Testing direction shifts under orthogonally invariant observations

Clément Pierquin [1 2] **Aurélien Bellet** [3] **Marc Tommasi** [2] **Matthieu Boussard** [1]

Abstract

We study high-dimensional shift testing when the statistician observes only an orthogonally invariant transformation of the shifted sample $Y_v = X + v$. Examples include norms or Gram matrices. Such observations discard the left angular orientation of Y_v, so when the noise X is itself $O(d)$-invariant, they cannot distinguish a shift v from any shift Uv. We ask when this geometric loss of information yields a dimension-dependent contraction of distinguishability. We answer this question through Fisher information. For the class of $O(d)$-invariant noise distributions, we prove that every invariant observation satisfies a Fisher-information contraction bound whose leading factor scales as $1/d$, up to constants depending only on the noise law. Then, we instantiate the result for Gaussian and Frobenius-radial Laplace distributions, and show how the Gaussian case improves existing analyses in *privacy amplification by synthetic data release*.

1. Introduction

Let $d, k \in \mathbb{N}^*$. Let $X \in \mathbb{R}^{d\times k} \sim \mu$ be a random matrix, and $v, w \in \mathbb{R}^{d\times k}$ be two deterministic matrices. Let K be a Markov kernel from $\mathbb{R}^{d\times k}$ to a measurable space $\mathcal{Z}$. We say that K is $O(d)$-invariant if for all $U \in O(d), y \in \mathbb{R}^{d\times k}$, $K(\cdot \mid Uy) = K(\cdot \mid y)$. To lighten notation, we write $f(Y)$ for the observation generated by the kernel, and we refer to f as an $O(d)$-invariant randomized map. Let the hypothesis test:

$$H_0 : Y \sim X + v, \qquad H_1 : Y \sim X + w,$$

and assume that the statistician does not observe Y. Instead, the statistician observes only a randomized $O(d)$-invariant statistic $f(Y)$. This setting appears whenever only rotation-free summary statistics are released, such as for example $\|X+v\|_F$, $(X+v)^\top(X+v)$ or $Z(X+v)$ when $Z \in \mathbb{R}^{n\times d}$ is an independent Gaussian matrix. It also applies to any data processing of these quantities, for example the first layer of an unbiased neural network for a Gaussian input. Such releases discard the left angular orientation of the realized sample.

Beyond the specific examples above, invariant observations arise naturally whenever the data are only meaningful up to a choice of coordinates. In high-dimensional learning problems, many transformations of the parameters or latent representations leave the observable behavior unchanged: permutations of hidden units, rescalings of homogeneous networks, rotations of latent spaces, and Gram-type summaries are common examples (Zhao et al., 2026). Orthogonal invariance is a particularly tractable instance of this phenomenon.

Direction shift testing is also central in differential privacy: output perturbation (Chaudhuri et al., 2011), noisy gradients (Abadi et al., 2016), and Gaussian mechanisms (Dwork and Roth, 2014) all reduce local distinguishability between neighboring datasets to testing shifted noise distributions. We return to this connection in Section 4. By the data processing inequality, such releases contract divergences, for example: $D_\alpha(f(X+v), f(X+w)) \le D_\alpha(X+v, X+w)$, where D_α is the Rényi divergence of order $\alpha > 1$. By leveraging $f(UY) = f(Y)$, we can obtain the following simple improved upper bound:

$$D_\alpha(f(X+v), f(X+w)) \le \inf_{U\in O(d)} D_\alpha(U(X+v), X+w).$$

The central question we ask in this paper is: *when can we obtain a dimension-dependent contraction*? In fact, when X is rotationally symmetric, an $O(d)$-invariant observation cannot distinguish v from Uv. Thus the testing problem is not naturally posed on the full parameter space $\mathbb{R}^{d\times k}$, but on the quotient by the left action of $O(d)$. For example, in the case $k = 1$, this quotient is one-dimensional, which means that only the radius matters, and the $d-1$ angular directions are not observable. This dimensional mismatch suggests that an invariant observation should contract divergences in a stronger way than data processing. It should contract them by a factor depending on d. The main question is to turn this

[1]Craft AI, Paris, France [2]Université de Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France [3]PreMeDICaL team, Inria, Idesp, Inserm, Université de Montpellier. Correspondence to: Clément Pierquin <clement.pierquin@craft-ai.fr>.

ICML 2026 Workshop on Hypothesis Testing, Seoul, South Korea, 2026. Copyright 2026 by the author(s).

geometric intuition into a quantitative Fisher-information bound. In this paper, we study such contraction properties for Fisher information. In other words, we aim to assess when the following inequality holds:

$$I_v(f(X+v)) \preceq \eta_{d,k}(X) I_v(X+v),$$

where $\eta_{d,k}(X) < 1$ and I_v is the Fisher information for the parameter v. More generally, we focus on general upper envelopes of $I_v(f(X+v))$, not only contractions. In Section 4, we also study contractions for Rényi divergences, yielding the contraction factor η^{syn}. We first study the case where X is $O(d)$ invariant, and prove that general upper envelopes can be obtained, yielding contraction bounds in the Gaussian case. Then, we show that these bounds can be leveraged to obtain *privacy amplification by synthetic data release* results, therefore improving upon the bounds of Pierquin et al. (2026). Finally, we discuss the potential contraction in non $O(d)$ settings. All proofs can be found in Appendix.

2. Setting and notation

Let $X \in \mathbb{R}^{d\times k}$ and let $f : \mathbb{R}^{d\times k} \to \mathcal{Z}$ be measurable and $O(d)$-invariant. We write:

$$P_v = \mathcal{L}(X+v), \qquad Q_v^f = \mathcal{L}(f(X+v)),$$

and assume that these distributions admit regular densities p_v and q_v. Define the location score: $S(X) := -\nabla \log p(X) = \nabla_v \log p_v(Y_v)$, where p is the density of X. The Fisher information of the released experiment in direction $H \in \mathbb{R}^{d\times k}$ is denoted by $I_f(v)[H] := \mathbb{E}_{Y\sim p_\theta}[(D_\theta \log Q_v^f(Y)[H])^2] = \mathbb{E}_{Y\sim p_\theta}\left[\left(\partial_t|_{t=0} \log Q_{v+tH}^f(Y)\right)^2\right]$.

Inspired by Pierquin et al. (2026), we leverage the following maximal invariant property of this problem in order to upper bound Fisher information.

