TV, we obtain

$$\begin{aligned}
\sup_{\theta\in\Theta} \mathsf{d}_{\mathsf{TV}}\big(\widetilde{\mathsf{K}}_{\mathsf{plugin}}(X_\theta), Y_\theta\big) &\geq \sup_{\theta\in\Theta} \sup_{t\in\mathbb{R}} \Big|\mathbb{P}\big\{\widetilde{\mathsf{K}}_{\mathsf{plugin}}(X_\theta) \leq t\big\} - \mathbb{P}\big\{Y_\theta \leq t\big\}\Big| \\
&\overset{\text{(i)}}{\geq} \sup_{\theta\in\Theta} \sup_{t\in\mathbb{R}} \left|\mathbb{P}\left\{\frac{\theta^2 + 2\theta Z + Z^2}{\tau^2} + G \leq t\right\} - \mathbb{P}\left\{\frac{\theta^2}{\tau^2} + G \leq t\right\}\right| \\
&\overset{\text{(ii)}}{\geq} \left|\mathbb{P}\left\{\frac{Z^2}{\tau^2} + G \leq 0\right\} - \frac{1}{2}\right|
\end{aligned}$$

where in step (i) we use that for $Z, G \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, we have

$$X_\theta \overset{(d)}{=} \theta + Z, \quad Y_\theta \overset{(d)}{=} \frac{\theta^2}{\tau^2} + G, \quad \widetilde{\mathsf{K}}_{\mathsf{plugin}}(X_\theta) \overset{(d)}{=} \frac{\theta^2 + 2\theta Z + Z^2}{\tau^2} + G,$$

and in step (ii) we evaluate the supremum at $t = \theta^2/\tau^2$ and $\theta = 0$. Continuing, we have

$$\mathbb{P}\bigg\{\frac{Z^2}{\tau^2} + G \le 0\bigg\} = \mathbb{E}_{Z\sim\mathcal{N}(0,1)}\bigg[\int_{-\infty}^{-Z^2/\tau^2} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x\bigg] = \frac{1}{2} - \mathbb{E}_{Z\sim\mathcal{N}(0,1)}\bigg[\int_{0}^{Z^2/\tau^2} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x\bigg].$$

Putting the two pieces together yields

$$\begin{aligned}
\sup_{\theta\in\Theta} \mathsf{d}_{\mathsf{TV}}\big(\mathsf{K}_{\mathsf{plugin}}(X_\theta), Y_\theta\big) &\ge \mathbb{E}_{Z\sim\mathcal{N}(0,1)}\bigg[\int_{0}^{Z^2/\tau^2} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x\bigg] \\
&= \int_{-\infty}^{\infty} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{0}^{t^2/\tau^2} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x\,\mathrm{d}t \\
&\ge \int_{1}^{\sqrt{2}} \frac{e^{-1}}{\sqrt{2\pi}}\mathrm{d}t \int_{0}^{1/\tau^2} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x = \frac{(\sqrt{2}-1)e^{-1}}{\sqrt{2\pi}} \int_{0}^{1/\tau^2} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x.
\end{aligned}$$

Continuing, we have

$$\int_{0}^{1/\tau^2} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x \ge \begin{cases} \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}}, & \text{if } \tau^2 \le 1 \\ \frac{1}{\tau^2}\frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}}, & \text{if } \tau^2 > 1. \end{cases}$$

Combining the two inequalities in the display above yields

$$\sup_{\theta\in\Theta} \mathsf{d}_{\mathsf{TV}}\big(\widetilde{\mathsf{K}}_{\mathsf{plugin}}(X_\theta), Y_\theta\big) \ge \frac{(\sqrt{2}-1)e^{-3/2}}{2\pi} \cdot \min\{1, \tau^{-2}\} \quad \text{for all } \tau \in \mathbb{R}.$$

Clearly, this implies that one must take $\tau^2 \gtrsim \epsilon^{-1}$ to guarantee ϵ-TV deficiency. □

B Proofs of supporting lemmas

In this section, we provide additional proofs deferred from the main text.

B.1 Proof of Lemma 2

By definition, we have

$$f^{(n)}(x) = \sum_{k=1}^{K} c_k f_k^{(n)}(x) \quad \text{for all } n \in \mathbb{N} \text{ and } x \in \mathbb{R}.$$

Consequently, applying the triangle inequality yields that for all $t \in \mathbb{R}$,

$$\begin{aligned}
\bigg(\sum_{j=1}^{n} \frac{|f^{(j)}(t)|}{(\widetilde{C}_f)^j j!}\bigg)^n &\le \bigg(\sum_{k=1}^{K} |c_k| \cdot \sum_{j=1}^{n} \frac{|f_k^{(j)}(t)|}{(\widetilde{C}_f)^j j!}\bigg)^n \\
&\le \bigg(\sum_{\ell=1}^{K} |c_\ell|\bigg)^n \sum_{k=1}^{K} \bigg(\frac{|c_k|}{\sum_{\ell=1}^{K}|c_\ell|}\bigg)^n \cdot \bigg(\sum_{j=1}^{n} \frac{|f_k^{(j)}(t)|}{(\widetilde{C}_f)^j j!}\bigg)^n \\
&\le \sum_{k=1}^{K} \bigg(\frac{|c_k|}{\sum_{\ell=1}^{K}|c_\ell|}\bigg)^n \cdot \bigg(\sum_{j=1}^{n} \frac{|f_k^{(j)}(t)|}{(\widetilde{C}_{f_k})^j j!}\bigg)^n,
\end{aligned}$$

