Error analysis of a compositional score-based algorithm for simulation-based inference

Camille Touron
Univ. Grenoble Alpes, Inria
CNRS, Grenoble INP, LJK, France
camille.touron@inria.fr

Gabriel V. Cardoso
Geostatistics team, Centre for geosciences and geoengineering
Mines Paris, PSL University, Fontaineableau, France
gabriel.victorino_cardoso@minesparis.psl.eu

Julyan Arbel
Univ. Grenoble Alpes, Inria
CNRS, Grenoble INP, LJK, France
julyan.arbel@inria.fr

Pedro L. C. Rodrigues
Univ. Grenoble Alpes, Inria
CNRS, Grenoble INP, LJK, France
pedro.rodrigues@inria.fr

Abstract

Simulation-based inference (SBI) has become a widely used framework in applied sciences for estimating the parameters of stochastic models that best explain experimental observations. A central question in this setting is how to effectively combine multiple observations in order to improve parameter inference and obtain sharper posterior distributions. Recent advances in score-based diffusion methods address this problem by constructing a compositional score, obtained by aggregating individual posterior scores within the diffusion process. While it is natural to suspect that the accumulation of individual errors may significantly degrade sampling quality as the number of observations grows, this important theoretical issue has so far remained unexplored. In this paper, we study the compositional score produced by the GAUSS algorithm of Linhart et al. (2024) and establish an upper bound on its mean squared error in terms of both the individual score errors and the number of observations. We illustrate our theoretical findings on a Gaussian example, where all analytical expressions can be derived in a closed form.

1 Introduction

Probabilistic approaches to inverse problems (Tarantola, 2005) aim at inferring parameters $\boldsymbol{\theta}$ of a stochastic model from its outputs $\boldsymbol{x}$ through the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$. Directly sampling from this posterior is often difficult in practice, since the associated likelihood $p(\boldsymbol{x}|\boldsymbol{\theta})$ is frequently intractable. To address this, simulation-based inference with conditional score-based modeling (Sharrock et al., 2024) approximates the posterior by learning a time-dependent conditional estimate $s_\phi(\boldsymbol{\theta}, \boldsymbol{x}, t)$ of the score of noisy versions of the posterior, $\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta}|\boldsymbol{x})$. Importantly, this training relies only on joint samples $(\boldsymbol{\theta}_i, \boldsymbol{x}_i) \sim p(\boldsymbol{\theta}, \boldsymbol{x})$ which can be obtained sequentially as $\boldsymbol{\theta}_i \sim \lambda(\boldsymbol{\theta})$, the prior, and then $\boldsymbol{x}_i|\boldsymbol{\theta}_i \sim p(\boldsymbol{x}|\boldsymbol{\theta}_i)$, thus circumventing the need to evaluate the likelihood (see Appendix A.1 for details). A central question in this line of work is how the denoising score matching mean squared error, denoted $\epsilon^2_{\mathrm{DSM}}$, affects the quality of samples obtained via a backward diffusion process (see Figure 1). Recent results by Gao et al. (2025) provide explicit bounds on the Wasserstein error of the approximate samples in terms of $\epsilon^2_{\mathrm{DSM}}$ and the discretization hyperparameters (number

Preprint. Under review.

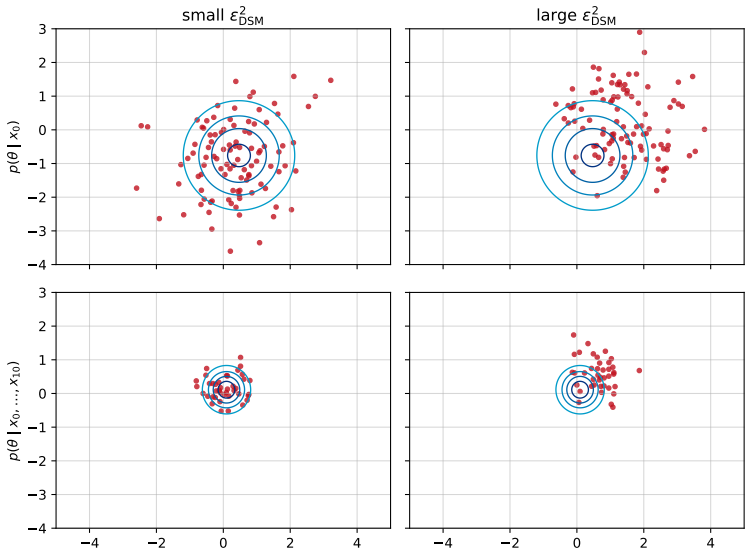

Figure 1: Contours represent the true target posterior distribution p (see Appendix A.2.1 for details), either conditioned on a single observation (top row) or eleven of them (bottom row): increasing the number of conditional observations sharpens the posterior and allows for better inference. Scattered points in red are samples from $\tilde{p}$ obtained by a diffusion process with an inexact score estimate with mean squared error ϵ^2_{DSM}: larger errors tend to bias the sampling and degrade its quality.

of steps, step size) used in the backward process. These results imply that one can reach a prescribed sampling accuracy (in Wasserstein distance) either by tuning the diffusion hyperparameters or by improving the accuracy of the score estimator.

We extend this setting to n IID observations, with the goal of inferring parameters from the posterior $p(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})$. In practice, additional observations improve inference, as the posterior concentrates with growing n (see Figure 1). Recent works (Linhart et al., 2024; Geffner et al., 2023; Arruda et al., 2025; Gloeckler et al., 2025) propose a compositional score approach: instead of directly approximating $\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})$ with a highly complex network, they aggregate individual posterior scores $s_\phi(\boldsymbol{\theta}, \boldsymbol{x}_j, t) \approx \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta}|\boldsymbol{x}_j)$. The resulting compositional score can then be used in backward diffusion or annealed Langevin dynamics to sample from the multi-observation posterior. Importantly, if we can bound the error of the compositional score estimate, then we can guarantee convergence of our sampling to the target multi-observation posterior in terms of Wasserstein distance (Gao et al., 2025).

However, while the error of each individual score estimate can be controlled during training time, the error of the aggregated compositional score arises from the accumulation of individual errors as n grows. To the best of our knowledge, this accumulation effect has not been theoretically analyzed so far. Here we address this gap: focusing on the compositional score (1) defined by the `GAUSS` algorithm of Linhart et al. (2024), we derive in Proposition 2 an upper bound on the mean squared error (MSE) of its estimate (2) as a function of the n individual score errors.

2 Background on compositional score estimation with `GAUSS`

Linhart et al. (2024) adopt the common assumption (Sohl-Dickstein et al., 2015) that the backward kernels of the diffused prior and the diffused individual posteriors can be well approximated by Gaussian distributions for all times $t \in [0, T]$ of a diffusion process, namely

$$\hat{q}^{\lambda}_{0|t}(\boldsymbol{\theta}_0|\boldsymbol{\theta}_t) = \mathcal{N}\big(\boldsymbol{\theta}_0; \mu_{t,\lambda}(\boldsymbol{\theta}_t), \Sigma_{t,\lambda}(\boldsymbol{\theta})\big) \quad \text{and} \quad \hat{q}_{0|t}(\boldsymbol{\theta}_0|\boldsymbol{\theta}_t, \boldsymbol{x}_j) = \mathcal{N}\big(\boldsymbol{\theta}_0; \mu_{t,j}(\boldsymbol{\theta}_t), \Sigma_{t,j}(\boldsymbol{\theta})\big),$$

with $j \in \{1, \ldots, n\}$. In addition, their `GAUSS` algorithm assumes that both the prior and each individual posterior are themselves Gaussian, $\mathcal{N}(\boldsymbol{\theta}; \mu_\lambda, \Sigma_\lambda)$ and $\mathcal{N}(\boldsymbol{\theta}; \mu_j, \Sigma_j)$. Under this assumption, the diffused covariances $\Sigma_{t,\lambda}$ and $\Sigma_{t,j}$ are fully determined and, importantly, no longer depend on $\boldsymbol{\theta}$. With these simplifications, the diffused compositional score takes the form

$$\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta}|\boldsymbol{x}_{1:n}) = \Lambda^{-1} \left(\sum_{j=1}^{n} \Sigma_{t,j}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta}|\boldsymbol{x}_j) - (n-1)\Sigma_{t,\lambda}^{-1} \nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta}) \right), \qquad (1)$$

where $\Lambda = \sum_j \Sigma_{t,j}^{-1} - (n-1)\Sigma_{t,\lambda}^{-1}$. Note that the compositional score (1) does not stem from the simple addition of the n individual scores, as it aims at correctly approximating the true score of a

noisy version of the multi-observation posterior for some specific noise level prescribed by a classical diffusion process. For most priors used in the SBI literature (Gaussian, Uniform, log-Normal), the diffused score $\nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta})$ has a closed-form expression and therefore introduces no error (Sharrock et al., 2024). In contrast, Equation (1) also involves the scores of the individual posteriors, which are only available through the approximations $s_\phi(\boldsymbol{\theta}, \boldsymbol{x}_j, t)$ and thus contribute to errors $\epsilon^2_{\mathrm{DSM},j}$. A further source of bias comes from the estimation of the covariance matrices Σ_λ and Σ_j. Covariance Σ_λ can be approximated by the empirical covariance of prior samples. However, for each posterior one must simulate a backward diffusion process using the approximate score $s_\phi(\boldsymbol{\theta}, \boldsymbol{x}_j, t)$ to generate samples from $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{x}_j) \approx p(\boldsymbol{\theta}|\boldsymbol{x}_j)$, and then compute the empirical covariance $\tilde{\Sigma}_j$. The estimator for the compositional score (1) is then written as:

$$s(\boldsymbol{\theta}, \boldsymbol{x}_{1:n}, t) = \tilde{\Lambda}^{-1} \left(\sum_{j=1}^{n} \tilde{\Sigma}_{t,j}^{-1} s_\phi(\boldsymbol{\theta}, \boldsymbol{x}_j, t) - (n-1)\tilde{\Sigma}_{t,\lambda}^{-1} \nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta}) \right). \tag{2}$$

In what follows, we derive an upper bound on the mean squared error of this estimator in terms of the individual score errors $\epsilon^2_{\mathrm{DSM},j}$, the errors incurred in estimating the precision matrices Σ_λ^{-1} and Σ_j^{-1}, and the number of observations n. We use $\|.\|$ to refer either to the Euclidean norm when applied to vectors, or spectral norm when applied to matrices.

3 Contributions

We first derive an upper bound on the error of the estimator for the precision matrices, noting that for the `GAUSS` algorithm the errors $\|\Sigma_{t,j}^{-1} - \tilde{\Sigma}_{t,j}^{-1}\|$ at each $t \in [0, T]$ are directly related to the initial estimation error $\|\Sigma_j^{-1} - \tilde{\Sigma}_j^{-1}\|$, which in turn depends on the sampling quality from each individual posterior. Since Gaussian distributions are smooth log-concave, we can use the aforementioned results from Gao et al. (2025) to tune the discretization hyperparameters of each backward diffusion process and drive the individual score error $\epsilon^2_{\mathrm{DSM},j}$ sufficiently low to achieve a prescribed small Wasserstein error. The following proposition (proof in Appendix A.3.2) provides a bound on the precision estimation error as a function of a fixed Wasserstein error η.

