

STEIN'S METHOD FOR THE WISHART DISTRIBUTION

Gabriel Bailly, Robert E. Gaunt, Frédéric Ouimet, Donald Richards, Rainer von Sachs

LIDAM Discussion Paper ISBA
2026 / 21

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

<https://uclouvain.be/en/research-institutes/lidam/isba/publication>

Stein’s method for the Wishart distribution

Gabriel Bailly^a, Robert E. Gaunt^b, Frédéric Ouimet^{c,*}, Donald Richards^d, Rainer von Sachs^a

^a*ISBA/LIDAM, UCLouvain, Louvain-La-Neuve, Belgium*

^b*The University of Manchester, Manchester, M13 9PL, UK*

^c*Université du Québec à Trois-Rivières, Trois-Rivières, QC G8Z 4M3, Canada*

^d*Penn State University, University Park, PA 16802, USA*

Abstract

In this work, we develop Stein’s method for the Wishart distribution on the cone of positive definite matrices. We establish the basic ingredients of a Wishart Stein framework: we derive an extended-generator-based Stein characterization from the Wishart diffusion process, identify the corresponding transition semigroup through the noncentral Wishart law, provide an explicit semigroup representation for the solution of the Stein equation, and obtain regularity estimates for the solution. The new methodology is demonstrated in four applications: (i) an order n^{-1} bound, for smooth test functions, for the Wishart approximation of uncentered group-mean scatter matrices in MANOVA; (ii) a quantitative multivariate Satterthwaite approximation; (iii) local/integrated De Bruijn identities and logarithmic Sobolev inequalities for the Wishart measure; and (iv) Stein’s method of moments for the shape and scale parameters, including structured scale estimation.

Keywords: Stein’s method, Wishart approximation, Wishart distribution, Wishart process

2020 MSC: Primary: 60F05 Secondary: 60H10, 60J60, 62E20, 62H10, 62H12

1. Introduction

Stein’s method is a powerful technique that has classically been applied to bound the distance between two probability distributions with respect to a probability metric. The method was introduced in Charles Stein’s seminal paper (Stein, 1972) in the context of normal approximation. Shortly afterwards, the method was extended to Poisson approximation (Chen, 1975), and the method has since been adapted to many of the most important univariate probability distributions, including the exponential (Chatterjee et al., 2011; Peköz and Röllin, 2011), gamma (Luk, 1994; Gaunt et al., 2017), variance-gamma (Gaunt, 2014) and α -stable distributions (Arras and Houdré, 2019b; Xu, 2019). Stein’s method has also been developed for multivariate distributions, most notably the multivariate normal (Barbour, 1990; Götze, 1991) and multivariate α -stable distributions (Arras and Houdré, 2019a; Chen et al., 2024), and a general theory of Stein’s method for multivariate distributions is beginning to emerge (Mijoule et al., 2023). Introductions to Stein’s method and some of its numerous applications throughout the mathematical sciences are given in the monographs (Chen et al., 2011; Nourdin and Peccati, 2012) and surveys (Ross, 2011; Anastasiou et al., 2023).

*Corresponding author. Email address: frederic.ouimet2@uqtr.ca

Whilst Stein’s method has reached an impressive level of maturity for univariate and multivariate distributions, a conspicuous absence in the literature is a systematic development for matrix-variate distributions. The first step towards the development of Stein’s method for matrix-variate distributions was taken in the recent work of [Gaunt et al. \(2026\)](#), which develops the basic framework of Stein’s method for matrix normal approximation. In this paper, we make a considerably more substantial step by extending Stein’s method to the Wishart distribution, arguably the most important matrix-variate distribution, with numerous applications in multivariate statistics, Bayesian analysis and random matrix theory.

The development of Stein’s method for the Wishart distribution is of particular methodological interest on the following grounds. Firstly, whilst there are instances of matrix normal approximations in which the problem must be dealt with at the matrix-variate level (see Section 4.2 of [Gaunt et al. \(2026\)](#) for an example), there have been instances in the literature in which a vectorization argument has been used to reduce a matrix normal approximation problem to a multivariate normal approximation, for which the powerful existing theory on Stein’s method for multivariate normal approximation can then be applied; see, e.g., [Mikulincer \(2022\)](#); [Nourdin and Zheng \(2022\)](#); [Dhoyer and Tudor \(2024\)](#). Such vectorization arguments are not possible for the Wishart distribution, which underlines the importance of a theory of Stein’s method for Wishart approximation.

Secondly, the basic Stein’s method framework for matrix normal approximation that was established by [Gaunt et al. \(2026\)](#) extended the framework of Stein’s method for multivariate normal approximation in a very natural manner with the proofs following similar lines to those employed in Stein’s method for multivariate normal approximation. However, the relatively clean extension of Stein’s method for multivariate normal approximation to matrix normal approximation is perhaps atypical. In contrast, extending Stein’s method for gamma approximation to the Wishart distribution is highly non-trivial, and the techniques we employ in setting up Stein’s method for the Wishart distribution may prove to be crucial in enabling researchers to extend Stein’s method to other matrix-variate distributions.

1.1. Review of Stein’s method for the gamma distribution

Before giving a summary of our contributions on Stein’s method for the Wishart distribution, it will be helpful to recall the essential ingredients of Stein’s method for gamma approximation. The starting point for this is the following *Stein characterization* of the gamma distribution; see [Diaconis and Zabell \(1991\)](#); [Luk \(1994\)](#); [Gaunt et al. \(2017\)](#). For $r, \lambda > 0$, if $\Gamma(r, \lambda)$ denotes the gamma distribution with density $f_{r,\lambda}(x) = \lambda^r x^{r-1} e^{-\lambda x} / \Gamma(r)$, $x > 0$, then a real-valued random variable X satisfies

$$X \sim \Gamma(r, \lambda) \quad \Leftrightarrow \quad \mathbb{E}[X f''(X) + (r - \lambda X) f'(X)] = 0 \quad \forall f \in C_{\mathcal{A}r}^2(0, \infty), \quad (1.1)$$

where $C_{\mathcal{A}r}^2(0, \infty)$ denotes the class of all twice differentiable functions $f : (0, \infty) \rightarrow \mathbb{R}$ for which the expectations $\mathbb{E}[|f'(X_\infty)|]$, $\mathbb{E}[|X_\infty f'(X_\infty)|]$, and $\mathbb{E}[|X_\infty f''(X_\infty)|]$ are finite for $X_\infty \sim \Gamma(r, \lambda)$. The Stein characterization (1.1) leads to the following *Stein equation* for the gamma distribution:

$$\mathcal{A}^\Gamma f(x) = h(x) - \mathbb{E}[h(X)], \quad (1.2)$$

where the *Stein operator* $\mathcal{A}^\Gamma$ is defined by

$$\mathcal{A}^\Gamma f(x) = x f''(x) + (r - \lambda x) f'(x), \quad (1.3)$$

$h : (0, \infty) \rightarrow \mathbb{R}$ is a real-valued test function, and $X_\infty \sim \Gamma(r, \lambda)$. On evaluating both sides of the Stein equation (1.2) at a random variable of interest X and taking expectations, we get that

$$\mathbb{E}[h(X)] - \mathbb{E}[h(X_\infty)] = \mathbb{E}[X f_h''(X) + (r - \lambda X) f_h'(X)], \quad (1.4)$$

where f_h solves the Stein equation (1.2). Thus, the problem of bounding the quantity $\mathbb{E}[h(X)] - \mathbb{E}[h(X_\infty)]$ is reduced to solving the Stein equation (1.2) and then bounding the right-hand side of (1.4), which is generally more tractable on account of the fact that this expectation only involves the random variable X . In order for this procedure to be effective, suitable bounds on the derivatives of the solution f_h are required.

The Stein equation (1.2) was obtained by Luk (1994) using the generator approach to Stein's method (Barbour, 1990; Götze, 1991) in which the Stein operator for the $\Gamma(r, \lambda)$ distribution is recognized as the generator of the well-known Cox–Ingersoll–Ross process, which is defined through the stochastic differential equation (SDE)

$$dX_t = (r - \lambda X_t) dt + \sqrt{2X_t} dB_t, \quad X_0 = x,$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion. By standard theory on Markov processes, it then follows that the solution to the Stein equation (1.2) is given by

$$f_h(x) = - \int_0^\infty \{ \mathcal{P}_t h(x) - \mathbb{E}[h(X)] \} dt, \quad (1.5)$$

where the transition semigroup operator $\mathcal{P}_t$ is given by $\mathcal{P}_t h(x) := \mathbb{E}[h(X_t) \mid X_0 = x]$, for $t \geq 0$. Provided that the test function h has sufficient regularity such that an interchange in the order of differentiation and integration is permissible, we get that

$$f_h^{(m)}(x) = - \int_0^\infty (\mathcal{P}_t h)^{(m)}(x) dt, \quad m \geq 1. \quad (1.6)$$

In order to obtain bounds on the derivatives of f_h , Luk (1994) derived an explicit formula for the transition semigroup $\mathcal{P}_t h(x)$ and its derivatives. It was shown in Lemma 2.4 of Luk (1994) that $X_t \mid \{X_0 = x\} \sim \sigma^2(t) \cdot \Gamma(r, \lambda, \theta_t x)$, where $\sigma^2(t) = 1 - e^{-\lambda t}$, $\theta_t = 2\lambda e^{-\lambda t} / \sigma^2(t)$, and $\Gamma(r, \lambda, \theta)$ denotes the noncentral gamma distribution with noncentrality parameter θ . It thus follows that

$$\mathcal{P}_t h(x) = \mathbb{E}[h(\sigma^2(t) Y)], \quad (1.7)$$

where $Y \sim \Gamma(r, \lambda, \theta_t x)$. An expression for the m -th order derivative of the transition semigroup $\mathcal{P}_t h(x)$ can then be obtained via direct calculation, with the result, found in Lemma 2.5 of Luk (1994), given by the simple formula:

$$(\mathcal{P}_t h)^{(m)}(x) = e^{-m\lambda t} \mathbb{E}[h^{(m)}(\sigma^2(t) Y_m)], \quad m \geq 1, \quad (1.8)$$

where $Y_m \sim \Gamma(r + m, \lambda, \theta_t x)$. If the m -th derivative of h exists and is bounded, then we have the bound $|(\mathcal{P}_t h)^{(m)}(x)| \leq e^{-m\lambda t} \|h^{(m)}\|_\infty$, where $\|\cdot\|_\infty$ is the supremum norm. When this estimate is applied to (1.6), it then follows from dominated convergence (see Theorem 2.6 of Luk (1994)) that

$$\|f_h^{(m)}\|_\infty \leq \frac{1}{m\lambda} \|h^{(m)}\|_\infty, \quad m \geq 1. \quad (1.9)$$

Under the weaker assumption that the $(m - 1)$ -th derivative of h exists and is bounded, [Gaunt \(2013, Theorem 2.19\)](#) obtained the following bound:

$$\|f_h^{(m)}\|_\infty \leq \left\{ \frac{\sqrt{2\pi} + e^{-1}}{\sqrt{r+m-1}} + \frac{2}{r+m-1} \right\} \|h^{(m-1)}\|_\infty, \quad m \geq 1, \quad (1.10)$$

where $h^{(0)} \equiv h$. One cannot hope for a bound of the form $\|f_h^{(m)}\| \leq C_{r,\lambda,m} \|h^{(m-2)}\|_\infty$, $m \geq 2$, for some constant $C_{r,\lambda,m} > 0$ due to a universal counterexample of [Eden and Viquez \(2015\)](#).

1.2. Summary of our contributions

We begin our development of Stein's method for the Wishart distribution by identifying the Wishart diffusion as the matrix-valued analog of the Cox–Ingersoll–Ross process. In [Proposition 3.1](#), we give the extended generator of the Wishart process, introduced by [Bru \(1991\)](#), and this generator provides the natural candidate for the Wishart Stein operator. We then provide the transition law of the Wishart process in [Proposition 3.2](#). In particular, for positive definite matrices Σ and W , write $\Sigma_t := (1 - e^{-2t})\Sigma$ and, for $t > 0$, let $\mathfrak{S}_t \sim \mathcal{W}_d(\alpha, I_d, e^{-2t}\Sigma_t^{-1/2}W\Sigma_t^{-1/2})$. Then the transition semigroup admits the explicit representation

$$\mathcal{P}_t^{\mathcal{W}} h(W) = \mathbb{E}[h(\Sigma_t^{1/2} \mathfrak{S}_t \Sigma_t^{1/2})],$$

which is the Wishart analog of the gamma semigroup representation [\(1.7\)](#) used by [Luk \(1994\)](#). This transition law also identifies the invariant law of the Wishart process as $\mathfrak{W}_\infty \sim \mathcal{W}_d(\alpha, \Sigma)$. The generator formulation is then used to establish the full Stein characterization in [Corollary 3.3](#). As noted in [Remark 3.1](#), when $d = 1$ the Wishart Stein operator reduces, up to the positive multiplicative factor 4Σ , to the classical gamma Stein operator [\(1.3\)](#).

The main technical obstacle in obtaining the semigroup solution of the Wishart Stein equation and its regularity estimates is the need to control the transition semigroup as a function of its starting point W . The difficulty is already visible in the scalar gamma case. If $f(u; r, \lambda, \theta_t x)$ denotes the noncentral gamma density appearing in [\(1.7\)](#), then the key transfer identity underlying [Lemma 2.5 of Luk \(1994\)](#) is

$$\frac{\partial}{\partial x} f(u; r, \lambda, \theta_t x) = \frac{\theta_t}{2} \{f(u; r+1, \lambda, \theta_t x) - f(u; r, \lambda, \theta_t x)\} = -\frac{\theta_t}{2\lambda} \frac{\partial}{\partial u} f(u; r+1, \lambda, \theta_t x), \quad (1.11)$$

where the two equalities follow from the standard Bessel identities obtained by specializing [Eqs. \(16.3.1\) and \(16.3.4\) of Askey and Olde Daalhuis \(2010\)](#) to $\gamma = r$, $n = 1$ and $p = q = 0$. We recover [\(1.8\)](#) by iterating the above transfer identity and integrating by parts: for any $m \geq 1$,

$$(\mathcal{P}_t h)^{(m)}(x) = \left(\frac{\theta_t \sigma^2(t)}{2\lambda} \right)^m \int_0^\infty h^{(m)}(\sigma^2(t)u) f(u; r+m, \lambda, \theta_t x) du = e^{-m\lambda t} \mathbb{E}[h^{(m)}(\sigma^2(t)Y_m)],$$

where $Y_m \sim \Gamma(r+m, \lambda, \theta_t x)$. The actual proof of [Lemma 2.5](#) by [Luk \(1994\)](#) exploits the Poisson mixture representation of the noncentral gamma density to justify differentiating under the integral sign and integrating by parts, and to apply the argument described above on each summand of the mixture.

The Wishart case is substantially more difficult. There is no comparably simple Poisson mixture representation for the noncentral Wishart density, and the scalar Bessel derivative identities

are replaced by determinantal derivative formulae for the Bessel function of matrix argument; see Eqs. (2.3) and (2.4) of [Herz \(1955\)](#). A direct matrix analog of Luk's proof is therefore not tractable. The key new observation of the present paper is that the transfer from derivatives in the noncentrality parameter to derivatives in the state variable, as in (1.11), can instead be obtained through Laplace transforms. This is carried out in two steps in Section 5. First, recall that for a sufficiently integrable function $g : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$, its Laplace transform is defined by

$$\mathcal{L}[g](T) := \int_{\mathcal{S}_{++}^d} \text{etr}(-TX) g(X) dX, \quad T \in \mathcal{S}_+^d, \quad (1.12)$$

where $\mathcal{S}_+^d$ denotes the cone of $d \times d$ real nonnegative definite matrices. Using this definition, Lemma 5.1 proves the following integration-by-parts Laplace calculus for polynomial differential operators $P(\nabla_X)$ built from the symmetric gradient ∇_X :

$$\mathcal{L}[P(\nabla_X) f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T) = P(T) \mathcal{L}[f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T), \quad T \in \mathcal{S}_+^d,$$

where $f_{\alpha, \Sigma, \Theta}^{\mathcal{W}}$ denotes the noncentral Wishart density defined in (2.2). Second, differentiating the noncentral Wishart Laplace transform with respect to the noncentrality parameter Θ produces the factor $T(I_d + 2T)^{-1}$, and the identity

$$T(I_d + 2T)^{-1} = \frac{T \text{adj}(I_d + 2T)}{|I_d + 2T|} = T \text{adj}(I_d + 2T) \mathcal{L}[f_{2, I_d, 0_{d \times d}}^{\mathcal{W}}](T)$$

converts this factor into a polynomial in T at the cost of shifting the shape parameter of the Wishart density by 2. Combining this with Lemma 5.1 gives Lemma 5.2, namely, for any pairs of indices $(i_1, j_1), \dots, (i_k, j_k) \in \{1, \dots, d\}^2$,

$$\left(\prod_{\ell=1}^k \nabla_{\Theta, i_\ell j_\ell} \right) f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) = (-1)^k \left(\prod_{\ell=1}^k \left\{ \nabla_X \text{adj}(I_d + 2\nabla_X) \right\}_{i_\ell j_\ell} \right) f_{\alpha+2k, I_d, \Theta}^{\mathcal{W}}(X).$$

For $k = 1$, this is the matrix transfer formula which generalizes (1.11). Notice that, unlike in the gamma case where each derivative in the noncentrality parameter is transferred to a first-order derivative in the state variable, each factor $\{\nabla_X \text{adj}(I_d + 2\nabla_X)\}_{ij}$ is a differential operator of order d , since $\text{adj}(I_d + 2\nabla_X)$ is a matrix polynomial in ∇_X of degree $d - 1$.

After an integration by parts on the cone $\mathcal{S}_{++}^d$, the operator in Lemma 5.2 is moved from the noncentral Wishart density onto the test function h , and we can prove Lemma 5.3, namely

$$\left(\prod_{\ell=1}^k \nabla_{\Lambda, i_\ell j_\ell} \right) (\mathcal{Q}_t h)(\Lambda) = e^{-2kt} \mathbb{E} \left[\left(\prod_{\ell=1}^k \mathcal{D}_{i_\ell j_\ell} \right) h(\mathfrak{Y}_{k,t}) \right], \quad \mathcal{D} = \nabla_X \text{adj}(I_d - 2\nabla_X),$$

where $\Lambda \in \mathcal{S}_{++}^d$, $\mathfrak{Y}_{k,t} \sim \mathcal{W}_d(\alpha+2k, I_d, e^{-2t}\Lambda)$, and $\mathcal{Q}_t$ is a scale-normalized analog of the transition semigroup operator $\mathcal{P}_t^{\mathcal{W}}$; see the definition in (3.15). This lemma is the technical core of the paper: it shows how derivatives of $(\mathcal{Q}_t h)(\Lambda)$ with respect to the entries of the noncentrality matrix Λ can be pushed through the noncentral Wishart transition kernel and transferred onto the test function h , thereby generalizing Lemma 2.5 of [Luk \(1994\)](#), or equivalently, Eq. (1.8).

This transfer mechanism first enters the proofs of the main results through Lemma 3.4. In Lemma 3.4, the transition semigroup is written as

$$(\mathcal{P}_t^{\mathcal{W}}h)(W) = (\mathcal{Q}_th_t)(\Sigma_t^{-1/2}W\Sigma_t^{-1/2}), \quad h_t(X) := h(\Sigma_t^{1/2}X\Sigma_t^{1/2}),$$

and the difference between $(\mathcal{Q}_th_t)(\Sigma_t^{-1/2}W\Sigma_t^{-1/2})$ and the central value $(\mathcal{Q}_th_t)(0)$ is controlled by differentiating along the line segment in the Λ variable. The case $k = 1$ of Lemma 5.3 gives the required derivative bound in terms of the $\mathcal{D}$ -seminorm of h_t . This decay estimate is the key input in Theorem 3.5, where we prove that the semigroup integral

$$f_h(W) := - \int_0^\infty \left\{ \mathcal{P}_t^{\mathcal{W}}h(W) - \mathbb{E}[h(\mathfrak{W}_\infty)] \right\} dt$$

is well defined and solves the Wishart Stein equation. This representation is the Wishart analog of the gamma solution formula (1.5).

Finally, Theorem 3.6 completes the analytic part of the Wishart Stein framework by deriving regularity estimates for the semigroup solution. Its proof applies Lemma 5.3 with $k = m$ to the scale-normalized transition semigroup, transfers the resulting derivatives onto the test function through products of the operators $\mathcal{D}_{t,\Sigma} := \nabla_W \text{adj}(I_d - 2\Sigma_t \nabla_W)$, and then integrates the resulting semigroup bounds in the representation of f_h . This yields the indexed partial derivative estimate (3.19) and the coordinate-free directional estimate (3.22). As noted in Remark 3.6, when $d = 1$ the bound (3.19) reduces to the classical bound (1.9) of Luk (1994), up to the 4Σ factor inherited from the relation between the one-dimensional Wishart and gamma Stein operators.

The form of the estimates in Theorem 3.6 is dictated by the transfer operator in Lemma 5.3. Since each entry of $\mathcal{D}_{t,\Sigma}$ is a differential operator of order at most d , the product of m such operators naturally leads to seminorms involving derivatives of h up to order md . Equivalently, the appearance of the md -th order regularity assumption is the cost of pushing m derivatives of the Wishart transition semigroup through the noncentral Wishart kernel and onto the test function. This derivative count is sharp for the transfer identity and for the $\mathcal{D}$ -seminorm bounds used in Theorem 3.6; see Remark 3.5. It does not, however, rule out the possibility that a different argument could save derivatives at the level of the solution of the Stein equation itself. In the gamma case, such a phenomenon is reflected in the bound (1.10), where one derivative of h is saved relative to (1.9). It will be the subject of future work to determine whether a Wishart generalization of (1.10), involving derivatives of h up to order $m(d - 1)$, can be attained. We expect that one cannot in general obtain bounds for the m -th order partial derivatives of the solution of the Wishart Stein equation under substantially weaker differentiability assumptions on h .

This phenomenon of bounds on derivatives of the solution of a Stein equation requiring existence of derivatives whose order increases as a function of the dimension d appears to be novel, although there is a precedent in the literature that stronger regularity conditions may need to be imposed in order to guarantee existence of derivatives of solutions of Stein equations as one moves from a Stein equation for a univariate distribution to its multivariate analog. Indeed, the classic bound $\|f_h^{(3)}\|_\infty \leq 2\|h'\|_\infty$ of Stein (1986) for the solution of the standard normal Stein equation $f''(x) - xf'(x) = h(x) - \mathbb{E}[h(Z)]$, for $Z \sim \mathcal{N}(0, 1)$, was shown by Raič (2004) to not carry over to the solution of the multivariate normal Stein equation, in that there exist Lipschitz test functions h for which the third-order partial derivatives of the solution of the multivariate normal Stein equation do not exist.

In Section 4, we provide several applications of our framework for Stein’s method for the Wishart distribution, with examples given that concern distributional approximation, information theory and parameter estimation. We begin by considering the Wishart approximation of uncentered group-mean scatter matrices in Multivariate Analysis of Variance (MANOVA). In Section 4.1, we derive an explicit order n^{-1} bound to quantify this classical distributional approximation. Our approach involves Wishart generalizations of two techniques that have proven to be highly effective in deriving chi-square approximations via Stein’s method (see, for example, Gaunt et al. (2017); Gaunt and Reinert (2023)), these being a connection between the Wishart and matrix normal Stein operators and symmetry considerations to attain the optimal n^{-1} rate. In Section 4.2, we derive explicit bounds to quantify the multivariate Satterthwaite approximation proposed by Tan and Gupta (1983), in which the distribution of a positive linear combination of independent Wishart random matrices is approximated by a single Wishart distribution with matching first two moments, a multivariate analog of the classical Satterthwaite approximation introduced by Satterthwaite (1941, 1946).

Whilst Stein’s method has traditionally been applied to derive distributional approximations, there has been a growing trend in recent years in which the method has found exciting applications in other areas such as information-theoretic inequalities (Ledoux et al., 2015) and computational statistics (Anastasiou et al., 2023). In our applications of Sections 4.3 and 4.4, we exhibit the utility of our Wishart Stein framework in such domains. In Section 4.3, we apply aspects of our generator approach formulation of Stein’s method for the Wishart distribution developed in Section 3 to derive local and integrated De Bruijn identities for the Wishart semigroup, and complement these results by deriving a logarithmic Sobolev inequality for the Wishart measure, thereby generalizing some of the main results of Arras and Swan (2017) given in the gamma setting. In Section 4.4, we utilize our Stein characterization of the Wishart distribution together with the recently introduced Stein’s method of moments of Ebner et al. (2025) to derive new closed-form estimators for the Wishart distribution. Our simulations suggest that our estimators improve on the classical moment estimators in terms of the relative Frobenius error and offer an attractive alternative to the maximum likelihood estimator, with the closed-form of the estimators offering benefits in terms of computational cost and amenability to theoretical analysis.

1.3. Outline of the paper

Section 2 introduces the definitions and notation regarding matrix spaces, the noncentral Wishart distribution, and infinitesimal and extended generators. Sections 3 and 4 present the main results and the applications, respectively. Section 5 establishes preliminary lemmas regarding the noncentral Wishart distribution that are needed in the proofs of the main results. Sections 6 and 7 provide the detailed proofs for the main results and the applications, respectively. Supplementary material for some of the applications is collected at the end of the paper.

2. Definitions and notation

Throughout, $[d] := \{1, \dots, d\}$ for $d \in \mathbb{N} \equiv \{1, 2, \dots\}$. Let $\mathcal{S}^d$, $\mathcal{S}_+^d$, and $\mathcal{S}_{++}^d$ denote the sets of real symmetric, nonnegative definite, and positive definite $d \times d$ matrices, respectively. Unless mentioned otherwise, these spaces are equipped with the Frobenius norm $\|\cdot\|_F$. Let $O(d)$ denote the orthogonal group of $d \times d$ matrices, consisting of all real matrices H such that $H^\top H = I_d$. For

any square matrix A , let $\text{tr}(A)$ be its trace, $\text{etr}(A) := \exp\{\text{tr}(A)\}$, and $|A|$ its determinant. For $S \in \mathcal{S}_+^d$, the matrix $S^{1/2}$ denotes the unique nonnegative definite square root, $\|S\|_2$ the spectral norm, and $\text{vecp}(S) = (S_{11}, S_{12}, S_{22}, \dots, S_{1d}, \dots, S_{dd})^\top$ its half-vectorization. The symbols $\mathbf{0}_d$, $0_{d \times d}$, and I_d denote the d -dimensional zero vector, the $d \times d$ zero matrix, and the $d \times d$ identity, respectively.