Proposition 1 (Maximal invariant reduction)**.** *Note $g(X) = X^\top X$. Let $v, H \in \mathbb{R}^{d\times k}$. Then, for any $O(d)$-invariant randomized map $f : \mathbb{R}^{d\times k} \mapsto \mathcal{Z}$, we have:*

$$I_f(v)[H] \le I_g(v)[H].$$

Therefore, every invariant release is a data processing of g. Hence the most informative invariant experiment corresponds to the observation of the Gram statistics $(X+v)^\top(X+v)$. Proving a Fisher bound for g automatically proves it for every invariant f. Therefore, during the rest of the paper, we consider $f = g$.

Also, we leverage the following property connecting Fisher information and conditional expectation of the score throughout the paper:

Proposition 2 (Conditional score identity, Adapted from Proposition 3.1 of (Delattre and Kuhn, 2019))**.** *Assume that X admits a sufficiently regular distribution. Then,*

$$I_f(v)[H] = \mathbb{E}\left[\left(\mathbb{E}[\langle S(X), H\rangle_{\mathrm{F}} \mid f(X+v)]\right)^2\right].$$

Equivalently, we have:

$$I_f(v)[H] = \sup_{\substack{F\in L^2(\sigma(f(X+v)))\\ \mathbb{E}[F^2]>0, \mathbb{E}[F]=0}} \frac{\mathbb{E}[F\langle S(X), H\rangle_F]^2}{\mathbb{E}[F^2]}.$$

where L_0^2 denotes centered square-integrable functions.

Hence Fisher information is small when the invariant release is nearly orthogonal, in L^2, to the score direction being tested. Thus bounding Fisher information reduces to bounding the correlations between score directions and invariant test functions. Leveraging Cauchy-Schwarz inequality, we obtain:

$$I_f(v)[H] \le \sup_{\substack{F\in L^2(\sigma(f(X+v)))\\ \mathbb{E}[F^2]>0, \mathbb{E}[F]=0}} \frac{\|\mathbb{E}[FS(X)]\|_F^2}{\mathbb{E}[F^2]} \|H\|_F^2.$$

Therefore, to upper bound the Fisher information, we will relate $\|\mathbb{E}[FS(X)]\|_F^2$ to $\mathbb{E}[F^2]$ for any test random variable F.

3. Releasing $f(X+v)$ when X is invariant

Throughout this section, X has a smooth $O(d)$-invariant density p on $\mathbb{R}^{d\times k}$: $UX \overset{d}{=} X$ for all $U \in O(d)$. The Gaussian law is one example, but the argument below also applies to other smooth invariant distributions, such as the Frobenius-radial Laplace distribution: $p(x) \propto \exp(-\beta\|x\|_F)$. The radial Laplace law is not C^2 at the origin, but the score argument remains valid by approximation. We give the details in Appendix A.3. We note $S(X) := -\nabla \log p(X)$ the score of p. Then, we define the second-order operator:

$$T_X := S(X) \otimes S(X) - DS(X),$$

where $(A\otimes B)C = \langle A, C\rangle_F B$. The likelihood ratio admits the following expansion:

$$\frac{p(x-h)}{p(x)} = 1 + \langle S(x), h\rangle_F + \frac{1}{2}\langle T_X h, h\rangle_F + o(\|h\|_F^2).$$

For $F \in L^2(\sigma(f(X+v))), \mathbb{E}[F^2] > 0, \mathbb{E}[F] = 0$, the first-order and second-order coefficients $\mathbb{E}[FS(X)]$ and $\mathbb{E}[FT_X]$ coincide with the first and second Hermite coefficients of the expansion of F when X is Gaussian.

Proposition 3. *Let $F \in L^2(\sigma(Y))$ such that $\mathbb{E}[F] = 0, \mathbb{E}[F^2] > 0$. We note:*

$$R(p) = \lambda_{\max}(\mathbb{E}[T_X \otimes T_X]),$$
$$L(p) = \lambda_{\max}(\mathbb{E}[S(X) \otimes S(X)]).$$

Assume that $0 < L(p) < +\infty, R(p) < +\infty$. Then, we have:

$$\|\mathbb{E}[FT_X]\|_{\mathrm{HS}}^2 \leq R(p)(\mathbb{E}[F^2] - \|E[FS(X)]\|_F^2/L(p)).$$

This upper bound relates $\|\mathbb{E}[FT_X]\|_{\mathrm{HS}}$ to $\|E[FS(X)]\|_F^2$ and $\mathbb{E}[F^2]$. Let $A = -A^\top$ be a skew-symmetric matrix. Since F does not change when rotating $X + v$, we can leverage this invariance to show that any first-order quantity is related to a second-order quantity, through the equality $\mathbb{E}[FT_X(Av)] = A\mathbb{E}[FS(X)]$. This equality constrains $\mathbb{E}[FS(X)]$. In fact, summing over the canonical matrices $A_{pq} = e_p e_q^\top - e_q e_p^\top$ yields:

$$\sum_{p<q} \|A_{pq}\mathbb{E}[FS(X)]\|_F^2 = (d-1)\|\mathbb{E}[FS(X)]\|_F^2.$$

On the other side of the equality, we have:

$$\sum_{p<q} \|\mathbb{E}[FT_X(A_{pq}v)]\|_F^2 \leq \|\mathbb{E}[FT_X]\|_{\mathrm{HS}}^2 \|v\|_F^2.$$

This creates the imbalance:

$$(d-1)\|\mathbb{E}[FS(X)]\|_F^2 \leq \|\mathbb{E}[FT_X]\|_{\mathrm{HS}}^2 \|v\|_F^2,$$

which is the source of dimension-dependent contraction. This consideration yields the following theorem:

Theorem 4. (Fisher contraction for $O(d)$-invariant distributions). *Let $d \geq 2$. Assume X has a smooth $O(d)$-invariant density. Assume that $0 < L(p) < +\infty, R(p) < +\infty$. Let f be any $O(d)$-invariant release. Note $\lambda(p) = R(p)/L(p)$. Then, for any $v, H \in \mathbb{R}^{d\times k}$,*

$$I_f(v)[H] \leq L(p)\frac{\lambda(p)\|v\|_F^2}{\lambda(p)\|v\|_F^2 + d - 1}\|H\|_F^2.$$

The full proof of this theorem can be found in Appendix A.2. We can instantiate this theorem for Gaussian distributions and matrix Laplace distributions for example:

Corollary 5. *Theorem 4 provides the following Fisher information upper envelopes:*