Proposition 1 (Precision matrix error)**.** *Choosing* $0 < \eta < \min\left(\sqrt{\frac{\|\Sigma\|}{2}}, f(\|\Sigma\|, \|\Sigma^{-1}\|)\right)$ *we get*

$$\mathcal{W}_2(p, \tilde{p}) \leq \eta \Rightarrow \|\tilde{\Sigma} - \Sigma\| \leq \gamma \Rightarrow \|\tilde{\Sigma}^{-1} - \Sigma^{-1}\| \leq \frac{\gamma \|\Sigma^{-1}\|^2}{1 - \gamma\|\Sigma^{-1}\|},$$

where

$$\gamma = \left(2\sqrt{2\|\Sigma\|}\eta + \eta^2(1+\sqrt{2})\right) \frac{\|\Sigma\|}{\|\Sigma\| - \eta\sqrt{2\|\Sigma\|}} > 0,$$

and function $f : \mathbb{R}_+^2 \to \mathbb{R}_+$ *is defined in Appendix A.3.2.*

We now propose an upper bound on the MSE between the compositional score (1) and its estimate (2) for any time $t \in [0, T]$ of the diffusion process as a function of the individual score errors, the precision estimation errors and the number of observations n. For simplicity, we assume a Gaussian prior, leading to closed-form formula for the corresponding scores and the precision matrices $\Sigma_{t,\lambda}^{-1}$ (a more general result is stated in Appendix A.5.2). We also assume that individual posterior score errors $\epsilon^2_{\mathrm{DSM},j}$(resp. precision errors) are bounded by the same constant $\epsilon^2_{\mathrm{DSM}}$ (resp. ϵ). Note that in practice, the error ϵ can be directly linked to the Wasserstein error η of each diffusion process, and thus indirectly to $\epsilon^2_{\mathrm{DSM}}$, as stated in Proposition 1. Finally, all constants, except ϵ, are time-dependent (proof in Appendix A.4).

Proposition 2 (Compositional score error)**.** *Let* η *be chosen as in Proposition 1 and denote*

$$M = \max_j \|\Sigma_{t,j}^{-1}\| \quad \textit{with} \quad L = \max_j \mathbb{E}_{\boldsymbol{\theta} \sim p_t(\cdot|\boldsymbol{x}_{1:n})} \left(\|\nabla \log p_t(\boldsymbol{\theta}|\boldsymbol{x}_j)\|^2 \right)$$

$$M_\lambda \geq \|\Sigma_{t,\lambda}^{-1}\| \quad \textit{with} \quad L_\lambda \geq \mathbb{E}_{\boldsymbol{\theta} \sim p_t(\cdot|\boldsymbol{x}_{1:n})} \left(\|\nabla \log \lambda_t(\boldsymbol{\theta})\|^2 \right).$$

Suppose that $\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}(\|\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta}|\boldsymbol{x}_j) - s_\phi(\boldsymbol{\theta}, \boldsymbol{x}_j, t)\|^2) \leq \epsilon^2_{\text{DSM}}$ *and* $\|\Sigma_j^{-1} - \tilde{\Sigma}_j^{-1}\| \leq \epsilon$ *for all* $j = 1, \dots, n$ *such that* $n\epsilon < \frac{1}{\|\Lambda^{-1}\|}$. *Then it holds*

$$\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}\left[\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}) - s(\boldsymbol{\theta}, \boldsymbol{x}_{1:n}, t)\|^2\right] \leq \left[(n-1)\|\Lambda^{-1}\|\sqrt{L_\lambda}\left(\frac{n\epsilon\|\Lambda^{-1}\|M_\lambda}{1-n\epsilon\|\Lambda^{-1}\|}\right)\right.$$
$$\left. + n\|\Lambda^{-1}\|(\sqrt{L} + \epsilon_{\text{DSM}})\left(\epsilon + \frac{n\epsilon\|\Lambda^{-1}\|(M+\epsilon)}{1-n\epsilon\|\Lambda^{-1}\|}\right) + \|\Lambda^{-1}\|nM\epsilon_{\text{DSM}}\right]^2.$$

If ϵ_{DSM} *is sufficiently small and accompanied by a proper choice of diffusion hyperparameters such that* $\mathcal{W}_2(\tilde{p}(\boldsymbol{\theta} \mid \boldsymbol{x}_j), p(\boldsymbol{\theta} \mid \boldsymbol{x}_j)) \leq \eta$ *for all* $j = 1, \dots, n$ *(Proposition 4 in Gao et al., 2025), then the precision estimation error* ϵ *can be further bounded using Proposition 1 and* η.

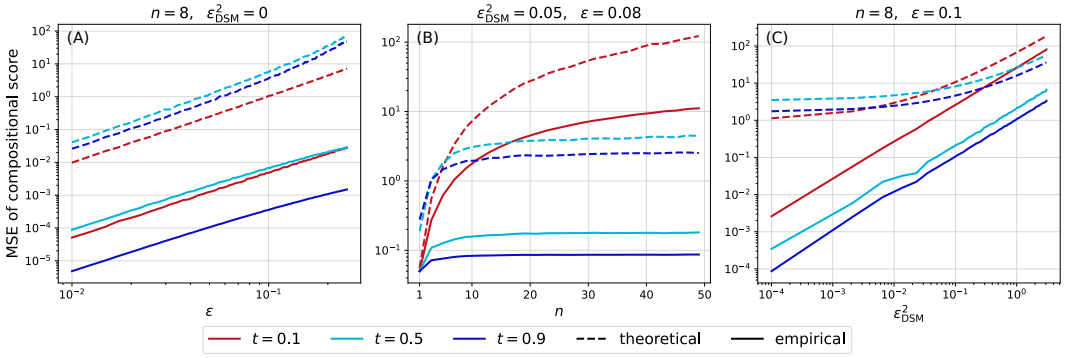

Figure 2: Solid lines stand for the evolution of the empirical MSE of the compositional score estimate (2) computed with the algorithm `GAUSS` at different times $t \in [0, 1]$ of a diffusion process for the 2D Gaussian example (see Appendix A.2.1). Dashed lines represent the evolution of the theoretical bound of the aforementioned compositional score error derived in Proposition 2. The empirical and theoretical evolutions are represented with respect to (A) the precision estimation error ϵ assuming exact individual scores are known, (B) the number of conditional observations n and (C) the individual score error ϵ^2_{DSM} that contributes both directly to the compositional score error and indirectly through the precision estimation. The choice of the fixed parameters, especially ϵ in panel (B) and (C), is discussed in detail in Appendix A.2.2.

Remark. In the assumptions of Proposition 2, we bound the individual score error in expectation over the multi-observation posterior, although one could argue that it would be more natural to bound in expectation over each individual posterior. This choice is questionable, but it ensures that individual score estimates provide good approximations on both the support of the multi-observation posterior, which is our real target, and on the support of each *individual* posterior (see Appendix A.6 for details).

Numerical illustrations. We showcase our theoretical bound on a Gaussian example (see Appendix A.2.1 for more details) where all true scores can be analytically computed and for which the Gaussianity assumption of both the backward kernels and the individual posteriors happens to be verified : in this setting, the compositional score (1) corresponds to the true score; therefore, no source of error stemming from external assumptions interferes with the analysis of the compositional MSE. We compare the evolution of our bound to that of the compositional score error obtained empirically, depending on the number of conditional observations n, the precision estimation error ϵ and the individual score error ϵ^2_{DSM}. When all individual scores are perfectly known ($\epsilon^2_{\text{DSM}} = 0$), the sole intrinsic source of error associated with the compositional score (2) stems from the precision estimation error ϵ, which itself depends only on the discretization hyperparameters of each individual diffusion process and the number of samples used to compute each empirical precision matrix. In such a case, Figure 2(A) shows that our upper bound successfully captures the global trend of the empirical error. As the number of conditional observations n grows, the empirical compositional score error and our theoretical bound surprisingly do not blow up but rather tend to stabilize more or less quickly depending on the diffusion time: $\|\Lambda^{-1}\|$ (appearing in both Equation (1) and our bound) indeed hides a dependence in $1/n$ that seems to counterbalance the simple accumulation of the n individual score errors. Our bound follows this trend especially for large n and/or advanced diffusion times as seen in Figure 2(B). Finally, Figure 2(C) shows that our theoretical bound well captures the global evolution of the empirical score error for individual score errors $\epsilon^2_{\text{DSM}} \geq 0.01$ but seems to

be a bit large for smaller values: the empirical evolution seems to be $\mathcal{O}(\epsilon_{\text{DSM}}^2)$ while our theoretical bound contains a constant term C and is thus in $\mathcal{O}(\epsilon_{\text{DSM}}^2 + \epsilon_{\text{DSM}} + C)$, which biases the evolution.

Conclusion and perspectives. We provide a bound on the mean squared error between the compositional score (1) and its estimate (2) as a function of the n individual score errors and the precision estimation error, that is specific to the algorithm. Note that we do not try to quantify the error made by the Gaussian approximations of backward kernels in our analysis, but rather provide a bound for a compositional score estimate computed according to the `GAUSS` method. Our upper bound could be used to derive convergence bounds for diffusion process using such compositional estimate or even tune the corresponding diffusion discretization hyperparatemeters to achieve a better sampling quality from the multi-observation posterior. Also, our analysis allows to better understand the effect of each source of errors involved in the compositional estimate (2) as well as the link between the precision estimation error and the individual score error.

It would be interesting to extend our error analysis to other methods for estimating precision matrices, in particular those keeping a dependence in $\boldsymbol{\theta}$ for such matrices (e.g. `JAC` algorithm, see Linhart et al., 2024): indeed, the algorithm `GAUSS` conveniently renders the precision matrices independent of $\boldsymbol{\theta}$ but at the cost of a strong Gaussianity assumption on each individual posterior. Note that our upper bound remains valid for these other types of estimation methods, if we consider bounding $\sup_{\boldsymbol{\theta}} \|\Sigma_{t,j}(\boldsymbol{\theta}) - \tilde{\Sigma}_{t,j}(\boldsymbol{\theta})\|_2$ and not $\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})}\|\Sigma_{t,j}(\boldsymbol{\theta}) - \tilde{\Sigma}_{t,j}(\boldsymbol{\theta})\|_2$ as it is done in Proposition 2 for individual score errors.

Aknowledgements

PLCR was supported by a national grant managed by the French National Research Agency (Agence Nationale de la Recherche) attributed to the SBI4C project of the MIAI AI Cluster, under the reference ANR-23-IACL-0006.