If B is an open subset of a finite-dimensional real Euclidean space, in particular if $B \subseteq \mathcal{S}^d$ is open in the relative topology, and $m \in \mathbb{N}_0$, let $C^m(B)$ be the class of real-valued functions $f : B \rightarrow \mathbb{R}$ that are m times continuously differentiable on B (all partial derivatives up to total order m exist and are continuous), and let $C_b^m(B)$ be the subclass for which all partial derivatives up to total order m , including the order-0 derivative f itself, are bounded on B . For a map $F : B \rightarrow E$, where E is a finite-dimensional real Euclidean space, $D^k F(x)$ denotes the k -th Fréchet derivative of F at x , viewed as a k -linear map on the ambient real vector space of B with values in E ; its evaluation in directions $u_1, \dots, u_k$ is written $D^k F(x)[u_1, \dots, u_k]$, with the convention $D^0 F = F$, and, when $B \subseteq \mathcal{S}^d$, the directions u_i lie in $\mathcal{S}^d$.

For any $d \in \mathbb{N}$, the multivariate gamma function Γ_d is defined by

$$\Gamma_d(\nu) := \int_{\mathcal{S}_{++}^d} \text{etr}(-X) |X|^{\nu-(d+1)/2} dX, \quad \text{Re}(\nu) > (d-1)/2, \quad (2.1)$$

which is a natural generalization to the cone $\mathcal{S}_{++}^d$ of the classical gamma function.

For any shape parameter $\alpha \in (d-1, \infty)$, any scale matrix $\Sigma \in \mathcal{S}_{++}^d$, and any noncentrality parameter $\Theta \in \mathbb{R}^{d \times d}$ such that $\Theta \Sigma^{-1} \in \mathcal{S}_+^d$, the density of the noncentral Wishart distribution, henceforth denoted $\mathcal{W}_d(\alpha, \Sigma, \Theta)$, is given, for all $X \in \mathcal{S}_{++}^d$, by

$$f_{\alpha, \Sigma, \Theta}^{\mathcal{W}}(X) := \frac{|X|^{\alpha/2-(d+1)/2} \text{etr}(-\Sigma^{-1}X/2)}{|2\Sigma|^{\alpha/2} \Gamma_d(\alpha/2)} \text{etr}(-\Theta/2) {}_0F_1(\alpha/2; \Theta \Sigma^{-1}X/4), \quad (2.2)$$

where ${}_0F_1$ denotes the Bessel function of matrix argument introduced by Herz (1955, Section 2). Whenever a random matrix $\mathfrak{X}$ of size $d \times d$ follows this distribution, one writes $\mathfrak{X} \sim \mathcal{W}_d(\alpha, \Sigma, \Theta)$. When $\Theta = 0_{d \times d}$, one recovers the (central) Wishart distribution, and one writes $\mathfrak{X} \sim \mathcal{W}_d(\alpha, \Sigma)$, omitting the noncentrality parameter.

Remark 2.1. For any shape parameter $\alpha \in (d-1, \infty)$, any scale matrix $\Sigma \in \mathcal{S}_{++}^d$, and any noncentrality parameter $\Theta \in \mathbb{R}^{d \times d}$ such that $\Theta \Sigma^{-1} \in \mathcal{S}_+^d$, the Laplace transform of $\mathfrak{W} \sim \mathcal{W}_d(\alpha, \Sigma, \Theta)$ is

$$\mathbb{E}[\text{etr}(-T\mathfrak{W})] = \frac{\text{etr}\{-T\Sigma(I_d + 2T\Sigma)^{-1}\Theta\}}{|I_d + 2T\Sigma|^{\alpha/2}}, \quad T \in \mathcal{S}_+^d;$$

see, e.g., Theorem 10.3.3 of Muirhead (1982).

For a matrix-variate Markov process $(\mathfrak{M}_t)_{t \geq 0}$ taking values in $\mathcal{S}_+^d$, the transition semigroup of operators $(\mathcal{P}_t)_{t \geq 0}$ is defined, for every measurable function f for which the expectation below is finite, by

$$\mathcal{P}_t f(M) := \mathbb{E}[f(\mathfrak{M}_t) \mid \mathfrak{M}_0 = M], \quad t \geq 0.$$

The corresponding infinitesimal generator of $(\mathfrak{M}_t)_{t \geq 0}$ is defined on its domain by

$$\mathcal{A}f(M) := \lim_{s \downarrow 0} \frac{\mathcal{P}_s f(M) - f(M)}{s},$$

provided that the limit exists. More generally, for a diffusion, we use the same notation for the extended generator: if f is sufficiently smooth and there exists a measurable function g such that $(f(\mathfrak{M}_t) - f(\mathfrak{M}_0) - \int_0^t g(\mathfrak{M}_s) ds)_{t \geq 0}$ is a local martingale, then we write $\mathcal{A}f = g$. When f belongs to the domain of the infinitesimal generator, the two notions agree. In this sense, Itô's formula identifies $\mathcal{A}f(M)$ with the drift term in the stochastic differential of $f(\mathfrak{M}_t)$. Such a characterization is central both to describing the dynamics of the Markov process $(\mathfrak{M}_t)_{t \geq 0}$ and to deriving Stein-type identities.

3. Main results

Following [Bru \(1991, Section 5.2\)](#), we define the Wishart process $(\mathfrak{W}_t)_{t \geq 0}$ as a $\mathcal{S}_{++}^d$ -valued generalization of the well-known Cox–Ingersoll–Ross process through the SDE

$$d\mathfrak{W}_t := 2(\alpha\Sigma - \mathfrak{W}_t)dt + \sqrt{2}(\mathfrak{W}_t^{1/2} d\mathfrak{B}_t \Sigma^{1/2} + \Sigma^{1/2} d\mathfrak{B}_t^\top \mathfrak{W}_t^{1/2}), \quad \mathfrak{W}_0 := W, \quad (3.1)$$

where $W, \Sigma \in \mathcal{S}_{++}^d$ are constant matrices, $\alpha > d - 1$ is a scalar, and $(\mathfrak{B}_t)_{t \geq 0}$ is a $d \times d$ matrix of independent standard Brownian motions.

The explicit expression for the extended generator of $(\mathfrak{W}_t)_{t \geq 0}$, denoted $\mathcal{A}^\mathcal{W}$, is derived in [Proposition 3.1](#) below. This result appears without proof in Eq. (5.12) of [Bru \(1991\)](#), under a different parametrization. It also follows from the much more general result of [Cuchiero et al. \(2011, Theorems 2.4 and 2.6\)](#), by taking $a = 2\Sigma$, $b = 2\alpha\Sigma$, $B(S) = -2S$, $c = 0$, $\gamma = 0_{d \times d}$, and $m(\cdot) = \mu(\cdot) = 0$, with the diffusion matrix parameter denoted here by a to avoid confusion with the shape parameter α . A short and self-contained proof is included in [Section 6](#) for completeness.

Proposition 3.1. *For any $f \in C^2(\mathcal{S}_{++}^d)$, the extended generator of $(\mathfrak{W}_t)_{t \geq 0}$ is given by*

$$\mathcal{A}^\mathcal{W}f(S) = 2\operatorname{tr}\{(\alpha\Sigma - S)\nabla f(S)\} + 4\operatorname{tr}\{S\nabla\Sigma\nabla f(S)\}, \quad S \in \mathcal{S}_{++}^d, \quad (3.2)$$

where $\nabla := (\frac{1}{2}(1 + \delta_{ij})\partial/\partial S_{ij})_{1 \leq i, j \leq d}$ denotes the $d \times d$ symmetric gradient, and δ_{ij} denotes the Kronecker delta.

According to [Theorem 4.20 of Pfaffel \(2012\)](#), under the assumption that $\alpha \in [d + 1, \infty) \cap \mathbb{N}$,

$$\mathfrak{W}_t \mid \{\mathfrak{W}_0 = W\} \sim \mathcal{W}_d(\alpha, \Sigma_t, e^{-2t}\Sigma_t^{-1}W), \quad \Sigma_t := (1 - e^{-2t})\Sigma. \quad (3.3)$$

The strengthened assumption on α ensures positive definiteness of the Wishart process.

The next proposition extends [\(3.3\)](#) to the full range $\alpha \in (d - 1, \infty)$. It identifies the transition law of the matrix Wishart process started from any $W \in \mathcal{S}_+^d$, gives its limiting law, and records the semigroup representation and invariance of $\mathcal{W}_d(\alpha, \Sigma)$ needed below.

Proposition 3.2. *Let $\alpha \in (d - 1, \infty)$ and $\Sigma \in \mathcal{S}_{++}^d$ be given, and set $\Sigma_t := (1 - e^{-2t})\Sigma$. For the matrix Wishart process defined in [\(3.1\)](#), with transition kernel on $\mathcal{S}_+^d$, for every initial state $W \in \mathcal{S}_+^d$ we have*

$$\mathfrak{W}_t \mid \{\mathfrak{W}_0 = W\} \sim \mathcal{W}_d(\alpha, \Sigma_t, e^{-2t}\Sigma_t^{-1}W), \quad t > 0, \quad (3.4)$$

$$\mathfrak{W}_t \mid \{\mathfrak{W}_0 = W\} \xrightarrow{\text{law}} \mathfrak{W}_\infty \sim \mathcal{W}_d(\alpha, \Sigma), \quad t \rightarrow \infty. \quad (3.5)$$

In particular, let $(\mathcal{P}_t^{\mathcal{W}})_{t \geq 0}$ be the transition semigroup with kernel $P_t(W, dY)$ on $\mathcal{S}_+^d$, so that, for every bounded Borel measurable function $h : \mathcal{S}_+^d \rightarrow \mathbb{R}$,

$$(\mathcal{P}_t^{\mathcal{W}}h)(W) := \int_{\mathcal{S}_+^d} h(Y) P_t(W, dY) = \mathbb{E}[h(\mathfrak{W}_t) \mid \mathfrak{W}_0 = W]. \quad (3.6)$$

When h is only defined on $\mathcal{S}_{++}^d$, we use the same notation for $t > 0$ and $W \in \mathcal{S}_+^d$ by choosing any bounded Borel extension of h to $\mathcal{S}_+^d$ since $P_t(W, \partial\mathcal{S}_+^d) = 0$. For $\mathfrak{S}_t \sim \mathcal{W}_d(\alpha, I_d, e^{-2t}\Sigma_t^{-1/2}W\Sigma_t^{-1/2})$, the above shows

$$\mathcal{P}_t^{\mathcal{W}}h(W) = \mathbb{E}[h(\Sigma_t^{1/2}\mathfrak{S}_t\Sigma_t^{1/2})]. \quad (3.7)$$

Moreover, given a probability measure μ on $\mathcal{S}_+^d$, the push-forward measure $\mu\mathcal{P}_t^{\mathcal{W}}$ is defined by

$$(\mu\mathcal{P}_t^{\mathcal{W}})(A) := \int_{\mathcal{S}_+^d} P_t(W, A) \mu(dW), \quad A \subseteq \mathcal{S}_+^d \text{ Borel.}$$

Hence, for $\xi := \mathcal{W}_d(\alpha, \Sigma)$, we have the invariance

$$\xi\mathcal{P}_t^{\mathcal{W}} = \xi. \quad (3.8)$$

This leads to the following Stein characterization for the Wishart distribution.

Corollary 3.3 (Stein characterization). *Let $\alpha \in (d-1, \infty)$ and $\Sigma \in \mathcal{S}_{++}^d$ be given. Then*

$$\mathfrak{W} \sim \mathcal{W}_d(\alpha, \Sigma) \quad \Leftrightarrow \quad \mathbb{E}[\mathcal{A}^{\mathcal{W}}f(\mathfrak{W})] = 0 \quad \forall f \in C_{\mathcal{A}^{\mathcal{W}}}^2(\mathcal{S}_{++}^d),$$

where, for $\mathfrak{W}_\infty \sim \mathcal{W}_d(\alpha, \Sigma)$,

$$\begin{aligned} C_{\mathcal{A}^{\mathcal{W}}}^2(\mathcal{S}_{++}^d) := \Big\{ f \in C^2(\mathcal{S}_{++}^d) : & \mathbb{E}[|\text{tr}\{\Sigma \nabla f(\mathfrak{W}_\infty)\}|] < \infty, \mathbb{E}[|\text{tr}\{\mathfrak{W}_\infty \nabla f(\mathfrak{W}_\infty)\}|] < \infty, \\ & \mathbb{E}[|\text{tr}\{\mathfrak{W}_\infty \nabla \Sigma \nabla f(\mathfrak{W}_\infty)\}|] < \infty, \mathbb{E}[|f(\mathfrak{W}_\infty)|] < \infty, \\ & \text{and } \mathbb{E}[\|\mathfrak{W}_\infty^{1/2} \nabla f(\mathfrak{W}_\infty) \Sigma^{1/2}\|_F^2] < \infty \Big\}. \end{aligned} \quad (3.9)$$

Remark 3.1. When $d = 1$, the space $\mathcal{S}_{++}^1$ is naturally identified with $(0, \infty)$ and the scale matrix Σ is a positive scalar. Under this identification, $\mathfrak{W} \sim \mathcal{W}_1(\alpha, \Sigma) \equiv \Gamma(r = \alpha/2, \lambda = 1/(2\Sigma))$, where the gamma distribution is parametrized as in (1.1). Moreover, (3.2) becomes

$$\mathcal{A}^{\mathcal{W}}f(x) = 2(\alpha\Sigma - x)f'(x) + 4\Sigma xf''(x) = 4\Sigma \{xf''(x) + (r - \lambda x)f'(x)\}.$$

Thus Corollary 3.3 reduces, up to the positive multiplicative factor 4Σ and a slightly more restrictive class of test functions, to the gamma Stein characterization (1.1).

Remark 3.2. Aside from the different class of test functions, the forward implication in Corollary 3.3 is a special case of the classical Stein–Haff identity for the Wishart distribution (Haff, 1979, Theorem 2.1). To compare notations, set $p = d$ and $k = \alpha$. If $\mathfrak{W} \sim \mathcal{W}_d(\alpha, \Sigma)$, then, under the regularity, integrability and boundary assumptions in that theorem, for a scalar function $h : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$ and a matrix-valued function $F : \mathcal{S}_{++}^d \rightarrow \mathbb{R}^{d \times d}$,

$$\begin{aligned} \mathbb{E}[h(\mathfrak{W})\text{tr}\{F(\mathfrak{W})\Sigma^{-1}\}] &= 2\mathbb{E}[h(\mathfrak{W})(\text{div}^*F_{(1/2)})(\mathfrak{W})] + 2\mathbb{E}\left[\text{tr}\left\{\frac{\partial h}{\partial S}(\mathfrak{W})F_{(1/2)}(\mathfrak{W})\right\}\right] \\ &\quad + \{\alpha - (d+1)\}\mathbb{E}[h(\mathfrak{W})\text{tr}\{\mathfrak{W}^{-1}F(\mathfrak{W})\}]. \end{aligned} \quad (3.10)$$

Here, $F_{(1/2)}(S)$ denotes the matrix obtained by multiplying the off-diagonal entries of $F(S)$ by $1/2$, and div^* denotes the matrix divergence, namely

$$(\text{div}^* F)(S) := \sum_{i,j=1}^d \frac{\partial F_{ij}(S)}{\partial S_{ij}}.$$

Taking $h = 1$ in (3.10) gives

$$\mathbb{E} [2(\text{div}^* F_{(1/2)})(\mathfrak{W}) + \{\alpha - (d+1)\} \text{tr}\{\mathfrak{W}^{-1} F(\mathfrak{W})\} - \text{tr}\{F(\mathfrak{W}) \Sigma^{-1}\}] = 0. \quad (3.11)$$

For a scalar function $f \in C^2(\mathcal{S}_{++}^d)$ for which this specialization is justified, take

$$F(S) = 2S \nabla f(S) \Sigma.$$

Using the symmetric-gradient convention from Proposition 3.1, so that $\nabla_{ij} S_{kl} = (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk})/2$, and using the symmetries of Σ and ∇ , one has

$$\begin{aligned} 4 \text{div}^* \{(S \nabla f(S) \Sigma)_{(1/2)}\} &= 4 \sum_{i,j,k=1}^d \nabla_{ij} \{S_{ik} (\nabla f(S) \Sigma)_{kj}\} \\ &= 4 \sum_{i,j,k=1}^d \frac{\delta_{jk} + \delta_{ij} \delta_{ik}}{2} (\nabla f(S) \Sigma)_{kj} + 4 \sum_{i,j,k,\ell=1}^d S_{ik} \Sigma_{\ell j} \nabla_{ij} \nabla_{k\ell} f(S) \\ &= 2(d+1) \text{tr}\{\Sigma \nabla f(S)\} + 4 \text{tr}\{S \nabla \Sigma \nabla f(S)\}. \end{aligned}$$

Substituting $F(S) = 2S \nabla f(S) \Sigma$ into (3.11) therefore yields

$$\mathbb{E} [2 \text{tr}\{(\alpha \Sigma - \mathfrak{W}) \nabla f(\mathfrak{W})\} + 4 \text{tr}\{\mathfrak{W} \nabla \Sigma \nabla f(\mathfrak{W})\}] = 0,$$

that is,

$$\mathbb{E} [\mathcal{A}^{\mathcal{W}} f(\mathfrak{W})] = 0.$$

Thus the forward implication in Corollary 3.3 is recovered from the Stein–Haff identity by restricting Haff’s arbitrary matrix field to the gradient-type field $F(S) = 2S \nabla f(S) \Sigma$. Conversely, the full Stein–Haff identity is not a formal consequence of the forward implication in Corollary 3.3, since it applies to arbitrary matrix fields $F(S)$ and scalar multipliers $h(S)$, whereas the generator identity tests only fields generated by scalar potentials f .

The fundamental differential operator matrix appearing in the sequel is

$$\mathcal{D} := \nabla_X \text{adj}(I_d - 2\nabla_X),$$

where ∇_X denotes the symmetric gradient with respect to the matrix variable X . This is the transfer operator behind Lemma 5.3: derivatives in the noncentrality parameter of the Wishart kernel are rewritten as applications of $\mathcal{D}$ to the test function. The seminorms built from $\mathcal{D}$, and later from its time-dependent analogue $\mathcal{D}_{t,\Sigma}$, are therefore the quantities that control the semigroup contraction estimate (Lemma 3.4), the absolute convergence of the Stein-solution representation (Theorem 3.5), and the derivative bounds for the Stein solution (Theorem 3.6).

The next lemma bounds the Wishart semigroup’s deviation from its stationary limit by leveraging the test function’s Lipschitz and $\mathcal{D}$ -regularity. This decay rate is crucial to obtain well-defined solutions to the Wishart Stein equation, as it ensures the absolute convergence of the integral in the semigroup representation presented in Theorem 3.5 below.

Lemma 3.4. *Let $\alpha > 3d - 3$ and $\Sigma \in \mathcal{S}_{++}^d$ be given. Let $(\mathfrak{W}_t)_{t \geq 0}$ be the Wishart process defined in (3.1), with transition semigroup $(\mathcal{P}_t^{\mathcal{W}})_{t \geq 0}$ and stationary limiting distribution $\xi = \mathcal{W}_d(\alpha, \Sigma)$. For any given $t > 0$, let $\Sigma_t = (1 - e^{-2t})\Sigma$ and define*

$$h_t(X) := h(\Sigma_t^{1/2} X \Sigma_t^{1/2}), \quad X \in \mathcal{S}_{++}^d,$$

with $h : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$ being bounded and Lipschitz. Assume that $h \in C^d(\mathcal{S}_{++}^d)$ and that for every differential monomial $\bar{\nabla}$ of order at most d in the entries of ∇_X , there exist constants $C_{\bar{\nabla}}, N_{\bar{\nabla}} \geq 0$ such that

$$|\bar{\nabla} h(X)| \leq C_{\bar{\nabla}}(1 + \|X\|_F^{N_{\bar{\nabla}}}), \quad X \in \mathcal{S}_{++}^d, \quad (3.12)$$

and assume that

$$M_t^{\mathcal{D}}(h) := \sup_{X \in \mathcal{S}_{++}^d} \sup_{\substack{U \in \mathcal{S}^d \\ \|U\|_F=1}} |\langle U, \mathcal{D} \rangle_F h_t(X)| < \infty. \quad (3.13)$$

Then, for every $W \in \mathcal{S}_{++}^d$, we have

$$\begin{aligned} |(\mathcal{P}_t^{\mathcal{W}} h)(W) - \mathbb{E}[h(\mathfrak{W}_\infty)]| &\leq \frac{e^{-2t}}{1 - e^{-2t}} M_t^{\mathcal{D}}(h) \|\Sigma^{-1/2} W \Sigma^{-1/2}\|_F \\ &\quad + e^{-2t} [h]_1 \|\Sigma\|_2 \mathbb{E}[\|\mathfrak{W}\|_F], \end{aligned} \quad (3.14)$$

where $\mathfrak{W} \sim \mathcal{W}_d(\alpha, I_d)$.

Define the smoothing operator

$$(\mathcal{Q}_t h)(\Lambda) := \int_{\mathcal{S}_{++}^d} h(X) f_{\alpha, I_d, e^{-2t}\Lambda}^{\mathcal{W}}(X) dX, \quad \Lambda \in \mathcal{S}_+^d, \quad t \geq 0. \quad (3.15)$$

By Proposition 3.2, for every bounded Borel measurable function $h : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$ and every $W \in \mathcal{S}_+^d$, we have

$$\mathcal{P}_t^{\mathcal{W}} h(W) = (\mathcal{Q}_t h_t)(\Lambda_t(W)), \quad t > 0,$$

where $\Lambda_t(W) := \Sigma_t^{-1/2} W \Sigma_t^{-1/2}$, $h_t(X) = h(\Sigma_t^{1/2} X \Sigma_t^{1/2})$, and $\Sigma_t = (1 - e^{-2t})\Sigma$. Theorem 3.5 below provides an explicit semigroup solution of the Wishart Stein equation for bounded Lipschitz test functions whose rescaled family $(h_t)_{t>0}$ has uniformly bounded first $\mathcal{D}$ -seminorm away from $t = 0$.

Theorem 3.5 (Solution of the Wishart Stein equation for C_b^d test functions). *Let $\alpha > 3d - 3$ and $\Sigma \in \mathcal{S}_{++}^d$ be given. Let $(\mathfrak{W}_t)_{t \geq 0}$ be the Wishart process in (3.1) with transition semigroup $(\mathcal{P}_t^{\mathcal{W}})_{t \geq 0}$, extended generator $\mathcal{A}^{\mathcal{W}}$ given by (3.2), and stationary limiting distribution $\xi = \mathcal{W}_d(\alpha, \Sigma)$ from Proposition 3.2. Let $h \in C_b^d(\mathcal{S}_{++}^d)$ be real-valued. Since $\mathcal{S}_{++}^d$ is convex, h is bounded and Lipschitz; let $[h]_1$ denote its minimum Lipschitz constant with respect to the Frobenius norm. Set*

$$M_{1,\Sigma}^{\mathcal{D}}(h) := \sup_{t \geq 1} \sup_{X \in \mathcal{S}_{++}^d} \sup_{\substack{U \in \mathcal{S}^d \\ \|U\|_F=1}} |\langle U, \mathcal{D} \rangle_F h_t(X)|, \quad \langle U, \mathcal{D} \rangle_F := \sum_{i,j=1}^d U_{ij} \mathcal{D}_{ij}.$$

Then $M_{1,\Sigma}^{\mathcal{D}}(h) < \infty$, and the following hold. The function

$$f_h(W) := - \int_0^\infty \left\{ \mathcal{P}_t^{\mathcal{W}} h(W) - \mathbb{E}[h(\mathfrak{W}_\infty)] \right\} dt \quad (3.16)$$

is well defined for every $W \in \mathcal{S}_{++}^d$ and solves the Wishart Stein equation

$$\mathcal{A}^W f_h(W) = h(W) - \mathbb{E}[h(\mathfrak{W}_\infty)], \quad W \in \mathcal{S}_{++}^d, \quad (3.17)$$

where $\mathfrak{W}_\infty \sim \mathcal{W}_d(\alpha, \Sigma)$. Moreover, if $\mathfrak{V} \sim \mathcal{W}_d(\alpha, I_d)$, then

$$|f_h(W)| \leq 2\|h\|_\infty + \frac{1}{2} \left\{ \frac{M_{1,\Sigma}^{\mathcal{P}}(h)}{1 - e^{-2}} \|\Sigma^{-1/2} W \Sigma^{-1/2}\|_F + [h]_1 \|\Sigma\|_2 \mathbb{E}[\|\mathfrak{V}\|_F] \right\}. \quad (3.18)$$

Remark 3.3. The condition $\alpha > 3d - 3$ is a sufficient condition imposed by the differentiation argument used in the proof of Theorem 3.5 and explained below, and is not a sharp existence condition for the Wishart process or for its transition law. Indeed, Proposition 3.2 works already in the larger range $\alpha > d - 1$. The stronger lower bound enters through Lemma 5.3, whose proof transfers derivatives in the noncentrality parameter to X -derivatives using Lemma 5.2 and the integration-by-parts estimate of Lemma 5.1. It is not required for the existence of the expectation on the right-hand side of (5.14); as shown in Remark 5.1, that expectation is finite under the natural condition $\alpha > d - 1$.

A single differentiation with respect to an entry of the noncentrality parameter Θ gives, at the level of Laplace transforms and up to sign, an entry of the matrix $T(I_d + 2T)^{-1}$. We then write

$$T(I_d + 2T)^{-1} = \frac{T \operatorname{adj}(I_d + 2T)}{|I_d + 2T|}.$$

This identity has two effects. The denominator $|I_d + 2T|^{-1}$ shifts the Wishart shape parameter from α to $\alpha + 2$, since it changes the determinant factor $|I_d + 2T|^{-\alpha/2}$ into $|I_d + 2T|^{-(\alpha+2)/2}$. The numerator $T \operatorname{adj}(I_d + 2T)$ is a polynomial of degree at most d in the entries of T , so Lemma 5.1 has to be applied to $f_{\alpha+2, I_d, \Theta}^W$ with $m = d$. Its condition is therefore $\alpha + 2 > d - 1 + 2d$. Equivalently, after dividing by 2, this is $(\alpha + 2)/2 > (d - 1)/2 + d$.

The term $(d - 1)/2$ is the usual base integrability threshold in the multivariate gamma integral on $\mathcal{S}_{++}^d$. For a shape parameter γ , the determinant factor in the Wishart density is $|X|^{\gamma/2 - (d+1)/2}$, and the matrix-gamma integrability condition near $\partial\mathcal{S}_+^d$ is $\gamma/2 > (d - 1)/2$. In the one-step argument of Lemma 5.2, the relevant shifted shape is $\gamma = \alpha + 2$, not α , because of the denominator above. The additional $+d$ in $(\alpha + 2)/2 > (d - 1)/2 + d$ accounts for the worst possible loss caused by the differential operator of order at most d coming from $T \operatorname{adj}(I_d + 2T)$. Heuristically, each X -derivative may lower by one the exponent of the determinant factor controlling the behavior near $\partial\mathcal{S}_+^d$. Thus, after up to d derivatives, the determinant factor behaves as if its exponent were $\gamma/2 - (d + 1)/2 - d$. Requiring this exponent to remain larger than -1 gives $\gamma/2 > (d - 1)/2 + d$, and with $\gamma = \alpha + 2$ this is exactly $\alpha > 3d - 3$.