- ***Gaussian case:*** *If $p = \mathcal{N}(0, I_d \otimes I_k)$, we have $S(X) = X$, $T_X = X \otimes X - I$, and therefore $R(p) = 2, L(p) = 1$ and:*

$$I_f(v)[H] \leq \frac{2\|v\|_F^2}{2\|v\|_F^2 + d - 1}\|H\|_F^2.$$

- ***Matrix Laplace case:*** *Assume that $dk \geq 3$. If $p(x) \propto \exp(-\beta\|x\|_F)$, we have: $L(p) = \frac{\beta^2}{dk}$ and $R(p) \leq \frac{6\beta^4}{d^2k^2}$. Therefore,*

$$I_f(v)[H] \leq \frac{\beta^2}{dk}\frac{6\beta^2\|v\|_F^2}{6\beta^2\|v\|_F^2 + dk(d-1)}\|H\|_F^2.$$

In the case where X is Gaussian, the Fisher information $I_f(v)$ associated to the observation of $f(X + v)$ can be compared to the Fisher information $I_v(X + v)$ associated to the observation of $X + v$. In fact, $I_v(X + v)[H] = \|H\|_F^2$, yielding the contraction coefficient:

$$\eta_{d,k} \leq \frac{2\|v\|_F^2}{2\|v\|_F^2 + d - 1}.$$

4. Application to privacy amplification by synthetic data release

Our bounds can be leveraged in the setting of privacy amplification by synthetic data release, improving upon the existing bounds of Pierquin et al. (2026).

Differential privacy (DP) (Dwork and Roth, 2014) is a standard framework for limiting the information leaked by machine-learning algorithms trained on sensitive data.

Definition 1 (Rényi Differential Privacy, (Mironov, 2017).). *Two datasets $\mathcal{D}, \mathcal{D}'$ of size n are adjacent if they differ in one data point. A randomized mechanism $\mathcal{M}$ satisfies (α, ε)-RDP if, for every adjacent pair $\mathcal{D}, \mathcal{D}'$,*

$$D_\alpha(\mathcal{M}(\mathcal{D})\|\mathcal{M}(\mathcal{D}')) \leq \varepsilon,$$

where, for $\alpha > 1$,

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1}\log \mathbb{E}_{x\sim Q}\left[\left(\frac{P(x)}{Q(x)}\right)^\alpha\right].$$

When $V \sim P$ and $W \sim Q$, we also write $D_\alpha(V, W) = D_\alpha(P\|Q)$.

Private models can be obtained either by perturbing the output of a learning algorithm (Chaudhuri et al., 2011; Zhang et al., 2017b; Lowy and Razaviyayn, 2024), or by adding noise during training, as in noisy gradient methods (Abadi et al., 2016). By the post-processing property of DP, any transformation of a private model remains private, making differentially private generative models (Zhang et al., 2017a; Xie et al., 2018; McKenna et al., 2019; Jordon et al., 2019; McKenna et al., 2021; Lee et al., 2022; Dockhorn et al., 2023; Bie et al., 2023) attractive. Model parameters are privately trained and synthetic data samples inherit privacy guarantees from the parameters.

Recent empirical work suggests that synthetic data may sometimes leak less information than what is predicted by

the privacy guarantee of the underlying generator (Annamalai et al., 2024; Houssiau et al., 2022) suggesting a possible privacy amplification effect. Pierquin et al. (2025) first studied this phenomenon formally for output perturbation in linear regression, and Pierquin et al. (2026) later showed that amplification can persist even when infinitely many synthetic samples are released for linear generators.

Classical DP synthetic data generation pipelines comprise two stages. First, a private algorithm $\mathcal{M}$ trained on a sensitive dataset $\mathcal{D}$ outputs generator parameters $V = \mathcal{M}(\mathcal{D})$. Then, conditionally on V, the generator samples latent variables $Z = (Z_1, \dots, Z_{n_{\text{syn}}}) \sim \mu^{\otimes n_{\text{syn}}}$ and releases: $f_V(Z) = (f_V(Z_1), \dots, f_V(Z_{n_{\text{syn}}}))$. For adjacent datasets $\mathcal{D}, \mathcal{D}'$, with $W = \mathcal{M}(\mathcal{D}')$, post-processing gives:

$$D_\alpha(f_V(Z), f_W(Z)) \leq D_\alpha(V, W).$$

Privacy amplification by synthetic data release amounts to proving a sharper inequality of the form

$$D_\alpha(f_V(Z), f_W(Z)) \leq \eta^{\text{syn}}_{\mathcal{D},\mathcal{D}'} D_\alpha(V, W), \qquad \eta^{\text{syn}}_{\mathcal{D},\mathcal{D}'} < 1,$$

where the gain comes from the additional sampling noise Z. For realistic generators, such as deep generative models, the divergence $D_\alpha(f_V(Z), f_W(Z))$ is generally intractable. We therefore study a linear proxy extending the setting of Pierquin et al. (2025). Specifically, we take

$$f_V(Z) = ZV, \qquad Z \in \mathbb{R}^{n_{\text{syn}} \times d}, \qquad V \in \mathbb{R}^{d \times k},$$

where n_{syn} is the number of synthetic samples, k the data dimension, and d the model width, and assume that Z has $\mathcal{N}(0, 1)$ iid entries. We assume that adjacent datasets yield deterministic parameters $v, w \in \mathbb{R}^{d \times k}$ through a bounded-sensitivity procedure, and privacy is enforced by Gaussian output perturbation. We note $V := X + v$, $W := X + w$, where X is a matrix with $\mathcal{N}(0, 1)$ iid entries. Therefore, the privacy loss associated to synthetic data release is:

$$D_\alpha(ZV, ZW) \leq \eta^{\text{syn}}_{\mathcal{D},\mathcal{D}'}(n_{\text{syn}}, d, k, v, w) D_\alpha(V, W),$$

with $\eta^{\text{syn}}_{\mathcal{D},\mathcal{D}'}(n_{\text{syn}}, d, k, v, w) < 1$, that we write η^{syn} for brevity. In order to prove amplification, Pierquin et al. (2026) embeds the distribution of ZV into a parametric family p_θ and computes an upper bound to the Fisher information $I(\theta)$ associated to this family. From there, they leverage the local relationship between Fisher information and Rényi divergences (van Erven and Harremos, 2014): for $\theta \in \mathbb{R}$, $D_\alpha(p_{\theta+\Delta}, p_\theta) \approx \frac{\alpha}{2} I(\theta)\Delta^2 + o(\Delta^2)$.