References

Arruda, J., Pandey, V., Sherry, C., Barroso, M., Intes, X., Hasenauer, J., and Radev, S. T. (2025). Compositional amortized inference for large-scale hierarchical Bayesian models. *arXiv preprint arXiv:2505.14429*.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Gao, X., Nguyen, H. M., and Zhu, L. (2025). Wasserstein convergence guarantees for a general class of score-based generative models. *Journal of Machine Learning Research*, 26(43):1–54.

Geffner, T., Papamakarios, G., and Mnih, A. (2023). Compositional score modeling for simulation-based inference. In *International Conference on Machine Learning*.

Gloeckler, M., Toyota, S., Fukumizu, K., and Macke, J. H. (2025). Compositional simulation-based inference for time series. In *International Conference on Learning Representations*.

Huggins, J. H., Campbell, T., Kasprzak, M., and Broderick, T. (2018). Practical bounds on the error of Bayesian posterior approximations: A nonasymptotic approach. *arXiv preprint arXiv:1809.09505*.

Li, A. and Huynh, C. (2022). Analysis of neural fragility: Bounding the norm of a rank-one perturbation matrix. *arXiv preprint arXiv:2202.07026*.

Linhart, J., Cardoso, G. V., Gramfort, A., Corff, S. L., and Rodrigues, P. L. (2024). Diffusion posterior sampling for simulation-based inference in tall data settings. *arXiv preprint arXiv:2404.07593*.

Sharrock, L., Simons, J., Liu, S., and Beaumont, M. (2024). Sequential neural score estimation: Likelihood-free inference with conditional score based diffusion models. In *International Conference on Machine Learning*.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

A Technical Appendices and Supplementary Material

A.1 Conditional score-based modelling

We consider a stochastic model $\mathcal{M}$, encoded in a simulator, that outputs observations $\boldsymbol{x}$ depending on some parameters $\boldsymbol{\theta}$. We encode knowledge about the parameter space through a prior distribution $\lambda(\boldsymbol{\theta})$ and consider the case where the likelihood $p(\boldsymbol{x} \mid \boldsymbol{\theta})$ of $\mathcal{M}$ is intractable. Given some specific observation $\boldsymbol{x}_0$, the goal of Bayesian inference is to sample from the posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{x}_0)$ relating the parameters to this specific observation. One way to avoid likelihood evaluations while obtaining the desired samples is to run a (conditional) diffusion process, which requires learning the score of all noisy versions of the posterior distribution $\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_0)$. As posterior samples are not directly accessible, conditional score-based modeling usually trains a conditional score estimate $s_\phi(\boldsymbol{\theta}, \boldsymbol{x}, t)$ by minimizing the following loss in ϕ on the **joint** model $p(\boldsymbol{\theta}, \boldsymbol{x}) = \lambda(\boldsymbol{\theta})p(\boldsymbol{x} \mid \boldsymbol{\theta})$:

$$\mathcal{L}(\phi) = \sum_{t=1}^{T} \gamma_t^2 \mathbb{E}_{\boldsymbol{\theta}\sim\lambda(\boldsymbol{\theta})} \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}\mid\boldsymbol{\theta})} \mathbb{E}_{\boldsymbol{\theta}_t\sim q_{t\mid 0}(\boldsymbol{\theta}_t\mid\boldsymbol{\theta})} \left[\| s_\phi(\boldsymbol{\theta}_t, \boldsymbol{x}, t) - \nabla \log q_{t\mid 0}(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}) \|^2 \right]$$

where γ_t are positive weights and $q_{t|0}$ denotes a forward diffusion kernel. This loss is the traditional denoising score matching loss for unconditional distributions (Sohl-Dickstein et al., 2015) averaged over the observation space. With this averaging operation it becomes easy to create a training dataset $(\boldsymbol{\theta}_i, \boldsymbol{x}_i) \sim p(\boldsymbol{\theta}, \boldsymbol{x})$ and use a Monte-Carlo approximation of $\mathcal{L}$. Also, the final score estimate s_ϕ becomes a valid score approximation for individual posteriors conditioned by any $\boldsymbol{x}$, i.e.:

$$s_\phi(\boldsymbol{\theta}, \boldsymbol{x}, t) \approx \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}) \quad \forall \boldsymbol{x} \sim p(\boldsymbol{x}), \forall t \in [0, T], \forall \boldsymbol{\theta} \sim p_t(\boldsymbol{\theta} \mid \boldsymbol{x}).$$

If we need to estimate the score of $p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_1)$ where $\boldsymbol{x}_1$ is a new observation different from $\boldsymbol{x}_0$, it is sufficient to evaluate our estimate at $\boldsymbol{x}_1$, i.e. $s_\phi(\boldsymbol{\theta}, \boldsymbol{x}_1, t)$ without having to train another score estimate from scratch.

A.2 Gaussian test case

A.2.1 Theoretical setting

We consider a Gaussian prior defined as $\lambda(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \mu_\lambda, \Sigma_\lambda)$ and a Gaussian simulator model (or likelihood) defined as $p(\boldsymbol{x} \mid \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\theta}, \Sigma)$. The goal of Bayesian inference is to determine the mean $\boldsymbol{\theta} \in \mathbb{R}^d$ of the Gaussian simulator model, given some specific observation $\boldsymbol{x}_0$. This boils down to sampling from the posterior distribution that is also Gaussian

$$p(\boldsymbol{\theta} \mid \boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{\theta}; \mu_{\text{post}}(\boldsymbol{x}_0), \Sigma_{\text{post}}),$$

where $\Sigma_{\text{post}} = (\Sigma^{-1} + \Sigma_\lambda^{-1})^{-1}$ and $\mu_{\text{post}}(\boldsymbol{x}_0) = \Sigma_{\text{post}}(\Sigma^{-1}\boldsymbol{x}_0 + \Sigma_\lambda^{-1}\mu_\lambda)$.
When we have multiple IID observations $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ (generated by the same $\boldsymbol{\theta}$), the multi-observation posterior remains Gaussian:

$$p(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}) = \mathcal{N}(\boldsymbol{\theta}; \mu_{\text{post}}(\boldsymbol{x}_{1:n}), \Sigma_{\text{post}}(n)),$$

where $\Sigma_{\text{post}}(n) = (n\Sigma^{-1} + \Sigma_\lambda^{-1})^{-1}$ and $\mu_{\text{post}}(\boldsymbol{x}_{1:n}) = \Sigma_{\text{post}}(n)\big(\sum_{i=1}^n \Sigma^{-1}\boldsymbol{x}_i + \Sigma_\lambda^{-1}\mu_\lambda\big)$. If we follow a Variance-Preserving diffusion process, then the forward transition kernels read as follows:

$$q_{t|0}(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_0) = \mathcal{N}(\boldsymbol{\theta}_t; \sqrt{\alpha_t}\boldsymbol{\theta}_0, (1-\alpha_t)I) \quad \forall t \in [0, T].$$

We can then analytically find the expressions of the individual posterior scores and prior scores over time as well as the true multi-observation posterior scores (see Appendix D in Linhart et al. (2024)):

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) &= -\big(\alpha_t\Sigma_{\text{post}} + (1-\alpha_t)I\big)^{-1}(\boldsymbol{\theta} - \sqrt{\alpha_t}\mu_{\text{post}}(\boldsymbol{x}_i)), \\
\nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta}) &= -\big(\alpha_t\Sigma_\lambda + (1-\alpha_t)I\big)^{-1}(\boldsymbol{\theta} - \sqrt{\alpha_t}\mu_\lambda), \\
\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}) &= -\big(\alpha_t\Sigma_{\text{post}}(n) + (1-\alpha_t)I\big)^{-1}(\boldsymbol{\theta} - \sqrt{\alpha_t}\mu_{\text{post}}(\boldsymbol{x}_{1:n})).
\end{aligned}$$

A.2.2 Numerical illustration

We propose a numerical illustration that aims at testing the quality of our upper bound with respect to one source of error, leaving the others fixed. We use the theoretical setting described in Appendix A.2.1 where the dimension $d = 2$.

In panel (A) of Figure 2, we fix the number of conditional observations n and assume exact individual scores ($\epsilon^2_{\text{DSM}} = 0$) : the sole source of error comes from the precision estimation error ϵ. In practice,

it can be impacted by the choice of discretization parameters during the backward process and the number of samples used to compute each empirical precision matrix.

In panel (B), we fix ϵ^2_{DSM} and let n grows from 1 to 50. The precision error ϵ used in the bound is fixed but chosen according to ϵ^2_{DSM} : we empirically compute the precision error of the 50 diffusion processes (one per conditional observation) and take the largest value.

In panel (C), we fix n and ϵ and study the evolution with respect to ϵ^2_{DSM}. Note that ϵ used in the theoretical bound corresponds to the largest precision error across all diffusion processes (each of them is specific to a conditional observation and uses a score estimate with a specific error level ϵ^2_{DSM}).

In Figure 2, we do not try to assess the quality of the precision bound given in Proposition 1 but rather the quality of the compositional score error given in Proposition 2 : in panel (A), we let ϵ runs over a specific range of values (that can be empirically encountered) and in the last two panels, we set the score errors (fixed or not) and make a fixed choice of discretization parameters, which led to Wasserstein errors during the individual diffusion processes slightly too high, thus not satisfying the assumptions of Proposition 1.

A.3 Proof of Proposition 1

A.3.1 Intermediate results

We first begin with some lemmas that will be useful to derive the proof of Proposition 1. The following lemma as well as the corresponding proof is inspired from Theorem 4.1 in Huggins et al. (2018).