where in the last step we use $\widetilde{C}_f = \sum_{\ell=1}^K |c_\ell| \cdot \max_{k\in[K]}\{\widetilde{C}_{f_k}\}$. By letting $t(x) = (\theta + \sqrt{2}\sigma x)/\tau$ and integrating over $x \in \mathbb{R}$ yields

$$\begin{aligned}\int_{\mathbb{R}} \left(\sum_{j=1}^n \frac{|f^{(j)}(t(x))|}{(\widetilde{C}_f)^j j!}\right)^n \frac{e^{-x^2}}{\sqrt{2\pi}}\mathrm{d}x &\le \sum_{k=1}^K \left(\frac{|c_k|}{\sum_{\ell=1}^K |c_\ell|}\right)^n \int_{\mathbb{R}} \left(\sum_{j=1}^n \frac{|f_k^{(j)}(t)|}{(\widetilde{C}_{f_k})^j j!}\right)^n \frac{e^{-x^2}}{\sqrt{2\pi}}\mathrm{d}x \\ &\overset{\text{(i)}}{\le} \sum_{k=1}^K \left(\frac{|c_k|}{\sum_{\ell=1}^K |c_\ell|}\right)^n \cdot (C_{f_k} n)! \\ &\le \Big(\max_{k\in[K]}\{C_{f_k}\} \cdot n\Big)!,\end{aligned}$$

where in step (i) we use the fact that f_k satisfies Assumption 3 with parameters $\widetilde{C}_{f_k}$ and C_{f_k}. This verifies that f satisfies Assumption 3. We next bound $L_f(n)$ and $\widetilde{L}_f(n,\tau)$. By definition (35), we obtain

$$\begin{aligned}L_f(n) = \sup_{|x|\le 1} \sum_{j=1}^n \frac{|f^{(j)}(x)|}{(\widetilde{C}_f)^j j!} &\le \sum_{k=1}^K |c_k| \sup_{|x|\le 1} \sum_{j=1}^n \frac{|f_k^{(j)}(x)|}{(\widetilde{C}_f)^j j!} \\ &\overset{\text{(i)}}{\le} \sum_{k=1}^K \frac{|c_k|}{\sum_{\ell=1}^K |c_\ell|} \sup_{|x|\le 1} \sum_{j=1}^n \frac{|f_k^{(j)}(x)|}{(\widetilde{C}_{f_k})^j j!} \\ &= \sum_{k=1}^K \frac{|c_k|}{\sum_{\ell=1}^K |c_\ell|} L_{f_k}(n),\end{aligned}$$

where in step (i) we use $\widetilde{C}_f = \sum_{\ell=1}^K |c_\ell| \cdot \max_{k\in[K]}\{\widetilde{C}_{f_k}\}$. Following identical steps yields the desired bound for $\widetilde{L}_f(n,\tau)$. □

B.2 Proof of Lemma 4

We prove each part separately.

Proof of part (i). Applying Faà di Bruno's formula as in Eq. (82) yields that for all $n \in \mathbb{N}$ and all $x, y \in \mathbb{R}$, we have

$$\nabla_x^{(n)} v(y;x) = \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \cdot n! \cdot \sum_{m_1,\ldots,m_n} \frac{(-1)^{m_1+\cdots+m_n}}{m_1!\, m_2! \cdots m_n!} \prod_{j=1}^n \left(\frac{\psi^{(j)}(\frac{y-x}{\tau})}{(-\tau)^j j!}\right)^{m_j}, \tag{140}$$

where the sum is over all n-tuples of nonnegative integers $(m_1, \ldots, m_n)$ satisfying $\sum_{j=1}^n j m_j = n$. Consequently, using definition (14) yields

$$\begin{aligned}\mathcal{S}_N^\star(y \mid x) &= \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \left(1 + \sum_{k=1}^N \frac{(-1)^k \sigma^{2k} (2k)!}{(2k)!!} \sum_{m_1,\ldots,m_{2k}} \frac{(-1)^{m_1+\cdots+m_{2k}}}{m_1!\, m_2! \cdots m_{2k}!} \prod_{j=1}^{2k} \left(\frac{\psi^{(j)}(\frac{y-x}{\tau})}{(-\tau)^j j!}\right)^{m_j}\right) \\ &\overset{\text{(i)}}{\ge} \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \left(1 - \sum_{k=1}^N \frac{\sigma^{2k}(2k)!}{(2k)!!} \frac{\lambda^{2k}}{\tau^{2k}} \sum_{m_1,\ldots,m_{2k}} \frac{1}{m_1!\, m_2! \cdots m_{2k}!} \prod_{j=1}^{2k} \left(\frac{\left|\psi^{(j)}(\frac{y-x}{\tau})\right|}{\lambda^j j!}\right)^{m_j}\right) \\ &\overset{\text{(ii)}}{\ge} \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \left(1 - \sum_{k=1}^N \frac{\sigma^{2k}(2k)!}{(2k)!!} \frac{\lambda^{2k}}{\tau^{2k}} \sum_{m_1,\ldots,m_{2k}} \left(\sum_{j=1}^{2k} \frac{\left|\psi^{(j)}(\frac{y-x}{\tau})\right|}{\lambda^j j!}\right)^{m_1+\cdots+m_{2k}}\right),\end{aligned}$$

where in step (i) we apply the triangle inequality and use $\sum_{j=1}^{2k} jm_j = 2k$ so that

$$\frac{\lambda^{2k}}{\tau^{2k}} \prod_{j=1}^{2k} \left(\frac{1}{\lambda^j}\right)^{m_j} = \prod_{j=1}^{2k} \left(\frac{1}{\tau^j}\right)^{m_j},$$

and in step (ii) we use the multinomial theorem. Continuing, note that we have $(y-x)/\tau \in \mathcal{A}(\psi, 2N, \lambda)$ (24) by assumption, and so we obtain

$$\sum_{j=1}^{2k} \frac{\left|\psi^{(j)}\left(\frac{y-x}{\tau}\right)\right|}{\lambda^j j!} \le 1 \quad \text{for all } k \in [N].$$