All $\lesssim$ bounds below are meant in this same local sense: for small $\Delta = \|v - w\|_F$, Pierquin et al. (2026) bound $D_\alpha(ZV, ZW)/D_\alpha(V, W)$ up to an $o(1)$ term, not a global finite-distance divergence contraction (see Section 3 of Pierquin et al. (2026) for more details). The value Δ here relates to the sensitivity of the output perturbation mechanism. This bound is relevant when the sensitivity is small. However, if parameters are obtained through empirical risk minimization Δ is of order $1/n$, the number of training samples (Chaudhuri et al., 2011). Therefore, the work of (Pierquin et al., 2026) has focused on deriving contraction bounds to the Fisher information only. In this context, they derived the following upper bound:

Theorem 6 (Adapted from Theorem 5.4 of (Pierquin et al., 2026))**.** *Let $v, w \in \mathbb{R}^{d \times k}$ and $Z \in \mathbb{R}^{n_{\text{syn}} \times d}$ have $\mathcal{N}(0, 1)$ iid entries, independent of X. Assume that $d \geq 3k$. Then,*

$$\eta^{\text{syn}} \lesssim \frac{2\max\{\|v\|_F, \|w\|_F\}^2}{2\max\{\|v\|_F, \|w\|_F\}^2 + \lfloor d/k \rfloor - 3}.$$

This proposition does not depend on the number of generated samples n_{syn}, and therefore implies amplification even when the number of released samples is large. However, it requires the input dimension d to be much larger than the output dimension k to imply meaningful amplification. We therefore improve upon these bounds through our analysis:

Theorem 7. *(Privacy amplification for linear generators). Let $d \geq 2$. Let $v, w \in \mathbb{R}^{d \times k}$ and $Z \in \mathbb{R}^{n_{\text{syn}} \times d}$ have $\mathcal{N}(0, 1)$ iid entries, independent of X. In the same local sense as Theorem 6,*

$$\eta^{\text{syn}} \lesssim \frac{2\max\{\|v\|_F, \|w\|_F\}^2}{2\max\{\|v\|_F, \|w\|_F\}^2 + d - 1}.$$

The full proof can be found in Appendix A.4. Our proof technique does not require $d \geq k$, and therefore improves upon the bounds of Pierquin et al. (2026), especially in settings where the input dimension is similar to the output dimension. Thus Theorem 7 directly improves the Fisher-to-Rényi coefficient of Pierquin et al. (2026), replacing $\lfloor d/k \rfloor - 3$ by $d - 1$. Note that Pierquin et al. (2026) also introduce a global criterion that allows non-asymptotic analysis without $o(1)$, and our bounds also translate to this context.

5. Releasing $f(X + v)$ when X is not invariant

In this section, we obtain upper bounds when X has an invariant component and a non-invariant component under an additive structure. Specifically, we write $X = Z + \xi$, where Z and ξ are independent, and we assume that Z is $O(d)$-invariant. This includes for example the anisotropic matrix Gaussian distribution. In this setting, the identity $\mathbb{E}[FT_X(Av)] = A\mathbb{E}[FS(X)]$ does not hold anymore, and we have to work around it.

Theorem 8. *(Fisher contraction for non $O(d)$-invariant distributions). Let $d \geq 2$. Assume that $0 < L(p) < +\infty, R(p) < +\infty$. Assume that $X = Z + \xi$, with $UZ \overset{d}{=} Z$ for all $U \in O(d)$, and Z independent of an arbitrary*

random matrix ξ. We note p the density of Z. Then every $O(d)$-invariant release satisfies:

$$I_f(v)[H] \leq L(p) \frac{\lambda(p)\mathbb{E}_\xi[\|v+\xi\|_F^2]}{\lambda(p)\mathbb{E}_\xi[\|v+\xi\|_F^2] + d - 1} \|H\|_F^2.$$

We instantiate this upper bound for anisotropic matrix Gaussian distributions in Appendix A.5. We believe that this argument can be extended to non-additive structures, for example when the distribution p of X disintegrates into an invariant and a non-invariant components. We reserve this analysis for future work.

6. Conclusion

In this paper, we studied the contraction of Fisher information under orthogonally invariant release of location families. We showed that when the base distribution is invariant, we can obtain dimension-dependent upper bounds. This analysis allows to obtain better Fisher information upper bounds than existing ones in the literature (Pierquin et al., 2026). These bounds also translate into privacy amplification by synthetic data release.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016). Deep learning with differential privacy. In *CCS*.

Annamalai, M. S. M. S., Ganev, G., and Cristofaro, E. D. (2024). "what do you want from theory alone?" experimenting with tight auditing of differentially private synthetic data generation. In *USENIX Security 24*, pages 4855–4871.

Bie, A., Kamath, G., and Zhang, G. (2023). Private GANs, revisited. *Transactions on Machine Learning Research*. Survey Certification.

Chaudhuri, K., Monteleoni, C., and Sarwate, A. D. (2011). Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12:1069–1109.

Delattre, M. and Kuhn, E. (2019). Estimating fisher information matrix in latent variable models based on the score function. *arXiv preprint arXiv:1909.06094*.

Dockhorn, T., Cao, T., Vahdat, A., and Kreis, K. (2023). Differentially Private Diffusion Models. *Transactions on Machine Learning Research*.

Dwork, C. and Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407.

Houssiau, F., Jordon, J., Cohen, S. N., Daniel, O., Elliott, A., Geddes, J., Mole, C., Rangel-Smith, C., and Szpruch, L. (2022). Tapas: a toolbox for adversarial privacy auditing of synthetic data. *SyntheticData4ML @ Neurips Workshop*.

Idier, J. and Collewet, G. (2014). Properties of Fisher information for Rician distributions and consequences in MRI. Research report.

Jordon, J., Yoon, J., and van der Schaar, M. (2019). Pategan: Generating synthetic data with differential privacy guarantees. In *International Conference on Learning Representations (ICLR)*.

Lee, J., Kim, M., Jeong, Y., and Ro, Y. (2022). Differentially private normalizing flows for synthetic tabular data generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7345–7353.

Lowy, A. and Razaviyayn, M. (2024). Output perturbation for differentially private convex optimization: Faster and more general. arXiv:2102.04704.

McKenna, R., Miklau, G., and Sheldon, D. (2021). Winning the nist contest: A scalable and general approach to differentially private synthetic data. *Journal of Privacy and Confidentiality*, 11(3).

McKenna, R., Sheldon, D., and Miklau, G. (2019). Graphical-model based estimation and inference for differential privacy. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 4435–4444.

Mironov, I. (2017). Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275.

Pierquin, C., Bellet, A., Tommasi, M., and Boussard, M. (2025). Privacy Amplification Through Synthetic Data: Insights from Linear Regression. In *ICML*.

Pierquin, C., Bellet, A., Tommasi, M., and Boussard, M. (2026). Privacy amplification persists under unlimited synthetic data release.

van Erven, T. and Harremos, P. (2014). Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820.

Xie, L., Lin, K., Wang, S., Wang, F., and Zhou, J. (2018). Differentially private generative adversarial network. arXiv:1802.06739.