Lemma 1.1. *Let p and $\tilde{p}$ be two probability distributions with corresponding covariance matrices Σ, respectively $\tilde{\Sigma}$. Suppose that $\mathcal{W}_2(p, \tilde{p}) \leq \eta$ with $0 < \eta < \sqrt{\frac{\|\Sigma\|}{2}}$. Then it holds*

$$\|\Sigma - \tilde{\Sigma}\| \leq \left(2\sqrt{2}\eta\|\Sigma\|^{\frac{1}{2}} + \eta^2(\sqrt{2}+1)\right)\left(\frac{\|\Sigma\|^{\frac{1}{2}}}{\|\Sigma\|^{\frac{1}{2}} - \sqrt{2}\eta}\right).$$

Proof. We start by assuming the latent dimension to be one, so that parameters are scalars. Suppose that $\boldsymbol{\theta} \sim p$ and $\tilde{\boldsymbol{\theta}} \sim \tilde{p}$ are distributed according to the optimal coupling for the 2-Wasserstein distance i.e. $\mathcal{W}_2^2(p, \tilde{p}) = \mathbb{E}|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}|^2$. Let μ (resp. $\tilde{\mu}$) be the mean of p (resp. $\tilde{p}$) and σ (resp. $\tilde{\sigma}$) be the standard deviation of p (resp. $\tilde{p}$). We assume $\mu = 0$ without loss of generality (otherwise, consider the variable $\boldsymbol{\theta} - \mu$) and $\mathcal{W}_2(p, \tilde{p}) \leq \eta$. Let $\zeta^2 = \mathbb{E}(\boldsymbol{\theta}^2) = \sigma^2$ and $\tilde{\zeta}^2 = \mathbb{E}(\tilde{\boldsymbol{\theta}}^2) = \tilde{\sigma}^2 + \tilde{\mu}^2$.

$$\begin{aligned}
|\zeta^2 - \tilde{\zeta}^2| &= |\mathbb{E}(\boldsymbol{\theta}^2 - \tilde{\boldsymbol{\theta}}^2)| = |\mathbb{E}[(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} + \tilde{\boldsymbol{\theta}})]| \\
&\leq \mathbb{E}[(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^2]^{\frac{1}{2}}\mathbb{E}[(\boldsymbol{\theta} + \tilde{\boldsymbol{\theta}})^2]^{\frac{1}{2}} \quad \text{by Cauchy–Schwarz} \\
&\leq \eta\mathbb{E}[(\boldsymbol{\theta} + \tilde{\boldsymbol{\theta}})^2]^{\frac{1}{2}} \quad \text{by optimal coupling and Wasserstein bound} \\
&\leq \eta\mathbb{E}[2(\boldsymbol{\theta}^2 + \tilde{\boldsymbol{\theta}}^2)]^{\frac{1}{2}} \quad \text{since } (a+b)^2 \leq 2a^2 + 2b^2 \\
&= 2^{\frac{1}{2}}\eta(\mathbb{E}[\boldsymbol{\theta}^2] + \mathbb{E}[\tilde{\boldsymbol{\theta}}^2])^{\frac{1}{2}} \\
&\leq 2^{\frac{1}{2}}\eta\big(\mathbb{E}[\boldsymbol{\theta}^2]^{\frac{1}{2}} + \mathbb{E}[\tilde{\boldsymbol{\theta}}^2]^{\frac{1}{2}}\big) \quad \text{since } \sqrt{a+b} \leq \sqrt{a} + \sqrt{b} \text{ for } a, b \geq 0 \\
&= 2^{\frac{1}{2}}\eta\big(\zeta + \tilde{\zeta}\big).
\end{aligned}$$

From Huggins et al. (2018), it is also shown that $|\mu - \tilde{\mu}| = |\tilde{\mu}| \leq \eta$. Then,

$$\begin{aligned}
|\sigma^2 - \tilde{\sigma}^2| &= |\zeta^2 - \tilde{\zeta}^2 + \tilde{\mu}^2| \leq |\zeta^2 - \tilde{\zeta}^2| + |\tilde{\mu}^2| \\
&\leq \sqrt{2}\eta\left(\zeta + \tilde{\zeta}\right) + \eta^2 = \sqrt{2}\eta\left(\sigma + \sqrt{\tilde{\sigma}^2 + \tilde{\mu}^2}\right) + \eta^2 \\
&\leq \sqrt{2}\eta\left(\sigma + \tilde{\sigma} + |\tilde{\mu}|\right) + \eta^2 \quad \text{since } \sqrt{a+b} \leq \sqrt{a} + \sqrt{b} \text{ for } a, b \geq 0 \\
&\leq \sqrt{2}\eta(\sigma + \tilde{\sigma} + \eta) + \eta^2 \\
&= \eta^2(\sqrt{2}+1) + 2^{\frac{3}{2}}\eta\max(\sigma, \tilde{\sigma}).
\end{aligned}$$

When the dimension is greater than one, we can use Lemma C.3 and Corollary C.2 of Huggins et al. (2018) to extend the previous result and get:

$$\|\Sigma - \tilde{\Sigma}\| \leq 2^{\frac{3}{2}}\eta\max(\|\Sigma\|^{\frac{1}{2}}, \|\tilde{\Sigma}\|^{\frac{1}{2}}) + \eta^2(\sqrt{2}+1).$$

We can pursue the computation to get rid of $\|\tilde{\Sigma}\|$. Indeed by triangular inequality we have:

$$\begin{aligned}\|\tilde{\Sigma}\|^{\frac{1}{2}} &\leq \Big(\|\Sigma\| + \|\tilde{\Sigma} - \Sigma\|\Big)^{\frac{1}{2}} \\ &\leq \|\Sigma\|^{\frac{1}{2}} + \frac{\|\tilde{\Sigma} - \Sigma\|}{2\|\Sigma\|^{\frac{1}{2}}} \quad \text{using} \quad \sqrt{a+b} \leq \sqrt{a} + \frac{b}{2\sqrt{a}} \quad \text{for} \quad a \geq 0,\ b \geq -a.\end{aligned}$$

So we get the following bound:

$$\begin{aligned}&\|\Sigma - \tilde{\Sigma}\| \leq 2^{\frac{3}{2}} \left(\|\Sigma\|^{\frac{1}{2}} + \frac{\|\tilde{\Sigma} - \Sigma\|}{2\|\Sigma\|^{\frac{1}{2}}}\right) \eta + \eta^2(\sqrt{2}+1) \\ &\Rightarrow \left(1 - \frac{\sqrt{2}}{\|\Sigma\|^{\frac{1}{2}}}\eta\right) \|\Sigma - \tilde{\Sigma}\| \leq 2\sqrt{2}\eta\|\Sigma\|^{\frac{1}{2}} + \eta^2(\sqrt{2}+1) \\ &\Rightarrow \|\Sigma - \tilde{\Sigma}\| \leq \Big(2\sqrt{2}\eta\|\Sigma\|^{\frac{1}{2}} + \eta^2(\sqrt{2}+1)\Big)\left(\frac{\|\Sigma\|^{\frac{1}{2}}}{\|\Sigma\|^{\frac{1}{2}} - \sqrt{2}\eta}\right) \quad \text{since} \quad \eta < \sqrt{\frac{\|\Sigma\|}{2}}.\end{aligned}$$

□

Lemma 1.2. *Suppose that we have $\|\tilde{\Sigma} - \Sigma\| \leq \gamma$ with $\gamma < \frac{1}{\|\Sigma^{-1}\|}$, then it holds*

$$\|\tilde{\Sigma}^{-1} - \Sigma^{-1}\| \leq \frac{\gamma\|\Sigma^{-1}\|^2}{1 - \gamma\|\Sigma^{-1}\|}.$$

Proof. Suppose that $\|\tilde{\Sigma} - \Sigma\| = \|E\| \leq \gamma$ and that $1/\gamma > \|\Sigma^{-1}\|$ then we can write the following:

$$\begin{aligned}\tilde{\Sigma}^{-1} &= (\Sigma + E)^{-1}, \\ &= \Big(\Sigma(I + \Sigma^{-1}E)\Big)^{-1}, \\ &= (I + \Sigma^{-1}E)^{-1}\Sigma^{-1}.\end{aligned}$$

Now we can write

$$\begin{aligned}\tilde{\Sigma}^{-1} - \Sigma^{-1} &= (I + \Sigma^{-1}E)^{-1}\Sigma^{-1} - \Sigma^{-1} \\ &= \Big((I + \Sigma^{-1}E)^{-1} - I\Big)\Sigma^{-1} \\ &= -(I + \Sigma^{-1}E)^{-1}\Sigma^{-1}E\Sigma^{-1} \quad \text{since} \quad I = (I + \Sigma^{-1}E)^{-1}(I + \Sigma^{-1}E).\end{aligned}$$

We will use the following lemma:

Lemma 1.3 (Li and Huynh, 2022)**.** *For any matrix, $A \in \mathcal{M}_n(\mathrm{C})$, with $\|A\| < 1$. The matrix $(I - A)$ is invertible and*

$$\|(I - A)^{-1}\| \leq \frac{1}{1 - \|A\|}.$$

Since $\|\Sigma^{-1}E\| \leq \|\Sigma^{-1}\|\|E\| \leq \gamma\|\Sigma^{-1}\| < 1$ by assumption, we can write

$$\|(I + \Sigma^{-1}E)^{-1}\| \leq \frac{1}{1 - \|\Sigma^{-1}E\|}.$$

As such, we can bound the norms as follows:

$$\begin{aligned}\|\tilde{\Sigma}^{-1} - \Sigma^{-1}\| &\leq \|(I + \Sigma^{-1}E)^{-1}\|\ \|\Sigma^{-1}\|\ \|E\|\ \|\Sigma^{-1}\| \\ &\leq \frac{\gamma\|\Sigma^{-1}\|^2}{1 - \gamma\|\Sigma^{-1}\|}.\end{aligned}$$

□

A.3.2 Proof of Proposition 1

We are now ready to derive a proof of Proposition 1 that we restate below.

Proposition 1 (Precision matrix error) *Choosing* $0 < \eta < \min\left(\sqrt{\frac{\|\Sigma\|}{2}}, f(\|\Sigma\|, \|\Sigma^{-1}\|)\right)$ *we get*

$$\mathcal{W}_2(p, \tilde{p}) \leq \eta \Rightarrow \|\tilde{\Sigma} - \Sigma\| \leq \gamma \Rightarrow \|\tilde{\Sigma}^{-1} - \Sigma^{-1}\| \leq \frac{\gamma\|\Sigma^{-1}\|^2}{1 - \gamma\|\Sigma^{-1}\|},$$

where

$$\gamma = \left(2\sqrt{2\|\Sigma\|}\eta + \eta^2(1+\sqrt{2})\right)\frac{\|\Sigma\|}{\|\Sigma\| - \eta\sqrt{2\|\Sigma\|}} > 0,$$

and function $f : \mathbb{R}_+^2 \to \mathbb{R}_+$ *is defined in the following proof.*

Proof. To improve readability, we denote $m = \sqrt{\|\Sigma\|}$. Thanks to Lemma 1.1 and since $\eta < \frac{m}{\sqrt{2}}$ we know that $\|\Sigma - \tilde{\Sigma}\| \leq \left(2\sqrt{2}\eta m + \eta^2(\sqrt{2}+1)\right)\left(\frac{m}{m-\sqrt{2}\eta}\right)$ if $\mathcal{W}_2(p, \tilde{p}) \leq \eta$.

Let $g(\eta) = \eta^2(\sqrt{2}+1) + 2\sqrt{2}\eta m - \frac{1}{\|\Sigma^{-1}\|}\left(1 - \frac{\sqrt{2}\eta}{m}\right) = \eta^2(\sqrt{2}+1) + \sqrt{2}\eta(2m + \frac{1}{m\|\Sigma^{-1}\|}) - \frac{1}{\|\Sigma^{-1}\|}$.