Consequently,

$$\begin{aligned}
\mathcal{S}_N^\star(y \mid x) &\ge \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \left(1 - \sum_{k=1}^{N} \frac{\sigma^{2k}(2k)!}{(2k)!!} \frac{\lambda^{2k}}{\tau^{2k}} \sum_{m_1,\ldots,m_{2k}} 1\right) \\
&\overset{\text{(i)}}{\ge} \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \left(1 - \sum_{k=1}^{N} \frac{\sigma^{2k}(2k)^k \lambda^{2k} e^{2k}}{\tau^{2k}}\right) \\
&\overset{\text{(ii)}}{\ge} \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \left(1 - \sum_{k=1}^{N} 4^{-k}\right) \ge 0.
\end{aligned}$$

Here in step (i), we use the fact there are at most e^{2k} terms in the summation and that $(2k)!/(2k)!! \le (2k)^k$. On the other hand, step (ii) holds under the assumption that $\tau \ge 8e\lambda\sigma N$, which is the condition in the lemma statement. This proves part (i) of Lemma 4.

Proof of part (ii). We first verify the first part of Eq. (83). Note that $\rho(t) = g(\psi(t))$, where $g(t) = e^{-t}$. Applying Faà di Bruno's formula yields

$$\rho^{(n)}(t) = \sum_{m_1,\ldots,m_n} \frac{n!}{m_1!\, m_2! \cdots m_n!} (-1)^{m_1+\cdots+m_n} e^{-\psi(t)} \prod_{j=1}^{n} \left(\frac{\psi^{(j)}(t)}{j!}\right)^{m_j}.$$

Applying the triangle inequality yields

$$\begin{aligned}
\left|\rho^{(n)}(t)\right| &\le \sum_{m_1,\ldots,m_n} \frac{n!}{m_1!\, m_2! \cdots m_n!} e^{-\psi(t)} \prod_{j=1}^{n} \left(\frac{|\psi^{(j)}(t)|}{j!}\right)^{m_j} \\
&\overset{\text{(i)}}{\le} n! \cdot \lambda^n \sum_{m_1,\ldots,m_n} \frac{e^{-\psi(t)}}{m_1!\, m_2! \cdots m_n!} \prod_{j=1}^{n} \left(\frac{|\psi^{(j)}(t)|}{\lambda^j j!}\right)^{m_j} \\
&\le n! \cdot \lambda^n \sum_{m_1,\ldots,m_n} e^{-\psi(t)} \left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{\lambda^j j!}\right)^{m_1+\cdots+m_n},
\end{aligned} \tag{141}$$

where in step (i) we use $\sum_{j=1}^{n} jm_j = n$ so that $\prod_{j=1}^{n} (\frac{1}{\lambda^j})^{m_j} = \lambda^{-n}$ and in step (ii) we use the multinomial theorem. By setting $\lambda = \widetilde{C}_\psi$, we obtain

$$\begin{aligned}
\int_{\mathbb{R}} \left|\rho^{(n)}(t)\right| \mathrm{d}t &\le n! \cdot \left(\widetilde{C}_\psi\right)^n \sum_{m_1,\ldots,m_n} \int_{\mathbb{R}} e^{-\psi(t)} \left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{\left(\widetilde{C}_\psi\right)^j j!}\right)^{m_1+\cdots+m_n} \mathrm{d}t \\
&\overset{\text{(i)}}{\le} n! \cdot \left(\widetilde{C}_\psi\right)^n \sum_{m_1,\ldots,m_n} (C_\psi n)! \overset{\text{(ii)}}{\le} n! \cdot \left(\widetilde{C}_\psi\right)^n e^n \cdot (C_\psi n)! < +\infty,
\end{aligned}$$

where in step (i) we use Assumption 2 and in step (ii) we use the fact that there are at most e^n terms in the summation. Since $|\rho^{(n)}(t)|$ is continuous and nonnegative, we must have $\lim_{t\uparrow+\infty}\rho^{(n)}(t) = \lim_{t\downarrow-\infty}\rho^{(n)}(t) = 0$. Consequently, we obtain for $n \geq 1$ that

$$\int_{\mathbb{R}} \rho^{(n)}(t)\mathrm{d}t = \int_{\mathbb{R}} \mathrm{d}\rho^{(n-1)}(t) = \Big[\rho^{(n-1)}(t)\Big]\Big|_{-\infty}^{+\infty} = 0.$$

This proves the first part of Eq. (83).

We now turn to prove the second part of Eq. (83). Using Eq. (140) and applying triangle inequality, we obtain for all $n \geq 1$ that

$$\begin{aligned}\big|\nabla_x^{(n)} v(y;x)\big| &\leq \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \cdot n! \cdot \sum_{m_1,\ldots,m_n} \frac{1}{m_1!\, m_2! \cdots m_n!} \prod_{j=1}^{n} \left(\frac{|\psi^{(j)}(\frac{y-x}{\tau})|}{\tau^j j!}\right)^{m_j} \\ &\leq \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \cdot n! \cdot \frac{\lambda^n}{\tau^n} \sum_{m_1,\ldots,m_n} \left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(\frac{y-x}{\tau})|}{\lambda^j j!}\right)^{m_1+\cdots+m_n}, \end{aligned} \tag{142}$$