Zhang, J., Cormode, G., Procopiuc, C. M., Srivastava, D., and Xiao, X. (2017a). Privbayes: Private data release via bayesian networks. *ACM Trans. Database Syst.*, 42(4).

Zhang, J., Zheng, K., Mou, W., and Wang, L. (2017b). Efficient private ERM for smooth objectives. In *IJCAI*.

Zhao, B., Walters, R., and Yu, R. (2026). Symmetry in neural network parameter spaces. *Transactions on Machine Learning Research*.

In this Appendix, we provide further discussion and detailed proofs.

A. Proofs

A.1. Proof of Proposition 9

Proposition 9. *Let $F \in L^2(\sigma(Y))$ such that $\mathbb{E}[F] = 0, \mathbb{E}[F^2] > 0$. We note:*

$$R(p) = \lambda_{\max}(\mathbb{E}[T_X \otimes T_X]),$$
$$L(p) = \lambda_{\max}(\mathbb{E}[S(X) \otimes S(X)]).$$

Assume that $0 < L(p) < +\infty, R(p) < +\infty$. Then, we have:

$$\|\mathbb{E}[FT_X]\|_{\mathrm{HS}}^2 \leq R(p)(\mathbb{E}[F^2] - \|E[FS(X)]\|_F^2/L(p)).$$

Proof. Let $\lambda = R(p)$ and $\mu = L(p)$. Since X is $O(d)$-invariant, we have $-X \overset{d}{=} X$. Then, $S(X)$ is odd and $T_{-X} = T_X$. Therefore, for any deterministic $A \in \mathbb{R}^{d\times k}$, we have: $\mathbb{E}[\langle A, S(X)\rangle_F T_X] = 0$. Then,

$$\mathbb{E}[FT_X] = \mathbb{E}[(F - \langle A, S(X)\rangle_F)T_X].$$

Now, we have:

$$\begin{aligned}
\|\mathbb{E}[FT_X]\|_{\mathrm{HS}}^2 &= \sup_{\|B\|_{\mathrm{HS}}=1} \langle \mathbb{E}[(F - \langle A, S(X)\rangle_F)T_X], B\rangle_{\mathrm{HS}}^2 \\
&= \sup_{\|B\|_{\mathrm{HS}}=1} \mathbb{E}[(F - \langle A, S(X)\rangle_F)\langle T_X, B\rangle_{\mathrm{HS}}]^2 \\
&\leq \mathbb{E}[(F - \langle A, S(X)\rangle_F)^2] \sup_{\|B\|_{\mathrm{HS}}=1} \mathbb{E}[\langle T_X, B\rangle_{\mathrm{HS}}^2] \\
&\leq \lambda \mathbb{E}[(F - \langle A, S(X)\rangle_F)^2].
\end{aligned}$$

Now, for $A = \mathbb{E}[FS(X)]/\mu$, and writing $\mathbb{E}[\langle A, S(X)\rangle_F^2] \leq \mu\|A\|_F^2$we have:

$$\begin{aligned}
\mathbb{E}[(F - \langle A, S(X)\rangle_F)^2] &= \mathbb{E}[F^2] - \frac{2}{\mu}\|\mathbb{E}[FS(X)]\|_F^2 + \frac{1}{\mu^2}\mathbb{E}[\langle \mathbb{E}[FS(X)], S(X)\rangle_F^2] \\
&\leq \mathbb{E}[F^2] - \frac{1}{\mu}\|\mathbb{E}[FS(X)]\|_F^2,
\end{aligned}$$

giving the desired result. □

A.2. Proof of Theorem 4

In order to prove this theorem, we will need the following Lemmas:

Lemma 10. *Assume that $F(X) = \phi((X+v)^\top(X+v))$. Let $A = -A^\top$ be a skew-symmetric matrix. Then,*

$$\mathbb{E}[FT_X(Av)] = A\mathbb{E}[FS(X)].$$

Proof. We let $M_F = \mathbb{E}[FT_X]$ and $b_F = \mathbb{E}[FS(X)]$. We let A be a skew symmetric matrix: $A^\top = -A$. We show that the orthogonal invariance provided by the Gram matrix operation enables to show $M_F(Av) = Ab_F$. To prove that, we define $R_t = e^{tA}$ and $T_t z = R_t(z+v) - v$, and leverage that F is invariant by T_t. Since F is $\sigma(Y)$-measurable, there exists a measurable function ϕ such that $F = \psi((X+v)^\top(X+v)) = \phi(X)$. Then,

$$\phi(T_t X) = \psi((T_t X + v)^\top(T_t X + v)) = \psi((X+v)^\top R_t^\top R_t(X+v)) = \psi((X+v)^\top(X+v)) = \phi(X).$$

Let $U \in \mathbb{R}^{d\times k}$ be a matrix. Then,

$$\mathbb{E}[\phi(T_t X)\langle U, S(X)\rangle_F] = \mathbb{E}[\phi(X)\langle U, S(X)\rangle_F].$$

The map T_t is an affine bijection with inverse T_{-t}. Since R_t is orthogonal and $\det R_t = 1$, the Jacobian is 1. Therefore for any integrable test function h, the change of variables $z \mapsto T_t z$ gives $\mathbb{E}[h(T_t X)] = \mathbb{E}[h(X)\rho_t(X)]$, with $\rho_t(X) = p(T_{-t}X)/p(X)$. Then,

$$\mathbb{E}[\phi(X)\langle U, S(X)\rangle_F] = \mathbb{E}[\phi(X)\langle U, S(T_{-t}X)\rangle_F \rho_t(X)].$$

Then, $\mathbb{E}[\phi(X)\langle U, S(T_{-t}X)\rangle_F \rho_t(X)]$ does not depend on t. We can differentiate at $t = 0$, since the integrand is differentiable and $T_{-t}X$ is affine in X up to smooth coefficients and p is assumed to be differentiable. Then, since

$$\langle S(X), AX\rangle_F = 0, \qquad DS(X)[AX] = AS(X),$$

we identify:

$$\langle U, M_F(Av)\rangle_F = \mathbb{E}[F(\langle Av, S(X)\rangle_F \langle U, S(X)\rangle_F - \langle DS(X)[Av], U\rangle_F)].$$

Therefore, we obtain:

$$\langle U, M_F(Av)\rangle_F = \mathbb{E}[F\langle U, AS(X)\rangle_F] = \langle U, A\,E[FS(X)]\rangle_F.$$

Since this equality holds for all $U \in \mathbb{R}^{d\times k}$, we have the desired result:

$$M_F(Av) = Ab_F.$$

□

Now, we can prove the theorem:

Theorem 4. (Fisher contraction for $O(d)$-invariant distributions). *Let $d \geq 2$. Assume X has a smooth $O(d)$-invariant density. Assume that $0 < L(p) < +\infty, R(p) < +\infty$. Let f be any $O(d)$-invariant release. Note $\lambda(p) = R(p)/L(p)$. Then, for any $v, H \in \mathbb{R}^{d\times k}$,*

$$I_f(v)[H] \leq L(p)\frac{\lambda(p)\|v\|_F^2}{\lambda(p)\|v\|_F^2 + d - 1}\|H\|_F^2.$$

Proof. Let $A_{pq} = E_{pq} - E_{qp}$, with $1 \leq p < q \leq d$ be skew-symmetric matrices.