Thus $\alpha > 3d - 3$ guarantees enough boundary decay to justify the integration by parts and to avoid boundary contributions in the proof. It should be viewed as a regularity assumption for the method used here, rather than as a proved necessary condition for the existence of the expectation in (5.14), for the existence of the solution of the Stein equation, or for derivative estimates obtained by a different method. It remains an open problem to extend Lemma 5.3, and the resulting Stein solution estimates, under the natural condition $\alpha > d - 1$.

Next, we derive bounds on the derivatives of the semigroup solution of the Wishart Stein equation. Define the following time-dependent differential operator matrix:

$$\mathcal{D}_{t,\Sigma} := \nabla_W \operatorname{adj}(I_d - 2\Sigma_t \nabla_W), \quad \Sigma_t = (1 - e^{-2t})\Sigma.$$

For any integer $p \geq 0$, let $\mathcal{B}_p^{\mathcal{D}, \Sigma}(\mathcal{S}_{++}^d)$ denote the class of functions $h : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$ such that, for every choice of index pairs $(i_1, j_1), \dots, (i_p, j_p) \in [d]^2$, the iterated derivatives $(\prod_{\ell=1}^p (\mathcal{D}_{t, \Sigma})_{i_\ell j_\ell}) h$ exist for every $t > 0$ and satisfy

$$\sup_{t>0} \left\| \left(\prod_{\ell=1}^p (\mathcal{D}_{t, \Sigma})_{i_\ell j_\ell} \right) h \right\|_\infty < \infty,$$

with the convention that the order-0 iterate is h itself.

Theorem 3.6 (Regularity of the semigroup solution of the Wishart Stein equation). *Let $m \in \mathbb{N}$, let $\alpha > 3d - 3$ and $\Sigma \in \mathcal{S}_{++}^d$. Let $h \in C_b^{md}(\mathcal{S}_{++}^d)$ be real-valued, and let f_h denote the semigroup solution of the Wishart Stein equation defined in (3.16). Then the following hold:*

(i) *If $h \in \mathcal{B}_m^{\mathcal{D}, \Sigma}(\mathcal{S}_{++}^d)$, then for any index pairs $(i_1, j_1), \dots, (i_m, j_m) \in [d]^2$,*

$$\left\| \left(\prod_{\ell=1}^m \nabla_{W, i_\ell j_\ell} \right) f_h \right\|_\infty \leq \frac{1}{2m} \sup_{t>0} \left\| \left(\prod_{\ell=1}^m (\mathcal{D}_{t, \Sigma})_{i_\ell j_\ell} \right) h \right\|_\infty. \quad (3.19)$$

(ii) *For any sufficiently smooth $g : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$, define*

$$\mathcal{M}_m(g) := \sup_{W \in \mathcal{S}_{++}^d} \sup_{\substack{U_1, \dots, U_m \in \mathcal{S}^d \\ \|U_1\|_F = \dots = \|U_m\|_F = 1}} |D^m g(W)[U_1, \dots, U_m]|. \quad (3.20)$$

Define also

$$\mathcal{M}_m^{\mathcal{D}, \Sigma}(h) := \sup_{t>0} \sup_{W \in \mathcal{S}_{++}^d} \sup_{\substack{U_1, \dots, U_m \in \mathcal{S}^d \\ \|U_1\|_F = \dots = \|U_m\|_F = 1}} \left| \left(\prod_{\ell=1}^m \langle U_\ell, \mathcal{D}_{t, \Sigma} \rangle_F \right) h(W) \right|, \quad (3.21)$$

where

$$\langle U, \mathcal{D}_{t, \Sigma} \rangle_F := \sum_{i, j=1}^d U_{ij} (\mathcal{D}_{t, \Sigma})_{ij}.$$

If $\mathcal{M}_m^{\mathcal{D}, \Sigma}(h) < \infty$, then

$$\mathcal{M}_m(f_h) \leq \frac{1}{2m} \mathcal{M}_m^{\mathcal{D}, \Sigma}(h). \quad (3.22)$$

Remark 3.4. The seminorms of h appearing in (3.19) and (3.21) can be bounded explicitly in terms of ordinary derivatives of h , at the cost of rather large constants depending on d and Σ . For $r \in \mathbb{N}$, set

$$\mathcal{N}_r(h) := \max_{1 \leq q \leq r} \mathcal{M}_q(h), \quad A_{d, \Sigma} := d! (1 + 2\sqrt{d} \|\Sigma\|_2)^{d-1}, \quad B_{d, \Sigma} := d A_{d, \Sigma},$$

where $\mathcal{M}_q$ is defined as in (3.20). Then, for any index pairs $(i_\ell, j_\ell) \in [d]^2$,

$$\sup_{t>0} \left\| \left(\prod_{\ell=1}^m (\mathcal{D}_{t, \Sigma})_{i_\ell j_\ell} \right) h \right\|_\infty \leq A_{d, \Sigma}^m \mathcal{N}_{md}(h), \quad (3.23)$$

and

$$\mathcal{M}_m^{\mathcal{D},\Sigma}(h) \leq B_{d,\Sigma}^m \mathcal{N}_{md}(h). \quad (3.24)$$

Consequently, (3.19) gives

$$\left\| \left(\prod_{\ell=1}^m \nabla_{W, i_\ell j_\ell} \right) f_h \right\|_\infty \leq \frac{A_{d,\Sigma}^m}{2m} \mathcal{N}_{md}(h),$$

and (3.22) gives

$$\mathcal{M}_m(f_h) \leq \frac{B_{d,\Sigma}^m}{2m} \mathcal{N}_{md}(h).$$

Indeed, the entries of $I_d - 2\Sigma_t \nabla_W$ have coefficient ℓ^1 -norm at most $1 + 2\sqrt{d}\|\Sigma\|_2$, uniformly in $t > 0$, because $\|\Sigma_t\|_2 \leq \|\Sigma\|_2$. Each entry of $\text{adj}(I_d - 2\Sigma_t \nabla_W)$ is a signed sum of $(d-1)!$ products of $d-1$ such entries. Multiplication by the outer matrix ∇_W and summation over one index give the factor d , and therefore each entry $(\mathcal{D}_{t,\Sigma})_{ij}$ is a constant-coefficient differential operator of order at most d whose total coefficient size is at most $A_{d,\Sigma}$. Iterating m such operators gives (3.23). For the directional seminorm, if $\|U\|_F = 1$, then $\sum_{i,j=1}^d |U_{ij}| \leq d$, so each directional operator $\langle U, \mathcal{D}_{t,\Sigma} \rangle_F$ has total coefficient size at most $B_{d,\Sigma}$, which gives (3.24).

The constants in (3.23) and (3.24) are included to make the dependence on d and Σ explicit and to remove the supremum over t from the test-function seminorms. The sharper seminorm $\mathcal{M}_m^{\mathcal{D},\Sigma}(h)$ is kept in the main statement because it tracks the actual differential operator generated by the Wishart kernel and can be substantially smaller than the right-hand side of (3.24).

Remark 3.5. The appearance of derivatives of h up to order md in Theorem 3.6 comes from the way derivatives of the Wishart transition semigroup are transferred from the initial state to the integration variable in the noncentral Wishart kernel. By Proposition 3.2, the initial state enters $(\mathcal{P}_t^W h)(W)$ through the noncentrality parameter $e^{-2t}\Lambda_t(W)$, where $\Lambda_t(W) = \Sigma_t^{-1/2} W \Sigma_t^{-1/2}$. Since $W \mapsto \Lambda_t(W)$ is linear, each derivative with respect to W gives one derivative with respect to the noncentrality parameter. Lemma 5.3 then transfers such a derivative to an X -differential operator:

$$\nabla_{\Lambda,ij}(\mathcal{Q}_t^{(\gamma)} g)(\Lambda) = e^{-2t} \mathcal{Q}_t^{(\gamma+2)}(\mathcal{D}_{ij}g)(\Lambda), \quad \mathcal{D} = \nabla_X \text{adj}(I_d - 2\nabla_X), \quad (3.25)$$

where $\mathcal{Q}_t^{(\gamma)}$ denotes the operator in (3.15) with shape parameter γ in place of α . The operator $\text{adj}(I_d - 2\nabla_X)$ is a polynomial of degree at most $d-1$ in the entries of ∇_X , and the additional factor ∇_X makes $\mathcal{D}$ an operator of order at most d . Thus one derivative in the initial state may cost up to d derivatives of the test function. Iterating (3.25) m times gives a product of m such operators. The resulting operator has order at most md , and the proof therefore requires enough regularity of h to control all the derivatives that can arise from this product. This is the reason for the hypotheses involving $C_b^{md}(\mathcal{S}_{++}^d)$, $\mathcal{B}_m^{\mathcal{D},\Sigma}(\mathcal{S}_{++}^d)$, and $\mathcal{M}_m^{\mathcal{D},\Sigma}(h)$.

This derivative loss is not merely an artifact of estimating too many terms separately. The product of m transfer operators obtained by iterating (3.25) has genuine order md in general: although there are cancellations in some entries of $\mathcal{D}$, these cancellations do not remove the highest-order part of the operator in general. For instance, in the scale-normalized case the diagonal entries of $\mathcal{D}$ retain order d , and iterating such entries gives operators of order md . Thus the md derivative count is sharp for the transfer identity and for the $\mathcal{D}$ -seminorm bounds used in Theorem 3.6. This does not by itself prove that md derivatives of h are necessary for the mere

existence of the m -th derivative of the solution of the Stein equation, since a different argument might avoid transferring all derivatives onto h through (3.25).

Remark 3.6. When $d = 1$, the adjugate in the definition of $\mathcal{D}_{t,\Sigma}$ is the adjugate of a 1×1 matrix and hence is equal to 1. Since the symmetric gradient is then the ordinary derivative, one has $\mathcal{D}_{t,\Sigma} = d/dx$, $t > 0$. Thus (3.19) becomes

$$\|f_h^{(m)}\|_\infty \leq \frac{1}{2m} \|h^{(m)}\|_\infty, \quad m \geq 1.$$

As in Remark 3.1, $\mathcal{W}_1(\alpha, \Sigma) \equiv \Gamma(r = \alpha/2, \lambda = 1/(2\Sigma))$ and $\mathcal{A}^\mathcal{W} = 4\Sigma \mathcal{A}^\Gamma$. Thus, if $f_h^\Gamma := 4\Sigma f_h$ denotes the corresponding solution of the gamma Stein equation, then

$$\|(f_h^\Gamma)^{(m)}\|_\infty \leq \frac{2\Sigma}{m} \|h^{(m)}\|_\infty = \frac{1}{m\lambda} \|h^{(m)}\|_\infty, \quad m \geq 1,$$

which is (1.9).

4. Applications

4.1. Quantitative Wishart approximation of uncentered group-mean scatter matrices

The Wishart distribution naturally arises in multivariate analysis of variance (MANOVA) as the exact finite-sample distribution of the uncentered group-mean scatter matrix when the underlying populations are Gaussian. Indeed, consider a balanced design with ν independent groups, where each group $k \in [\nu]$ consists of n independent observations $\mathbf{Z}_{k,i} \in \mathbb{R}^d$ drawn from a normal distribution $\mathcal{N}_d(\mathbf{0}_d, \Sigma)$ with $\nu \in \mathbb{N} \cap [d, \infty)$ and $\Sigma \in \mathcal{S}_{++}^d$. If we define the standardized group means as $\mathbf{X}_k^{(n)} := n^{-1/2} \sum_{i=1}^n \mathbf{Z}_{k,i}$, the matrix composed of these vectors,

$$\mathfrak{X}^{(n)} := [\mathbf{X}_1^{(n)}, \dots, \mathbf{X}_\nu^{(n)}]^\top,$$

follows a matrix normal distribution $\mathcal{N}_{\nu \times d}(0_{\nu \times d}, I_\nu \otimes \Sigma)$. Its density with respect to the Lebesgue measure is given, for all $X \in \mathbb{R}^{\nu \times d}$, by $\phi_{0_{\nu \times d}, I_\nu, \Sigma}(X) = (2\pi)^{-\nu d/2} |\Sigma|^{-\nu/2} \text{etr}\{-\frac{1}{2}X\Sigma^{-1}X^\top\}$; see, e.g., Gupta and Nagar (2000, Theorem 2.2.1). The uncentered group-mean scatter matrix $(\mathfrak{X}^{(n)})^\top \mathfrak{X}^{(n)}$ then has a Wishart distribution, namely

$$(\mathfrak{X}^{(n)})^\top \mathfrak{X}^{(n)} \sim \mathcal{W}_d(\nu, \Sigma);$$

see, e.g., Muirhead (1982, Theorem 10.3.2).

When the underlying raw observations $\mathbf{Z}_{k,i}$ have mean zero and covariance Σ but are not exactly Gaussian, the multivariate central limit theorem implies that the standardized group means $\mathbf{X}_k^{(n)}$ converge to a Gaussian distribution as the group size $n \rightarrow \infty$. By the continuous mapping theorem, the scatter matrix converges in law to a Wishart distribution for any fixed number of groups ν . Instead of applying the Wishart Stein equation directly, we quantify this convergence by applying Stein's method for matrix normal approximation (Gaunt et al., 2026) to the matrix of standardized group means and then by pushing the result forward through the quadratic map $X \mapsto X^\top X$.

The connection between the two Stein operators is explicit and is isolated in the following lemma. In the mean-zero case considered here, the matrix normal Stein operator of [Gaunt et al. \(2026\)](#), specialized to row covariance $\Psi = I_\nu$ and column covariance Σ , is

$$\mathcal{A}^{\text{OU}} g(X) = -\text{tr}\{X^\top \nabla_X g(X)\} + \text{tr}\{\Sigma \nabla_X^\top \nabla_X g(X)\}, \quad X \in \mathbb{R}^{\nu \times d}, \quad (4.1)$$

where $\nabla_X = (\nabla_{X,ar})_{a \in [\nu], r \in [d]}$ denotes the ordinary rectangular gradient, with $\nabla_{X,ar} = \partial/\partial X_{ar}$.

Lemma 4.1 (Wishart pullback of the matrix normal Stein operator). *Let $\nu \in [d, \infty) \cap \mathbb{N}$, $\Sigma \in \mathcal{S}_{++}^d$, and let $f : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$ be smooth. Set*

$$\mathcal{R}_{\nu,d} := \{X \in \mathbb{R}^{\nu \times d} : \text{rank}(X) = d\}.$$

The set $\mathcal{R}_{\nu,d}$ is open, and its complement has Lebesgue measure zero. For $X \in \mathcal{R}_{\nu,d}$, define $g_f(X) = f(X^\top X)$. Then, pointwise on $\mathcal{R}_{\nu,d}$,

$$(\mathcal{A}^{\text{OU}} g_f)(X) = (\mathcal{A}_{\nu,\Sigma}^{\mathcal{W}} f)(X^\top X), \quad X \in \mathcal{R}_{\nu,d}, \quad (4.2)$$

where ∇f denotes the symmetric gradient in the variable S and

$$(\mathcal{A}_{\nu,\Sigma}^{\mathcal{W}} f)(S) = 2 \text{tr}\{(\nu \Sigma - S) \nabla f(S)\} + 4 \text{tr}\{S \nabla \Sigma \nabla f(S)\}, \quad S \in \mathcal{S}_{++}^d.$$

For the proof of Proposition 4.2 below, we use a standard positive-definite regularization of the quadratic pullback. This regularization is not meant to extend the unregularized Wishart generator to the boundary of $\mathcal{S}_{++}^d$; rather, it gives a rectangular test function whose derivatives are defined on all of $\mathbb{R}^{\nu \times d}$, which is needed for the Taylor expansions along line segments in $\mathbb{R}^{\nu \times d}$. With f as in Lemma 4.1, for $\varepsilon > 0$, set

$$g_{f,\varepsilon}(X) := f(X^\top X + \varepsilon I_d), \quad S_\varepsilon(X) := X^\top X + \varepsilon I_d, \quad X \in \mathbb{R}^{\nu \times d}.$$

Then the regularized identity corresponding to (4.2) is, for every $X \in \mathbb{R}^{\nu \times d}$,

$$\begin{aligned} (\mathcal{A}^{\text{OU}} g_{f,\varepsilon})(X) &= 2 \text{tr}\{(\nu \Sigma - X^\top X) \nabla f(S_\varepsilon(X))\} + 4 \text{tr}\{X^\top X \nabla \Sigma \nabla f(S_\varepsilon(X))\} \\ &= (\mathcal{A}_{\nu,\Sigma}^{\mathcal{W}} f)(S_\varepsilon(X)) + 2\varepsilon \text{tr}\{\nabla f(S_\varepsilon(X))\} - 4\varepsilon \text{tr}\{I_d \nabla \Sigma \nabla f(S_\varepsilon(X))\}. \end{aligned} \quad (4.3)$$

Since $\mathcal{S}_{++}^d$ is convex, every $h \in C_b^1(\mathcal{S}_{++}^d)$ is Lipschitz on $\mathcal{S}_{++}^d$ and admits a unique bounded continuous extension to $\mathcal{S}_+^d$. In the statement of the proposition below, $h(\mathfrak{W}_\nu^{(n)})$ is understood through this extension whenever $\mathfrak{W}_\nu^{(n)}$ is possibly singular. We make use of the connection (4.3) to generalize Theorem 3.1 of [Gaunt et al. \(2017\)](#) to the Wishart setting.

Proposition 4.2 (Quantitative Wishart approximation for MANOVA). *Let $\nu \in [d, \infty) \cap \mathbb{N}$ satisfy $\nu > 3d - 3$, and $\Sigma \in \mathcal{S}_{++}^d$. Let $\mathfrak{W} \sim \mathcal{W}_d(\nu, \Sigma)$, and let $\mathfrak{Z}_1, \dots, \mathfrak{Z}_n$ be independent and identically distributed (iid) random matrices in $\mathbb{R}^{\nu \times d}$ satisfying $\mathbb{E}[\mathfrak{Z}_1] = 0_{\nu \times d}$, $\mathbb{E}[\text{vec}(\mathfrak{Z}_1) \text{vec}(\mathfrak{Z}_1^\top)^\top] = I_\nu \otimes \Sigma$, and assume that $\mathbb{E}[\|\mathfrak{Z}_1\|_F^8] < \infty$. For $n \in \mathbb{N}$, set*

$$\mathfrak{X}^{(n)} := \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathfrak{Z}_i, \quad \mathfrak{W}_\nu^{(n)} := (\mathfrak{X}^{(n)})^\top \mathfrak{X}^{(n)}.$$

Let $h \in C_b^{5d}(\mathcal{S}_{++}^d)$ be real-valued and assume that $\sum_{m=1}^5 \mathcal{M}_m^{\mathcal{D}, \Sigma}(h) < \infty$, where $\mathcal{M}_m^{\mathcal{D}, \Sigma}(h)$ is defined in (3.21). Define

$$\alpha_j := r_j + s_j \sum_{(a,r),(b,s),(c,t) \in [\nu] \times [d]} |\mathbb{E}[(\mathfrak{Z}_1)_{ar}(\mathfrak{Z}_1)_{bs}(\mathfrak{Z}_1)_{ct}]|, \quad j \in \{2, 3, 4, 5\},$$

where

$$(r_2, r_3, r_4, r_5) := (2, 54, 1547, 0), \quad (s_2, s_3, s_4, s_5) := (2, 85, 6891, 629\,799).$$

Then, for all $n \geq 2$,

$$\begin{aligned} |\mathbb{E}[h(\mathfrak{W}_\nu^{(n)})] - \mathbb{E}[h(\mathfrak{W})]| &\leq \frac{1}{n} \times (\nu d)^8 (1 + K) \left(1 \vee \max_{(a,r) \in [\nu] \times [d]} \mathbb{E}[|(\mathfrak{Z}_1)_{ar}|^8] \right) \\ &\quad \times \left\{ \alpha_2 \mathcal{M}_2^{\mathcal{D}, \Sigma}(h) + \alpha_3 \mathcal{M}_3^{\mathcal{D}, \Sigma}(h) + \alpha_4 \mathcal{M}_4^{\mathcal{D}, \Sigma}(h) + \alpha_5 \mathcal{M}_5^{\mathcal{D}, \Sigma}(h) \right\}, \end{aligned}$$

where $K := 8\pi \|\Sigma^{-1}\|_\infty \{1 + \|\Sigma\|_2^3 \nu d (\nu d + 2)(\nu d + 4)\}$ and $\|\Sigma^{-1}\|_\infty := \max_{r \in [d]} \sum_{s=1}^d |(\Sigma^{-1})_{rs}|$.

4.2. Quantitative multivariate Satterthwaite approximation

We now turn to the multivariate Satterthwaite approximation, in which the distribution of a positive linear combination of independent Wishart random matrices is approximated by a single Wishart distribution with matching first two moments. Such linear combinations of Wishart random matrices arise in many problems in multivariate statistics: MANOVA when the usual homoscedastic, independent-error covariance structure is relaxed but can be expressed as a Kronecker product (Tan and Gupta, 1983; Naik and Rao, 2001; Mortarino, 2005), matrix quadratic forms (Khatri, 1966; Singull and Koski, 2012), and robustness studies involving multivariate normal mixture distributions (Tan, 1980). This is the multivariate analog of the classical Satterthwaite approximation introduced by Satterthwaite (1941, 1946), in which the distribution of a linear combination of chi-square random variables is approximated by a single gamma distribution with matching first two moments, providing a tractable sidestep to an otherwise difficult to handle distribution.

Khatri (1971) developed exact series representations, whose truncations yield approximations, for the density function of positive definite quadratic forms in normal vectors. In the central case, these quadratic forms are equivalent to positive linear combinations of independent rank-one Wishart random matrices with a common scale matrix and, by grouping equal coefficients, to sums of independent Wishart random matrices with integer shape parameters and proportional scale matrices. While such truncated series approximations appear to be more accurate than a Satterthwaite approximation, they do not directly apply to the heterogeneous setting considered below, where $\mathfrak{G}_j \sim \mathcal{W}_d(\alpha_j, \Sigma_j)$ may have arbitrary integer shape parameters and non-proportional scale matrices. Furthermore, theoretically assessing the closeness between the true distribution and the truncated series representation remains an open question. On the other hand, Stein's method is tailored to treating this problem in the case of approximation by a single Wishart distribution.

Let $N \in \mathbb{N}$, $\alpha_1, \dots, \alpha_N \in \mathbb{N}$, and let $\mathfrak{G}_1, \dots, \mathfrak{G}_N$ be independent random matrices such that $\mathfrak{G}_j \sim \mathcal{W}_d(\alpha_j, \Sigma_j)$ for some fixed $\Sigma_1, \dots, \Sigma_N \in \mathcal{S}_{++}^d$. Here, for integer α_j , the law $\mathcal{W}_d(\alpha_j, \Sigma_j)$

is understood through the standard Gaussian representation, possibly singular when $\alpha_j < d$. Define

$$\mathfrak{T} := \sum_{j=1}^N \mathfrak{G}_j,$$

such that

$$\mathbb{E}[\mathfrak{T}] = \sum_{j=1}^N \alpha_j \Sigma_j \equiv \bar{\Sigma}, \quad \text{Var}\{\text{vecp}(\mathfrak{T})\} = 2 \sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j),$$

where we use the notation

$$\mathcal{V}(W) := B_d^\top (W \otimes W) B_d,$$

and B_d denotes the $d^2 \times d(d+1)/2$ transition matrix such that $\text{vecp}(X) = B_d^\top \text{vec}(X)$, $X \in \mathcal{S}^d$; see, e.g., [Gupta and Nagar \(2000, Definition 2.5.1 and Eq. \(1.2.11\)\)](#). Introduce the Wishart random matrix $\mathfrak{W}_{\text{Sat}} \sim \mathcal{W}_d(\nu, \tilde{\Sigma})$ matching the first two moments of $\mathfrak{T}$:

$$\begin{aligned} \mathbb{E}[\mathfrak{W}_{\text{Sat}}] &= \nu \tilde{\Sigma} = \bar{\Sigma}, \\ \text{Var}\{\text{vecp}(\mathfrak{W}_{\text{Sat}})\} &= \frac{2}{\nu} \mathcal{V}(\bar{\Sigma}) = 2 \sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j). \end{aligned} \quad (4.4)$$

Matching the first moment yields $\tilde{\Sigma} = \nu^{-1} \bar{\Sigma}$, and it remains to find ν by matching the second moment. However, in general, Eq. (4.4) is an overdetermined matrix equation for the single scalar ν and need not admit an exact solution. An exact scalar solution exists in the proportional-scale case $\Sigma_j = \lambda_j \Sigma$ with $\lambda_j > 0$ for all $j \in [N]$; then ν is uniquely determined by

$$m_1 := \sum_{j=1}^N \lambda_j \alpha_j, \quad m_2 := \sum_{j=1}^N \lambda_j^2 \alpha_j, \quad \nu = \frac{m_1^2}{m_2}, \quad \tilde{\Sigma} = \frac{m_2}{m_1} \Sigma.$$

On the other hand, when the condition $\Sigma_j = \lambda_j \Sigma$ does not hold, alternative methods are required to circumvent solving (4.4) for ν . Using the Stein equation associated with the approximating law $\mathcal{W}_d(\nu, \tilde{\Sigma}_\nu)$, where $\tilde{\Sigma}_\nu = \nu^{-1} \bar{\Sigma}$ and ν is the proposed effective degrees of freedom, we can motivate a particular choice of ν by comparing the upper bounds on the discrepancy in (4.5). Some of these methods are presented and numerically compared in Section S.1 of the [Supplementary material](#).

Proposition 4.3 (Quantitative multivariate Satterthwaite approximation). *In the setting described above, assume that we have $\sum_{j=1}^N \alpha_j \geq d+1$. Also, fix $\nu > 3d-3$ and set $\mathfrak{W}_\nu \sim \mathcal{W}_d(\nu, \tilde{\Sigma}_\nu)$ with $\tilde{\Sigma}_\nu := \nu^{-1} \bar{\Sigma}$, so that, in particular, $\mathbb{E}[\mathfrak{W}_\nu] = \bar{\Sigma} = \mathbb{E}[\mathfrak{T}]$. Let $h \in C_b^{2d}(\mathcal{S}_{++}^d)$ and assume that $\mathcal{M}_1^{\mathcal{D}, \tilde{\Sigma}_\nu}(h) + \mathcal{M}_2^{\mathcal{D}, \tilde{\Sigma}_\nu}(h) < \infty$ in Theorem 3.6. Then,*

$$|\mathbb{E}[h(\mathfrak{T})] - \mathbb{E}[h(\mathfrak{W}_\nu)]| \leq \frac{d+1}{2} \mathcal{M}_2^{\mathcal{D}, \tilde{\Sigma}_\nu}(h) \sum_{j=1}^N \alpha_j \text{tr}(\Sigma_j) \|\Sigma_j - \tilde{\Sigma}_\nu\|_F. \quad (4.5)$$

Remark 4.1. While the proof of this result relies on Gaussian integration by parts for each $\mathfrak{G}_j$, the results from Section 3 are essential in the approximation step to derive the upper bound. Indeed, in general, the selected ν is not an integer.