Let $\gamma = \left(2\sqrt{2}\eta m + \eta^2(\sqrt{2}+1)\right)\left(\frac{m}{m-\sqrt{2}\eta}\right) > 0$. If we want $\gamma < \frac{1}{\|\Sigma^{-1}\|}$, we need to solve (in η) the inequality $g(\eta) < 0$. Let $\Delta = 2\left(2m + \frac{1}{m\|\Sigma^{-1}\|}\right)^2 + \frac{4(\sqrt{2}+1)}{\|\Sigma^{-1}\|}$ be the discriminant of g. As it is nonnegative, g admits exactly 2 roots η_- and η_+ and $g(\eta) < 0$ for $\eta \in (\eta_-, \eta_+)$. After simple computations, we get:

$$\eta_\pm = \frac{-\sqrt{2}(2m + \frac{1}{m\|\Sigma^{-1}\|}) \pm \sqrt{\Delta}}{2(1+\sqrt{2})},$$

with $\eta_- \leq 0 \leq \eta_+$. As η should be positive, it is sufficient to select $\eta < \eta_+ = \frac{-\sqrt{2}\left(2\sqrt{\|\Sigma\|} + \frac{1}{\sqrt{\|\Sigma\|\|\Sigma^{-1}\|}}\right) + \sqrt{\Delta}}{2(1+\sqrt{2})} := f(\|\Sigma\|, \|\Sigma^{-1}\|)$ to solve the inequality. This ensures that $\gamma < \frac{1}{\|\Sigma^{-1}\|}$. Therefore, we can use Lemma 1.2 to get the last inequality of the statement. □

A.4 Time-independence of precision errors

We show here that the precision estimation error ϵ is not time dependent, contrary to all other constants described in Proposition 2. The algorithm `GAUSS` assumes that all individual posteriors are Gaussian of the form $p(\boldsymbol{\theta} \mid \boldsymbol{x}_j) = \mathcal{N}(\mu_j(\boldsymbol{\theta}), \Sigma_j)$ for all $j = 1, \ldots, n$. We consider a Variance-Preserving diffusion process that noises these individual posteriors along a forward path using the following Gaussian kernels $q_{t|0}(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}) = \mathcal{N}(\sqrt{\alpha_t}\boldsymbol{\theta}, (1-\alpha_t)I)$. Then known results (e.g. Equation 2.115 in Bishop, 2006) ensure that $p(\boldsymbol{\theta} \mid \boldsymbol{\theta}_t, \boldsymbol{x}_j) = \mathcal{N}(\mu_{t,j}, \Sigma_{t,j})$ where $\Sigma_{t,j} = (\Sigma_j^{-1} + \frac{\alpha_t}{1-\alpha_t}I)^{-1}$. Thus,

$$\begin{aligned}
\|\tilde{\Sigma}_{t,j}^{-1} - \Sigma_{t,j}^{-1}\| &= \|\tilde{\Sigma}_j^{-1} + \frac{\alpha_t}{1-\alpha_t}I - (\Sigma_j^{-1} + \frac{\alpha_t}{1-\alpha_t}I)\| \\
&= \|\tilde{\Sigma}_j^{-1} - \Sigma_j^{-1}\| \\
&\leq \epsilon \quad \text{by assumption on precision estimation error.}
\end{aligned}$$

This shows that $\|\tilde{\Sigma}_{t,j}^{-1} - \Sigma_{t,j}^{-1}\|$ is bounded uniformly in time by the initial precision estimation error ϵ.

A.5 Proof of Proposition 2 - detailed version

A.5.1 Proof of intermediate results on compositional score error

Lemma 2.1. *Suppose that* $\|\tilde{\Sigma}_{t,i}^{-1} - \Sigma_{t,i}^{-1}\| \leq \epsilon$ *for* $i = 1, \ldots, n$ *and* $\|\tilde{\Sigma}_{t,\lambda}^{-1} - \Sigma_{t,\lambda}^{-1}\| \leq \epsilon_\lambda$ *then it holds*

$$\|\Lambda - \tilde{\Lambda}\| \leq (n-1)\epsilon_\lambda + n\epsilon.$$

Proof. Recall that

$$\Lambda = (1-n)\Sigma_{t,\lambda}^{-1} + \sum_{i=1}^{n} \Sigma_{t,i}^{-1},$$

where we omit the dependence of Λ in $\boldsymbol{\theta}$ since no covariance matrix depends explicitly in $\boldsymbol{\theta}$ with the `GAUSS` algorithm. Then,

$$\begin{aligned}
\|\Lambda - \tilde{\Lambda}\| &= \left\| (1-n)\left(\Sigma_{t,\lambda}^{-1} - \tilde{\Sigma}_{t,\lambda}^{-1}\right) + \sum_{i=1}^{n}\left(\Sigma_{t,i}^{-1} - \tilde{\Sigma}_{t,i}^{-1}\right) \right\| \\
&\leq (n-1)\|\Sigma_{t,\lambda}^{-1} - \tilde{\Sigma}_{t,\lambda}^{-1}\| + \sum_{i=1}^{n} \|\Sigma_{t,i}^{-1} - \tilde{\Sigma}_{t,i}^{-1}\| \quad \text{since } n \geq 1 \\
&\leq (n-1)\epsilon_\lambda + n\epsilon.
\end{aligned}$$

□

Corollary 1. *Let $\epsilon > 0$ and $\epsilon_\lambda > 0$ such that $(n-1)\epsilon_\lambda + n\epsilon < \frac{1}{\|\Lambda^{-1}\|}$. Then,*

$$\|\tilde{\Lambda}^{-1} - \Lambda^{-1}\| \leq \frac{((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|^2}{1 - ((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}.$$

Proof. We only apply Lemma 1.2 with the correct assumptions. □

In the following, we set $s_\phi(\boldsymbol{\theta}, \boldsymbol{x}_j, t) := s_j(\boldsymbol{\theta}, t)$ and let $s_\lambda(\boldsymbol{\theta}, t) \approx \nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta})$ for more clarity (recall that in this detailed version of Proposition 2 we assume that prior scores are unknown and need to be approximated). Thus, the linear combination of individual posterior scores involved in the compositional score expression (1) and its estimate in (2) reads as follows:

$$\begin{aligned}
\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) &= (1-n)\Sigma_{t,\lambda}^{-1}\nabla \log \lambda_t + \sum_{i=1}^{n} \Sigma_{t,i}^{-1} \nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i), \\
\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) &= (1-n)\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t) + \sum_{i=1}^{n} \tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t).
\end{aligned}$$

Lemma 2.2. *Suppose that*

- $\|\tilde{\Sigma}_{t,\lambda}^{-1} - \Sigma_{t,\lambda}^{-1}\| \leq \epsilon_\lambda$ *and* $\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}(\|s_\lambda(\boldsymbol{\theta}, t) - \nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta})\|^2) \leq \epsilon_{\text{DSM},\lambda}^2$,
- $\|\tilde{\Sigma}_{t,j}^{-1} - \Sigma_{t,j}^{-1}\| \leq \epsilon$ *and* $\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}(\|s_j(\boldsymbol{\theta}, t) - \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid x_j)\|^2) \leq \epsilon_{\text{DSM}}^2$,

for all $j = 1, \ldots, n$. Then it holds

$$\begin{aligned}
&\mathbb{E}_{p_t(\cdot|\boldsymbol{x}_{1:n})}(\|\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \Gamma(\boldsymbol{\theta}\, t; \boldsymbol{x}_{1:n})\|^2) \\
&\qquad \leq \left[(n-1)\left(\epsilon_\lambda \sqrt{\mathbb{E}_{p_t(\cdot|\boldsymbol{x}_{1:n})}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\text{DSM},\lambda}\right)\right. \\
&\qquad \left. + \sum_{i=1}^{n}\left(\epsilon \sqrt{\mathbb{E}_{p_t(\cdot|\boldsymbol{x}_{1:n})}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\|\epsilon_{\text{DSM}}\right)\right]^2.
\end{aligned}$$

Proof. We first show intermediate results that will be useful for the proof. In the following the expectations are taken over $\boldsymbol{\theta}$ drawn from the diffused multi-observation posterior $p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n})$ and we will only use $\mathbb{E}_{\boldsymbol{\theta}}$.

Result (R1):

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}\left(\left\|(1-n)\left(\tilde{\Sigma}_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)-\Sigma_{t,\lambda}^{-1}\nabla_{\boldsymbol{\theta}}\log\lambda_t(\boldsymbol{\theta})\right)\right\|^2\right)\\
&\quad\leq(1-n)^2\mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)-\Sigma_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)\|+\|\Sigma_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)-\Sigma_{t,\lambda}^{-1}\nabla_{\boldsymbol{\theta}}\log\lambda_t(\boldsymbol{\theta})\|\right)^2\\
&\qquad\text{triangular inequality}\\
&\quad=(1-n)^2\left(\mathbb{E}_{\boldsymbol{\theta}}\|\tilde{\Sigma}_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)-\Sigma_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)\|^2+\mathbb{E}_{\boldsymbol{\theta}}\|\Sigma_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)-\Sigma_{t,\lambda}^{-1}\nabla_{\boldsymbol{\theta}}\log\lambda_t(\boldsymbol{\theta})\|^2\right)\\
&\qquad+2(1-n)^2\mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)-\Sigma_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)\|\|\Sigma_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)-\Sigma_{t,\lambda}^{-1}\nabla_{\boldsymbol{\theta}}\log\lambda_t(\boldsymbol{\theta})\|\right)\\
&\qquad\text{expand the square}\\
&\quad\leq(1-n)^2\|\tilde{\Sigma}_{t,\lambda}^{-1}-\Sigma_{t,\lambda}^{-1}\|^2\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta},t)\|^2)+(1-n)^2\|\Sigma_{t,\lambda}^{-1}\|^2\mathbb{E}_{\boldsymbol{\theta}}\|s_\lambda(\boldsymbol{\theta},t)-\nabla_{\boldsymbol{\theta}}\log\lambda_t(\boldsymbol{\theta})\|^2\\
&\qquad+2(1-n)^2\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|\tilde{\Sigma}_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)-\Sigma_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)\|^2)}\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|\Sigma_{t,\lambda}^{-1}s_\lambda(\boldsymbol{\theta},t)-\Sigma_{t,\lambda}^{-1}\nabla_{\boldsymbol{\theta}}\log\lambda_t(\boldsymbol{\theta})\|^2)}\\
&\qquad\text{norm inequality and Cauchy–Schwarz inequality}\\
&\quad\leq(1-n)^2\epsilon_\lambda^2\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta},t)\|^2)+(1-n)^2\|\Sigma_{t,\lambda}^{-1}\|^2\epsilon_{\mathrm{DSM},\lambda}^2\\
&\qquad+2(1-n)^2\|\tilde{\Sigma}_{t,\lambda}^{-1}-\Sigma_{t,\lambda}^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta},t)\|^2)}\|\Sigma_{t,\lambda}^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta},t)-\nabla_{\boldsymbol{\theta}}\log\lambda_t(\boldsymbol{\theta})\|^2)}\\
&\quad\leq(1-n)^2\epsilon_\lambda^2\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta},t)\|^2)+(1-n)^2\|\Sigma_{t,\lambda}^{-1}\|^2\epsilon_{\mathrm{DSM},\lambda}^2\\
&\qquad+2(1-n)^2\epsilon_\lambda\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta},t)\|^2)}\|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\mathrm{DSM},\lambda}\\
&\quad=(1-n)^2\left(\epsilon_\lambda\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta},t)\|^2)}+\|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\mathrm{DSM},\lambda}\right)^2. \qquad \text{(R1)}
\end{aligned}
$$

In the following results, we will first consider n precision errors $\epsilon_1,\ldots,\epsilon_n$ that we will finally bound by the same constant ϵ at the end of the proof.