Then, by Lemma 10, we have: $M_F(A_{pq}v) = A_{pq}b_F$.

Therefore, by summing over all the combinations,

$$\sum_{1\leq p<q\leq d} \|M_F(A_{pq}v)\|_F^2 = \sum_{1\leq p<q\leq d} \|A_{pq}b_F\|_F^2.$$

Also,

$$\sum_{1\leq p<q\leq d} \|A_{pq}\mathbb{E}[S(X)F]\|_F^2 = \sum_{1\leq p<q\leq d} \|\mathbb{E}[S(X)F]_{p,\cdot}\|_2^2 + \|\mathbb{E}[S(X)F]_{q,\cdot}\|_2^2.$$

Each row appears exactly $d-1$ times in the sum. Therefore,

$$\sum_{1\leq p<q\leq d} \|A_{pq}\mathbb{E}[S(X)F]\|_F^2 = (d-1)\|\mathbb{E}[S(X)F]\|_F^2.$$

Also, for all $v, w \in \mathbb{R}^{d\times k}$, $\sum_{1\leq p<q\leq d}\langle w, A_{pq}v\rangle_F^2 \leq \|w\|_F^2\|v\|_F^2$. In fact,

$$\begin{aligned}
\sum_{1\le p<q\le d} \langle w, A_{pq} v\rangle_F^2 &= \sum_{1\le p<q\le d} \left(\sum_{i=1}^{d}\sum_{j=1}^{k}\sum_{l=1}^{d} w_{ij}(A_{pq})_{il} v_{lj} \right)^2 \\
&= \sum_{1\le p<q\le d} \left(\sum_{i=1}^{d}\sum_{j=1}^{k}\sum_{l=1}^{d} w_{ij}(\delta_{ip}\delta_{lq} - \delta_{iq}\delta_{lp}) v_{lj} \right)^2 \\
&= \sum_{1\le p<q\le d} \left(\sum_{j=1}^{k} w_{pj} v_{qj} - w_{qj} v_{pj} \right)^2 \\
&= \sum_{1\le p<q\le d} \left(\langle w_p, v_q\rangle - \langle v_p, w_q\rangle \right)^2
\end{aligned}$$

Then, by Minkovski inequality,

$$\left(\sum_{1\le p<q\le d} \left(\langle w_p, v_q\rangle - \langle v_p, w_q\rangle \right)^2 \right)^{1/2} \le \sum_{j=1}^{k} \left(\sum_{1\le p<q\le d} (w_{pj} v_{qj} - w_{qj} v_{pj})^2 \right)^{1/2}.$$

By Lagrange's identity:

$$\sum_{1\le p<q\le d} (w_{pj} v_{qj} - w_{qj} v_{pj})^2 = \|w_{\cdot,j}\|^2 \|v_{\cdot,j}\|^2 - \langle w_{\cdot,j}, v_{\cdot,j}\rangle^2$$

Therefore,

$$\begin{aligned}
\left(\sum_{1\le p<q\le d} \left(\langle w_p, v_q\rangle - \langle v_p, w_q\rangle \right)^2 \right)^{1/2} &\le \sum_{j=1}^{k} \left(\|w_{\cdot,j}\|^2 \|v_{\cdot,j}\|^2 - \langle w_{\cdot,j}, v_{\cdot,j}\rangle^2 \right)^{1/2} \\
&\le \sum_{j=1}^{k} \|w_{\cdot,j}\| \|v_{\cdot,j}\| \le \|w\|_F \|v\|_F.
\end{aligned}$$

Now, let (e_i) be a basis of $\mathbb{R}^{d\times k}$. By symmetry of M_F:

$$\begin{aligned}
\sum_{1\le p<q\le d} \|M_F(A_{pq} v)\|_F^2 &= \sum_{j=1}^{dk} \sum_{1\le p<q\le d} \langle e_j, M_F(A_{pq} v)\rangle_F^2 \\
&= \sum_{j=1}^{dk} \sum_{1\le p<q\le d} \langle M_F^*(e_j), A_{pq} v\rangle_F^2 \\
&\le \sum_{j=1}^{dk} \|M_F^*(e_j)\|_F^2 \|v\|_F^2 \\
&\le \|M_F\|_{\mathrm{HS}}^2 \|v\|_F^2.
\end{aligned}$$

Leveraging Proposition 9, we obtain:

$$(d-1)\|b_F\|_F^2 \le \|M_F\|_F^2 \|v\|_F^2 \le \|v\|_F^2 R(p)(\mathbb{E}[F^2] - \|b_F\|_F^2 / L(p)).$$

Then,

$$\|b_F\|_F^2 \le \frac{R(p)\|v\|_F^2}{R(p)/L(p)\|v\|_F^2 + d - 1} \mathbb{E}[F^2].$$

We leverage our variational representation of Fisher information:

$$I_f(v)[H] = \sup_{F \in L^2(\sigma(Y)), \mathbb{E}[F]=0, \mathbb{E}[F^2]>0} \frac{\mathbb{E}[F\langle H, S(X)\rangle_F]^2}{\mathbb{E}[F^2]}$$

Then, Let $F \in L^2(\sigma(Y))$ such that $\mathbb{E}[F] = 0$ and $\mathbb{E}[F^2] > 0$, and:

$$\|\mathbb{E}[S(X)F]\|_F^2 \leq L(p) \frac{\lambda(p)\|v\|_F^2}{\lambda(p)\|v\|_F^2 + d - 1} \mathbb{E}[F^2].$$

Moreover, by the Cauchy-Schwarz inequality, $\mathbb{E}[F\langle H, S(X)\rangle_F]^2 \leq \|H\|_F^2 \|\mathbb{E}[FS(X)]\|_F^2$, yielding the result. □

A.3. Proof of Corollary 5

Corollary 11. *Theorem 4 provides the following Fisher information upper envelopes:*