Remark 4.2. Since $\tilde{\Sigma}_\nu = \nu^{-1}\bar{\Sigma} \rightarrow 0_{d \times d}$ as $\nu \rightarrow \infty$, the operator $\mathcal{D}_{t, \tilde{\Sigma}_\nu} = \nabla_W \text{adj}(I_d - 2\tilde{\Sigma}_{\nu, t} \nabla_W)$ converges formally to the ordinary symmetric gradient ∇_W . Consequently, although the bound in Proposition 4.3 is stated for test functions in $C_b^{2d}(\mathcal{S}_{++}^d)$ through the seminorm $\mathcal{M}_2^{\mathcal{D}, \tilde{\Sigma}_\nu}(h)$, in the regime $\nu \rightarrow \infty$ this seminorm becomes asymptotically equivalent to the usual second-order seminorm $\mathcal{M}_2(h)$. Thus the bound asymptotically behaves like a control in a smooth distance based on bounded second-order derivatives.

4.3. De Bruijn identities and logarithmic Sobolev inequalities for the Wishart measure

We begin this section by obtaining a local De Bruijn identity for the Wishart measure, which we state in Proposition 4.4. Our result provides a natural generalization of the local De Bruijn identity for the gamma measure of Theorem 1 of Arras and Swan (2017) that was proved under minimal conditions. Previously, in the context of the Gamma calculus of Bakry and Émery (1985) and Bakry et al. (2014), a local De Bruijn identity had been known to hold under a density assumption; see Proposition 5.2.2 of Bakry (1996).

Fix $\alpha \in (d-1, \infty)$ and $\Sigma \in \mathcal{S}_{++}^d$, and write $\xi = \mathcal{W}_d(\alpha, \Sigma)$ for the stationary law from Proposition 3.2. Let μ be a probability measure on $\mathcal{S}_{++}^d$, which we suppose is absolutely continuous with respect to the Wishart measure ξ with Radon–Nikodym density $g \in C_b^2(\mathcal{S}_{++}^d)$. That is,

$$\mu(dX) = g(X) \xi(dX).$$

For $t \geq 0$, define the evolved law and its density with respect to ξ by

$$\mu_t := \mu \mathcal{P}_t^{\mathcal{W}}, \quad g_t := \frac{d\mu_t}{d\xi}.$$

Here the measure μ_t is the Wishart generalisation of the law of the random variable X_τ given in Definition 1 of Arras and Swan (2017). It should, however, be noted that Theorem 1 of Arras and Swan (2017) is stated without a density assumption. We also define the relative entropy and the Wishart Fisher information by

$$\begin{aligned} \text{Ent}(\mu_t \parallel \xi) &:= \int_{\mathcal{S}_{++}^d} g_t(X) \log(g_t(X)) \xi(dX), \\ J_{\mathcal{W}, \Sigma}(\mu_t \mid \xi) &:= \int_{\mathcal{S}_{++}^d} \Gamma_{\mathcal{W}, \Sigma}(\log(g_t))(X) \mu_t(dX), \end{aligned}$$

(see, e.g., Bakry et al., 2014, Eq. (5.1.5) and (5.1.6)) where the Wishart *carré du champ* operator $\Gamma_{\mathcal{W}, \Sigma}$ is formally defined, for every $X \in \mathcal{S}_{++}^d$, by

$$\begin{aligned} \Gamma_{\mathcal{W}, \Sigma}(\varphi)(X) &:= \Gamma_{\mathcal{W}, \Sigma}(\varphi, \varphi)(X), \\ \Gamma_{\mathcal{W}, \Sigma}(\varphi, \psi)(X) &:= \frac{1}{2} \{ \mathcal{A}^{\mathcal{W}}(\varphi\psi)(X) - \varphi(X) \mathcal{A}^{\mathcal{W}}\psi(X) - \psi(X) \mathcal{A}^{\mathcal{W}}\varphi(X) \}. \end{aligned}$$

Let $\varphi, \psi \in C_b^2(\mathcal{S}_{++}^d)$. We now use the extended generator identity (3.2) to obtain a useful representation of the Wishart *carré du champ* operator that we will employ in our proof. From the extended generator identity (3.2) and the Leibniz rule, the drift term for $f = \varphi\psi$ is

$$2 \text{tr}\{(\alpha\Sigma - X) \nabla(\varphi\psi)(X)\} = 2 \varphi(X) \text{tr}\{(\alpha\Sigma - X) \nabla\psi(X)\} + 2 \psi(X) \text{tr}\{(\alpha\Sigma - X) \nabla\varphi(X)\},$$

and the diffusion term is

$$\begin{aligned} 4 \operatorname{tr}\{X \nabla \Sigma \nabla(\varphi \psi)(X)\} &= 4 \varphi(X) \operatorname{tr}\{X \nabla \Sigma \nabla \psi(X)\} + 4 \psi(X) \operatorname{tr}\{X \nabla \Sigma \nabla \varphi(X)\} \\ &\quad + 8 \operatorname{tr}\{X \nabla \varphi(X) \Sigma \nabla \psi(X)\}. \end{aligned}$$

Here the two cross terms in the second-order Leibniz rule are equal by the symmetry of $\nabla \varphi(X)$ and $\nabla \psi(X)$. Consequently,

$$\mathcal{A}^{\mathcal{W}}(\varphi \psi)(X) = \varphi(X) \mathcal{A}^{\mathcal{W}} \psi(X) + \psi(X) \mathcal{A}^{\mathcal{W}} \varphi(X) + 8 \operatorname{tr}\{X \nabla \varphi(X) \Sigma \nabla \psi(X)\}, \quad (4.6)$$

and thus,

$$\Gamma_{\mathcal{W}, \Sigma}(\varphi, \psi)(X) = 4 \operatorname{tr}\{X \nabla \varphi(X) \Sigma \nabla \psi(X)\}.$$

In particular, for all $X \in \mathcal{S}_{++}^d$, $\Gamma_{\mathcal{W}, \Sigma}(\varphi)(X) = 4 \|X^{1/2} \nabla \varphi(X) \Sigma^{1/2}\|_F^2 \geq 0$, and

$$J_{\mathcal{W}, \Sigma}(\mu_t \mid \xi) = \int_{\mathcal{S}_{++}^d} 4 \operatorname{tr}\{X \nabla \log(g_t)(X) \Sigma \nabla \log(g_t)(X)\} \mu_t(dX) \geq 0.$$

Proposition 4.4 (Local De Bruijn identity for the Wishart measure). *Assume that there exist constants $m_g, M_g \in (0, \infty)$ such that $m_g \leq g(X) \leq M_g$ for all $X \in \mathcal{S}_{++}^d$. Then, for every $t > 0$, we have $g_t = \mathcal{P}_t^{\mathcal{W}} g$ and*

$$\frac{d}{dt} \operatorname{Ent}(\mu_t \mid \xi) = -J_{\mathcal{W}, \Sigma}(\mu_t \mid \xi). \quad (4.7)$$

In particular, the map $t \mapsto \operatorname{Ent}(\mu_t \mid \xi)$ is nonincreasing. Equivalently, after the change of variable $\tau = e^{-2t}$ and the definition $\mu_\tau := \mu \mathcal{P}_{-(1/2)\log(\tau)}^{\mathcal{W}}$, we have

$$\frac{d}{d\tau} \operatorname{Ent}(\mu_\tau \mid \xi) = \frac{1}{2\tau} J_{\mathcal{W}, \Sigma}(\mu_\tau \mid \xi), \quad \tau \in (0, 1). \quad (4.8)$$

Remark 4.3 (On the assumptions in Proposition 4.4). The boundedness assumption in Proposition 4.4 is a simple sufficient condition for the two analytic justifications needed in the proof. More precisely, the proof only requires the following two assumptions.

(A) For every $t > 0$, all integrals below are finite and differentiation under the integral sign in the entropy functional is valid, namely

$$\frac{d}{dt} \operatorname{Ent}(\mu_t \mid \xi) = \int_{\mathcal{S}_{++}^d} (1 + \log(g_t(X))) \mathcal{A}^{\mathcal{W}} g_t(X) \xi(dX).$$

(B) For every $t > 0$, the functions g_t and $\log(g_t)$ satisfy the two integration-by-parts identities:

$$\int_{\mathcal{S}_{++}^d} \mathcal{A}^{\mathcal{W}} g_t(X) \xi(dX) = 0$$

and

$$\int_{\mathcal{S}_{++}^d} \log(g_t(X)) \mathcal{A}^{\mathcal{W}} g_t(X) \xi(dX) = -4 \int_{\mathcal{S}_{++}^d} \operatorname{tr}\{X \nabla g_t(X) \Sigma \nabla \log(g_t)(X)\} \xi(dX).$$

A simple sufficient condition for Assumptions (A) and (B) is precisely the boundedness condition used in Proposition 4.4, namely that there exist constants $m_g, M_g \in (0, \infty)$ such that $m_g \leq g(X) \leq M_g$ for all $X \in \mathcal{S}_{++}^d$. Indeed, since $\mathcal{P}_t^{\mathcal{W}}$ is Markov, this condition gives $m_g \leq g_t(X) \leq M_g$ for all $X \in \mathcal{S}_{++}^d$ and $t \geq 0$, so $\log(g_t)$ is bounded. Moreover, the regularity of the Wishart semigroup on $C_b^2(\mathcal{S}_{++}^d)$ gives bounded first and second derivatives of g_t on compact time intervals contained in $(0, \infty)$. Since the coefficients of $\mathcal{A}^{\mathcal{W}}$ grow at most linearly and ξ has finite first moment, dominated convergence justifies Assumption (A), while Assumption (B) follows from the invariant Wishart integration-by-parts formula. Thus Proposition 4.4 remains valid under Assumptions (A) and (B) alone. The same analytic conditions are also the ones needed for the forthcoming Proposition 4.5, since the proof of that result integrates (4.7) from Proposition 4.4.

The next proposition considers the integrated counterpart of the local identity. The first part of the proposition generalizes the integrated De Bruijn identity for the gamma measure given by Theorem 18 of Arras and Swan (2017). The second part, which turns this relative entropy formula into an entropy jump statement when the initial law matches the stationary Wishart law in mean and log-determinant, generalizes the differential entropy gap identity for the gamma case that was treated in Remark 19 of Arras and Swan (2017), which also required a matching mean and logarithmic moment condition.

Proposition 4.5 (Integrated De Bruijn identity for the Wishart measure and entropy jump). *Under the assumptions of Proposition 4.4, we have*

$$\text{Ent}(\mu \parallel \xi) = \int_0^\infty J_{\mathcal{W}, \Sigma}(\mu_t \mid \xi) dt = \int_0^1 \frac{1}{2\tau} J_{\mathcal{W}, \Sigma}(\mu_\tau \mid \xi) d\tau. \quad (4.9)$$

Assume furthermore that μ is the law of a random matrix $\mathfrak{X}$ and that this law has Lebesgue density $f_{\mathfrak{X}}$, that is, $\mu(dX) = f_{\mathfrak{X}}(X) dX$. Equivalently, if f_ξ denotes the Lebesgue density of ξ , then the initial relative density is $g(X) = f_{\mathfrak{X}}(X)/f_\xi(X)$. Assume also that $\mathfrak{X}$ has finite differential entropy

$$H(\mathfrak{X}) := - \int_{\mathcal{S}_{++}^d} f_{\mathfrak{X}}(X) \log(f_{\mathfrak{X}}(X)) dX.$$

If $E[\mathfrak{X}] = E[\mathfrak{W}_\infty] = \alpha \Sigma$ and $E[\log(|\mathfrak{X}|)] = E[\log(|\mathfrak{W}_\infty|)] \in \mathbb{R}$ with $\mathfrak{W}_\infty \sim \xi$, then

$$H(\mathfrak{W}_\infty) - H(\mathfrak{X}) = \int_0^\infty J_{\mathcal{W}, \Sigma}(\mu_t \mid \xi) dt = \int_0^1 \frac{1}{2\tau} J_{\mathcal{W}, \Sigma}(\mu_\tau \mid \xi) d\tau. \quad (4.10)$$

We now derive a logarithmic Sobolev inequality for the Wishart measure with $\alpha \geq d$, thereby generalizing the logarithmic Sobolev inequality for the gamma case; see Bakry (1996) and Proposition 2 of Arras and Swan (2017) for an alternative proof. The gamma logarithmic Sobolev inequality is stated for $\alpha \geq 1$ (if one uses the same parametrization for the gamma distribution as we do for the Wishart distribution), and so our condition $\alpha \geq d$ reduces exactly to this condition in the univariate gamma setting.

For a probability measure $\mu \ll \xi$ with density $g = d\mu/d\xi$, we define the Wishart Fisher information in the weak sense by

$$J_{\mathcal{W}, \Sigma}(\mu \mid \xi) := 16 \int_{\mathcal{S}_{++}^d} \text{tr}\{X \nabla \sqrt{g}(X) \Sigma \nabla \sqrt{g}(X)\} \xi(dX),$$

whenever $\sqrt{g}$ belongs to the corresponding weighted Sobolev domain, and we set $J_{\mathcal{W},\Sigma}(\mu \mid \xi) = +\infty$ otherwise. For smooth positive g , this is equivalent to

$$J_{\mathcal{W},\Sigma}(\mu \mid \xi) = \int_{\mathcal{S}_{++}^d} 4 \operatorname{tr}\{X \nabla \log(g)(X) \Sigma \nabla \log(g)(X)\} \mu(dX).$$

Proposition 4.6 (Logarithmic Sobolev inequality for the Wishart measure). *Let $\alpha \in [d, \infty)$, $\Sigma \in \mathcal{S}_{++}^d$ and $\xi = \mathcal{W}_d(\alpha, \Sigma)$. Let $\mu \ll \xi$ with density $g = d\mu/d\xi$. If $J_{\mathcal{W},\Sigma}(\mu \mid \xi) < \infty$, then*

$$\operatorname{Ent}(\mu \parallel \xi) \leq \frac{1}{2} J_{\mathcal{W},\Sigma}(\mu \mid \xi). \quad (4.11)$$

In the gamma case, Theorem 3 of [Arras and Swan \(2017\)](#) provides an HSI inequality (that connects entropy H , Stein discrepancy S and Fisher information I), which improves on the classical logarithmic Sobolev inequality (their Proposition 2), and provides a gamma analogue of the Gaussian HSI inequality of [Ledoux et al. \(2015\)](#) which improved upon the classical Gaussian logarithmic Sobolev inequality of [Gross \(1975\)](#). Given that we have established natural Wishart generalizations of the local De Bruijn identity and logarithmic Sobolev inequality for the gamma case, it is natural to ask for a Wishart HSI inequality, which we leave as an open problem.

Open problem. Generalize the HSI inequality of [Arras and Swan \(2017, Theorem 3\)](#) to our Wishart setting.

4.4. Estimation of the parameters of a Wishart distribution

Let $\alpha \in (d-1, \infty)$ and $\Sigma \in \mathcal{S}_{++}^d$. Given a random sample $\mathfrak{W}^{(1)}, \dots, \mathfrak{W}^{(n)} \stackrel{\text{iid}}{\sim} \mathcal{W}_d(\alpha, \Sigma)$ and any function $g : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$, define the averaging operator

$$\overline{g(\mathfrak{W})} = \frac{1}{n} \sum_{k=1}^n g(\mathfrak{W}^{(k)}).$$

We apply a method of moments (MOM) to the Stein characterization found in Corollary 3.3 by replacing the expectation with the empirical average.

Suppose first that α is known. This suggests estimating Σ by imposing, for a chosen collection of test functions $f \in C_{\mathcal{A}\mathcal{W}}^2(\mathcal{S}_{++}^d)$, the empirical Stein equations

$$\overline{\mathcal{A}\mathcal{W}f(\mathfrak{W})} = \overline{2 \operatorname{tr}\{(\alpha \widehat{\Sigma} - \mathfrak{W}) \nabla f(\mathfrak{W})\}} + \overline{4 \operatorname{tr}\{\mathfrak{W} \nabla \widehat{\Sigma} \nabla f(\mathfrak{W})\}} = 0. \quad (4.12)$$

For symmetric $U \in \mathcal{S}^d$, define the linear probe

$$f_U(S) = \operatorname{tr}(US), \quad S \in \mathcal{S}_{++}^d. \quad (4.13)$$

Proposition 4.7 (Known shape parameter). *The Stein's method-of-moments estimator of Σ based on the linear probes in (4.13) is*

$$\widehat{\Sigma} = \frac{1}{\alpha n} \sum_{k=1}^n \mathfrak{W}^{(k)}.$$

Remark 4.4. The estimator $\widehat{\Sigma}$ coincides with the classical method-of-moments estimator obtained from the first moment identity $\mathbb{E}[\mathfrak{W}^{(1)}] = \alpha\Sigma$. Nonlinear probes would produce other Stein's method-of-moments estimators, but they are not expected to improve on $\widehat{\Sigma}$ when α is known.

Suppose now that α is unknown. Then linear probes only identify the product $M := \alpha\Sigma$. Indeed, rewriting (3.2) in terms of (M, α) gives

$$\mathcal{A}_{M,\alpha}^{\mathcal{W}} f(S) = 2 \operatorname{tr}\{(M - S)\nabla f(S)\} + \frac{4}{\alpha} \operatorname{tr}\{S\nabla M\nabla f(S)\}, \quad S \in \mathcal{S}_{++}^d.$$

Accordingly, one seeks a pair $(\widehat{M}, \widehat{\alpha})$ satisfying, for suitable test functions f ,

$$\overline{\mathcal{A}_{\widehat{M},\widehat{\alpha}}^{\mathcal{W}} f(\mathfrak{W})} = 2 \operatorname{tr}\{(\widehat{M} - \mathfrak{W})\nabla f(\mathfrak{W})\} + \frac{4}{\widehat{\alpha}} \operatorname{tr}\{\mathfrak{W}\nabla \widehat{M}\nabla f(\mathfrak{W})\} = 0. \quad (4.14)$$

Following Example 2.4 of Ebner et al. (2025), which treats both the classical moment choice and a logarithmic Stein's method-of-moments choice for the gamma distribution, we consider the following matrix analogs. Since the Wishart Stein operator is written in extended-generator form, these scalar gamma test functions are implemented through scalar potentials whose gradients are proportional to S and equal to $\log(S)$, respectively:

$$q(S) = \operatorname{tr}(S^2), \quad \ell(S) = \operatorname{tr}(S \log(S) - S), \quad S \in \mathcal{S}_{++}^d, \quad (4.15)$$

where $\log(S)$ denotes the symmetric matrix logarithm. For $x, y > 0$, let

$$\log^{[1]}(x, y) := \begin{cases} \frac{\log(x) - \log(y)}{x - y}, & x \neq y, \\ 1/x, & x = y. \end{cases}$$

If $S = H\Lambda H^\top$, with $H \in O(d)$ and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_d)$, let $M^{(S)} := H^\top M H$ denote the representation of M in the eigenbasis of S , and define

$$\mathcal{J}_S(M) := \frac{1}{2} \operatorname{tr}(M) + \frac{1}{2} \sum_{i,j=1}^d \lambda_i M_{jj}^{(S)} \log^{[1]}(\lambda_j, \lambda_i).$$

Proposition 4.8 (Unknown shape parameter). *Let $n \geq 2$ and $\widehat{M} = \frac{1}{n} \sum_{k=1}^n \mathfrak{W}^{(k)}$. The Stein's method-of-moments estimators based on the linear probes in (4.13) together with q and ℓ in (4.15), respectively, are given by*

$$\widehat{\alpha}_{\text{quad}} = \frac{\{\operatorname{tr}(\widehat{M})\}^2 + \operatorname{tr}(\widehat{M}^2)}{\operatorname{tr}(\mathfrak{W}^2) - \operatorname{tr}(\widehat{M}^2)}, \quad \widehat{\Sigma}_{\text{quad}} = \frac{\widehat{M}}{\widehat{\alpha}_{\text{quad}}}, \quad (4.16)$$

$$\widehat{\alpha}_{\log} = \frac{2 \overline{\mathcal{J}_{\mathfrak{W}}(\widehat{M})}}{\operatorname{tr}\{\mathfrak{W} \log(\mathfrak{W})\} - \operatorname{tr}\{\widehat{M} \log(\mathfrak{W})\}}, \quad \widehat{\Sigma}_{\log} = \frac{\widehat{M}}{\widehat{\alpha}_{\log}}, \quad (4.17)$$

provided that the denominators of $\widehat{\alpha}_{\text{quad}}$ and $\widehat{\alpha}_{\log}$ are nonzero.

Remark 4.5. The quadratic estimator $(\hat{\alpha}_{\text{quad}}, \hat{\Sigma}_{\text{quad}})$ coincides with the classical method-of-moments estimator obtained by matching the first matrix moment $\mathbb{E}[\mathfrak{W}^{(1)}] = \alpha \Sigma$ and the scalar second moment

$$\mathbb{E}[\text{tr}\{(\mathfrak{W}^{(1)})^2\}] = \alpha(\alpha + 1)\text{tr}(\Sigma^2) + \alpha\text{tr}(\Sigma)^2 = \text{tr}(M^2) + \frac{1}{\alpha} [\{\text{tr}(M)\}^2 + \text{tr}(M^2)].$$

The logarithmic estimator $(\hat{\alpha}_{\log}, \hat{\Sigma}_{\log})$ is included as a matrix analog of the logarithmic Stein’s method-of-moments choice considered for the gamma distribution in [Ebner et al. \(2025\)](#).

Remark 4.6. The estimators $(\hat{\alpha}_{\text{quad}}, \hat{\Sigma}_{\text{quad}})$ and $(\hat{\alpha}_{\log}, \hat{\Sigma}_{\log})$ are unconstrained moment estimators. When their denominators are positive, the corresponding estimates of α are positive and the estimates of Σ belong to $\mathcal{S}_{++}^d$, but $\hat{\alpha}_{\text{quad}}$ and $\hat{\alpha}_{\log}$ need not be larger than $d - 1$ for every finite sample. If an estimator taking values in the Wishart parameter space is required, one may impose the constraint $\alpha > d - 1$ by truncation or by constrained moment estimation.

Remark 4.7. A numerical comparison of $(\hat{\alpha}_{\log}, \hat{\Sigma}_{\log})$, $(\hat{\alpha}_{\text{quad}}, \hat{\Sigma}_{\text{quad}})$ and $(\hat{\alpha}_{\text{MLE}}, \hat{\Sigma}_{\text{MLE}})$ is summarized in Table 4.1. A detailed description of the experiment, together with the full median and interquartile range values for the relative Frobenius error $\|\hat{\Sigma} - \Sigma_0\|_F / \|\Sigma_0\|_F$ of each estimator with respect to a target Σ_0 , is relegated to Section S.2.1 of the [Supplementary material](#). The maximum likelihood estimator (MLE) attains the smallest median error most often, the logarithmic Stein’s method-of-moments estimator is the second strongest competitor, and the quadratic, classical moment estimator is least often the best. The same qualitative ranking is observed for the interquartile ranges. Thus, while the MLE is the strongest competitor overall, the logarithmic Stein estimator often comes reasonably close and improves on the quadratic moment estimator. It also has the advantage of the closed-form expression in (4.17), which makes it simpler to compute and more directly amenable to theoretical analysis than the MLE.

Table 4.1: Counts, over the 100 parameter configurations, of how often each estimator attains the smallest median and the smallest interquartile range in Table S.2, respectively.

Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR
log-MOM	15	10	14	19	13	18	16	11
quad-MOM	0	0	1	1	0	0	0	1
MLE	85	90	85	80	87	82	84	88

Returning to the case where α is known, by choosing M linear probes

$$f_{U_1}(S) = \text{tr}(U_1 S), \dots, f_{U_M}(S) = \text{tr}(U_M S), \quad S \in \mathcal{S}_{++}^d,$$

and feeding them into (4.12), we obtain a family of Stein-type moment equations that produce multiple unbiased estimators of linear functionals of Σ . When we restrict Σ to a linear structured subspace, these equations can be projected by least squares onto that subspace, as detailed next.

Proposition 4.9 (Projection onto a structured subspace). *Suppose that the true scale matrix $\Sigma \in \mathcal{S}_{++}^d$ admits the linear representation*

$$\Sigma = \Sigma(\beta^*) = \sum_{j=1}^p \beta_j^* B_j$$

for some $\beta^* \in \mathbb{R}^p$ and fixed symmetric templates $\{B_j\}_{j=1}^p \subseteq \mathcal{S}^d$. Let $\{U_m\}_{m=1}^M \subseteq \mathcal{S}^d$ be given, and for the corresponding linear probes $f_{U_m}(S) := \text{tr}(U_m S)$, form the structured Stein moment equations

$$\text{tr}\{\Sigma(\beta)U_m\} = y_{n,m}, \quad y_{n,m} := \frac{1}{\alpha n} \sum_{k=1}^n \text{tr}(\mathfrak{W}^{(k)}U_m), \quad m \in \{1, \dots, M\}. \quad (4.18)$$

Let $C \in \mathbb{R}^{M \times p}$ be the design matrix with entries $C_{mj} := \text{tr}(B_j U_m)$, and let $\mathbf{y}_n := (y_{n,1}, \dots, y_{n,M})^\top$. If C has full column rank, then the least-squares solution

$$\hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^p} \|C\beta - \mathbf{y}_n\|_2^2 = (C^\top C)^{-1} C^\top \mathbf{y}_n$$

is strongly consistent for the true coefficient vector β^* , as $n \rightarrow \infty$. In particular, the structured estimator $\hat{\Sigma}_n := \Sigma(\hat{\beta}_n) = \sum_{j=1}^p \hat{\beta}_{n,j} B_j$ converges almost surely to Σ and solves the projected Stein moment equations (4.18) in the least-squares sense.

Remark 4.8. The estimator $\hat{\Sigma}_n$ in Proposition 4.9 is the unconstrained least-squares projection onto the linear span of the templates $\{B_j\}_{j=1}^p$. It need not belong to $\mathcal{S}_{++}^d$ for every finite sample unless positive definiteness is imposed separately. Since $\mathcal{S}_{++}^d$ is open and $\hat{\Sigma}_n \rightarrow \Sigma \in \mathcal{S}_{++}^d$ almost surely, it is nevertheless positive definite eventually almost surely.