Result (R2):

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}}\log p_t(\boldsymbol{\theta}\mid\boldsymbol{x}_i)\|^2\right)\\
&\quad\leq\mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}s_i(\boldsymbol{\theta},t)\|+\|\Sigma_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}}\log p_t(\boldsymbol{\theta}\mid\boldsymbol{x}_i)\|\right)^2\\
&\qquad\text{triangular inequality}\\
&\quad=\mathbb{E}_{\boldsymbol{\theta}}\|\tilde{\Sigma}_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}s_i(\boldsymbol{\theta},t)\|^2+\mathbb{E}_{\boldsymbol{\theta}}\|\Sigma_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}}\log p_t(\boldsymbol{\theta}\mid\boldsymbol{x}_i)\|^2\\
&\qquad+2\mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}s_i(\boldsymbol{\theta},t)\|\|\Sigma_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}}\log p_t(\boldsymbol{\theta}\mid\boldsymbol{x}_i)\|\right)\\
&\quad\leq\|\tilde{\Sigma}_{t,i}^{-1}-\Sigma_{t,i}^{-1}\|^2\mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta},t)\|^2+\|\Sigma_{t,i}^{-1}\|^2\mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta},t)-\nabla_{\boldsymbol{\theta}}\log p_t(\boldsymbol{\theta}\mid\boldsymbol{x}_i)\|^2\\
&\qquad+2\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|\tilde{\Sigma}_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}s_i(\boldsymbol{\theta},t)\|^2)}\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|\Sigma_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}}\log p_t(\boldsymbol{\theta}\mid\boldsymbol{x}_i)\|^2)}\\
&\quad\leq\epsilon_i^2\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta},t)\|^2)+\|\Sigma_{t,i}^{-1}\|^2\epsilon_{\mathrm{DSM}}^2\\
&\qquad+2\|\tilde{\Sigma}_{t,i}^{-1}-\Sigma_{t,i}^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta},t)\|^2)}\|\Sigma_{t,i}^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta},t)-\nabla_{\boldsymbol{\theta}}\log p_t(\boldsymbol{\theta}\mid\boldsymbol{x}_i)\|^2)}\\
&\quad\leq\epsilon_i^2\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta},t)\|^2)+\|\Sigma_{t,i}^{-1}\|^2\epsilon_{\mathrm{DSM}}^2\\
&\qquad+2\epsilon_i\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta},t)\|^2)}\|\Sigma_{t,i}^{-1}\|\epsilon_{\mathrm{DSM}}\\
&\quad=\left(\epsilon_i\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta},t)\|^2)}+\|\Sigma_{t,i}^{-1}\|\epsilon_{\mathrm{DSM}}\right)^2. \qquad \text{(R2)}
\end{aligned}
$$

Result (R3):

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}\left(\|\sum_{i=1}^n\tilde{\Sigma}_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}}\log p_t(\boldsymbol{\theta}\mid\boldsymbol{x}_i)\|^2\right)\\
&\quad=\mathbb{E}_{\boldsymbol{\theta}}\sum_{i=1}^n\|\tilde{\Sigma}_{t,i}^{-1}s_i(\boldsymbol{\theta},t)-\Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}}\log p_t(\boldsymbol{\theta}\mid\boldsymbol{x}_i)\|^2\ \text{ expand the squared norm}
\end{aligned}
$$

$$+ \mathbb{E}_{\boldsymbol{\theta}} \sum_{i=1}^{n} \sum_{j \neq i} \left\langle \tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i), \tilde{\Sigma}_{t,j}^{-1} s_j(\boldsymbol{\theta}, t) - \Sigma_{t,j}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid x_j) \right\rangle$$

$$= \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{\theta}} \| \tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) \|^2 \quad \text{linearity of expectation}$$

$$+ \sum_{i=1}^{n} \sum_{j \neq i} \mathbb{E}_{\boldsymbol{\theta}} \left\langle \tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i), \tilde{\Sigma}_{t,j}^{-1} s_j(\boldsymbol{\theta}, t) - \Sigma_{t,j}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid x_j) \right\rangle$$

$$\leq \sum_{i=1}^{n} \left(\epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\| \epsilon_{\text{DSM}} \right)^2 \quad \text{by Result (R2)}$$

$$+ \sum_{i=1}^{n} \sum_{j \neq i} \mathbb{E}_{\boldsymbol{\theta}} \| \tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) \| \| \tilde{\Sigma}_{t,j}^{-1} s_j(\boldsymbol{\theta}, t) - \Sigma_{t,j}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid x_j) \|$$

using Cauchy–Schwarz inequality

$$\leq \sum_{i=1}^{n} \left(\epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\| \epsilon_{\text{DSM}} \right)^2$$

$$+ \sum_{i=1}^{n} \sum_{j \neq i} \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|^2)} \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|\tilde{\Sigma}_{t,j}^{-1} s_j(\boldsymbol{\theta}, t) - \Sigma_{t,j}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid x_j)\|^2)}$$

Cauchy–Schwarz a second time

$$\leq \sum_{i=1}^{n} \left(\epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\| \epsilon_{\text{DSM}} \right)^2$$

$$+ \sum_{i=1}^{n} \sum_{j \neq i} \left(\epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\| \epsilon_{\text{DSM}} \right) \left(\epsilon_j \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_j(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,j}^{-1}\| \epsilon_{\text{DSM}} \right)$$

by Result (R2)

$$= \left(\sum_{i=1}^{n} \epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\| \epsilon_{\text{DSM}} \right)^2. \tag{R3}$$

We are now ready to derive the proof of Lemma 2.2.

$$\mathbb{E}_{\boldsymbol{\theta}} \left(\|\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 \right)$$

$$= \mathbb{E}_{\boldsymbol{\theta}} \left\| (1-n) \left(\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t) - \Sigma_{t,\lambda}^{-1} \nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta}) \right) + \sum_{i=1}^{n} \left(\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) \right) \right\|^2$$

$$= \mathbb{E}_{\boldsymbol{\theta}} \left(\left\| (1-n) \left(\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t) - \Sigma_{t,\lambda}^{-1} \nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta}) \right) \right\|^2 \right)$$

$$+ \mathbb{E}_{\boldsymbol{\theta}} \left(\left\| \sum_{i=1}^{n} \left(\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) \right) \right\|^2 \right)$$

$$+ 2\mathbb{E}_{\boldsymbol{\theta}} \left\langle (1-n) \left(\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t) - \Sigma_{t,\lambda}^{-1} \nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta}) \right), \sum_{i=1}^{n} \left(\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1} \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) \right) \right\rangle$$

expanding the squared norm

$$\leq (1-n)^2 \left(\epsilon_\lambda \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,\lambda}^{-1}\| \epsilon_{\text{DSM},\lambda} \right)^2$$

$$+ \left(\sum_{i=1}^{n} \epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} \| + \|\Sigma_{t,i}^{-1}\| \epsilon_{\text{DSM}} \right)^2$$

$$+ 2\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{\theta}} \left\langle (1-n)\left(\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t) - \Sigma_{t,\lambda}^{-1}\nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta})\right), \tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\right\rangle$$

by Results (R1) and (R3) and linearity of expectation and scalar product

$$\leq (1-n)^2 \left(\epsilon_\lambda \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\text{DSM},\lambda}\right)^2$$
$$+ \left(\sum_{i=1}^{n} \epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)}\| + \|\Sigma_{t,i}^{-1}\|\epsilon_{\text{DSM}}\right)^2$$
$$+ 2\sum_{i=1}^{n}(n-1)\mathbb{E}_{\boldsymbol{\theta}}\|\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t) - \Sigma_{t,\lambda}^{-1}\nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta})\|\|\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|$$

by Cauchy–Schwarz $\langle \boldsymbol{x}, \mathbf{y}\rangle \leq |\langle \boldsymbol{x}, \mathbf{y}\rangle| \leq \|\boldsymbol{x}\|\|\mathbf{y}\|$ and $n \geq 1$

$$\leq (1-n)^2 \left(\epsilon_\lambda \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\text{DSM},\lambda}\right)^2$$
$$+ \left(\sum_{i=1}^{n} \epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)}\| + \|\Sigma_{t,i}^{-1}\|\epsilon_{\text{DSM}}\right)^2$$
$$+ 2(n-1)\sum_{i=1}^{n} \sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t) - \Sigma_{t,\lambda}^{-1}\nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta})\|^2}\sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t) - \Sigma_{t,i}^{-1}\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|^2}$$

by Cauchy–Schwarz $\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$

$$\leq (1-n)^2 \left(\epsilon_\lambda \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\text{DSM},\lambda}\right)^2$$
$$+ \left(\sum_{i=1}^{n} \epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\|\epsilon_{\text{DSM}}\right)^2$$
$$+ 2(n-1)\sum_{i=1}^{n} \left(\epsilon_\lambda \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\text{DSM},\lambda}\right)\left(\epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\|\epsilon_{\text{DSM}}\right)$$

by Results (R1) and (R2)

$$= \left((n-1)\left(\epsilon_\lambda \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\text{DSM},\lambda}\right) + \sum_{i=1}^{n}\left(\epsilon_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\|\epsilon_{\text{DSM}}\right)\right)^2$$
$$\leq \left((n-1)\left(\epsilon_\lambda \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\text{DSM},\lambda}\right) + \sum_{i=1}^{n}\left(\epsilon \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,i}^{-1}\|\epsilon_{\text{DSM}}\right)\right)^2$$

since we assume that $\epsilon_i \leq \epsilon \; \forall \; i$.

□

Lemma 2.3. *Suppose that*

- $\|\tilde{\Sigma}_{t,\lambda}^{-1} - \Sigma_{t,\lambda}^{-1}\| \leq \epsilon_\lambda$ *and for all* $j = 1, \ldots, n$ $\|\tilde{\Sigma}_{t,j}^{-1} - \Sigma_{t,j}^{-1}\| \leq \epsilon_j$
- $s_\lambda(\boldsymbol{\theta}, t)$ $\left(\textit{resp. } s_j(\boldsymbol{\theta}, t)\right)$ *is a score estimate of* $\nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta})$ $\left(\textit{resp. } \nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta} \mid x_j)\right)$

Then it holds

$$\mathbb{E}_{p_t(\cdot|\boldsymbol{x}_{1:n})}(\|\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2) \leq \left((n-1)(\|\Sigma_{t,\lambda}^{-1}\| + \epsilon_\lambda)\sqrt{\mathbb{E}_{p_t(\cdot|\boldsymbol{x}_{1:n})}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)}\right.$$
$$\left. + \sum_{i=1}^{n}(\|\Sigma_{t,i}^{-1}\| + \epsilon_i)\sqrt{\mathbb{E}_{p_t(\cdot|\boldsymbol{x}_{1:n})}(\|s_i(\boldsymbol{\theta}, t)\|^2)}\right)^2.$$

Proof. Here again the expectations are taken over $\boldsymbol{\theta}$ under the diffused multi-observation posterior $p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n})$.