where in the last step we use $(\lambda/\tau)^n \prod_{j=1}^n (\frac{1}{\lambda^j})^{m_j} = \prod_{j=1}^n (\frac{1}{\tau^j})^{m_j}$ and then apply the multinomial theorem. Integrating over y yields

$$\begin{aligned}&\int_{\frac{y-x}{\tau}\in\mathcal{A}^{\complement}(\psi,2N,\lambda)} \big|\nabla_x^{(n)} v(y;x)\big|\mathrm{d}y \\ &\leq n! \cdot \frac{\lambda^n}{\tau^n} \sum_{m_1,\ldots,m_n} \int_{\frac{y-x}{\tau}\in\mathcal{A}^{\complement}(\psi,2N,\lambda)} \frac{1}{\tau} e^{-\psi(\frac{y-x}{\tau})} \left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(\frac{y-x}{\tau})|}{\lambda^j j!}\right)^{m_1+\cdots+m_n} \mathrm{d}y \\ &\overset{\text{(i)}}{\leq} n! \cdot \frac{\lambda^n}{\tau^n} \sum_{m_1,\ldots,m_n} \int_{\mathcal{A}^{\complement}(\psi,2N,\lambda)} e^{-\psi(t)} \left(\sum_{j=1}^{n} \frac{|\psi^{(j)}(t)|}{\lambda^j j!}\right)^{m_1+\cdots+m_n} \mathrm{d}t \\ &\leq n! \cdot \frac{\lambda^n}{\tau^n} \cdot e^n \cdot \mathsf{Err}(\psi, 2N, \lambda),\end{aligned}$$

where in step (i) we use the change of variables $t = (y-x)/\tau$. This proves Eq. (83). □

B.3 Proof of Lemma 5

Note that by definition (33) and Eq. (99), we have $v(y;x) = \phi(g(x))$. Applying Faà di Bruno's formula yields

$$\begin{aligned}\nabla_x^{(n)} v(y;x) &= \sum_{m_1,\ldots,m_n} \frac{n!}{m_1!\, m_2! \cdots m_n!} \cdot \phi^{(m_1+\cdots+m_n)}(g(x)) \cdot \prod_{j=1}^{n} \left(\frac{g^{(j)}(x)}{j!}\right)^{m_j} \\ &\overset{\text{(i)}}{=} n! \cdot \sum_{m_1,\ldots,m_n} \frac{(-1)^{m_1+\cdots+m_n}}{m_1!\, m_2! \cdots m_n!} \cdot \phi(g(x)) \cdot H_{m_1+\cdots+m_n}(g(x)) \cdot \prod_{j=1}^{n} \left(\frac{f^{(j)}(x/\tau)}{2\tau^j j!}\right)^{m_j} \\ &\overset{\text{(ii)}}{=} \frac{(\widetilde{C}_f)^n}{\tau^n} \cdot n! \cdot \sum_{m_1,\ldots,m_n} \frac{(-1)^{m_1+\cdots+m_n}}{m_1!\, m_2! \cdots m_n!} \cdot \phi(g(x)) \cdot H_{m_1+\cdots+m_n}(g(x)) \cdot \prod_{j=1}^{n} \left(\frac{f^{(j)}(x/\tau)}{2(\widetilde{C}_f)^j j!}\right)^{m_j},\end{aligned}$$

where in all steps the summation is over all nonnegative integers $(m_1, \cdots, m_n)$ satisfying $\sum_{j=1}^n j m_j = n$. In step (i), we use the definition of Hermite polynomials (6) and in step (ii), we use $\prod_{j=1}^n (\frac{1}{\tau^j})^{m_j} = \frac{(\widetilde{C}_f)^n}{\tau^n} \prod_{j=1}^n (\frac{1}{(\widetilde{C}_f)^j})^{m_j}$. Using the definition of $h_j(x)$ in Eq. (99) and combining the pieces then yields Eq. (100).

To prove the second claim of the lemma, note that from Ineq. (62a), we have

$$\begin{aligned}\left|\phi(g(x)) \cdot H_{m_1+\cdots+m_n}(g(x))\right| &\le \frac{e^{-g(x)^2/2}}{\sqrt{2\pi}}\pi^{1/4} \cdot (m_1+\cdots+m_n)! \cdot 2^{m_1+\cdots+m_n} \\ &\le e^{-g(x)^2/2} \cdot 2^n \cdot (m_1+\cdots+m_n)!.\end{aligned}$$

Now applying the triangle inequality to the display before yields

$$\begin{aligned}\left|\nabla_x^{(n)} v(y;x)\right| &\le e^{-\frac{g(x)^2}{2}} \cdot \frac{(2\widetilde{C}_f)^n}{\tau^n} \cdot n! \sum_{m_1,\ldots,m_n} \frac{(m_1+\cdots+m_n)!}{m_1!\,m_2!\cdots m_n!} \prod_{j=1}^n |h_j(x)|^{m_j} \\ &\overset{\text{(i)}}{\le} e^{-\frac{g(x)^2}{2}} \cdot \frac{(2\widetilde{C}_f)^n}{\tau^n} \cdot n! \sum_{m_1,\ldots,m_n} \left(\sum_{j=1}^n |h_j(x)|\right)^{m_1+\cdots+m_n} \\ &\le e^{-\frac{g(x)^2}{2}} \cdot \frac{(2\widetilde{C}_f)^n}{\tau^n} \cdot n! \cdot e^n \cdot \left[\left(\sum_{j=1}^n |h_j(x)|\right)^n \vee 1\right]\end{aligned}$$

where in step (i) we use the multinomial theorem so that

$$\frac{(m_1+\cdots+m_n)!}{m_1!\,m_2!\cdots m_n!} \prod_{j=1}^n |h_j(x)|^{m_j} \le \left(\sum_{j=1}^n |h_j(x)|\right)^{m_1+\cdots+m_n},$$

and in the last step we use there are at most e^n in the summation and $m_1+\cdots+m_n \le n$. This concludes the proof. □

B.4 Proof of Lemma 7

We prove each part separately. Recall the shorthand $h_j(x)$, $g(x)$, and $\phi(x)$ from Eq. (99).