- ***Gaussian case:*** *If $p = \mathcal{N}(0, I_d \otimes I_k)$, we have $S(X) = X$, $T_X = X \otimes X - I$, and therefore $R(p) = 2, L(p) = 1$ and:*

$$I_f(v)[H] \leq \frac{2\|v\|_F^2}{2\|v\|_F^2 + d - 1} \|H\|_F^2.$$

- ***Matrix Laplace case:*** *Assume that $dk \geq 3$. If $p(x) \propto \exp(-\beta\|x\|_F)$, we have: $L(p) = \frac{\beta^2}{dk}$ and $R(p) \leq \frac{6\beta^4}{d^2k^2}$. Therefore,*

$$I_f(v)[H] \leq \frac{\beta^2}{dk} \frac{6\beta^2\|v\|_F^2}{6\beta^2\|v\|_F^2 + dk(d-1)} \|H\|_F^2.$$

Proof. For Gaussian distributions, the proof is elementary. In fact, the score function writes $S(X) = X$, and the second-order tensor writes $T_X = X \otimes X - I$. Therefore, we have: $L(p) = 1, R(p) = 2$, yielding the result. For the Matrix Laplace case, we refer to the following corollary. □

Corollary 12 (Frobenius-radial Laplace noise)**.** *Let $m = dk$, and let $X \in \mathbb{R}^{d\times k} \simeq \mathbb{R}^m$ have density:*

$$p_\beta(x) = Z_{\beta,m}^{-1} \exp(-\beta\|x\|_F), \qquad \beta > 0.$$

Assume $m \geq 3$. Then,

$$L(p_\beta) = \frac{\beta^2}{m}, \qquad R(p_\beta) \leq \frac{\beta^4}{m^2}\kappa_m,$$

where:

$$\kappa_m := \frac{2m(m^2 - m - 1)}{(m+2)(m-1)(m-2)} + \frac{m}{m-2} \leq 6.$$

Therefore,

$$I_f(v)[H] \leq \frac{\beta^2}{m} \frac{\beta^2\kappa_m\|v\|_F^2}{\beta^2\kappa_m\|v\|_F^2 + m(d-1)} \|H\|_F^2.$$

Proof. We identify $\mathbb{R}^{d\times k}$ with $\mathbb{R}^m$, $m = dk$, and write

$$R = \|X\|_F, \qquad U = \frac{X}{\|X\|_F}$$

Then U is uniform on $\mathbb{S}^{m-1}$ and independent of R. We let $\rho := \beta R \sim \Gamma(m, 1)$. The location score is:

$$S(X) = -\nabla \log p_\beta(X) = \beta U.$$

Therefore, $\mathbb{E}[S(X) \otimes S(X)] = \beta^2\mathbb{E}[U \otimes U] = \frac{\beta^2}{m}I$, and $L(p_\beta) = \frac{\beta^2}{m}$. Also, for $x \neq 0$,

$$DS(x)[H] = \frac{\beta}{\|x\|_F}(I - UU^\top)H.$$

Thus,

$$T_X = S(X) \otimes S(X) - DS(X) = \left(\beta^2 + \frac{\beta}{R}\right) UU^\top - \frac{\beta}{R} I = \beta^2 \left[\left(1 + \frac{1}{\rho}\right) UU^\top - \frac{1}{\rho} I\right].$$

Let A be any matrix with $\|A\|_{\mathrm{HS}} = 1$. Since T_X is symmetric, only the symmetric part of A contributes to $\langle A, T_X \rangle_{\mathrm{HS}}$, so we may assume without loss of generality that $A = A^\top$. Write:

$$a := \mathrm{tr}(A), \qquad B := A - \frac{a}{m} I.$$

Then $\mathrm{tr}(B) = 0$ is traceless and:

$$\|B\|_{\mathrm{HS}}^2 + \frac{a^2}{m} = \|A\|_{\mathrm{HS}}^2 = 1.$$

Therefore,

$$\langle A, T_X \rangle_{\mathrm{HS}} = \beta^2 \left[\left(1 + \frac{1}{\rho}\right) U^\top B U + a \left(\frac{1}{m} - \frac{m-1}{m\rho}\right)\right].$$

The cross term vanishes because U is independent of ρ and $\mathbb{E}_U[U^\top B U] = \mathrm{tr}(B)/m = 0$.

Hence:

$$\mathbb{E}[\langle A, T_X \rangle_{\mathrm{HS}}^2] = \beta^4 \mathbb{E}\left[\left(1 + \frac{1}{\rho}\right)^2\right] \mathbb{E}[(U^\top B U)^2] + \beta^4 a^2 \mathbb{E}\left[\left(\frac{1}{m} - \frac{m-1}{m\rho}\right)^2\right].$$

Using standard moments,

$$\mathbb{E}[(U^\top B U)^2] = \frac{\mathrm{tr}(B)^2 + 2\|B\|_{\mathrm{HS}}^2}{m(m+2)} = \frac{2\|B\|_{\mathrm{HS}}^2}{m(m+2)},$$

and the Gamma moments of ρ, we get:

$$\mathbb{E}\left[\left(1 + \frac{1}{\rho}\right)^2\right] = 1 + \frac{2}{m-1} + \frac{1}{(m-1)(m-2)} = \frac{m^2 - m - 1}{(m-1)(m-2)}.$$

Moreover,

$$\mathbb{E}\left[\left(\frac{1}{m} - \frac{m-1}{m\rho}\right)^2\right] = \frac{1}{m^2(m-2)}.$$

Therefore,

$$\mathbb{E}[\langle A, T_X \rangle_{\mathrm{HS}}^2] \leq \beta^4 \frac{2(m^2 - m - 1)}{m(m+2)(m-1)(m-2)} \|B\|_{\mathrm{HS}}^2 + \beta^4 \frac{a^2}{m^2(m-2)}.$$

Since $\|B\|_{\mathrm{HS}}^2 \leq 1$ and $a^2 \leq m$, this gives the upper bound

$$\mathbb{E}[\langle A, T_X \rangle_{\mathrm{HS}}^2] \leq \beta^4 \left[\frac{2(m^2 - m - 1)}{m(m+2)(m-1)(m-2)} + \frac{1}{m(m-2)}\right].$$

Taking the supremum over $\|A\|_{\mathrm{HS}} = 1$, we obtain $R(p_\beta) \leq \frac{\beta^4}{m^2} \kappa_m$, where

$$\kappa_m := \frac{2m(m^2 - m - 1)}{(m+2)(m-1)(m-2)} + \frac{m}{m-2} \leq 6.$$

In particular, $\kappa_m \overset{m \to \infty}{\to} 3$.