Remark 4.9. A numerical comparison of the projected Stein's method-of-moments estimator $\hat{\Sigma}_n$ and the naive (unstructured) Stein's method-of-moments estimator $\tilde{\Sigma}_n$ is summarized in Table 4.2. A detailed description of the experiment, together with the full median and interquartile range values for the relative Frobenius error of $\hat{\Sigma}$ and the corresponding error ratios, is relegated to Section S.2.2 of the [Supplementary material](#). For the randomly generated compound-symmetry configurations considered there, the projected Stein estimator yields the smaller median relative Frobenius error for every structured configuration and every sample size. The naive estimator more often yields the smaller interquartile range, but the median-error comparison clearly supports the gain from using the structural information in the Stein moment equations. The naive estimator is the natural unstructured benchmark. The unstructured MLE is not displayed separately, since α_0 is known and therefore coincides with $\alpha_0^{-1} \hat{M} = \tilde{\Sigma}_n$.

Table 4.2: Winner counts for the structured Wishart simulation. For each sample size, the table counts among the 100 structured configurations how often each estimator attains the smallest median error and the smallest interquartile range.

Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR
proj-SMOM	100	9	100	9	100	10	100	9
naive-SMOM	0	91	0	91	0	90	0	91

5. Preliminary lemmas

This section shows how to differentiate the scale-normalized version of the Wishart transition semigroup ($\mathcal{Q}_t$ in (3.15)) with respect to the entries of the matrix $\Lambda = (\Lambda_{ij})$ appearing in the

noncentrality parameter of the noncentral Wishart kernel $f_{\alpha, I_d, e^{-2t}\Lambda}^{\mathcal{W}}$. The derivatives of $(\mathcal{Q}_t h)(\Lambda)$ with respect to Λ_{ij} can be “pushed through” the transition kernel and transferred onto h using an appropriate differential operator; see Lemma 5.3. This lemma is key in proving the existence of the semigroup solution of the Wishart Stein equation in Theorem 3.5 and its regularity in Theorem 3.6. Before we prove Lemma 5.3, two preliminary lemmas are needed; see Lemmas 5.1 and 5.2 below.

Lemma 5.1 provides an expression for Laplace transforms of products of partial derivatives of the noncentral Wishart density (recall that the Laplace transform $\mathcal{L}[g]$ of a sufficiently integrable function $g : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$ is given by (1.12)). Lemma 5.2 shows how partial derivatives of the noncentral Wishart density with respect to the noncentrality parameter transfer to partial derivatives with respect to the state variable.

Lemma 5.1. *Let $m \in \mathbb{N}$, $\alpha > d - 1 + 2m$, and $\Theta \in \mathcal{S}_+^d$. Then, for all $T \in \mathcal{S}_+^d$ and all index pairs $(i_\ell, j_\ell) \in [d]^2$,*

$$\mathcal{L} \left[\left(\prod_{\ell=1}^m \nabla_{i_\ell j_\ell} \right) f_{\alpha, I_d, \Theta}^{\mathcal{W}} \right] (T) = \left(\prod_{\ell=1}^m T_{i_\ell j_\ell} \right) \mathcal{L}[f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T). \quad (5.1)$$

More generally, if $P(U)$ is a scalar polynomial of degree at most m in the entries of a symmetric $d \times d$ matrix U , and $P(\nabla)$ denotes the differential operator obtained by replacing U_{ij} by ∇_{ij} , then

$$\mathcal{L}[P(\nabla) f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T) = P(T) \mathcal{L}[f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T). \quad (5.2)$$

Proof of Lemma 5.1. Let $T \in \mathcal{S}_+^d$ be given. Throughout the proof, $\nabla \equiv \nabla_X$. For any $i, j \in [d]$, note that

$$\nabla_{ij} \operatorname{etr}(-TX) = -T_{ij} \operatorname{etr}(-TX), \quad T, X \in \mathcal{S}^d.$$

By repeated integration by parts for which all boundary terms vanish, we have

$$\int_{\mathcal{S}_{++}^d} \operatorname{etr}(-TX) \left(\prod_{\ell=1}^m \nabla_{i_\ell j_\ell} \right) f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) \, dX = \left(\prod_{\ell=1}^m T_{i_\ell j_\ell} \right) \int_{\mathcal{S}_{++}^d} \operatorname{etr}(-TX) f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) \, dX, \quad (5.3)$$

which proves (5.1). Since $\alpha > d - 1 + 2m \geq d - 1 + 2r$ for any $r \leq m$, this identity holds for any product of r derivatives. The claim (5.2) follows by linearity and the fact that any polynomial of degree at most m is a finite linear combination of monomials of degree at most m .

It remains to justify (5.3) rigorously. First, the heuristic is the following. Set

$$a := \frac{\alpha}{2}, \quad \beta := \frac{\alpha - d - 1}{2}, \quad B_\Theta(X) := {}_0F_1 \left(a; \frac{\Theta^{1/2} X \Theta^{1/2}}{4} \right).$$

The density $f_{\alpha, I_d, \Theta}^{\mathcal{W}}$ can be written as

$$f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) = c_{\alpha, \Theta} |X|^\beta \operatorname{etr}(-X/2) B_\Theta(X), \quad X \in \mathcal{S}_{++}^d, \quad (5.4)$$

for an appropriate constant $c_{\alpha, \Theta} > 0$. Each differentiation of $f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X)$ can lower the power of $|X|$ by at most one, so since $\beta > m - 1$, derivatives of order at most $m - 1$ should vanish on $\partial \mathcal{S}_+^d$, while derivatives of order m should still be locally integrable there. The only additional

issue is the noncentral factor $B_\Theta(X)$: one must show that its derivatives do not grow too quickly at infinity. We prove below that every differential monomial $\bar{\nabla} = \prod_{\ell=1}^r \nabla_{i_\ell j_\ell}$ of order at most m (i.e., $r \leq m$) in the entries of ∇ satisfies, for appropriate constants $c_\Theta, C_{\bar{\nabla}, \Theta} > 0$,

$$|\bar{\nabla} B_\Theta(X)| \leq C_{\bar{\nabla}, \Theta} \exp\{c_\Theta \sqrt{1 + \text{tr}(X)}\}, \quad X \in \mathcal{S}_+^d, \quad (5.5)$$

and this stretched-exponential growth is dominated by the factor $\text{etr}(-X/2)$. It is enough to control $B_\Theta(Z)$ for complex symmetric matrices Z .

For a partition $\kappa = (\kappa_1, \dots, \kappa_d) \vdash k$ (meaning $\kappa_1 \geq \dots \geq \kappa_d \geq 0$ and $\kappa_1 + \dots + \kappa_d = k$), set $\kappa_{d+1} := 0$. By James' integral formula for zonal polynomials (see, e.g., [Richards, 2010](#), Eq. 35.4.3), we have for real symmetric Y ,

$$C_\kappa(Y) = C_\kappa(I_d) \int_{O(d)} \prod_{j=1}^d |(HYH^\top)_j|^{\kappa_j - \kappa_{j+1}} dH,$$

where dH is the normalized Haar probability measure on $O(d)$, and A_j denotes the $j \times j$ top-left corner of A . Both sides are polynomials in the entries of Y , so the identity extends to complex symmetric Z by polynomial identity. Therefore, for any complex symmetric Z ,

$$|C_\kappa(Z)| \leq C_\kappa(I_d) \prod_{j=1}^d \|Z\|_2^{j(\kappa_j - \kappa_{j+1})} = C_\kappa(I_d) \|Z\|_2^k.$$

Next, since $\alpha > d - 1 + 2m$ and $m \geq 1$, we have $a > (d + 1)/2$, hence

$$(a)_\kappa = \prod_{j=1}^d \left(a - \frac{j-1}{2}\right)_{\kappa_j} \geq \prod_{j=1}^d \kappa_j! = \frac{k!}{\binom{k}{\kappa_1, \dots, \kappa_d}} \geq \frac{k!}{\sum_{x_1 + \dots + x_d = k} \binom{k}{x_1, \dots, x_d}} = \frac{k!}{d^k}.$$

Combining the last two inequalities yields

$$\left| \frac{1}{(a)_\kappa k!} C_\kappa \left(\frac{\Theta^{1/2} Z \Theta^{1/2}}{4} \right) \right| \leq \frac{d^k C_\kappa(I_d)}{(k!)^2} \left(\frac{\|\Theta\|_2 \|Z\|_2}{4} \right)^k = \frac{d^k C_\kappa(I_d)}{(k!)^2} \left(\frac{\sqrt{\|\Theta\|_2 \|Z\|_2}}{2} \right)^{2k}.$$

Therefore, summing over $k \in \mathbb{N}_0$ and $\kappa \vdash k$, and using $\sum_{\kappa \vdash k} C_\kappa(I_d) = \{\text{tr}(I_d)\}^k = d^k$, we get

$$|B_\Theta(Z)| \leq \sum_{k=0}^{\infty} \frac{1}{(k!)^2} \left(\frac{d \sqrt{\|\Theta\|_2 \|Z\|_2}}{2} \right)^{2k} \leq \exp \left\{ d \sqrt{\|\Theta\|_2 \|Z\|_2} \right\} \leq \exp \left\{ c_\Theta \sqrt{\|Z\|_F} \right\}, \quad (5.6)$$

for some constant $c_\Theta \in (0, \infty)$ that depends only on d and Θ . In particular, the Bessel function B_Θ is entire on $\mathcal{S}^d \oplus i\mathcal{S}^d$ (the complexification of $\mathcal{S}^d$); cf. [Herz \(1955, p. 486\)](#).

Let $N := d(d + 1)/2$, choose the standard orthonormal basis of $\mathcal{S}^d$ for the Frobenius inner product (namely, $\{\mathbf{e}_i \mathbf{e}_i^\top\}_{i=1}^d \cup \{(\mathbf{e}_i \mathbf{e}_j^\top + \mathbf{e}_j \mathbf{e}_i^\top)/\sqrt{2}\}_{1 \leq i < j \leq d}$, where $\mathbf{e}_i$ is the i -th standard basis vector of $\mathbb{R}^d$), and identify $\mathcal{S}^d \oplus i\mathcal{S}^d$ with $\mathbb{C}^N$. Under this identification, B_Θ becomes an entire function on $\mathbb{C}^N$. Fix any multi-index $\boldsymbol{\eta} \in \mathbb{N}_0^N$ with $|\boldsymbol{\eta}| \leq m$. By Cauchy's integral formula on the unit polydisc (see, e.g., [Krantz, 2001](#), Theorem 1.2.2),

$$\frac{\partial^{|\boldsymbol{\eta}|}}{\partial w_1^{\eta_1} \dots \partial w_N^{\eta_N}} B_\Theta(\mathbf{w}) = \frac{\eta_1! \dots \eta_N!}{(2\pi i)^N} \int_{|z_1 - w_1|=1} \dots \int_{|z_N - w_N|=1} \frac{B_\Theta(\mathbf{z})}{\prod_{\nu=1}^N (z_\nu - w_\nu)^{\eta_\nu + 1}} dz_1 \dots dz_N,$$

hence

$$\left| \frac{\partial^{|\eta|}}{\partial w_1^{\eta_1} \dots \partial w_N^{\eta_N}} B_\Theta(\mathbf{w}) \right| \leq \eta_1! \dots \eta_N! \sup_{|z_1 - w_1| \leq 1, \dots, |z_N - w_N| \leq 1} |B_\Theta(\mathbf{z})|.$$

On this polydisc, $\|\mathbf{z}\|_2 \leq \|\mathbf{w}\|_2 + \sqrt{N}$, so the previous bound and the subadditivity of the square root yield

$$\left| \frac{\partial^{|\eta|}}{\partial w_1^{\eta_1} \dots \partial w_N^{\eta_N}} B_\Theta(\mathbf{w}) \right| \leq C_{\eta, \Theta} \exp\{c_\Theta \sqrt{1 + \|\mathbf{w}\|_2}\},$$

with $C_{\eta, \Theta} := \eta_1! \dots \eta_N! \exp\{c_\Theta N^{1/4}\}$. Since each entry of the symmetric gradient ∇ is a constant multiple of a coordinate derivative in these variables, it follows, after changing the constant $C_{\eta, \Theta}$, that every differential monomial $\bar{\nabla}$ of order at most m in the entries of ∇ satisfies

$$|\bar{\nabla} B_\Theta(X)| \leq C_{\bar{\nabla}, \Theta} \exp\{c_\Theta \sqrt{1 + \|X\|_F}\} \leq C_{\bar{\nabla}, \Theta} \exp\{c_\Theta \sqrt{1 + \text{tr}(X)}\}, \quad X \in \mathcal{S}_+^d, \quad (5.7)$$

for some other constant $C_{\bar{\nabla}, \Theta} \in (0, \infty)$, as claimed in (5.5).

Next, we record the cone-integrability estimate that will be used repeatedly: for every $\gamma > -1$, $M \in [0, \infty)$ and $c \in (0, \infty)$,

$$\int_{\mathcal{S}_{++}^d} (1 + \text{tr}(X))^M |X|^\gamma \text{etr}(-cX) dX < \infty. \quad (5.8)$$

This is an immediate consequence of the convergence of the multivariate gamma integral (2.1).

Now let $\bar{\nabla}$ be any differential monomial of order $r \leq m$ in the entries of ∇ . Using Leibniz' rule together with $\nabla |X|^q = q \text{adj}(X) |X|^{q-1}$, an induction on r shows that

$$\bar{\nabla} |X|^\beta = \sum_{s=0}^r p_{\bar{\nabla}, s}(X) |X|^{\beta-s},$$

where $p_{\bar{\nabla}, s}$ are scalar-valued polynomials. Therefore, every derivative $\bar{\nabla} f_{\alpha, I_d, \Theta}^\mathcal{W}(X)$ of order $r \leq m$ is a finite linear combination of terms of the form

$$p(X) |X|^{\beta-s} \text{etr}(-X/2) \psi(X), \quad s \leq r,$$

where p is a scalar-valued polynomial and ψ is a derivative of B_Θ of order at most r . Given that $X \in \mathcal{S}_+^d$ implies $|X_{ij}| \leq \sqrt{X_{ii} X_{jj}} \leq \text{tr}(X)$ for all $i, j \in [d]$, every generic polynomial of degree M ,

$$p(X) = \sum_{|\alpha| \leq M} c_\alpha \prod_{1 \leq i \leq j \leq d} X_{ij}^{\alpha_{ij}},$$

satisfies

$$|p(X)| \leq \left(\sum_{|\alpha| \leq M} |c_\alpha| \right) (1 + \text{tr}(X))^M, \quad X \in \mathcal{S}_+^d.$$

Combining this with the bound (5.7) on the derivatives of B_Θ yields, for every $X \in \mathcal{S}_{++}^d$,

$$\begin{aligned} |\bar{\nabla} f_{\alpha, I_d, \Theta}^\mathcal{W}(X)| &\leq C_{\bar{\nabla}} (1 + \text{tr}(X))^{M_{\bar{\nabla}}} \sum_{s=0}^r |X|^{\beta-s} \exp \left\{ -\frac{1}{2} \text{tr}(X) + c_\Theta \sqrt{1 + \text{tr}(X)} \right\} \\ &\leq C'_{\bar{\nabla}} (1 + \text{tr}(X))^{M'_{\bar{\nabla}}} |X|^{\beta-r} \text{etr}(-X/4), \end{aligned} \quad (5.9)$$

for appropriate constants $C_{\overline{\nabla}}, C'_{\overline{\nabla}} > 0$ and $M_{\overline{\nabla}}, M'_{\overline{\nabla}} \in \mathbb{N}$.

This estimate gives exactly the required boundary behavior. If $r \leq m-1$, then $\beta - r > 0$, so $\beta - s \geq \beta - r > 0$ for all $s \leq r$. Since the derivatives of B_{Θ} are continuous on all of $\mathcal{S}^d$, every derivative of order at most $m-1$ extends continuously to $\mathcal{S}^d_+$ and vanishes on $\partial\mathcal{S}^d_+$. If $r = m$, then $\beta - m > -1$, so derivatives of order m are locally integrable near $\partial\mathcal{S}^d_+$ by the cone-integrability estimate (5.8). In particular, $\overline{\nabla} f_{\alpha, I_d, \Theta}^{\mathcal{W}} \in L^1(\mathcal{S}^d_{++})$.

We now justify the repeated integration by parts (5.3) by a localization argument. For the fixed index pairs $(i_{\ell}, j_{\ell}) \in [d]^2$ and any $A \subseteq [m]$, set

$$\overline{\nabla}_A := \prod_{\ell \in A} \nabla_{i_{\ell} j_{\ell}},$$

with the convention that $\overline{\nabla}_{\emptyset}$ is the identity operator. Choose $\chi, \zeta \in C^{\infty}((0, \infty))$ such that

$$\begin{aligned} \chi(t) &= 0 \text{ for } t \in (0, 1], & \chi(t) &= 1 \text{ for } t \in [2, \infty), \\ \zeta(t) &= 1 \text{ for } t \in (0, 1], & \zeta(t) &= 0 \text{ for } t \in [2, \infty). \end{aligned}$$

For $\varepsilon \in (0, 1)$ and $R > 1$, define

$$\omega_{\varepsilon, R}(X) := \chi\left(\frac{|X|}{\varepsilon}\right) \zeta\left(\frac{\text{tr}(X)}{R}\right), \quad \phi_T(X) := \text{etr}(-TX), \quad X \in \mathcal{S}^d_{++}.$$

Then $\omega_{\varepsilon, R} \in C^{\infty}(\mathcal{S}^d_{++})$, $\omega_{\varepsilon, R}(X) \rightarrow 1$ pointwise on $\mathcal{S}^d_{++}$ as $\varepsilon \downarrow 0$ and $R \uparrow \infty$, and $\omega_{\varepsilon, R} \phi_T$ has compact support in $\mathcal{S}^d_{++}$. Indeed, on the support of $\omega_{\varepsilon, R}$, we have $|X| \geq \varepsilon$ and $\text{tr}(X) \leq 2R$, so if $\lambda_1(X) \geq \dots \geq \lambda_d(X) > 0$ are the eigenvalues of X , then

$$[\lambda_d(X), \lambda_1(X)] = \left[\frac{|X|}{\prod_{i=1}^{d-1} \lambda_i(X)}, \lambda_1(X) \right] \subseteq \left[\frac{\varepsilon}{(2R)^{d-1}}, 2R \right] \subseteq (0, \infty).$$

Since $\omega_{\varepsilon, R} \phi_T$ is compactly supported in the interior of $\mathcal{S}^d \cong \mathbb{R}^{d(d+1)/2}$, ordinary Euclidean integration by parts gives

$$\int_{\mathcal{S}^d_{++}} \omega_{\varepsilon, R}(X) \phi_T(X) \overline{\nabla}_{[m]} f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) dX = (-1)^m \int_{\mathcal{S}^d_{++}} f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) \overline{\nabla}_{[m]} (\omega_{\varepsilon, R} \phi_T)(X) dX. \quad (5.10)$$

Expanding $\overline{\nabla}_{[m]}(\omega_{\varepsilon, R} \phi_T)$ by Leibniz' rule, we have

$$\overline{\nabla}_{[m]}(\omega_{\varepsilon, R} \phi_T) = \omega_{\varepsilon, R} \overline{\nabla}_{[m]} \phi_T + \sum_{\emptyset \neq A \subseteq [m]} (\overline{\nabla}_A \omega_{\varepsilon, R}) (\overline{\nabla}_{A^c} \phi_T),$$

where $A^c = [m] \setminus A$. Since

$$\overline{\nabla}_{[m]} \phi_T(X) = (-1)^m \left(\prod_{\ell=1}^m T_{i_{\ell} j_{\ell}} \right) \phi_T(X), \quad X \in \mathcal{S}^d_{++},$$

we obtain

$$\int_{\mathcal{S}^d_{++}} \omega_{\varepsilon, R}(X) \phi_T(X) \overline{\nabla}_{[m]} f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) dX = \left(\prod_{\ell=1}^m T_{i_{\ell} j_{\ell}} \right) \int_{\mathcal{S}^d_{++}} \omega_{\varepsilon, R}(X) \phi_T(X) f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) dX + E_{\varepsilon, R},$$

where the error term $E_{\varepsilon,R}$ satisfies

$$|E_{\varepsilon,R}| \leq C_T \sum_{\emptyset \neq A \subseteq [m]} \int_{\mathcal{S}_{++}^d} f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) |\bar{\nabla}_A \omega_{\varepsilon,R}(X)| dX,$$

for a suitable constant $C_T \in (0, \infty)$.

Each term in $\bar{\nabla}_A \omega_{\varepsilon,R}$ contains either a derivative of $\chi(|X|/\varepsilon)$, hence is supported on $\{\varepsilon \leq |X| \leq 2\varepsilon\}$, or a derivative of $\zeta(\text{tr}(X)/R)$, hence is supported on $\{R \leq \text{tr}(X) \leq 2R\}$, or both. Since derivatives of $|X|$ are polynomials and derivatives of $\text{tr}(X)$ are constants, there exist constants $C_A, M_A > 0$ such that

$$|\bar{\nabla}_A \omega_{\varepsilon,R}(X)| \leq C_A (1 + \text{tr}(X))^{M_A} \left[\sum_{q=1}^{|A|} |X|^{-q} \mathbf{1}_{\{\varepsilon \leq |X| \leq 2\varepsilon\}} + \mathbf{1}_{\{R \leq \text{tr}(X) \leq 2R\}} \right].$$

Here we used that every derivative landing on $\chi(|X|/\varepsilon)$ produces a factor ε^{-1} , and on the support of such a derivative we have $\varepsilon \leq |X| \leq 2\varepsilon$, so $\varepsilon^{-q} \leq 2^q |X|^{-q}$. Also, from the expression of the noncentral Wishart density in (5.4), and the bound (5.6) on the matrix Bessel function,

$$f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) \leq C_{\Theta} (1 + \text{tr}(X))^{M_{\Theta}} |X|^{\beta} \text{etr}(-X/4), \quad X \in \mathcal{S}_{++}^d,$$

for some constants $C_{\Theta} \in (0, \infty)$ and $M_{\Theta} \in \mathbb{N}$. Thus, the boundary-shell contribution to $E_{\varepsilon,R}$ is dominated by a constant multiple of

$$(1 + \text{tr}(X))^M |X|^{\beta-q} \text{etr}(-X/4), \quad 1 \leq q \leq m,$$

for some constant $M \in \mathbb{N}$, which is integrable because $\beta - m > -1$. The outer-shell contribution to $E_{\varepsilon,R}$ is dominated by

$$(1 + \text{tr}(X))^M |X|^{\beta} \text{etr}(-X/4),$$

which is also integrable. Since the corresponding shell indicators converge pointwise to 0 as $\varepsilon \downarrow 0$ and $R \uparrow \infty$, dominated convergence and (5.8) yield

$$E_{\varepsilon,R} \rightarrow 0.$$

Moreover, $\bar{\nabla}_{[m]} f_{\alpha, I_d, \Theta}^{\mathcal{W}}, f_{\alpha, I_d, \Theta}^{\mathcal{W}} \in L^1(\mathcal{S}_{++}^d)$ by (5.8) and (5.9), so dominated convergence gives

$$\begin{aligned} \int_{\mathcal{S}_{++}^d} \omega_{\varepsilon,R}(X) \phi_T(X) \bar{\nabla}_{[m]} f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) dX &\rightarrow \int_{\mathcal{S}_{++}^d} \phi_T(X) \bar{\nabla}_{[m]} f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) dX, \\ \int_{\mathcal{S}_{++}^d} \omega_{\varepsilon,R}(X) \phi_T(X) f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) dX &\rightarrow \int_{\mathcal{S}_{++}^d} \phi_T(X) f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) dX. \end{aligned}$$

Letting $\varepsilon \downarrow 0$ and $R \uparrow \infty$ in (5.10) completes the proof of (5.3) and the lemma. $\square$

Lemma 5.2. *Let $\alpha > 3d - 3$ be given, and let $\text{adj}(\cdot)$ denote the adjugate matrix operator. For any $k \in \mathbb{N}$, $(i_1, j_1), \dots, (i_k, j_k) \in [d]^2$, and $\Theta, X \in \mathcal{S}_{++}^d$, we have*

$$\left(\prod_{\ell=1}^k \nabla_{\Theta, i_{\ell} j_{\ell}} \right) f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) = (-1)^k \left(\prod_{\ell=1}^k \left\{ \nabla_X \text{adj}(I_d + 2\nabla_X) \right\}_{i_{\ell} j_{\ell}} \right) f_{\alpha+2k, I_d, \Theta}^{\mathcal{W}}(X),$$

where ∇_{Θ} acts on the noncentrality parameter Θ and ∇_X acts on the state variable X .

Proof of Lemma 5.2. By Remark 2.1 with $\Sigma = I_d$,

$$\mathcal{L}[f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T) = \frac{\text{etr}\{-T(I_d + 2T)^{-1}\Theta\}}{|I_d + 2T|^{\alpha/2}}, \quad T \in \mathcal{S}_+^d. \quad (5.11)$$

The interchange of $\nabla_{\Theta, ij}$ with the Laplace integral is justified locally uniformly in $\Theta \in \mathcal{S}_{++}^d$ by the same Bessel bounds used in the proof of Lemma 5.1; for suitable constants $C, M > 0$,

$$|\nabla_{\Theta, ij} f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X)| \leq C(1 + \text{tr}(X))^M |X|^{(\alpha-d-1)/2} \text{etr}(-X/4),$$

which is integrable under the present assumption on α . Differentiating (5.11) under the integral sign with respect to the (i, j) component of Θ gives

$$\begin{aligned} \mathcal{L}[\nabla_{\Theta, ij} f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T) &= \nabla_{\Theta, ij} \mathcal{L}[f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T) \\ &= -(T(I_d + 2T)^{-1})_{ij} \mathcal{L}[f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T), \quad i, j \in [d]. \end{aligned}$$

We can write

$$T(I_d + 2T)^{-1} = \frac{T \text{adj}(I_d + 2T)}{|I_d + 2T|}.$$

Moreover, by Remark 2.1 again,

$$|I_d + 2T|^{-1} \mathcal{L}[f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T) = \mathcal{L}[f_{\alpha+2, I_d, \Theta}^{\mathcal{W}}](T).$$

Hence, for each $(i, j) \in [d]^2$,

$$\mathcal{L}[\nabla_{\Theta, ij} f_{\alpha, I_d, \Theta}^{\mathcal{W}}](T) = -(T \text{adj}(I_d + 2T))_{ij} \mathcal{L}[f_{\alpha+2, I_d, \Theta}^{\mathcal{W}}](T).$$

Every entry of the matrix $U \text{adj}(I_d + 2U)$ is a polynomial of degree at most d in the entries of U . Since $\alpha > 3d - 3$, we have $\alpha + 2 > d - 1 + 2d$, so Lemma 5.1 applies to $f_{\alpha+2, I_d, \Theta}^{\mathcal{W}}$ with $m = d$:

$$\mathcal{L}\left[-\{\nabla_X \text{adj}(I_d + 2\nabla_X)\}_{ij} f_{\alpha+2, I_d, \Theta}^{\mathcal{W}}\right](T) = -(T \text{adj}(I_d + 2T))_{ij} \mathcal{L}[f_{\alpha+2, I_d, \Theta}^{\mathcal{W}}](T).$$

Thus the two sides in the next display have the same Laplace transform for every $T \in \mathcal{S}_+^d$. By uniqueness of Laplace transforms on $\mathcal{S}_+^d$, they are equal for a.e. $X \in \mathcal{S}_{++}^d$; since both sides are continuous on $\mathcal{S}_{++}^d$, the equality holds for every $X \in \mathcal{S}_{++}^d$:

$$\nabla_{\Theta, ij} f_{\alpha, I_d, \Theta}^{\mathcal{W}}(X) = -\{\nabla_X \text{adj}(I_d + 2\nabla_X)\}_{ij} f_{\alpha+2, I_d, \Theta}^{\mathcal{W}}(X), \quad X \in \mathcal{S}_{++}^d. \quad (5.12)$$

The claim follows by iterating (5.12) k times, using that the families of operators $\nabla_{\Theta, ij}$ and $\{\nabla_X \text{adj}(I_d + 2\nabla_X)\}_{ij}$ commute among themselves and with each other. Since the shape parameter α increases by 2 at each step, the baseline assumption $\alpha > 3d - 3$ guarantees that the same argument applies throughout the iteration. $\square$

Lemma 5.3 is a scale-normalized Wishart analog of Lemma 2.5 of Luk (1994).