Proof of Lemma 7(a). From Eq. (62c), we obtain for all $n \in \mathbb{Z}_{\ge 0}$, $n \ge 1$, and $x \in \mathbb{R}$ that

$$|H_n(x)| \le n! \cdot \frac{n}{2} \cdot (1 \vee 2|x|)^n.$$

Consequently, if $|g(x)| = |f(x/\tau) - y|/2 \le \frac{\sqrt{\tau}}{2}$, then for $\tau \ge 4$ and $1 \le m_1+\cdots+m_n \le n$, we have

$$\begin{aligned}\left|H_{m_1+\cdots+m_n}(g(x))\right| &\le \frac{m_1+\cdots+m_n}{2} \cdot (m_1+\cdots+m_n)! \cdot \tau^{\frac{m_1+\cdots+m_n}{2}} \\ &\le \frac{n}{2} \cdot (m_1+\cdots+m_n)! \cdot \tau^{\frac{n}{2}}.\end{aligned}$$

Using the bound in the above display along with Eq. (100), we obtain for $n \ge 1$ that

$$\begin{aligned}\left|\nabla_x^{(n)} v(y;x)\right| &\le \phi(g(x)) \cdot \frac{n}{2} \cdot \tau^{\frac{n}{2}} \cdot n! \cdot \frac{(\widetilde{C}_f)^n}{\tau^n} \sum_{m_1,\ldots,m_n} \frac{(m_1+\cdots+m_n)!}{m_1!\,m_2!\cdots m_n!} \cdot \prod_{j=1}^n \left|h_j(x)\right|^{m_j} \\ &\le \phi(g(x)) \cdot \frac{n}{2} \cdot \tau^{\frac{n}{2}} \cdot n! \cdot \frac{(\widetilde{C}_f)^n}{\tau^n} \cdot e^n \cdot \left[\left(\sum_{j=1}^n |h_j(x)|\right)^n \vee 1\right], \qquad (143)\end{aligned}$$

where in the last step we use the multinomial theorem so that

$$\frac{(m_1+\cdots+m_n)!}{m_1!\,m_2!\cdots m_n!} \cdot \prod_{j=1}^n \left|h_j(x)\right|^{m_j} \le \left(\sum_{j=1}^n |h_j(x)|\right)^{m_1+\cdots+m_n} \le \left(\sum_{j=1}^n |h_j(x)|\right)^n \vee 1,$$

and we also use there are at most e^n terms in the summation (see Ineq. (69)). Continuing, by using the signed kernel definition (14) for $\sigma = 1$, we obtain for $|x| \le \tau$ that

$$\begin{aligned}
\mathcal{S}_N^\star(y \mid x) &= \sum_{k=0}^{N} \frac{(-1)^k}{(2k)!!} \nabla_x^{(2k)} v(y; x) \\
&\ge v(y; x) - \sum_{k=1}^{N} \frac{1}{(2k)!!} \big|\nabla_x^{(2k)} v(y; x)\big| \\
&\overset{\text{(i)}}{\ge} \phi(g(x)) - \sum_{k=1}^{N} \frac{1}{(2k)!!} \cdot \phi(g(x)) \cdot \frac{2k}{2} \cdot (2k)! \cdot \tau^k \cdot \left(\frac{\widetilde{C}_f}{\tau}\right)^{2k} e^{2k} \left[\left(\sum_{j=1}^{2k} |h_j(x)|\right) \vee 1\right]^{2k} \\
&\overset{\text{(ii)}}{\ge} \phi(g(x)) \cdot \left(1 - \sum_{k=1}^{N} \frac{\left(e^3 \big(\widetilde{C}_f\big)^2 \big(L_f(2k) \vee 1\big) 2k\right)^k}{\tau^k}\right) \\
&\overset{\text{(iii)}}{\ge} \phi(g(x)) \cdot \left(1 - \sum_{k=1}^{N} \frac{1}{2^k}\right) \ge 0.
\end{aligned}$$

In step (i) above, we use Ineq. (143) for $n = 2k$. In step (ii), we use the inequalities $k \le e^k$ and $(2k)!/(2k)!! = (2k-1)!! \le (2k)^k$, and also that

$$\sum_{j=1}^{2k} |h_j(x)| \le L_f(2k) \quad \text{for all } |x| \le \tau,$$

which in turn follows from combining Eq. (35) and Eq. (99). Step (iii) holds if

$$\tau \ge 2e^3 \big(\widetilde{C}_f\big)^2 \big(L_f(2N) \vee 1\big) 2N \ge 2e^3 \big(\widetilde{C}_f\big)^2 \big(L_f(2k) \vee 1\big) 2k \quad \text{for all } k \in [N],$$

as assumed in the lemma. This concludes the proof of part (a).