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2\right) &= \mathbb{E}_{\boldsymbol{\theta}}\left(\|(1-n)\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t) + \sum_{i=1}^{n} \tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t)\|^2\right) \\
&= (1-n)^2 \mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t)\|^2\right) + \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t)\|^2\right) \\
&\quad \text{expanding the squared norm} \\
&\quad + 2(1-n)\mathbb{E}_{\boldsymbol{\theta}}\left\langle \tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t), \sum_{i=1}^{n} \tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t)\right\rangle \\
&= (1-n)^2 \mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t)\|^2\right) + \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t)\|^2\right) \\
&\quad + 2(1-n)\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{\theta}}\left\langle \tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t), \tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t)\right\rangle \\
&\quad \text{linearity of scalar product} \\
&\leq (1-n)^2 \|\tilde{\Sigma}_{t,\lambda}^{-1}\|^2 \mathbb{E}_{\boldsymbol{\theta}}\left(\|s_\lambda(\boldsymbol{\theta}, t)\|^2\right) + \sum_{i=1}^{n} \|\tilde{\Sigma}_{t,i}^{-1}\|^2 \mathbb{E}_{\boldsymbol{\theta}}\left(\|s_i(\boldsymbol{\theta}, t)\|^2\right) \\
&\quad + 2(n-1)\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{\theta}}\left(\|\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t)\| \|\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t)\|\right) \\
&\quad \text{Cauchy–Schwarz and } n \geq 1 \\
&\leq (1-n)^2 \|\tilde{\Sigma}_{t,\lambda}^{-1}\|^2 \mathbb{E}_{\boldsymbol{\theta}}\|s_\lambda(\boldsymbol{\theta}, t)\|^2 + \sum_{i=1}^{n} \|\tilde{\Sigma}_{t,i}^{-1}\|^2 \mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta}, t)\|^2 \\
&\quad + 2(n-1)\sum_{i=1}^{n} \sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|\tilde{\Sigma}_{t,\lambda}^{-1} s_\lambda(\boldsymbol{\theta}, t)\|^2} \sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|\tilde{\Sigma}_{t,i}^{-1} s_i(\boldsymbol{\theta}, t)\|^2} \\
&\quad \text{Cauchy–Schwarz inequality} \\
&\leq (1-n)^2 \|\tilde{\Sigma}_{t,\lambda}^{-1}\|^2 \mathbb{E}_{\boldsymbol{\theta}}\|s_\lambda(\boldsymbol{\theta}, t)\|^2 + \sum_{i=1}^{n} \|\tilde{\Sigma}_{t,i}^{-1}\|^2 \mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta}, t)\|^2 \\
&\quad + 2(n-1)\|\tilde{\Sigma}_{t,\lambda}^{-1}\| \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} \sum_{i=1}^{n} \|\tilde{\Sigma}_{t,i}^{-1}\| \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} \\
&\leq (1-n)^2 \|\tilde{\Sigma}_{t,\lambda}^{-1}\|^2 \mathbb{E}_{\boldsymbol{\theta}}\|s_\lambda(\boldsymbol{\theta}, t)\|^2 + \left(\sum_{i=1}^{n} \|\tilde{\Sigma}_{t,i}^{-1}\| \sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta}, t)\|^2}\right)^2 \\
&\quad + 2(n-1)\|\tilde{\Sigma}_{t,\lambda}^{-1}\| \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} \sum_{i=1}^{n} \|\tilde{\Sigma}_{t,i}^{-1}\| \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)} \\
&\quad \text{since } \sum a_i^2 \leq \left(\sum a_i\right)^2 \text{ when } a_i \geq 0 \\
&= \left((n-1)\|\tilde{\Sigma}_{t,\lambda}^{-1}\| \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \sum_{i=1}^{n} \|\tilde{\Sigma}_{t,i}^{-1}\| \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)}\right)^2 \\
&\leq \Big[(n-1)(\|\Sigma_{t,\lambda}^{-1}\| + \epsilon_\lambda) \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} \\
&\quad + \sum_{i=1}^{n} (\|\Sigma_{t,i}^{-1}\| + \epsilon_i) \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)}\Big]^2
\end{aligned}$$

using triangular inequality to bound $\|\tilde{\Sigma}_{t,\lambda}^{-1}\|$ and $\|\tilde{\Sigma}_{t,i}^{-1}\|$.

□

A.5.2 Proof of detailed version of Proposition 2

The final expressions of the multi-observation score (1) and its estimate (2) are the following:

$$\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}) = \Lambda^{-1}\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}),$$
$$s(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) = \tilde{\Lambda}^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}),$$

where $\Lambda = (1-n)\Sigma_{t,\lambda}^{-1} + \sum_{i=1}^{n}\Sigma_{t,i}^{-1}$ and $\tilde{\Lambda} = (1-n)\tilde{\Sigma}_{t,\lambda}^{-1} + \sum_{i=1}^{n}\tilde{\Sigma}_{t,i}^{-1}$.
We propose here a more detailed version of Proposition 2 than what was presented in the main text: we do not assume a Gaussian prior, and thus need to estimate both the prior score $\nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta})$ and the precision matrix Σ_{λ}^{-1}.

Proposition 2 (Compositional score error - detailed version) *Let η (resp. η_λ) be chosen as in Proposition 1 and denote*

$$M = \max_j \|\Sigma_{t,j}^{-1}\| \quad \textit{with} \quad L = \max_j \mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}\left(\|\nabla \log p_t(\boldsymbol{\theta}|\boldsymbol{x}_j)\|^2\right)$$
$$M_\lambda \geq \|\Sigma_{t,\lambda}^{-1}\| \quad \textit{with} \quad L_\lambda \geq \mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}\left(\|\nabla \log \lambda_t(\boldsymbol{\theta})\|^2\right).$$

Suppose that

$$\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}(\|\nabla_{\boldsymbol{\theta}} \log p_t(\boldsymbol{\theta}|\boldsymbol{x}_j) - s_\phi(\boldsymbol{\theta}, \boldsymbol{x}_j, t)\|^2) \leq \epsilon_{\text{DSM}}^2, \textit{ for all} \quad j = 1, \dots, n$$
$$\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}(\|\nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta}) - s_\lambda(\boldsymbol{\theta}, t)\|^2) \leq \epsilon_{\text{DSM},\lambda}^2$$
$$\|\tilde{\Sigma}_j^{-1} - \Sigma_j^{-1}\| \leq \epsilon, \textit{ for all} \quad j = 1, \dots, n$$
$$\|\tilde{\Sigma}_\lambda^{-1} - \Sigma_\lambda^{-1}\| \leq \epsilon_\lambda,$$

with $(n-1)\epsilon_\lambda + n\epsilon < \frac{1}{\|\Lambda^{-1}\|}$. Then it holds

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}\left[\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}) - s(\boldsymbol{\theta}, \boldsymbol{x}_{1:n}, t)\|^2\right] \\
&\leq \Bigg[(n-1)\|\Lambda^{-1}\|(\sqrt{L_\lambda} + \epsilon_{\text{DSM},\lambda})\left(\epsilon_\lambda + \frac{((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|(M_\lambda + \epsilon_\lambda)}{1 - ((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}\right) \\
&\quad + n\|\Lambda^{-1}\|(\sqrt{L} + \epsilon_{\text{DSM}})\left(\epsilon + \frac{((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|(M + \epsilon)}{1 - ((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}\right) \\
&\quad + \|\Lambda^{-1}\|((n-1)M_\lambda\epsilon_{\text{DSM},\lambda} + nM\epsilon_{\text{DSM}}\Bigg]^2.
\end{aligned}$$

If ϵ_{DSM} (resp. $\epsilon_{\text{DSM},\lambda}$) is sufficiently small and accompanied by a proper choice of diffusion hyperparameters such that $\mathcal{W}_2(\tilde{p}(\boldsymbol{\theta} \mid \boldsymbol{x}_j), p(\boldsymbol{\theta} \mid \boldsymbol{x}_j)) \leq \eta$ for all $j = 1, \dots, n$ (resp. $\mathcal{W}_2(\tilde{\lambda}(\boldsymbol{\theta}), \lambda(\boldsymbol{\theta})) \leq \eta_\lambda$) (Proposition 4 in Gao et al., 2025), then the precision estimation error ϵ (resp. ϵ_λ for the prior precision error) can be further bounded using Proposition 1 and η (resp. η_λ).

Proof. To simplify notations in the proofs, we write the expectation taken over the diffused multi-observation posterior $\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}$ simply by $\mathbb{E}_{\boldsymbol{\theta}}$, and we use $s_i(\boldsymbol{\theta}, t) := s_\phi(\boldsymbol{\theta}, \boldsymbol{x}_i, t)$ and $s_\lambda(\boldsymbol{\theta}, t) \approx \nabla_{\boldsymbol{\theta}} \log \lambda_t(\boldsymbol{\theta})$.