Lemma 5.3. *Let $\alpha > 3d - 3$, $t \geq 0$, and $k \in \mathbb{N}$. Suppose that $h \in C^{kd}(\mathcal{S}_{++}^d)$, and that for every differential monomial $\bar{\nabla}$ of order at most kd in the entries of ∇_X , including the order-zero monomial, there exist constants $C_{\bar{\nabla}}, N_{\bar{\nabla}} \geq 0$ such that*

$$|\bar{\nabla}h(X)| \leq C_{\bar{\nabla}}(1 + \|X\|_F^{N_{\bar{\nabla}}}), \quad X \in \mathcal{S}_{++}^d. \quad (5.13)$$

Then, for every $(i_1, j_1), \dots, (i_k, j_k) \in [d]^2$ and $\Lambda \in \mathcal{S}_{++}^d$,

$$\left(\prod_{\ell=1}^k \nabla_{\Lambda, i_\ell j_\ell} \right) (\mathcal{Q}_t h)(\Lambda) = e^{-2kt} \mathbb{E} \left[\left(\prod_{\ell=1}^k \mathcal{D}_{i_\ell j_\ell} \right) h(\mathfrak{Y}_{k,t}) \right], \quad (5.14)$$

where $\mathcal{D} = \nabla_X \operatorname{adj}(I_d - 2\nabla_X)$, $\mathfrak{Y}_{k,t} \sim \mathcal{W}_d(\alpha + 2k, I_d, e^{-2t}\Lambda)$ and is defined in (3.15).

Proof of Lemma 5.3. For a temporary shape parameter $\gamma > 3d - 3$, define

$$\mathcal{Q}_t^{(\gamma)} g(\Lambda) := \int_{\mathcal{S}_{++}^d} g(X) f_{\gamma, I_d, e^{-2t}\Lambda}^{\mathcal{W}}(X) dX, \quad \Lambda \in \mathcal{S}_{++}^d,$$

so that $\mathcal{Q}_t = \mathcal{Q}_t^{(\alpha)}$. We first prove the one-step identity

$$\nabla_{\Lambda, ij}(\mathcal{Q}_t^{(\gamma)} g)(\Lambda) = e^{-2t} \mathcal{Q}_t^{(\gamma+2)}(\mathcal{D}_{ij} g)(\Lambda), \quad (5.15)$$

whenever $g \in C^d(\mathcal{S}_{++}^d)$ and all differential monomials of order at most d in the entries of ∇_X , including the order-zero monomial, applied to g have polynomial growth.

Fix $\Lambda \in \mathcal{S}_{++}^d$ and choose a compact neighborhood $K \subseteq \mathcal{S}_{++}^d$ of $e^{-2t}\Lambda$. By Lemma 5.2 with $k = 1$, for each $\Theta \in K$ the quantity $\nabla_{\Theta, ij} f_{\gamma, I_d, \Theta}^{\mathcal{W}}$ can be written as a constant-coefficient X -differential operator of order at most d applied to $f_{\gamma+2, I_d, \Theta}^{\mathcal{W}}$. The estimates established in the proof of Lemma 5.1 are uniform for $\Theta \in K$, so after multiplication by $g(X)$ the integrands below are dominated by a single $L^1(\mathcal{S}_{++}^d)$ function of X . Hence differentiation under the integral sign is justified by dominated convergence.

By the chain rule and differentiating under the integral sign, we obtain

$$\nabla_{\Lambda, ij}(\mathcal{Q}_t^{(\gamma)} g)(\Lambda) = e^{-2t} \int_{\mathcal{S}_{++}^d} g(X) [\nabla_{\Theta, ij} f_{\gamma, I_d, \Theta}^{\mathcal{W}}(X)]_{\Theta=e^{-2t}\Lambda} dX.$$

Applying Lemma 5.2, which is permissible since $\gamma > 3d - 3$, gives

$$\nabla_{\Lambda, ij}(\mathcal{Q}_t^{(\gamma)} g)(\Lambda) = -e^{-2t} \int_{\mathcal{S}_{++}^d} g(X) \mathcal{A}_{ij} f_{\gamma+2, I_d, e^{-2t}\Lambda}^{\mathcal{W}}(X) dX,$$

where $\mathcal{A}_{ij} := \{\nabla_X \operatorname{adj}(I_d + 2\nabla_X)\}_{ij}$.

The operator $\mathcal{A}_{ij}$ is a constant-coefficient differential operator of order at most d . Given that $\gamma > 3d - 3$, we have $(\gamma + 2 - d - 1)/2 > d - 1$. Exactly as in the proof of Lemma 5.1, this implies that derivatives of $f_{\gamma+2, I_d, e^{-2t}\Lambda}^{\mathcal{W}}$ of order at most $d - 1$ vanish on $\partial\mathcal{S}_{++}^d$, while derivatives of order at most d are locally integrable there. Together with the polynomial growth assumption on the derivatives of g and the bounds in (5.9) for the derivatives of $f_{\gamma+2, I_d, e^{-2t}\Lambda}^{\mathcal{W}}$, this justifies the integration by parts below.

Since the entries of ∇_X are constant-coefficient commuting differential operators, the formal adjoint of $\mathcal{A}_{ij}$ with respect to Lebesgue measure is obtained by replacing ∇_X with $-\nabla_X$, namely

$$\mathcal{A}_{ij}^* = \{(-\nabla_X)\text{adj}(I_d - 2\nabla_X)\}_{ij} = -\mathcal{D}_{ij}.$$

Therefore,

$$\begin{aligned}\nabla_{\Lambda, ij}(\mathcal{Q}_t^{(\gamma)} g)(\Lambda) &= -e^{-2t} \int_{\mathcal{S}_{++}^d} \{\mathcal{A}_{ij}^* g(X)\} f_{\gamma+2, I_d, e^{-2t}\Lambda}^{\mathcal{W}}(X) dX \\ &= e^{-2t} \int_{\mathcal{S}_{++}^d} \{\mathcal{D}_{ij} g(X)\} f_{\gamma+2, I_d, e^{-2t}\Lambda}^{\mathcal{W}}(X) dX,\end{aligned}$$

which proves (5.15).

We now iterate (5.15). Set

$$g_0 := h, \quad g_r := \left(\prod_{\ell=1}^r \mathcal{D}_{i_\ell j_\ell} \right) h, \quad \gamma_r := \alpha + 2r, \quad r \in \{0, \dots, k\}.$$

Since each $\mathcal{D}_{ij}$ is a constant-coefficient differential operator of order at most d , the function g_r belongs to $C^{(k-r)d}(\mathcal{S}_{++}^d)$, and all its differential monomials of order at most $(k-r)d$, including the order-zero monomial, have polynomial growth. Moreover, $\gamma_r > 3d - 3$ for every $r \in \{0, \dots, k\}$. Thus the one-step identity (5.15) can be applied successively to the pairs (γ_r, g_r) for $r \in \{0, \dots, k-1\}$.

We prove by induction on r that

$$\left(\prod_{\ell=1}^r \nabla_{\Lambda, i_\ell j_\ell} \right) (\mathcal{Q}_t^{(\alpha)} h)(\Lambda) = e^{-2rt} \mathcal{Q}_t^{(\alpha+2r)} g_r(\Lambda), \quad r \in \{0, \dots, k\}. \quad (5.16)$$

The claim is trivial for $r = 0$. If it holds for some $r \in \{0, \dots, k-1\}$, then applying (5.15) with $\gamma = \alpha + 2r$ and $g = g_r$ gives

$$\begin{aligned}\left(\prod_{\ell=1}^{r+1} \nabla_{\Lambda, i_\ell j_\ell} \right) (\mathcal{Q}_t^{(\alpha)} h)(\Lambda) &= e^{-2rt} \nabla_{\Lambda, i_{r+1} j_{r+1}} \mathcal{Q}_t^{(\alpha+2r)} g_r(\Lambda) \\ &= e^{-2(r+1)t} \mathcal{Q}_t^{(\alpha+2r+2)} (\mathcal{D}_{i_{r+1} j_{r+1}} g_r)(\Lambda) \\ &= e^{-2(r+1)t} \mathcal{Q}_t^{(\alpha+2(r+1))} g_{r+1}(\Lambda),\end{aligned}$$

where we used the fact that constant-coefficient differential operators commute. This completes the induction and proves (5.16).

Taking $r = k$ yields

$$\left(\prod_{\ell=1}^k \nabla_{\Lambda, i_\ell j_\ell} \right) (\mathcal{Q}_t h)(\Lambda) = e^{-2kt} \int_{\mathcal{S}_{++}^d} \left\{ \left(\prod_{\ell=1}^k \mathcal{D}_{i_\ell j_\ell} \right) h(X) \right\} f_{\alpha+2k, I_d, e^{-2t}\Lambda}^{\mathcal{W}}(X) dX,$$

which is the claimed expression (5.14). □

Remark 5.1. The expectation on the right-hand side of (5.14) is finite under the weaker condition $\alpha > d - 1$. Indeed, set

$$G(X) := \left(\prod_{\ell=1}^k \mathcal{D}_{i_\ell j_\ell} \right) h(X).$$

Each entry of $\mathcal{D}$ is a constant-coefficient differential operator of order at most d , hence G is a finite linear combination of differential monomials of order at most kd in the entries of ∇_X . By (5.13), there exist constants $C, N < \infty$ such that

$$|G(X)| \leq C(1 + \|X\|_F^N), \quad X \in \mathcal{S}_{++}^d.$$

Let $\beta := \alpha + 2k$ and $\Theta := e^{-2t}\Lambda$. If $\alpha > d - 1$, then $\beta > d - 1$, so $\mathfrak{Y}_{k,t} \sim \mathcal{W}_d(\beta, I_d, \Theta)$ is well defined. The moment generating form of the noncentral Wishart transform, obtained from Remark 2.1 on its natural domain, gives, for every $0 < s < 1/2$,

$$\mathbb{E}[\exp\{s \operatorname{tr}(\mathfrak{Y}_{k,t})\}] = (1 - 2s)^{-\beta d/2} \exp\left\{\frac{s}{1 - 2s} \operatorname{tr}(\Theta)\right\} < \infty.$$

Since $\mathfrak{Y}_{k,t} \in \mathcal{S}_+^d$ almost surely, $\|\mathfrak{Y}_{k,t}\|_F \leq \operatorname{tr}(\mathfrak{Y}_{k,t})$, and therefore $\mathbb{E}[\|\mathfrak{Y}_{k,t}\|_F^N] < \infty$. Consequently, $\mathbb{E}[|G(\mathfrak{Y}_{k,t})|] < \infty$. Thus the condition $\alpha > 3d - 3$ in Lemma 5.3 is not an integrability condition for the expectation in (5.14); it is used in the present proof of the transfer identity.

6. Proofs of the main results

6.1. Proof of Proposition 3.1

For any $f \in C^2(\mathcal{S}_{++}^d)$, the extended generator of $(\mathfrak{W}_t)_{t \geq 0}$ arises from Itô's formula as

$$\mathcal{A}^{\mathcal{W}} f(S) = \underbrace{2 \operatorname{tr}\{(\alpha \Sigma - S) \nabla f(S)\}}_{\text{drift term}} + \underbrace{\frac{1}{2} \sum_{i,j,k,\ell=1}^d \frac{(\mathrm{d}\mathfrak{W}_t^{(\text{diff})})_{ij} \bullet (\mathrm{d}\mathfrak{W}_t^{(\text{diff})})_{k\ell}}{\mathrm{d}t} \nabla_{ij} \nabla_{k\ell} f(S)}_{\text{diffusion term}},$$

where $\bullet$ denotes the Itô product, and

$$\mathrm{d}\mathfrak{W}_t^{(\text{diff})} := \sqrt{2} (\mathrm{d}\mathfrak{X}_t + \mathrm{d}\mathfrak{X}_t^\top), \quad \mathrm{d}\mathfrak{X}_t := S^{1/2} \mathrm{d}\mathfrak{B}_t \Sigma^{1/2}.$$

Since $(\mathrm{d}\mathfrak{X}_t)_{ij} = \sum_{p,q=1}^d (S^{1/2})_{ip} (\mathrm{d}\mathfrak{B}_t)_{pq} (\Sigma^{1/2})_{qj}$, the identity $(\mathrm{d}\mathfrak{B}_t)_{pq} \bullet (\mathrm{d}\mathfrak{B}_t)_{p'q'} = \delta_{pp'} \delta_{qq'} \mathrm{d}t$ yields

$$\begin{aligned} (\mathrm{d}\mathfrak{X}_t)_{ij} \bullet (\mathrm{d}\mathfrak{X}_t)_{k\ell} &= \sum_{p,q,p',q'=1}^d (S^{1/2})_{ip} (S^{1/2})_{kp'} (\Sigma^{1/2})_{qj} (\Sigma^{1/2})_{q'\ell} \{(\mathrm{d}\mathfrak{B}_t)_{pq} \bullet (\mathrm{d}\mathfrak{B}_t)_{p'q'}\} \\ &= \sum_{p,q=1}^d (S^{1/2})_{ip} (S^{1/2})_{kp} (\Sigma^{1/2})_{qj} (\Sigma^{1/2})_{q\ell} \mathrm{d}t \\ &= S_{ik} \Sigma_{j\ell} \mathrm{d}t. \end{aligned}$$

It follows that the quadratic covariation between the (i, j) and (k, ℓ) entries of $\mathrm{d}\mathfrak{W}_t^{(\text{diff})}$ is

$$\begin{aligned} (\mathrm{d}\mathfrak{W}_t^{(\text{diff})})_{ij} \bullet (\mathrm{d}\mathfrak{W}_t^{(\text{diff})})_{k\ell} &= 2 \{(\mathrm{d}\mathfrak{X}_t)_{ij} + (\mathrm{d}\mathfrak{X}_t)_{ji}\} \bullet \{(\mathrm{d}\mathfrak{X}_t)_{k\ell} + (\mathrm{d}\mathfrak{X}_t)_{\ell k}\} \\ &= 2 \{S_{ik} \Sigma_{j\ell} + S_{i\ell} \Sigma_{jk} + S_{jk} \Sigma_{i\ell} + S_{j\ell} \Sigma_{ik}\} \mathrm{d}t. \end{aligned}$$

Using the symmetry of ∇ , the diffusion part of $\mathcal{A}^W f(S)$ becomes

$$\begin{aligned}
& \sum_{i,j,k,\ell=1}^d \{S_{ik}\Sigma_{j\ell} + S_{i\ell}\Sigma_{jk} + S_{jk}\Sigma_{i\ell} + S_{j\ell}\Sigma_{ik}\} \nabla_{ij} \nabla_{k\ell} f(S) \\
&= \left\{ \sum_{i,k=1}^d S_{ik} \sum_{j,\ell=1}^d \nabla_{ij} \Sigma_{j\ell} \nabla_{\ell k} + \sum_{i,\ell=1}^d S_{i\ell} \sum_{j,k=1}^d \nabla_{ij} \Sigma_{jk} \nabla_{k\ell} \right. \\
&\quad \left. + \sum_{j,k=1}^d S_{jk} \sum_{i,\ell=1}^d \nabla_{ji} \Sigma_{i\ell} \nabla_{\ell k} + \sum_{j,\ell=1}^d S_{j\ell} \sum_{i,k=1}^d \nabla_{ji} \Sigma_{ik} \nabla_{k\ell} \right\} f(S) \\
&= \left\{ \sum_{i,k=1}^d S_{ik} (\nabla \Sigma \nabla)_{ik} + \sum_{i,\ell=1}^d S_{i\ell} (\nabla \Sigma \nabla)_{i\ell} + \sum_{j,k=1}^d S_{jk} (\nabla \Sigma \nabla)_{jk} + \sum_{j,\ell=1}^d S_{j\ell} (\nabla \Sigma \nabla)_{j\ell} \right\} f(S) \\
&= 4 \operatorname{tr}\{S \nabla \Sigma \nabla f(S)\}.
\end{aligned}$$

Consequently, the extended generator of $(\mathfrak{W}_t)_{t \geq 0}$ is

$$\mathcal{A}^W f(S) = 2 \operatorname{tr}\{(\alpha \Sigma - S) \nabla f(S)\} + 4 \operatorname{tr}\{S \nabla \Sigma \nabla f(S)\},$$

as claimed.

6.2. Proof of Proposition 3.2

Fix $T, W \in \mathcal{S}_+^d$. The proof is divided into four steps. In Step 1, we identify the $\mathcal{S}_+^d$ -valued affine extension of the Wishart process as an affine diffusion and write down the corresponding Riccati equations. In Step 2, we solve these equations explicitly and identify the conditional transition law. In Step 3, we pass to the limiting law and obtain the semigroup representation. In Step 4, we prove the invariance of $\mathcal{W}_d(\alpha, \Sigma)$ by comparing Laplace transforms.

Step 1: Affine transform formula and Riccati equations. In the notation of Cuchiero et al. (2011, Theorems 2.4 and 2.6), with the diffusion matrix parameter denoted here by a to avoid confusion with the shape parameter α , take

$$a = 2\Sigma, \quad b = 2\alpha\Sigma, \quad B(S) = -2S, \quad c = 0, \quad \gamma = 0_{d \times d}, \quad m(\cdot) = \mu(\cdot) = 0.$$

Given that $\alpha > d - 1$, we have $b - (d - 1)a = 2(\alpha - d + 1)\Sigma \in \mathcal{S}_{++}^d$, so the constant drift admissibility condition in their Definition 2.3 is satisfied. The linear drift $B(S) = -2S$ is of the form $HS + SH^\top$ with $H = -I_d$, and the remaining admissibility conditions in their Definition 2.3 are immediate. Thus Theorem 2.4 of Cuchiero et al. (2011) yields an affine process on $\mathcal{S}_+^d$ with the affine transform formula below. Since $c = 0$, $\gamma = 0_{d \times d}$ and $m(\cdot) = \mu(\cdot) = 0$, the process is conservative. Moreover, by Theorem 2.6 of Cuchiero et al. (2011), it is an affine diffusion whose semimartingale representation is exactly (3.1). Its affine transform formula gives

$$\mathbb{E}[\operatorname{etr}(-T\mathfrak{W}_t) \mid \mathfrak{W}_0 = W] = \exp\{-\varphi_t(T) - \operatorname{tr}(\Psi_t(T)W)\}, \quad (6.1)$$

where $\varphi_t(T)$ and $\Psi_t(T)$ solve the generalized Riccati equations

$$\frac{d}{dt} \varphi_t(T) = 2\alpha \operatorname{tr}(\Sigma \Psi_t(T)), \quad \varphi_0(T) = 0, \quad (6.2)$$

and

$$\frac{d}{dt}\Psi_t(T) = -2\Psi_t(T) - 4\Psi_t(T)\Sigma\Psi_t(T), \quad \Psi_0(T) = T. \quad (6.3)$$

Step 2: Solution of the Riccati equations and transition law. Recall that $\Sigma_t = (1 - e^{-2t})\Sigma$, and set

$$R_t := I_d + 2T\Sigma_t, \quad \varphi_t(T) := \frac{\alpha}{2} \log(|R_t|), \quad \Psi_t(T) := e^{-2t}R_t^{-1}T.$$

Note that $|R_t| = |I_d + 2T^{1/2}\Sigma_t T^{1/2}| > 0$, so $\varphi_t(T)$ is well defined. First,

$$\Psi_t(T) = e^{-2t}(I_d + 2T\Sigma_t)^{-1}T = e^{-2t}T^{1/2}(I_d + 2T^{1/2}\Sigma_t T^{1/2})^{-1}T^{1/2} \in \mathcal{S}_+^d.$$

Since $\frac{d}{dt}R_t = 4e^{-2t}T\Sigma$, direct differentiation gives

$$\frac{d}{dt}\varphi_t(T) = \frac{\alpha}{2} \text{tr} \left(R_t^{-1} \frac{d}{dt}R_t \right) = 2\alpha e^{-2t} \text{tr}(R_t^{-1}T\Sigma) = 2\alpha \text{tr}(\Sigma\Psi_t(T)),$$

and

$$\frac{d}{dt}\Psi_t(T) = -2e^{-2t}R_t^{-1}T - e^{-2t}R_t^{-1} \left(\frac{d}{dt}R_t \right) R_t^{-1}T = -2\Psi_t(T) - 4\Psi_t(T)\Sigma\Psi_t(T).$$

Also, we have $R_0 = I_d$, $\varphi_0(T) = 0$ and $\Psi_0(T) = T$. Therefore, the displayed functions solve (6.2) and (6.3). Substituting them into (6.1) yields

$$\mathbb{E}[\text{etr}(-T\mathfrak{W}_t) \mid \mathfrak{W}_0 = W] = \frac{\text{etr}\{-e^{-2t}(I_d + 2T\Sigma_t)^{-1}TW\}}{|I_d + 2T\Sigma_t|^{\alpha/2}}. \quad (6.4)$$

Let $\Theta_t := e^{-2t}\Sigma_t^{-1}W$. Since $\Theta_t\Sigma_t^{-1} = e^{-2t}\Sigma_t^{-1}W\Sigma_t^{-1} \in \mathcal{S}_+^d$ and $\alpha > d - 1$, the noncentral Wishart law $\mathcal{W}_d(\alpha, \Sigma_t, \Theta_t)$ is well defined. Its Laplace transform is

$$\frac{\text{etr}\{-T\Sigma_t(I_d + 2T\Sigma_t)^{-1}\Theta_t\}}{|I_d + 2T\Sigma_t|^{\alpha/2}}, \quad T \in \mathcal{S}_+^d.$$

Since $T\Sigma_t$ commutes with $(I_d + 2T\Sigma_t)^{-1}$, we have

$$T\Sigma_t(I_d + 2T\Sigma_t)^{-1}\Theta_t = e^{-2t}(I_d + 2T\Sigma_t)^{-1}TW.$$

Thus the Laplace transform of $\mathcal{W}_d(\alpha, \Sigma_t, \Theta_t)$ agrees with (6.4). By uniqueness of Laplace transforms on $\mathcal{S}_+^d$,

$$\mathfrak{W}_t \mid \{\mathfrak{W}_0 = W\} \sim \mathcal{W}_d(\alpha, \Sigma_t, e^{-2t}\Sigma_t^{-1}W), \quad t > 0,$$

which is (3.4). Since $\alpha > d - 1$ and $\Sigma_t \in \mathcal{S}_{++}^d$, the density formula (2.2) also gives

$$P_t(W, \partial\mathcal{S}_+^d) = 0, \quad t > 0, \quad W \in \mathcal{S}_+^d.$$

Step 3: Limiting law and semigroup representation. Letting $t \rightarrow \infty$ in (6.4), and using $\Sigma_t \rightarrow \Sigma$ and $e^{-2t}(I_d + 2T\Sigma_t)^{-1}T \rightarrow 0_{d \times d}$, we obtain

$$\lim_{t \rightarrow \infty} \mathbb{E}[\text{etr}(-T\mathfrak{W}_t) \mid \mathfrak{W}_0 = W] = |I_d + 2T\Sigma|^{-\alpha/2}, \quad T \in \mathcal{S}_+^d.$$

The right-hand side is the Laplace transform of $\mathcal{W}_d(\alpha, \Sigma)$. Hence, by Lévy's continuity theorem for Laplace transforms on $\mathcal{S}_+^d$,

$$\mathfrak{W}_t \mid \{\mathfrak{W}_0 = W\} \xrightarrow{\text{law}} \mathfrak{W}_\infty \sim \mathcal{W}_d(\alpha, \Sigma), \quad t \rightarrow \infty,$$

which proves (3.5). Then (3.7) follows immediately because $\Sigma_t^{1/2} \mathfrak{S}_t \Sigma_t^{1/2} \sim \mathcal{W}_d(\alpha, \Sigma_t, e^{-2t} \Sigma_t^{-1} W)$ by a rescaling of the Laplace transform; see, e.g., Lemma 2.1 of Genest et al. (2026).

Step 4: Invariance. To prove (3.8), it suffices to compare Laplace transforms. For any $T \in \mathcal{S}_+^d$, by (6.4),

$$\begin{aligned} \int_{\mathcal{S}_+^d} \text{etr}(-TY) (\xi \mathcal{P}_t^{\mathcal{W}})(dY) &= \int_{\mathcal{S}_+^d} \mathbb{E}[\text{etr}(-T\mathfrak{W}_t) \mid \mathfrak{W}_0 = W] \xi(dW) \\ &= \frac{1}{|I_d + 2T\Sigma_t|^{\alpha/2}} \int_{\mathcal{S}_+^d} \text{etr}\{-U_t(T)W\} \xi(dW), \end{aligned}$$

where $U_t(T) := e^{-2t}(I_d + 2T\Sigma_t)^{-1}T = \Psi_t(T) \in \mathcal{S}_+^d$. Applying the Wishart Laplace transform to ξ gives

$$\int_{\mathcal{S}_+^d} \text{etr}\{-U_t(T)W\} \xi(dW) = \frac{1}{|I_d + 2U_t(T)\Sigma|^{\alpha/2}}.$$

Since $\Sigma_t = (1 - e^{-2t})\Sigma$, the matrices $T\Sigma_t$ and $T\Sigma$ commute. Therefore,

$$\begin{aligned} |I_d + 2U_t(T)\Sigma| &= |I_d + 2e^{-2t}(I_d + 2T\Sigma_t)^{-1}T\Sigma| \\ &= |(I_d + 2T\Sigma_t + 2e^{-2t}T\Sigma)(I_d + 2T\Sigma_t)^{-1}| = \frac{|I_d + 2T\Sigma|}{|I_d + 2T\Sigma_t|}. \end{aligned}$$

Substituting this into the previous display yields

$$\int_{\mathcal{S}_+^d} \text{etr}(-TY) (\xi \mathcal{P}_t^{\mathcal{W}})(dY) = \frac{1}{|I_d + 2T\Sigma_t|^{\alpha/2}} \left(\frac{|I_d + 2T\Sigma_t|}{|I_d + 2T\Sigma|} \right)^{\alpha/2} = |I_d + 2T\Sigma|^{-\alpha/2}.$$

This is the Laplace transform of ξ . By uniqueness of Laplace transforms on $\mathcal{S}_+^d$, we get $\xi \mathcal{P}_t^{\mathcal{W}} = \xi$, which is (3.8). This concludes the proof.