To avoid non-smoothness at 0, we can apply the same considerations to:

$$p_{\beta,\varepsilon}(x) \propto \exp\left(-\beta \sqrt{\|x\|_F^2 + \varepsilon^2}\right),$$

and then let $\varepsilon \downarrow 0$. The corresponding first and second scores converge in $L^2(p_{\beta,\varepsilon})$ to the expressions above whenever $m > 2$, so the upper bounds converge to the claimed bounds for $L(p_\beta)$ and $R(p_\beta)$. □

A.4. Proof of Theorem 7

We restate the local criterion relating Rényi divergences to Fisher information:

Proposition 13 (Local relationship between Rényi divergences and Fisher information)**.** *Let $\{P_\theta, \theta \in \Theta \subset \mathbb{R}\}$ be a family of probability distributions. Let $\theta \in \Theta$. Let $\Delta > 0$ such that $\theta + \Delta \in \Theta$. Then, if the family $\{P_\theta\}$ is sufficiently regular,*

$$D_\alpha(P_{\theta+\Delta}, P_\theta) = \frac{\alpha}{2} I(\theta)\Delta^2 + o(\Delta^2).$$

Therefore, we can prove our Theorem:

Theorem 7. *(Privacy amplification for linear generators). Let $d \geq 2$. Let $v, w \in \mathbb{R}^{d\times k}$ and $Z \in \mathbb{R}^{n_{\text{syn}}\times d}$ have $\mathcal{N}(0,1)$ iid entries, independent of X. In the same local sense as Theorem 6,*

$$\eta^{\text{syn}} \lesssim \frac{2\max\{\|v\|_F, \|w\|_F\}^2}{2\max\{\|v\|_F, \|w\|_F\}^2 + d - 1}.$$

Proof. Let $v, w \in \mathbb{R}^{d\times k}$. If $v = w$, the claim is trivial. Assume that $\Delta := \|v - w\|_F > 0$ and define:

$$u := \frac{v - w}{\Delta}, \qquad \theta_s := w + su, \qquad s \in [0, \Delta].$$

Let $P_s := \mathcal{L}((X + \theta_s)^\top (X + \theta_s))$. Then, $P_0 = \mathcal{L}(W^\top W)$ and $P_\Delta = \mathcal{L}(V^\top V)$. The family $\{P_s,\ s \in [0, \Delta]\}$ is a one-dimensional parametric family. Its scalar Fisher information is $\tilde{I}(s) := \mathbb{E}_{Y\sim P_s}[(\partial_s \log p_s(Y))^2]$, and by the chain rule, $\tilde{I}(s) = I(\theta_s)[u]$, where $I(\theta)[H]$ denotes the directional Fisher information of the non-central family. Applying Proposition 13 and leveraging Corollary 5, we obtain:

$$D_\alpha(V^\top V, W^\top W) \leq \frac{\alpha}{2}\left(\frac{2R(v,w)^2}{2R(v,w)^2 + d - 1 + o(1)}\right)\Delta^2,$$

where $R(v, w) = \max\{\|v\|_F, \|w\|_F\}$. Since $D_\alpha(V, W) = \frac{\alpha}{2}\Delta^2$ for Gaussian output perturbation, this gives the claimed local ratio. Finally, ZV is an $O(d)$-invariant randomized observation of V, so the same bound applies by data processing from the Gram experiment.

Note that this upper bound can be optimized by leveraging $\inf_{U\in O(d)} \|Uv - w\|_F$ instead of $\|v - w\|_F$.

□

Remark 14. *In the case $k = 1$, (Idier and Collewet, 2014) give the following lower bound on Fisher information, for the amplitude parameter θ:*

$$I(\theta) \geq \frac{2\|\theta\|^2}{2\|\theta\|^2 + d}.$$

Therefore, we can also extend our results to a lower local envelope on the Rényi divergence $D_\alpha(V^\top V, W^\top W)$ for the case $k = 1$:

$$D_\alpha(V^\top V, W^\top W) \geq \frac{\alpha}{2}\frac{2\min\{\|v\|, \|w\|\}^2}{2\min\{\|v\|, \|w\|\}^2 + d}(\|v\| - \|w\|)^2 + o((\|v\| - \|w\|)^2).$$

This considerations show that our approach is quite tight.

A.5. Proof of Theorem 8

Theorem 8. *(Fisher contraction for non $O(d)$-invariant distributions). Let $d \geq 2$. Assume that $0 < L(p) < +\infty, R(p) < +\infty$. Assume that $X = Z + \xi$, with $UZ \overset{d}{=} Z$ for all $U \in O(d)$, and Z independent of an arbitrary random matrix ξ. We note p the density of Z. Then every $O(d)$-invariant release satisfies:*

$$I_f(v)[H] \leq L(p)\frac{\lambda(p)\mathbb{E}_\xi[\|v + \xi\|_F^2]}{\lambda(p)\mathbb{E}_\xi[\|v + \xi\|_F^2] + d - 1}\|H\|_F^2.$$

Proof. We assume that $X = Z + \xi$. Then, by the data processing inequality, the Fisher information related to the observation of $f(X+v)$ is less than the Fisher information associated to the release of the couple $(f(X+v), \xi)$. Conditional on ξ, X is $O(d)$-invariant with shift $\xi + v$. Then, we leverage the chain rule for Fisher information:

$$\begin{aligned} I_f(v)[H] \leq I_{f,\xi}(v)[H] &\leq \mathbb{E}_\xi[I_{f|\xi}(v)[H]] \\ &\leq L(p)\mathbb{E}_\xi\left[\frac{\lambda(p)\|v+\xi\|_F^2}{\lambda(p)\|v+\xi\|_F^2 + d - 1}\right]\|H\|_F^2 \\ &\leq L(p)\frac{\lambda(p)\mathbb{E}_\xi[\|v+\xi\|_F^2]}{\lambda(p)\mathbb{E}_\xi[\|v+\xi\|_F^2] + d - 1}\|H\|_F^2, \end{aligned}$$

by leveraging Jensen's inequality, yielding the result. □

We can apply these bounds for anisotropic Gaussian noise. In fact, when $X \sim \mathcal{MN}_{d\times k}(0, \Sigma, I_k)$, X is not $O(d)$-invariant whenever $\Sigma \neq I_d$. Fortunately, noting $\tau^2 < \lambda_{\min(\Sigma)}$, we have $X = \tau N + \zeta_\tau$, where $\zeta \sim \mathcal{MN}_{d\times k}(0, \Gamma_\tau, I_k)$ and $\Gamma_\tau := \Sigma - \tau^2 I_d \succeq 0$, for which we can apply Theorem 8. When $\Sigma \approx I_d$, we can therefore obtain contraction upper bounds while X is not exactly $O(d)$-invariant.