Recall that precision errors are time independent (see Appendix A.4), i.e. for all $t \in [0, T]$:

$$\|\tilde{\Sigma}_{t,j}^{-1} - \Sigma_{t,j}^{-1}\| \leq \epsilon_j \leq \epsilon \quad \forall j = 1, \dots, n \quad \text{and} \quad \|\tilde{\Sigma}_{t,\lambda}^{-1} - \Sigma_{t,\lambda}^{-1}\| \leq \epsilon_\lambda.$$

We have

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}) - s(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\theta}}\|\Lambda^{-1}\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \tilde{\Lambda}^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 \\
&\leq \mathbb{E}_{\boldsymbol{\theta}}\left(\|\Lambda^{-1}\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\| + \|\Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \tilde{\Lambda}^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|\right)^2 \\
&\quad \text{using triangular inequality} \\
&= \mathbb{E}_{\boldsymbol{\theta}}\|\Lambda^{-1}\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 + \mathbb{E}_{\boldsymbol{\theta}}\|\Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \tilde{\Lambda}^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 \\
&\quad + 2\mathbb{E}_{\boldsymbol{\theta}}\left(\|\Lambda^{-1}\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|\|\Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \tilde{\Lambda}^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|\right) \\
&\quad \text{expanding the square} \\
&\leq \mathbb{E}_{\boldsymbol{\theta}}\|\Lambda^{-1}\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 + \mathbb{E}_{\boldsymbol{\theta}}\|\Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \tilde{\Lambda}^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 \\
&\quad + 2\sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|\Lambda^{-1}\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2}\sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|\Lambda^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \tilde{\Lambda}^{-1}\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2} \\
&\quad \text{by Cauchy–Schwarz} \\
&\leq \|\Lambda^{-1}\|^2\mathbb{E}_{\boldsymbol{\theta}}\|\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 + \|\Lambda^{-1} - \tilde{\Lambda}^{-1}\|^2\mathbb{E}_{\boldsymbol{\theta}}\|\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 \\
&\quad + 2\|\Lambda^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2)}\|\Lambda^{-1} - \tilde{\Lambda}^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2)} \\
&= \left(\|\Lambda^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|\Gamma(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n}) - \tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2} + \|\Lambda^{-1} - \tilde{\Lambda}^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|\tilde{\Gamma}(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2}\right)^2 \\
&\leq \left[\|\Lambda^{-1}\|\left((n-1)\left(\epsilon_\lambda\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\mathrm{DSM},\lambda}\right) + \sum_{i=1}^{n}\left(\epsilon\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)}\| + \|\Sigma_{t,i}^{-1}\|\epsilon_{\mathrm{DSM}}\right)\right)\right. \\
&\left.+\|\Lambda^{-1} - \tilde{\Lambda}^{-1}\|\left((n-1)\left(\|\Sigma_{t,\lambda}^{-1}\| + \epsilon_\lambda\right)\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)} + \sum_{i=1}^{n}\left(\|\Sigma_{t,i}^{-1}\| + \epsilon_i\right)\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)}\right)\right]^2 \\
&\quad \text{using Lemma 2.2 and Lemma 2.3} \\
&\leq \left[(n-1)\|\Lambda^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)}\left(\epsilon_\lambda + \frac{((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}{1 - ((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}(\|\Sigma_{t,\lambda}^{-1}\| + \epsilon_\lambda)\right)\right. \\
&\quad + \sum_{i=1}^{n}\|\Lambda^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)}\left(\epsilon + \frac{((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}{1 - ((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}(\|\Sigma_{t,i}^{-1}\| + \epsilon)\right) \\
&\quad \left.+\|\Lambda^{-1}\|\left((n-1)\|\Sigma_{t,\lambda}^{-1}\|\epsilon_{\mathrm{DSM},\lambda} + \sum_{i=1}^{n}\|\Sigma_{t,i}^{-1}\|\epsilon_{\mathrm{DSM}}\right)\right]^2 \\
&\quad \text{using Corollary 1 and assuming} \quad \epsilon_i \leq \epsilon \quad \forall i.
\end{aligned}
$$

By assumption, we have $M = \max_i \|\Sigma_{t,i}^{-1}\|$ and $\|\Sigma_{t,\lambda}^{-1}\| \leq M_\lambda$. This gives:

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}) - s(\boldsymbol{\theta}, t; \boldsymbol{x}_{1:n})\|^2 \\
&\quad \leq \left[(n-1)\|\Lambda^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_\lambda(\boldsymbol{\theta}, t)\|^2)}\left(\epsilon_\lambda + \frac{((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}{1 - ((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}(M_\lambda + \epsilon_\lambda)\right)\right. \\
&\qquad + \sum_{i=1}^{n}\|\Lambda^{-1}\|\sqrt{\mathbb{E}_{\boldsymbol{\theta}}(\|s_i(\boldsymbol{\theta}, t)\|^2)}\left(\epsilon + \frac{((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}{1 - ((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}(M + \epsilon)\right) \\
&\qquad \left.+\|\Lambda^{-1}\|\left((n-1)M_\lambda\epsilon_{\mathrm{DSM},\lambda} + nM\epsilon_{\mathrm{DSM}}\right)\right]^2.
\end{aligned}
$$

As $\mathbb{E}_{\boldsymbol{\theta}}\left(\|\nabla \log \lambda_t(\boldsymbol{\theta})\|^2\right) \leq L_\lambda$ and $L = \max_i \mathbb{E}_{\boldsymbol{\theta}}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|^2$, then it holds for all $i = 1, \ldots, n$:

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta}, t)\|^2 &\leq \mathbb{E}_{\boldsymbol{\theta}}[(\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\| + \|s_i(\boldsymbol{\theta}, t) - \nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|)^2] \\
&\leq \mathbb{E}_{\boldsymbol{\theta}}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|^2 + \mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta}, t) - \nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|^2
\end{aligned}
$$

$$
\begin{aligned}
&+ 2\mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta},t) - \nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\| \\
&\leq \mathbb{E}_{\boldsymbol{\theta}}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|^2 + \mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta},t) - \nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|^2 \\
&\quad + 2\sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|s_i(\boldsymbol{\theta},t) - \nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|^2}\sqrt{\mathbb{E}_{\boldsymbol{\theta}}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)\|^2} \\
&\quad \text{by Cauchy–Schwarz inequality} \\
&\leq L + \epsilon_{\text{DSM}}^2 + 2\sqrt{L}\epsilon_{\text{DSM}} \quad \text{by assumption on score matching error} \\
&= (\sqrt{L} + \epsilon_{\text{DSM}})^2.
\end{aligned}
$$

Similarly,

$$
\mathbb{E}_{\boldsymbol{\theta}}\|s_\lambda(\boldsymbol{\theta},t)\|^2 \leq (\sqrt{L_\lambda} + \epsilon_{\text{DSM},\lambda})^2.
$$

Thus we finally get:

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_{1:n})}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}) - s(\boldsymbol{\theta},t;\boldsymbol{x}_{1:n})\|^2 \\
&\quad \leq \Bigg[(n-1)\|\Lambda^{-1}\|(\sqrt{L_\lambda} + \epsilon_{\text{DSM},\lambda})\left(\epsilon_\lambda + \frac{(n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}{1 - ((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}(M_\lambda + \epsilon_\lambda)\right) \\
&\qquad + n\|\Lambda^{-1}\|(\sqrt{L} + \epsilon_{\text{DSM}})\left(\epsilon + \frac{(n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}{1 - ((n-1)\epsilon_\lambda + n\epsilon)\|\Lambda^{-1}\|}(M + \epsilon)\right) \\
&\qquad + \|\Lambda^{-1}\|\left((n-1)M_\lambda\epsilon_{\text{DSM},\lambda} + nM\epsilon_{\text{DSM}}\right)\Bigg]^2 \\
&\quad = \Bigg[(n-1)\|\Lambda^{-1}\|\sqrt{L_\lambda}\left(\frac{n\epsilon\|\Lambda^{-1}\|M_\lambda}{1 - n\epsilon\|\Lambda^{-1}\|}\right) \\
&\qquad + n\|\Lambda^{-1}\|(\sqrt{L} + \epsilon_{\text{DSM}})\left(\epsilon + \frac{n\epsilon\|\Lambda^{-1}\|}{1 - n\epsilon\|\Lambda^{-1}\|}(M + \epsilon)\right) \\
&\qquad + \|\Lambda^{-1}\|nM\epsilon_{\text{DSM}}\Bigg]^2 \\
&\qquad \text{if we assume } \epsilon_{\text{DSM},\lambda} = \epsilon_\lambda = 0 \text{ as in Proposition 2 in the main section.}
\end{aligned}
$$

□

A.6 Discussion on measure choice

In Proposition 2, all the expectations in the assumptions are taken under the **multi-observation** (diffused) posterior distribution $p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n})$ even though the variables inside the expectations come from the (diffused) prior $\lambda_t(\boldsymbol{\theta})$ or the **individual** (diffused) posteriors $p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)$. This measure choice is particularly visible for the individual score error $\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) - s_i(\boldsymbol{\theta},t)\|^2 \leq \epsilon_{\text{DSM}}^2$, which is usually computed in expectation over the individual posterior $p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i)$. But we can link individual posterior to multi-observation posterior as follows:

$$
\begin{aligned}
p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) &= \int p(\boldsymbol{\theta}_0 \mid \boldsymbol{x}_i) q_{t|0}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0) d\boldsymbol{\theta}_0 \quad \text{by definition of the individual diffused posterior} \\
&= \int q_{t|0}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0) \int p(\boldsymbol{\theta}_0, \boldsymbol{x}_{1:i-1}, \boldsymbol{x}_{i+1:n} \mid \boldsymbol{x}_i) dx_{1:i-1} dx_{i+1:n} d\boldsymbol{\theta}_0 \quad \text{by marginalisation} \\
&= \int q_{t|0}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0) \int p(\boldsymbol{\theta}_0 \mid \boldsymbol{x}_{1:n}) p(\boldsymbol{x}_{1:i-1}, \boldsymbol{x}_{i+1:n} \mid \boldsymbol{x}_i) dx_{1:i-1} dx_{i+1:n} d\boldsymbol{\theta}_0 \\
&= \int p(\boldsymbol{x}_{1:i-1}, \boldsymbol{x}_{i+1:n} \mid \boldsymbol{x}_i) \int p(\boldsymbol{\theta}_0 \mid \boldsymbol{x}_{1:n}) q_{t|0}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0) d\boldsymbol{\theta}_0 dx_{1:i-1} dx_{i+1:n} \quad \text{by Fubini theorem} \\
&= \int p(\boldsymbol{x}_{1:i-1}, \boldsymbol{x}_{i+1:n} \mid \boldsymbol{x}_i) p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_{1:n}) dx_{1:i-1} dx_{i+1:n} \quad \text{by definition of the diffused posterior.}
\end{aligned}
$$

Thus we can apply the double expectation equality and get for any measurable function f:

$$
\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_i)}(f(\boldsymbol{\theta})) = \mathbb{E}_{\boldsymbol{x}_{1:i-1},\boldsymbol{x}_{i+1:n}\sim p(.|\boldsymbol{x}_i)}\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})}(f(\boldsymbol{\theta})).
$$

When $f(\boldsymbol{\theta}) = \|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) - s_i(\boldsymbol{\theta},t)\|^2$ for any $i = 1, \ldots, n$ we have a bound for the individual score error:

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\cdot|\boldsymbol{x}_i)}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) - s_i(\boldsymbol{\theta},t)\|^2 &= \mathbb{E}_{\boldsymbol{x}_{1:i-1},\boldsymbol{x}_{i+1:n}\sim p(.|\boldsymbol{x}_i)}\mathbb{E}_{\boldsymbol{\theta}\sim p_t(\boldsymbol{\theta}|\boldsymbol{x}_{1:n})}\|\nabla \log p_t(\boldsymbol{\theta} \mid \boldsymbol{x}_i) - s_i(\boldsymbol{\theta},t)\|^2 \\
&\leq \mathbb{E}_{\boldsymbol{x}_{1:i-1},\boldsymbol{x}_{i+1:n}\sim p(.|\boldsymbol{x}_i)}(\epsilon_{\text{DSM}}^2) \quad \text{by assumption on the individual score error} \\
&= \epsilon_{\text{DSM}}^2.
\end{aligned}
$$

So if we manage to bound the individual score error in average under the **multi-observation** (diffused) posterior for a specific time t in the diffusion process, then this error will be bounded in average under any **individual** (diffused) posterior (by the same constant).