Proof of Lemma 7(b) Using Ineq. (101) and noting that $g(x) = (f(x/\tau) - y)/2$ (Eq. (99)), we obtain

$$\begin{aligned}
&\int_{f(\frac{x}{\tau})+\sqrt{\tau}}^{+\infty} \big|\nabla_x^{(2k)} v(y; x)\big| \mathrm{d}y + \int_{-\infty}^{f(\frac{x}{\tau})-\sqrt{\tau}} \big|\nabla_x^{(2k)} v(y; x)\big| \mathrm{d}y \\
&\le (2k)! \cdot e^{2k} \cdot \left(\frac{2\widetilde{C}_f}{\tau}\right)^{2k} \cdot \left[\left(\sum_{j=1}^{2k} \big|h_j(x)\big|\right)^{2k} \vee 1\right] \cdot 2 \int_{\sqrt{\tau}}^{+\infty} e^{-t^2/8} \mathrm{d}t \\
&\overset{\text{(i)}}{\le} 2\sqrt{2\pi} \exp(-\tau/8) \cdot (2k)! \cdot \left(\frac{2e\widetilde{C}_f}{\tau}\right)^{2k} \cdot \big(L_f(2k) \vee 1\big)^{2k},
\end{aligned}$$

where in step (i) we use the fact that $|x| \le \tau$ as well as the definitions of $h_j(x)$ (99) and $L_f(n)$ (35). In particular, these yield

$$\sum_{j=1}^{2k} \big|h_j(x)\big| \le L_f(2k) \quad \text{and} \quad 2 \int_{\sqrt{\tau}}^{+\infty} e^{-t^2/8} \mathrm{d}t \le 2\sqrt{2\pi} \exp(-\tau/8).$$

Proof of Lemma 7(c) Using the closed-form expression for $\nabla_x^{(n)} v(y;x)$ (Eq. (100)) and noting that $g(x) = (f(x/\tau) - y)/2$ (Eq. (99)), we obtain

$$\int_{\mathbb{R}} \nabla_x^{(n)} v(y;x) \mathrm{d}y = \sum_{m_1,\dots,m_n} \frac{n! \cdot (-1)^{m_1+\cdots+m_n}}{m_1! \, m_2! \cdots m_n!} \left(\frac{\widetilde{C}_f}{\tau}\right)^n \prod_{j=1}^{n} \Big(h_j(x)/2\Big)^{m_j} 2 \int_{\mathbb{R}} H_{m_1+\cdots+m_n}(t) \cdot \phi(t) \mathrm{d}t = 0.$$

In the last step, we use the facts that $m_1 + \cdots + m_n \geq 1$ and $H_0(x) = 1$ along with the orthogonal property of Hermite polynomials. In particular, these imply that

$$\int_{\mathbb{R}} H_{m_1+\cdots+m_n}(t) \cdot \phi(t) \mathrm{d}t = \frac{1}{\sqrt{2\pi}} \big\langle H_{m_1+\cdots+m_n}(\cdot), H_0(\cdot) \big\rangle = 0.$$

This concludes the proof of the final part of the lemma. □

B.5 Proof of Lemma 8

Note that there are two differences between the tensor $\widetilde{T}$ in Eq. (133) and the tensor T in Eq. (43): (i) the additive Gaussian noise $\widetilde{\zeta}$ is asymmetric, and (ii) the spike $\widetilde{v}$ is discrete. We now construct two procedures $\mathcal{R}_1, \mathcal{R}_2 : \mathbb{R}^{n^{\otimes s}} \to \mathbb{R}^{n^{\otimes s}}$, each designed to resolve one of these two differences. Our overall reduction is the composition of the two, i.e., $\mathcal{R}\big(\widetilde{T}\big) := \mathcal{R}_2 \circ \mathcal{R}_1\big(\widetilde{T}\big)$.

Construction of reduction $\mathcal{R}_1$ to make instance symmetric. We let

$$\mathcal{R}_1\big(\widetilde{T}\big) = \frac{1}{s!} \sum_{\pi \in \in \mathfrak{S}_s} \big(\widetilde{T}\big)^{\pi},$$

where $\big(\widetilde{T}\big)^{\pi} \in \mathbb{R}^{n^{\otimes s}}$ denotes the permutation of the tensor $\widetilde{T}$ so that

$$\big(\widetilde{T}\big)^{\pi}_{i_1,\dots,i_s} = \big(\widetilde{T}\big)^{\pi}_{i_{\pi(1)},\dots,i_{\pi(s)}} \quad \text{for all indices} \quad i_1, \cdots, i_s \in [n].$$

Consequently, by the definition of $\widetilde{T}$ in Eq. (133), we obtain

$$\mathcal{R}_1\big(\widetilde{T}\big) = \beta(\sqrt{n}\widetilde{v})^{\otimes s} + \zeta, \quad \textit{where} \quad \widetilde{v} \sim \mathsf{Unif}\big(\big\{\tfrac{-1}{\sqrt{n}}, \tfrac{1}{\sqrt{n}}\big\}^n\big)$$

and the noise tensor ζ is now symmetric and satisfies

$$\zeta \overset{(d)}{=} \frac{1}{s!} \sum_{\pi \in \in \mathfrak{S}_s} G^{\pi},$$

where $G \in \mathbb{R}^{n^{\otimes s}}$ is an asymmetric tensor with entries that are $\mathsf{i.i.d.}\ \mathcal{N}(0,1)$, and $G^{\pi} \in \mathbb{R}^{n^{\otimes s}}$ satisfies $G^{\pi}_{i_1,\dots,i_s} = G_{i_{\pi(1)},\dots,i_{\pi(s)}}$ for all indices $i_1, \cdots, i_s \in [n]$.