6.3. Proof of Corollary 3.3

Let $(\mathfrak{W}_t^\infty)_{t \geq 0}$ be a stationary Wishart process with invariant distribution $\mathcal{W}_d(\alpha, \Sigma)$, and write $\mathfrak{W}_\infty \sim \mathcal{W}_d(\alpha, \Sigma)$ for its one-time marginal distribution. For any given $f \in C_{\mathcal{A}^{\mathcal{W}}}^2(\mathcal{S}_{++}^d)$, define the local martingale $M_t^f := f(\mathfrak{W}_t^\infty) - f(\mathfrak{W}_0^\infty) - \int_0^t \mathcal{A}^{\mathcal{W}} f(\mathfrak{W}_s^\infty) ds$, $t \geq 0$. The integrability assumptions in (3.9) and stationarity imply that $(M_t^f)_{t \geq 0}$ is in fact a martingale. Hence, $\mathbb{E}[M_t^f] = 0$. By stationarity, $\mathfrak{W}_s^\infty \sim \mathfrak{W}_\infty$ for every $s \geq 0$, and by the invariance of the Wishart law in Proposition 3.2, we have $\mathbb{E}[f(\mathfrak{W}_t^\infty)] = \mathbb{E}[f(\mathfrak{W}_0^\infty)] = \mathbb{E}[f(\mathfrak{W}_\infty)]$ and $\mathbb{E}[\mathcal{A}^{\mathcal{W}} f(\mathfrak{W}_s^\infty)] = \mathbb{E}[\mathcal{A}^{\mathcal{W}} f(\mathfrak{W}_\infty)]$. Therefore, $\mathbb{E}[\mathcal{A}^{\mathcal{W}} f(\mathfrak{W}_\infty)] = 0$, which proves the forward implication.

We now prove the converse implication. Assume that the Stein identities $\mathbb{E}[\mathcal{A}^{\mathcal{W}} f(\mathfrak{W})] = 0$ hold for the $\mathcal{S}_{++}^d$ -valued random matrix $\mathfrak{W} \sim \mathcal{W}_d(\alpha, \Sigma)$ and all $f \in C_{\mathcal{A}^{\mathcal{W}}}^2(\mathcal{S}_{++}^d)$. Fix $T \in \mathcal{S}_{++}^d$, and set $f_T(S) := \text{etr}(-TS)$, $S \in \mathcal{S}_{++}^d$, and

$$L(T) := \mathbb{E}[\text{etr}(-T\mathfrak{W})].$$

We have $f_T \in C_{\mathcal{A}^w}^2(\mathcal{S}_{++}^d)$ because of the exponential decay. Also, we have $\nabla f_T(S) = -T f_T(S)$ and $\nabla \Sigma \nabla f_T(S) = T \Sigma T f_T(S)$. Hence, by Proposition 3.1 and the assumption $\mathbb{E}[\mathcal{A}^w f_T(\mathfrak{W})] = 0$,

$$\mathbb{E}[\text{tr}\{\mathfrak{W}T\} \text{etr}(-T\mathfrak{W})] + 2 \mathbb{E}[\text{tr}\{\mathfrak{W}T\Sigma T\} \text{etr}(-T\mathfrak{W})] = \alpha \text{tr}(\Sigma T) L(T). \quad (6.5)$$

For $H \in \mathcal{S}^d$, dominated convergence gives $DL(T)[H] = -\mathbb{E}[\text{tr}\{H\mathfrak{W}\} \text{etr}(-T\mathfrak{W})]$. Consequently, by (6.5),

$$DL(T)[-(T + 2T\Sigma T)] = \alpha \text{tr}(\Sigma T) L(T), \quad T \in \mathcal{S}_{++}^d. \quad (6.6)$$

Now define

$$T_t := (e^t T^{-1} + 2(e^t - 1)\Sigma)^{-1}, \quad t \geq 0.$$

Then $T_t \in \mathcal{S}_{++}^d$, $T_0 = T$, and $T_t \rightarrow 0_{d \times d}$ as $t \rightarrow \infty$. Since $T_t^{-1} = e^t T^{-1} + 2(e^t - 1)\Sigma$, differentiating $T_t T_t^{-1} = I_d$ gives $(\frac{d}{dt} T_t) T_t^{-1} + T_t (\frac{d}{dt} T_t^{-1}) = 0_{d \times d}$, and thus

$$\frac{d}{dt} T_t = -T_t \left(\frac{d}{dt} T_t^{-1} \right) T_t = -T_t (T_t^{-1} + 2\Sigma) T_t = -(T_t + 2T_t \Sigma T_t).$$

It follows from (6.6) that

$$\frac{d}{dt} \log L(T_t) = \frac{DL(T_t)[\frac{d}{dt} T_t]}{L(T_t)} = \alpha \text{tr}(\Sigma T_t).$$

By bounded convergence, $L(T_t) \rightarrow 1$ as $t \rightarrow \infty$. Therefore,

$$\log L(T) = -\alpha \int_0^\infty \text{tr}(\Sigma T_t) dt.$$

Using the change of variable $u = 1 - e^{-t}$ ($dt = e^t du$), we obtain

$$\begin{aligned} \int_0^\infty \text{tr}(\Sigma T_t) dt &= \int_0^1 \text{tr}\{\Sigma(T^{-1} + 2u\Sigma)^{-1}\} du = \frac{1}{2} \int_0^1 \text{tr}\{(T^{-1} + 2u\Sigma)^{-1}(2\Sigma)\} du \\ &= \frac{1}{2} [\log |T^{-1} + 2u\Sigma|]_{u=0}^{u=1} = \frac{1}{2} \log \left(\frac{|T^{-1} + 2\Sigma|}{|T^{-1}|} \right) = \frac{1}{2} \log |I_d + 2T\Sigma|, \end{aligned}$$

where the third equality follows from Jacobi's formula, $\frac{d}{du} \log |A(u)| = \text{tr}\{A(u)^{-1} A'(u)\}$, with $A(u) = T^{-1} + 2u\Sigma$. Thus

$$L(T) = |I_d + 2T\Sigma|^{-\alpha/2}, \quad T \in \mathcal{S}_{++}^d.$$

By bounded convergence, this extends to every $T \in \mathcal{S}_+^d$. By Remark 2.1 and the uniqueness of Laplace transforms on $\mathcal{S}_+^d$, we get $\mathfrak{W} \stackrel{\text{law}}{=} \mathfrak{W}_\infty \sim \mathcal{W}_d(\alpha, \Sigma)$. This concludes the proof.

6.4. Proof of Lemma 3.4

Fix $t > 0$ and $W \in \mathcal{S}_{++}^d$. Note that $\alpha > 3d - 3$ by assumption, so $\alpha > d - 1$. Therefore, by Proposition 3.2, we have

$$(\mathcal{P}_t^w h)(W) = \mathbb{E}[h(\Sigma_t^{1/2} \mathfrak{S}_t \Sigma_t^{1/2})] = (\mathcal{Q}_t h_t)(\Lambda_t),$$

where the operator $\mathcal{Q}_t$ is defined in (3.15), $\mathfrak{S}_t \sim \mathcal{W}_d(\alpha, I_d, e^{-2t}\Lambda_t)$, and $\Lambda_t = \Sigma_t^{-1/2}W\Sigma_t^{-1/2}$. Also, if $\mathfrak{V} \sim \mathcal{W}_d(\alpha, I_d)$, then $(\mathcal{Q}_t h_t)(0) = \mathbb{E}[h(\Sigma_t^{1/2}\mathfrak{V}\Sigma_t^{1/2})]$ and $\mathbb{E}[h(\mathfrak{W}_\infty)] = \mathbb{E}[h(\Sigma^{1/2}\mathfrak{V}\Sigma^{1/2})]$. Hence,

$$\begin{aligned} |(\mathcal{P}_t^\mathcal{W} h)(W) - \mathbb{E}[h(\mathfrak{W}_\infty)]| &\leq |(\mathcal{Q}_t h_t)(\Lambda_t) - (\mathcal{Q}_t h_t)(0)| \\ &\quad + |\mathbb{E}[h(\Sigma_t^{1/2}\mathfrak{V}\Sigma_t^{1/2})] - \mathbb{E}[h(\Sigma^{1/2}\mathfrak{V}\Sigma^{1/2})]|. \end{aligned} \quad (6.7)$$

We begin by bounding the first term on the right-hand side of (6.7). The map $X \mapsto \Sigma_t^{1/2}X\Sigma_t^{1/2}$ is linear and maps $\mathcal{S}_{++}^d$ into $\mathcal{S}_{++}^d$, so $h_t \in C^d(\mathcal{S}_{++}^d)$. By the chain rule, each differential monomial of order at most d in the entries of ∇_X applied to h_t is a finite linear combination of differential monomials of the same order applied to h at $\Sigma_t^{1/2}X\Sigma_t^{1/2}$, with coefficients depending only on Σ_t . Since $\|\Sigma_t^{1/2}X\Sigma_t^{1/2}\|_F \leq \|\Sigma_t\|_2\|X\|_F$, the assumption (3.12) implies that h_t satisfies the regularity and growth assumptions required in Lemma 5.3. Let $\Lambda \in \mathcal{S}_{++}^d$ and $U \in \mathcal{S}^d$. By the standard identification between Fréchet derivatives and directional derivatives,

$$D(\mathcal{Q}_t h_t)(\Lambda)[U] = \langle U, \nabla_\Lambda \rangle_F (\mathcal{Q}_t h_t)(\Lambda) = \sum_{i,j=1}^d U_{ij} \nabla_{\Lambda, ij} (\mathcal{Q}_t h_t)(\Lambda).$$

By applying Lemma 5.3 with $k = 1$, we obtain

$$D(\mathcal{Q}_t h_t)(\Lambda)[U] = e^{-2t} \mathbb{E}[\langle U, \mathcal{D} \rangle_F h_t(\mathfrak{Y}_{1,t})],$$

where $\mathfrak{Y}_{1,t} \sim \mathcal{W}_d(\alpha + 2, I_d, e^{-2t}\Lambda)$. Therefore, from the definition of $M_t^\mathcal{D}(h)$ in (3.13), it follows that

$$|D(\mathcal{Q}_t h_t)(\Lambda)[U]| \leq e^{-2t} M_t^\mathcal{D}(h) \|U\|_F.$$

Now apply the fundamental theorem of calculus along the line segment $s \mapsto s\Lambda_t$, $0 \leq s \leq 1$, to obtain

$$(\mathcal{Q}_t h_t)(\Lambda_t) - (\mathcal{Q}_t h_t)(0) = \int_0^1 D(\mathcal{Q}_t h_t)(s\Lambda_t)[\Lambda_t] ds,$$

and hence

$$\begin{aligned} |(\mathcal{Q}_t h_t)(\Lambda_t) - (\mathcal{Q}_t h_t)(0)| &\leq e^{-2t} M_t^\mathcal{D}(h) \|\Lambda_t\|_F \\ &= \frac{e^{-2t}}{1 - e^{-2t}} M_t^\mathcal{D}(h) \|\Sigma^{-1/2}W\Sigma^{-1/2}\|_F. \end{aligned} \quad (6.8)$$

We now bound the second term on the right-hand side of (6.7). Given that $\Sigma_t^{1/2}\mathfrak{V}\Sigma_t^{1/2} = (1 - e^{-2t})\Sigma^{1/2}\mathfrak{V}\Sigma^{1/2}$, we have, by the Lipschitz property of h ,

$$\begin{aligned} |\mathbb{E}[h(\Sigma_t^{1/2}\mathfrak{V}\Sigma_t^{1/2})] - \mathbb{E}[h(\Sigma^{1/2}\mathfrak{V}\Sigma^{1/2})]| &\leq [h]_1 \mathbb{E}[\|\Sigma_t^{1/2}\mathfrak{V}\Sigma_t^{1/2} - \Sigma^{1/2}\mathfrak{V}\Sigma^{1/2}\|_F] \\ &= e^{-2t} [h]_1 \mathbb{E}[\|\Sigma^{1/2}\mathfrak{V}\Sigma^{1/2}\|_F] \\ &\leq e^{-2t} [h]_1 \|\Sigma^{1/2}\|_2^2 \mathbb{E}[\|\mathfrak{V}\|_F] \\ &= e^{-2t} [h]_1 \|\Sigma\|_2 \mathbb{E}[\|\mathfrak{V}\|_F], \end{aligned}$$

where $[h]_1$ denotes the minimum Lipschitz constant of h with respect to the Frobenius norm. Combining this bound with (6.7) and (6.8) proves (3.14). This concludes the proof.

6.5. Proof of Theorem 3.5

We adapt the standard Markov semigroup argument; the key input is the semigroup decay bound in Lemma 3.4.

Let $\widehat{h}(W) := \mathbb{E}[h(\mathfrak{W}_\infty)] - h(W)$ for $W \in \mathcal{S}_{++}^d$. Then we have $\|\widehat{h}\|_\infty \leq 2\|h\|_\infty$ and $[\widehat{h}]_1 = [h]_1$. Also, for $t > 0$, $\widehat{h}_t(X) := h(\Sigma_t^{1/2} X \Sigma_t^{1/2}) - h_t(X)$, and therefore, since $\mathcal{D}$ annihilates constants,

$$M_{1,\Sigma}^{\mathcal{D}}(\widehat{h}) = M_{1,\Sigma}^{\mathcal{D}}(h).$$

We first verify that the assumptions of Lemma 3.4 are available uniformly for $t \geq 1$. Since $h \in C_b^d(\mathcal{S}_{++}^d)$, the regularity and growth assumption (3.12) holds with $N_{\overline{\nabla}} = 0$. Fix a differential monomial $\overline{\nabla}$ of order $m \leq d$ in the entries of ∇_X . By the chain rule, $\overline{\nabla} h_t(X)$ is a finite linear combination of differential monomials of order m applied to h at $\Sigma_t^{1/2} X \Sigma_t^{1/2}$, with coefficients depending only on Σ_t . Since $\|\Sigma_t^{1/2}\|_2^2 = \|\Sigma_t\|_2 = (1 - e^{-2t})\|\Sigma\|_2 \leq \|\Sigma\|_2$ for $t \geq 1$, and since $h \in C_b^d(\mathcal{S}_{++}^d)$, there exists a finite constant $C_{\overline{\nabla},\Sigma}$ such that

$$|\overline{\nabla} h_t(X)| \leq C_{\overline{\nabla},\Sigma}, \quad X \in \mathcal{S}_{++}^d, \quad t \geq 1.$$

Moreover, each entry of $\mathcal{D} = \nabla_X \text{adj}(I_d - 2\nabla_X)$ is a finite linear combination, with constants depending only on d , of differential monomials of orders at most d in the entries of ∇_X . Hence

$$M_{1,\Sigma}^{\mathcal{D}}(h) < \infty.$$

The same conclusions hold with h replaced by $\widehat{h}$.

The function defined in (3.16) can thus be rewritten as

$$f_h(W) = \int_0^\infty (\mathcal{P}_t^{\mathcal{W}} \widehat{h})(W) dt.$$

Fix $W \in \mathcal{S}_{++}^d$. For $t \in (0, 1]$, we simply have

$$|(\mathcal{P}_t^{\mathcal{W}} \widehat{h})(W)| \leq \|\widehat{h}\|_\infty \leq 2\|h\|_\infty.$$

For $t \geq 1$, Lemma 3.4 applied to $\widehat{h}$ yields

$$\begin{aligned} |(\mathcal{P}_t^{\mathcal{W}} \widehat{h})(W)| &= |\mathbb{E}[h(\mathfrak{W}_\infty)] - (\mathcal{P}_t^{\mathcal{W}} h)(W)| \\ &\leq e^{-2t} \left\{ \frac{M_{1,\Sigma}^{\mathcal{D}}(h)}{1 - e^{-2}} \|\Sigma^{-1/2} W \Sigma^{-1/2}\|_F + [h]_1 \|\Sigma\|_2 \mathbb{E}[\|\mathfrak{W}\|_F] \right\}, \end{aligned}$$

where $\mathfrak{W} \sim \mathcal{W}_d(\alpha, I_d)$. Hence,

$$\begin{aligned} |f_h(W)| &\leq \int_0^1 |(\mathcal{P}_t^{\mathcal{W}} \widehat{h})(W)| dt + \int_1^\infty |(\mathcal{P}_t^{\mathcal{W}} \widehat{h})(W)| dt \\ &\leq 2\|h\|_\infty + \left(\int_1^\infty e^{-2t} dt \right) \left\{ \frac{M_{1,\Sigma}^{\mathcal{D}}(h)}{1 - e^{-2}} \|\Sigma^{-1/2} W \Sigma^{-1/2}\|_F + [h]_1 \|\Sigma\|_2 \mathbb{E}[\|\mathfrak{W}\|_F] \right\} \\ &\leq 2\|h\|_\infty + \frac{1}{2} \left\{ \frac{M_{1,\Sigma}^{\mathcal{D}}(h)}{1 - e^{-2}} \|\Sigma^{-1/2} W \Sigma^{-1/2}\|_F + [h]_1 \|\Sigma\|_2 \mathbb{E}[\|\mathfrak{W}\|_F] \right\}, \end{aligned}$$

which proves (3.18). In particular, f_h is well defined pointwise.

We now verify that f_h belongs to the domain of $\mathcal{A}^{\mathcal{W}}$ and satisfies the Wishart Stein equation (3.17). Fix $s > 0$ and $W \in \mathcal{S}_{++}^d$. By the Markov property, the semigroup property, and Fubini's theorem, we have

$$(\mathcal{P}_s^{\mathcal{W}} f_h)(W) = \mathbb{E} \left[\int_0^\infty (\mathcal{P}_t^{\mathcal{W}} \widehat{h})(\mathfrak{W}_s) dt \mid \mathfrak{W}_0 = W \right] = \int_0^\infty (\mathcal{P}_{t+s}^{\mathcal{W}} \widehat{h})(W) dt, \quad (6.9)$$

provided that

$$\int_0^\infty \mathbb{E} [|(\mathcal{P}_t^{\mathcal{W}} \widehat{h})(\mathfrak{W}_s)| \mid \mathfrak{W}_0 = W] dt < \infty.$$

This integrability is immediate on $(0, 1]$ from the bound $|(\mathcal{P}_t^{\mathcal{W}} \widehat{h})(\mathfrak{W}_s)| \leq 2\|h\|_\infty$. On $[1, \infty)$, since $\mathfrak{W}_s \in \mathcal{S}_{++}^d$ almost surely by Proposition 3.2, Lemma 3.4 gives

$$\begin{aligned} & \mathbb{E} [|(\mathcal{P}_t^{\mathcal{W}} \widehat{h})(\mathfrak{W}_s)| \mid \mathfrak{W}_0 = W] \\ & \leq e^{-2t} \left\{ \frac{M_{1,\Sigma}^{\mathcal{P}}(h)}{1 - e^{-2}} \mathbb{E} [\| \Sigma^{-1/2} \mathfrak{W}_s \Sigma^{-1/2} \|_F \mid \mathfrak{W}_0 = W] + [h]_1 \|\Sigma\|_2 \mathbb{E} [\|\mathfrak{W}\|_F] \right\}. \end{aligned}$$

It therefore remains to check that, for $\mathfrak{W}_s \mid \{\mathfrak{W}_0 = W\} \sim \mathcal{W}_d(\alpha, \Sigma_s, e^{-2s} \Sigma_s^{-1} W)$, we have

$$\mathbb{E} [\| \Sigma^{-1/2} \mathfrak{W}_s \Sigma^{-1/2} \|_F \mid \mathfrak{W}_0 = W] < \infty.$$

Since $\Sigma^{-1/2} \mathfrak{W}_s \Sigma^{-1/2} \in \mathcal{S}_+^d$, we have $\| \Sigma^{-1/2} \mathfrak{W}_s \Sigma^{-1/2} \|_F \leq \text{tr}(\Sigma^{-1/2} \mathfrak{W}_s \Sigma^{-1/2})$, and thus

$$\begin{aligned} \mathbb{E} [\| \Sigma^{-1/2} \mathfrak{W}_s \Sigma^{-1/2} \|_F \mid \mathfrak{W}_0 = W] & \leq \text{tr} (\Sigma^{-1/2} \mathbb{E} [\mathfrak{W}_s \mid \mathfrak{W}_0 = W] \Sigma^{-1/2}) \\ & = \alpha \text{tr}(\Sigma^{-1/2} \Sigma_s \Sigma^{-1/2}) + e^{-2s} \text{tr}(\Sigma^{-1/2} W \Sigma^{-1/2}) \\ & = \alpha d(1 - e^{-2s}) + e^{-2s} \text{tr}(\Sigma^{-1} W) < \infty. \end{aligned}$$

Therefore the application of Fubini's theorem is justified in (6.9).

Changing variables $u := t + s$ gives

$$(\mathcal{P}_s^{\mathcal{W}} f_h)(W) = \int_s^\infty (\mathcal{P}_u^{\mathcal{W}} \widehat{h})(W) du = f_h(W) - \int_0^s (\mathcal{P}_u^{\mathcal{W}} \widehat{h})(W) du.$$

Therefore,

$$\frac{(\mathcal{P}_s^{\mathcal{W}} f_h)(W) - f_h(W)}{s} = -\frac{1}{s} \int_0^s (\mathcal{P}_u^{\mathcal{W}} \widehat{h})(W) du. \quad (6.10)$$

Since $\widehat{h}$ is bounded and continuous, and since $\mathfrak{W}_u \rightarrow W$ almost surely as $u \downarrow 0$ by continuity of the Wishart process paths, dominated convergence yields

$$\lim_{u \downarrow 0} (\mathcal{P}_u^{\mathcal{W}} \widehat{h})(W) = \widehat{h}(W).$$

Letting $s \downarrow 0$ in (6.10) therefore gives

$$\mathcal{A}^{\mathcal{W}} f_h(W) = \lim_{s \downarrow 0} -\frac{1}{s} \int_0^s (\mathcal{P}_u^{\mathcal{W}} \widehat{h})(W) du = -\widehat{h}(W) = h(W) - \mathbb{E}[h(\mathfrak{W}_\infty)],$$

which is precisely (3.17). This completes the proof.

6.6. Proof of Theorem 3.6

For $t > 0$, write $\Lambda_t(W) = \Sigma_t^{-1/2} W \Sigma_t^{-1/2}$ and $h_t(X) = h(\Sigma_t^{1/2} X \Sigma_t^{1/2})$. By Proposition 3.2,

$$(\mathcal{P}_t^{\mathcal{W}} h)(W) = (\mathcal{Q}_t h_t)(\Lambda_t(W)), \quad W \in \mathcal{S}_{++}^d.$$

We first check that Lemma 5.3 can be applied with $k = m$ to the family $(h_t)_{t>0}$. Fix a differential monomial $\bar{\nabla}$ of order $q \leq md$ in the entries of ∇_X . By the chain rule, $\bar{\nabla} h_t(X)$ is a finite linear combination of differential monomials of order q applied to h at $\Sigma_t^{1/2} X \Sigma_t^{1/2}$, with coefficients depending only on Σ_t . Since $\|\Sigma_t\|_2 = (1 - e^{-2t})\|\Sigma\|_2 \leq \|\Sigma\|_2$ for $t > 0$, and since $h \in C_b^{md}(\mathcal{S}_{++}^d)$, there exists a finite constant $C_{\bar{\nabla}, \Sigma}$ such that

$$|\bar{\nabla} h_t(X)| \leq C_{\bar{\nabla}, \Sigma}, \quad X \in \mathcal{S}_{++}^d, \quad t > 0.$$

Thus the growth condition (5.13) holds for h_t with $N_{\bar{\nabla}} = 0$. Since $\alpha > 3d - 3$, Lemma 5.3 is applicable with $k = m$.

We begin with a formula for the directional derivatives of $(\mathcal{P}_t^{\mathcal{W}} h)(W)$. Let $U_1, \dots, U_m \in \mathcal{S}^d$, and let $\mathfrak{Y}_{m,t,W} \sim \mathcal{W}_d(\alpha + 2m, I_d, e^{-2t} \Lambda_t(W))$. Since Λ_t is linear,

$$D^m(\mathcal{P}_t^{\mathcal{W}} h)(W)[U_1, \dots, U_m] = D^m(\mathcal{Q}_t h_t)(\Lambda_t(W))[\Sigma_t^{-1/2} U_1 \Sigma_t^{-1/2}, \dots, \Sigma_t^{-1/2} U_m \Sigma_t^{-1/2}].$$

By the directional form of Lemma 5.3,

$$D^m(\mathcal{Q}_t h_t)(\Lambda)[V_1, \dots, V_m] = e^{-2mt} \mathbb{E} \left[\left(\prod_{\ell=1}^m \langle V_\ell, \mathcal{D} \rangle_F \right) h_t(\mathfrak{Y}_{m,t}^\Lambda) \right],$$

where $\mathfrak{Y}_{m,t}^\Lambda \sim \mathcal{W}_d(\alpha + 2m, I_d, e^{-2t} \Lambda)$ and $\mathcal{D} = \nabla_X \text{adj}(I_d - 2\nabla_X)$.