Construction of reduction $\mathcal{R}_2$ to spread the signal spike. We next construct the second procedure $\mathcal{R}_2$, which transforms the tensor $\mathcal{R}_1\big(\widetilde{T}\big)$ into one whose spike v is uniformly distributed on the unit sphere and whose additive Gaussian noise remains independent of v and symmetric Gaussian. For notational convenience, we let

$$\overline{T} := \mathcal{R}_1\big(\widetilde{T}\big) = \beta(\sqrt{n}\widetilde{v})^{\otimes s} + \zeta.$$

We first review some definitions. For a tensor $T \in \mathbb{R}^{n^{\otimes s}}$, define its vectorization as

$$\mathsf{vec}(T) = \sum_{i_1=1}^{n} \cdots \sum_{i_s=1}^{n} T_{i_1,\ldots,i_s} \times e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_s},$$

where e_i is i-th standard basis vector in $\mathbb{R}^n$ and $\otimes$ denotes the Kronecker product. For any permutation $\pi : [s] \to [s]$, define $P_\pi \in \mathbb{R}^{n^s \times n^s}$ by its action on the Kronecker basis:

$$P_\pi\big(e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_s}\big) \;=\; e_{i_{\pi^{-1}(1)}} \otimes e_{i_{\pi^{-1}(2)}} \otimes \cdots \otimes e_{i_{\pi^{-1}(s)}}, \quad i_1, \ldots, i_s \in [n].$$

Since P_π maps the standard basis of $\mathbb{R}^{n^s}$ bijectively to itself, it is a permutation matrix.

We now construct the reduction $\mathcal{R}_2 : \mathbb{R}^{n^{\otimes s}} \to \mathbb{R}^{n^{\otimes s}}$: Let $U \in \mathbb{R}^{n \times n}$ be a unitary matrix that is drawn uniformly at random from the set $\{U \in \mathbb{R}^{n \times n} : UU^\top = I_n\}$. Given $\overline{T} \in \mathbb{R}^{n^{\otimes s}}$, construct $\mathcal{R}_2\big(\overline{T}\big)$ such that

$$\mathsf{vec}\big(\mathcal{R}_2\big(\overline{T}\big)\big) = (U \otimes U \otimes \cdots \otimes U)\, \mathsf{vec}\big(\overline{T}\big).$$

The tensor $\mathcal{R}_2\big(\overline{T}\big)$ by suitably stacking this vector into the appropriate tensor dimension.

The reduction achieves zero-deficiency. To verify $\mathcal{R}_2\big(\overline{T}\big) \overset{(d)}{=} T$, it is equivalent to verify $\mathsf{vec}\big(\mathcal{R}_2\big(\overline{T}\big)\big) \overset{(d)}{=} \mathsf{vec}(T)$. Note that we have by definition,

$$\begin{aligned}
\mathsf{vec}\big(\mathcal{R}_2\big(\overline{T}\big)\big) &= (U \otimes U \otimes \cdots \otimes U)\, \mathsf{vec}\big(\overline{T}\big) \\
&= (U \otimes U \otimes \cdots \otimes U)\, \Big(\beta \times (\sqrt{n}\widetilde{v}) \otimes (\sqrt{n}\widetilde{v}) \cdots \otimes (\sqrt{n}\widetilde{v}) + \mathsf{vec}\big(\zeta\big)\Big) \\
&= \beta \times (\sqrt{n}U\widetilde{v}) \otimes (\sqrt{n}U\widetilde{v}) \otimes \cdots \otimes (\sqrt{n}U\widetilde{v}) + (U \otimes U \otimes \cdots \otimes U)\, \mathsf{vec}\big(\zeta\big).
\end{aligned}$$

We now claim two equalities, whose proofs we defer to the end:

$$\mathsf{vec}\big(\zeta\big) = \frac{1}{s!} \sum_{\pi \in \mathfrak{S}_s} P_\pi\, \mathsf{vec}\big(G\big) \quad \text{and} \tag{144a}$$

$$(U \otimes U \otimes \cdots \otimes U)\, \mathsf{vec}\big(\zeta\big) = \frac{1}{s!} \sum_{\pi \in \mathfrak{S}_s} P_\pi\, (U \otimes U \otimes \cdots \otimes U)\, \mathsf{vec}\big(G\big). \tag{144b}$$

Putting the pieces together yields

$$\mathsf{vec}\big(\mathcal{R}_2\big(\overline{T}\big)\big) = \beta \times (\sqrt{n}U\widetilde{v}) \otimes (\sqrt{n}U\widetilde{v}) \otimes \cdots \otimes (\sqrt{n}U\widetilde{v}) + \frac{1}{\sqrt{s!}} \sum_{\pi \in \mathfrak{S}_s} P_\pi\, (U \otimes U \otimes \cdots \otimes U)\, \mathsf{vec}\big(G\big), \tag{145}$$

and by definition of T (43) and Eq. (144a), we obtain

$$\mathsf{vec}\big(T\big) = \beta \times (\sqrt{n}v) \otimes (\sqrt{n}v) \otimes \cdots \otimes (\sqrt{n}v) + \frac{1}{s!} \sum_{\pi \in \mathfrak{S}_s} P_\pi\, \mathsf{vec}\big(G\big). \tag{146}$$

We now compare Eq. (145) and Eq. (146) to verify that $\mathsf{vec}\big(\mathcal{R}\big(\overline{T}\big)\big) \overset{(d)}{=} \mathsf{vec}(T)$. First, note that $\widetilde{v} \sim \mathsf{Unif}\big(\big\{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\big\}^n\big)$, U is a unitary matrix that is selected uniformly at random and is independent

of $\widetilde{v}$, and $v \sim \mathsf{Unif}\big(\mathbb{S}^{n-1}\big)$. We have $U\widetilde{v} \overset{(d)}{=} v$. So the first terms in the RHS of Eq. (145) and Eq. (146) are equally distributed. Second, note that $\mathsf{vec}\big(G\big) \sim \mathcal{N}(0, I_{n^s})$ and we have

$$\begin{aligned}
&\mathbb{E}\big[(U \otimes U \otimes \cdots \otimes U)\ \mathsf{vec}\big(G\big)\big] = 0 \quad \text{and} \\
&\mathbb{E}\Big[\big((U \otimes U \otimes \cdots \otimes U)\ \mathsf{vec}\big(G\big)\big)^\top \big((U \otimes U \otimes \cdots \otimes U)\ \mathsf{vec}\big(G\big)\big)\Big] \\
&= \mathbb{E}\Big[\big(\mathsf{vec}\big(G\big)\big)^\top \big(U^\top \otimes U^\top \otimes \cdots \otimes U^\top\big)\big(U \otimes U \otimes \cdots \otimes U\big)\ \mathsf{vec}\big(G\big)\Big] \\
&= \mathbb{E}\Big[\big(\mathsf{vec}\big(G\big)\big)^\top \big((U^\top U) \otimes (U^\top U) \otimes \cdots \otimes (U^\top U)\big)\ \mathsf{vec}\big(G\big)\Big] \\
&= \mathbb{E}\Big[\big(\mathsf{vec}\big(G\big)\big)^\top\ \mathsf{vec}\big(G\big)\Big] = I_{n^s}.
\end{aligned}$$

Consequently, we obtain

$$(U \otimes U \otimes \cdots \otimes U)\ \mathsf{vec}\big(G\big) \sim \mathcal{N}(0, I_{n^s}).$$

This proves that the second terms in the RHS of Eq. (145) and Eq. (146) are equally distributed. Therefore, we have verified $\mathsf{vec}\big(\mathcal{R}\big(\overline{T}\big)\big) \overset{(d)}{=} \mathsf{vec}(T)$. It remains to prove the claimed equalities (144).

Proof of Eq. (144a) Note that we have

$$\mathsf{vec}\big(\zeta\big) = \frac{1}{s!} \sum_{\pi \in \mathfrak{S}_s} \mathsf{vec}\big(G^\pi\big).$$

By by definition, we have

$$\begin{aligned}
\mathsf{vec}\big(G^\pi\big) &= \sum_{i_1=1}^{n} \cdots \sum_{i_s=1}^{n} G_{i_{\pi(1)},\ldots,i_{\pi(s)}}\ e_{i_1} \otimes e_{i_1} \otimes \cdots \otimes e_{i_s} \\
&\overset{\text{(i)}}{=} \sum_{i_{\pi(1)}=1}^{n} \cdots \sum_{i_{\pi(s)}=1}^{n} G_{i_{\pi(1)},\ldots,i_{\pi(s)}}\ e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_s} \\
&\overset{\text{(ii)}}{=} \sum_{j_1=1}^{n} \cdots \sum_{j_s=1}^{n} G_{j_1,\ldots,j_s}\ e_{j_{\pi^{-1}(1)}} \otimes e_{j_{\pi^{-1}(2)}} \otimes \cdots \otimes e_{j_{\pi^{-1}(s)}} \\
&\overset{\text{(iii)}}{=} \sum_{j_1=1}^{n} \cdots \sum_{j_s=1}^{n} G_{j_1,\ldots,j_s}\ P_\pi\Big(e_{j_1} \otimes e_{j_2} \otimes \cdots \otimes e_{j_s}\Big) \\
&= P_\pi\ \mathsf{vec}\big(G\big),
\end{aligned}$$

where in step (i) we reorder the summation sequence, in step (ii) we rename the indices by letting

$$j_1 = i_{\pi(1)}, \quad j_2 = i_{\pi(2)}, \quad \cdots, \quad j_s = i_{\pi(s)}$$

so that

$$i_1 = j_{\pi^{-1}(1)}, \quad i_2 = j_{\pi^{-1}(2)}, \quad \cdots, \quad i_s = j_{\pi^{-1}(s)},$$

and in step (iii) we use the definition of P_π. Summing over permutations completes the proof.

Proof of Eq. (144b) We have

$$(U \otimes U \otimes \cdots \otimes U)\, \mathsf{vec}\big(\zeta\big) = \frac{1}{s!} \sum_{\pi \in \mathfrak{S}_s} (U \otimes U \otimes \cdots \otimes U)\, P_\pi\, \mathsf{vec}\big(G\big).$$

Note that the operators $(U \otimes U \otimes \cdots \otimes U)$ and P_π commute, since

$$\begin{aligned}
(U \otimes U \otimes \cdots \otimes U)\, P_\pi\, (e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_s}) &= (U \otimes U \otimes \cdots \otimes U)\, \big(e_{i_{\pi^{-1}(1)}} \otimes e_{i_{\pi^{-1}(2)}} \otimes \cdots \otimes e_{i_{\pi^{-1}(s)}}\big) \\
&= \big(U e_{i_{\pi^{-1}(1)}}\big) \otimes \big(U e_{i_{\pi^{-1}(2)}}\big) \otimes \cdots \otimes \big(U e_{i_{\pi^{-1}(s)}}\big) \\
&= P_\pi \Big(\big(U e_{i_1}\big) \otimes \big(U e_{i_2}\big) \otimes \cdots \otimes \big(U e_{i_s}\big)\Big) \\
&= P_\pi\, (U \otimes U \otimes \cdots \otimes U)\, (e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_s}).
\end{aligned}$$

Consequently, we have

$$(U \otimes U \otimes \cdots \otimes U)\, \mathsf{vec}\big(\zeta\big) = \frac{1}{s!} \sum_{\pi \in \mathfrak{S}_s} P_\pi\, (U \otimes U \otimes \cdots \otimes U)\, \mathsf{vec}\big(G\big),$$

as claimed. This concludes the proof. □