Fix $U \in \mathcal{S}^d$ and set $Y = \Sigma_t^{1/2} X \Sigma_t^{1/2}$. Since $\Sigma_t^{1/2}$ is constant and symmetric, the chain rule gives $\nabla_X = \Sigma_t^{1/2} \nabla_Y \Sigma_t^{1/2}$ when both sides act on $h_t(X) = h(\Sigma_t^{1/2} X \Sigma_t^{1/2})$. Hence

$$\mathcal{D} h_t(X) = (\Sigma_t^{1/2} \nabla_Y \Sigma_t^{1/2}) \text{adj}(I_d - 2\Sigma_t^{1/2} \nabla_Y \Sigma_t^{1/2}) h(Y).$$

Note that

$$\begin{aligned} \text{adj}(I_d - 2\Sigma_t^{1/2} \nabla_Y \Sigma_t^{1/2}) &= \text{adj}(\Sigma_t^{-1/2} (I_d - 2\Sigma_t \nabla_Y) \Sigma_t^{1/2}) \\ &= \text{adj}(\Sigma_t^{1/2}) \text{adj}(I_d - 2\Sigma_t \nabla_Y) \text{adj}(\Sigma_t^{-1/2}) \\ &= \Sigma_t^{-1/2} \text{adj}(I_d - 2\Sigma_t \nabla_Y) \Sigma_t^{1/2}, \end{aligned}$$

where the scalar determinant factors cancel in the last equality. Consequently,

$$\mathcal{D} h_t(X) = \Sigma_t^{1/2} \left\{ \nabla_Y \text{adj}(I_d - 2\Sigma_t \nabla_Y) h(Y) \right\} \Sigma_t^{1/2} = \Sigma_t^{1/2} (\mathcal{D}_{t,\Sigma} h)(Y) \Sigma_t^{1/2}.$$

Taking Frobenius inner products,

$$\langle \Sigma_t^{-1/2} U \Sigma_t^{-1/2}, \mathcal{D} \rangle_F h_t(X) = \langle \Sigma_t^{-1/2} U \Sigma_t^{-1/2}, \mathcal{D} h_t(X) \rangle_F = \langle U, (\mathcal{D}_{t,\Sigma} h)(Y) \rangle_F = \langle U, \mathcal{D}_{t,\Sigma} h(Y) \rangle_F.$$

Since all involved operators have constant coefficients and commute, iteration yields

$$\left(\prod_{\ell=1}^m \langle \Sigma_t^{-1/2} U_\ell \Sigma_t^{-1/2}, \mathcal{D} \rangle_F \right) h_t(X) = \left(\prod_{\ell=1}^m \langle U_\ell, \mathcal{D}_{t,\Sigma} \rangle_F \right) h(\Sigma_t^{1/2} X \Sigma_t^{1/2}).$$

If we now define $\mathfrak{Z}_{m,t,W} := \Sigma_t^{1/2} \mathfrak{Y}_{m,t,W} \Sigma_t^{1/2}$, then

$$D^m(\mathcal{P}_t^{\mathcal{W}}h)(W)[U_1, \dots, U_m] = e^{-2mt} \mathbb{E} \left[\left(\prod_{\ell=1}^m \langle U_\ell, \mathcal{D}_{t,\Sigma} \rangle_F \right) h(\mathfrak{Z}_{m,t,W}) \right]. \quad (6.11)$$

We now prove (3.19). Fix index pairs $(i_1, j_1), \dots, (i_m, j_m) \in [d]^2$. For $1 \leq i, j \leq d$, define

$$H^{(ij)} := \frac{1}{2}(\mathbf{e}_i \mathbf{e}_j^\top + \mathbf{e}_j \mathbf{e}_i^\top) \in \mathcal{S}^d.$$

The matrix $\mathcal{D}_{t,\Sigma}$ is symmetric as a matrix of commuting differential operators. Indeed, with $A = \nabla_W$ and $B = 2\Sigma_t$, the identity $A \operatorname{adj}(I_d - BA) = \operatorname{adj}(I_d - AB)A$ gives

$$\mathcal{D}_{t,\Sigma} = \nabla_W \operatorname{adj}(I_d - 2\Sigma_t \nabla_W) = \operatorname{adj}(I_d - 2\nabla_W \Sigma_t) \nabla_W,$$

and taking transposes gives $\mathcal{D}_{t,\Sigma}^\top = \mathcal{D}_{t,\Sigma}$. By the symmetric-gradient convention,

$$\langle H^{(ij)}, \nabla_W \rangle_F = \nabla_{W,ij} \quad \text{and} \quad \langle H^{(ij)}, \mathcal{D}_{t,\Sigma} \rangle_F = (\mathcal{D}_{t,\Sigma})_{ij}.$$

Applying (6.11) with $U_\ell = H^{(i_\ell j_\ell)}$, we obtain

$$\left(\prod_{\ell=1}^m \nabla_{W,i_\ell j_\ell} \right) (\mathcal{P}_t^{\mathcal{W}}h)(W) = e^{-2mt} \mathbb{E} \left[\left(\prod_{\ell=1}^m (\mathcal{D}_{t,\Sigma})_{i_\ell j_\ell} \right) h(\mathfrak{Z}_{m,t,W}) \right].$$

Hence

$$\left| \left(\prod_{\ell=1}^m \nabla_{W,i_\ell j_\ell} \right) (\mathcal{P}_t^{\mathcal{W}}h)(W) \right| \leq e^{-2mt} \sup_{s>0} \left\| \left(\prod_{\ell=1}^m (\mathcal{D}_{s,\Sigma})_{i_\ell j_\ell} \right) h \right\|_\infty.$$

Since $\mathbb{E}[h(\mathfrak{W}_\infty)]$ is constant in W , its derivatives vanish. The right-hand side is integrable in t , so dominated convergence allows differentiation under the integral sign in (3.16), giving

$$\left(\prod_{\ell=1}^m \nabla_{W,i_\ell j_\ell} \right) f_h(W) = - \int_0^\infty \left(\prod_{\ell=1}^m \nabla_{W,i_\ell j_\ell} \right) (\mathcal{P}_t^{\mathcal{W}}h)(W) dt.$$

Therefore,

$$\begin{aligned} \left| \left(\prod_{\ell=1}^m \nabla_{W,i_\ell j_\ell} \right) f_h(W) \right| &\leq \int_0^\infty e^{-2mt} dt \sup_{s>0} \left\| \left(\prod_{\ell=1}^m (\mathcal{D}_{s,\Sigma})_{i_\ell j_\ell} \right) h \right\|_\infty \\ &= \frac{1}{2m} \sup_{s>0} \left\| \left(\prod_{\ell=1}^m (\mathcal{D}_{s,\Sigma})_{i_\ell j_\ell} \right) h \right\|_\infty, \end{aligned}$$

which is exactly (3.19).

Next, we prove (3.22). Assume that $\mathcal{M}_m^{\mathcal{D},\Sigma}(h) < \infty$. For unit vectors $U_1, \dots, U_m \in \mathcal{S}^d$, (6.11) gives

$$|D^m(\mathcal{P}_t^{\mathcal{W}}h)(W)[U_1, \dots, U_m]| \leq e^{-2mt} \mathcal{M}_m^{\mathcal{D},\Sigma}(h).$$

Again, the right-hand side is integrable in t , so dominated convergence yields

$$D^m f_h(W)[U_1, \dots, U_m] = - \int_0^\infty D^m(\mathcal{P}_t^{\mathcal{W}}h)(W)[U_1, \dots, U_m] dt.$$

Hence, whenever $\|U_1\|_F = \dots = \|U_m\|_F = 1$,

$$|D^m f_h(W)[U_1, \dots, U_m]| \leq \int_0^\infty e^{-2mt} dt \mathcal{M}_m^{\mathcal{D}, \Sigma}(h) = \frac{1}{2m} \mathcal{M}_m^{\mathcal{D}, \Sigma}(h).$$

Taking the suprema over $W \in \mathcal{S}_{++}^d$ and over $\|U_1\|_F = \dots = \|U_m\|_F = 1$ yields (3.22). This completes the proof.

7. Proofs of the application results

7.1. Proof of Lemma 4.1

The set $\mathcal{R}_{\nu, d}$ is open because it is the inverse image of $\mathcal{S}_{++}^d$ under the continuous map $X \mapsto X^\top X$. Its complement has Lebesgue measure zero because it is contained in the zero set of any fixed $d \times d$ minor, viewed as a nonzero polynomial on $\mathbb{R}^{\nu \times d}$. Let $X \in \mathcal{R}_{\nu, d}$, and set $S := X^\top X$. By the chain rule, $\nabla_X g_f(X) = 2X \nabla f(S)$, so the drift part of (4.1) is

$$-\text{tr}\{X^\top \nabla_X g_f(X)\} = -2 \text{tr}\{X^\top X \nabla f(S)\} = -2 \text{tr}\{S \nabla f(S)\}.$$

It remains to compute the diffusion part of (4.1). In coordinates, $\nabla_X g_f(X) = 2X \nabla f(S)$ gives

$$\nabla_{X, ar} g_f(X) = 2 \sum_{u=1}^d X_{au} \nabla_{ur} f(S).$$

Differentiating once more with respect to X_{as} yields

$$\nabla_{X, as} \nabla_{X, ar} g_f(X) = 2 \nabla_{sr} f(S) + 2 \sum_{u=1}^d X_{au} \nabla_{X, as} \nabla_{ur} f(S).$$

Since $S_{pq} = \sum_{b=1}^\nu X_{bp} X_{bq}$, we have $\nabla_{X, as} S_{pq} = \delta_{ps} X_{aq} + \delta_{qs} X_{ap}$. Using the symmetric-gradient convention in the S variable, it follows that

$$\nabla_{X, as} \nabla_{ur} f(S) = \sum_{p, q=1}^d (\delta_{ps} X_{aq} + \delta_{qs} X_{ap}) \nabla_{pq} \nabla_{ur} f(S) = 2 \sum_{v=1}^d X_{av} \nabla_{sv} \nabla_{ur} f(S).$$

Consequently,

$$\nabla_{X, as} \nabla_{X, ar} g_f(X) = 2 \nabla_{sr} f(S) + 4 \sum_{u, v=1}^d X_{au} X_{av} \nabla_{sv} \nabla_{ur} f(S).$$

Substituting this identity into the second term of (4.1), we obtain

$$\begin{aligned} \text{tr}\{\Sigma \nabla_X^\top \nabla_X g_f(X)\} &= \sum_{a=1}^\nu \sum_{r, s=1}^d \Sigma_{rs} \nabla_{X, ar} \nabla_{X, as} g_f(X) \\ &= 2\nu \text{tr}\{\Sigma \nabla f(S)\} + 4 \sum_{u, v=1}^d S_{uv} \sum_{r, s=1}^d \Sigma_{rs} \nabla_{sv} \nabla_{ur} f(S) \\ &= 2\nu \text{tr}\{\Sigma \nabla f(S)\} + 4 \text{tr}\{S \nabla \Sigma \nabla f(S)\}. \end{aligned}$$

Combining the drift and diffusion terms gives, for $X \in \mathcal{R}_{\nu, d}$,

$$\mathcal{A}^{\text{OU}} g_f(X) = 2 \text{tr}\{(\nu \Sigma - X^\top X) \nabla f(X^\top X)\} + 4 \text{tr}\{X^\top X \nabla \Sigma \nabla f(X^\top X)\},$$

which is (4.2).

7.2. Proof of Proposition 4.2

Let $\mathfrak{N} \sim \mathcal{N}_{\nu \times d}(0_{\nu \times d}, I_{\nu} \otimes \Sigma)$ and note that $\mathfrak{W} \stackrel{\text{law}}{=} \mathfrak{N}^{\top} \mathfrak{N} \sim \mathcal{W}_d(\nu, \Sigma)$. Let f_h denote the solution of the Wishart Stein equation associated with h . Throughout the proof, we write $\mathcal{I} := [\nu] \times [d]$ and, for $i \in [n]$, $A, B \in \mathcal{I}$ and $\theta \in [0, 1]$, set $Z_{i,A} := (\mathfrak{Z}_i)_A$, $Z_A := Z_{1,A}$, $\sigma_{A,B} := \mathbb{E}[Z_A Z_B]$, and

$$\mathfrak{X}_{-i}^{(n)} := \mathfrak{X}^{(n)} - \frac{1}{\sqrt{n}} \mathfrak{Z}_i, \quad \mathfrak{X}_{i,\theta}^{(n)} := \mathfrak{X}_{-i}^{(n)} + \frac{\theta}{\sqrt{n}} \mathfrak{Z}_i.$$

For $\varepsilon > 0$, set

$$g_{h,\varepsilon}(X) := f_h(X^{\top} X + \varepsilon I_d), \quad S_{\varepsilon}(X) := X^{\top} X + \varepsilon I_d, \quad X \in \mathbb{R}^{\nu \times d}.$$

Since $S_{\varepsilon}(\mathfrak{X}^{(n)}) \in \mathcal{S}_{++}^d$ almost surely, the Wishart Stein equation and (4.3) give

$$\begin{aligned} \mathbb{E}[h(S_{\varepsilon}(\mathfrak{X}^{(n)}))] - \mathbb{E}[h(\mathfrak{W})] &= \mathbb{E}[(\mathcal{A}_{\nu,\Sigma}^{\mathcal{W}} f_h)(S_{\varepsilon}(\mathfrak{X}^{(n)}))] \\ &= \mathbb{E}[\mathcal{A}^{\text{OU}} g_{h,\varepsilon}(\mathfrak{X}^{(n)})] - 2\varepsilon \mathbb{E}[\text{tr}\{\nabla f_h(S_{\varepsilon}(\mathfrak{X}^{(n)}))\}] \\ &\quad + 4\varepsilon \mathbb{E}[\text{tr}\{I_d \nabla \Sigma \nabla f_h(S_{\varepsilon}(\mathfrak{X}^{(n)}))\}]. \end{aligned}$$

Moreover, by the assumption $\sum_{m=1}^5 \mathcal{M}_m^{\mathcal{D},\Sigma}(h) < \infty$ and (3.22) in Theorem 3.6, applied with $m \in \{1, 2\}$, we have $\mathcal{M}_1(f_h) < \infty$ and $\mathcal{M}_2(f_h) < \infty$. Therefore, there is a finite constant C_h , independent of n and $\varepsilon \in (0, 1]$, such that

$$|\mathbb{E}[h(S_{\varepsilon}(\mathfrak{X}^{(n)}))] - \mathbb{E}[h(\mathfrak{W})] - \mathbb{E}[\mathcal{A}^{\text{OU}} g_{h,\varepsilon}(\mathfrak{X}^{(n)})]| \leq C_h \varepsilon. \quad (7.1)$$

Fix $\varepsilon \in (0, 1]$. Using the coordinate form of the matrix normal Ornstein–Uhlenbeck operator, we have

$$\mathbb{E}[\mathcal{A}^{\text{OU}} g_{h,\varepsilon}(\mathfrak{X}^{(n)})] = \mathbb{E} \left[\sum_{A,B \in \mathcal{I}} \sigma_{A,B} \partial_A \partial_B g_{h,\varepsilon}(\mathfrak{X}^{(n)}) \right] - \mathbb{E} \left[\sum_{A \in \mathcal{I}} (\mathfrak{X}^{(n)})_A \partial_A g_{h,\varepsilon}(\mathfrak{X}^{(n)}) \right]. \quad (7.2)$$

We next record derivative bounds for $g_{h,\varepsilon}$ that are uniform in ε . We have

$$D(X^{\top} X + \varepsilon I_d)[U] = X^{\top} U + U^{\top} X, \quad D^2(X^{\top} X + \varepsilon I_d)[U, V] = U^{\top} V + V^{\top} U,$$

and $D^k(X^{\top} X + \varepsilon I_d) = 0$ for $k \geq 3$. Hence, if $\|U_1\|_F = \dots = \|U_k\|_F = 1$, then

$$\|D(X^{\top} X + \varepsilon I_d)[U_j]\|_F \leq 2\|X\|_F, \quad \|D^2(X^{\top} X + \varepsilon I_d)[U_i, U_j]\|_F \leq 2.$$

By Faà di Bruno's formula, $D^k g_{h,\varepsilon}(X)$ is a sum over partitions of $\{1, \dots, k\}$ into blocks of size one and two. If ℓ is the number of two-blocks, then the number of such partitions is $k!/(2^{\ell} \ell! (k-2\ell)!)$, and the corresponding term is bounded by $2^{k-\ell} \|X\|_F^{k-2\ell} \mathcal{M}_{k-\ell}(f_h)$. Consequently, for

$$\mathcal{G}_{k,h}(x) := \sum_{\ell=0}^{\lfloor k/2 \rfloor} \frac{2^{k-2\ell} k!}{\ell! (k-2\ell)!} x^{k-2\ell} \mathcal{M}_{k-\ell}(f_h), \quad (7.3)$$

we have

$$\sup_{\substack{U_1, \dots, U_k \in \mathbb{R}^{\nu \times d} \\ \|U_1\|_F = \dots = \|U_k\|_F = 1}} |D^k g_{h,\varepsilon}(X)[U_1, \dots, U_k]| \leq \mathcal{G}_{k,h}(\|X\|_F). \quad (7.4)$$

We now expand the two terms in (7.2). Since $\mathfrak{X}_{-i}^{(n)}$ and $\mathfrak{Z}_i$ are independent, $\mathbb{E}[Z_{i,A}] = 0$, and $\mathbb{E}[Z_{i,A}Z_{i,B}] = \sigma_{A,B}$, a Taylor expansion of $\partial_A g_{h,\varepsilon}$ around $\mathfrak{X}_{-i}^{(n)}$ gives

$$\begin{aligned} \sum_{A \in \mathcal{I}} \mathbb{E}[(\mathfrak{X}^{(n)})_A \partial_A g_{h,\varepsilon}(\mathfrak{X}^{(n)})] &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{A \in \mathcal{I}} \mathbb{E} \left[Z_{i,A} \partial_A g_{h,\varepsilon} \left(\mathfrak{X}_{-i}^{(n)} + \frac{1}{\sqrt{n}} \mathfrak{Z}_i \right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{A,B \in \mathcal{I}} \sigma_{A,B} \mathbb{E}[\partial_A \partial_B g_{h,\varepsilon}(\mathfrak{X}_{-i}^{(n)})] \\ &\quad + \frac{1}{2n^{3/2}} \sum_{i=1}^n \sum_{A,B,C \in \mathcal{I}} \mathbb{E}[Z_A Z_B Z_C] \mathbb{E}[\partial_A \partial_B \partial_C g_{h,\varepsilon}(\mathfrak{X}_{-i}^{(n)})] + R_{1,\varepsilon}, \end{aligned}$$

where, by (7.4),

$$|R_{1,\varepsilon}| \leq \frac{1}{6n} \sum_{A,B,C,D \in \mathcal{I}} \mathbb{E} \left[|Z_A Z_B Z_C Z_D| \mathcal{G}_{4,h}(\|\mathfrak{X}_{1,\theta_1}^{(n)}\|_F) \right],$$

for some $\theta_1 \in [0, 1]$. Similarly, a Taylor expansion of $\partial_A \partial_B g_{h,\varepsilon}$ around $\mathfrak{X}_{-i}^{(n)}$ gives

$$\sum_{A,B \in \mathcal{I}} \sigma_{A,B} \mathbb{E}[\partial_A \partial_B g_{h,\varepsilon}(\mathfrak{X}^{(n)})] = \frac{1}{n} \sum_{i=1}^n \sum_{A,B \in \mathcal{I}} \sigma_{A,B} \mathbb{E}[\partial_A \partial_B g_{h,\varepsilon}(\mathfrak{X}_{-i}^{(n)})] + R_{2,\varepsilon},$$

where, by (7.4),

$$|R_{2,\varepsilon}| \leq \frac{1}{2n} \sum_{A,B,C,D \in \mathcal{I}} |\sigma_{A,B}| \mathbb{E} \left[|Z_C Z_D| \mathcal{G}_{4,h}(\|\mathfrak{X}_{1,\theta_2}^{(n)}\|_F) \right],$$

for some $\theta_2 \in [0, 1]$. Also, a Taylor expansion of $\partial_A \partial_B \partial_C g_{h,\varepsilon}$ around $\mathfrak{X}^{(n)}$ gives

$$\begin{aligned} \frac{1}{2n^{3/2}} \sum_{i=1}^n \sum_{A,B,C \in \mathcal{I}} \mathbb{E}[Z_A Z_B Z_C] \mathbb{E}[\partial_A \partial_B \partial_C g_{h,\varepsilon}(\mathfrak{X}_{-i}^{(n)})] \\ = \frac{1}{2\sqrt{n}} \sum_{A,B,C \in \mathcal{I}} \mathbb{E}[Z_A Z_B Z_C] \mathbb{E}[\partial_A \partial_B \partial_C g_{h,\varepsilon}(\mathfrak{X}^{(n)})] + R_{3,\varepsilon}, \end{aligned}$$

where, by (7.4) and the assumption that $\mathfrak{Z}_1, \dots, \mathfrak{Z}_n$ are identically distributed,

$$|R_{3,\varepsilon}| \leq \frac{1}{2n} \sum_{A,B,C,D \in \mathcal{I}} |\mathbb{E}[Z_A Z_B Z_C]| \mathbb{E} \left[|Z_D| \mathcal{G}_{4,h}(\|\mathfrak{X}_{1,\theta_3}^{(n)}\|_F) \right],$$

for some $\theta_3 \in [0, 1]$. Combining the preceding displays yields

$$\begin{aligned} |\mathbb{E}[\mathcal{A}^{\text{OU}} g_{h,\varepsilon}(\mathfrak{X}^{(n)})]| &\leq |R_{1,\varepsilon}| + |R_{2,\varepsilon}| + |R_{3,\varepsilon}| \\ &\quad + \frac{1}{2\sqrt{n}} \sum_{A,B,C \in \mathcal{I}} |\mathbb{E}[Z_A Z_B Z_C]| |\mathbb{E}[\partial_A \partial_B \partial_C g_{h,\varepsilon}(\mathfrak{X}^{(n)})]|. \end{aligned} \quad (7.5)$$

It remains to bound the third-derivative term using symmetry considerations. Fix $A, B, C \in \mathcal{I}$. The function $g_{h,\varepsilon}$ is even because $g_{h,\varepsilon}(-Y) = g_{h,\varepsilon}(Y)$. Hence $\partial_A \partial_B \partial_C g_{h,\varepsilon}$ is odd. Since $\mathfrak{N}$ is centered matrix normal,

$$\mathbb{E}[\partial_A \partial_B \partial_C g_{h,\varepsilon}(\mathfrak{N})] = 0.$$

By (7.4), the function $Y \mapsto \partial_A \partial_B \partial_C g_{h,\varepsilon}(Y)$ has polynomial growth. We use Theorem 3.4 of [Gaunt et al. \(2026\)](#) through smooth compactly supported truncations: the theorem is first applied to the truncated functions and then the truncation is removed by dominated convergence, using (7.4) and the finiteness of Gaussian moments. This yields a solution $\Psi_{A,B,C}^{(\varepsilon)}$ of the matrix normal Stein equation

$$\sum_{D,E \in \mathcal{I}} \sigma_{D,E} \partial_D \partial_E \Psi_{A,B,C}^{(\varepsilon)}(Y) - \sum_{D \in \mathcal{I}} Y_D \partial_D \Psi_{A,B,C}^{(\varepsilon)}(Y) = \partial_A \partial_B \partial_C g_{h,\varepsilon}(Y), \quad Y \in \mathbb{R}^{\nu \times d}.$$

Equivalently, $\Psi_{A,B,C}^{(\varepsilon)}$ is the semigroup solution

$$\Psi_{A,B,C}^{(\varepsilon)}(Y) = - \int_0^\infty \mathbb{E} \left[\partial_A \partial_B \partial_C g_{h,\varepsilon} \left(e^{-t} Y + \sqrt{1 - e^{-2t}} \mathfrak{Y}' \right) \right] dt,$$

where $\mathfrak{Y}'$ is an independent copy of $\mathfrak{Y}$. The same domination justifies the differentiations under the integral sign below.

Arguing as we did to derive the bound (7.5), but expanding only up to third-order partial derivatives, we obtain the following:

$$|\mathbb{E}[\partial_A \partial_B \partial_C g_{h,\varepsilon}(\mathfrak{X}^{(n)})]| \leq |R_{4,\varepsilon}^{A,B,C}| + |R_{5,\varepsilon}^{A,B,C}|, \quad (7.6)$$

where

$$|R_{4,\varepsilon}^{A,B,C}| \leq \frac{1}{2\sqrt{n}} \sum_{D,E,F \in \mathcal{I}} \mathbb{E} \left[|Z_D Z_E Z_F| |\partial_D \partial_E \partial_F \Psi_{A,B,C}^{(\varepsilon)}(\mathfrak{X}_{1,\theta_4}^{(n)})| \right], \quad (7.7)$$

$$|R_{5,\varepsilon}^{A,B,C}| \leq \frac{1}{\sqrt{n}} \sum_{D,E,F \in \mathcal{I}} |\sigma_{D,E}| \mathbb{E} \left[|Z_F| |\partial_D \partial_E \partial_F \Psi_{A,B,C}^{(\varepsilon)}(\mathfrak{X}_{1,\theta_5}^{(n)})| \right], \quad (7.8)$$

for some $\theta_4, \theta_5 \in [0, 1]$.

Differentiating twice under the integral sign and then using Gaussian integration by parts to take the third derivative (using a similar calculation to the one used in the proof of inequality (3.14) of [Gaunt et al. \(2026\)](#)) gives

$$\begin{aligned} & \partial_D \partial_E \partial_F \Psi_{A,B,C}^{(\varepsilon)}(Y) \\ &= - \int_0^\infty \frac{e^{-3t}}{\sqrt{1 - e^{-2t}}} \sum_{Q \in \mathcal{I}} (I_\nu \otimes \Sigma^{-1})_{F,Q} \mathbb{E}[(\mathfrak{Y}')_Q \partial_A \partial_B \partial_C \partial_D \partial_E g_{h,\varepsilon}(e^{-t} Y + \sqrt{1 - e^{-2t}} \mathfrak{Y}')] dt. \end{aligned}$$

Note that $\int_0^\infty e^{-3t} (1 - e^{-2t})^{-1/2} dt = \pi/4$. Also, since $e^{-t} \leq 1$ and $\sqrt{1 - e^{-2t}} \leq 1$, we have

$$\|e^{-t} Y + \sqrt{1 - e^{-2t}} \mathfrak{Y}'\|_F^p \leq (\|Y\|_F + \|\mathfrak{Y}'\|_F)^p \leq (1 + \|Y\|_F)^p (1 + \|\mathfrak{Y}'\|_F)^p, \quad p \geq 0.$$

Consequently, we have the bound

$$|\partial_D \partial_E \partial_F \Psi_{A,B,C}^{(\varepsilon)}(Y)| \leq M \mathcal{G}_{5,h}(1 + \|Y\|_F),$$

where $\|\Sigma^{-1}\|_\infty = \|(I_\nu \otimes \Sigma^{-1})\|_\infty \equiv \max_{F \in \mathcal{I}} \sum_{Q \in \mathcal{I}} |(I_\nu \otimes \Sigma^{-1})_{F,Q}|$, and

$$M := \frac{\pi}{4} \|\Sigma^{-1}\|_\infty \max_{\substack{p \in \{1,3,5\} \\ Q \in \mathcal{I}}} \mathbb{E} [|(\mathfrak{Y}')_Q| (1 + \|\mathfrak{Y}'\|_F)^p].$$

It follows from (7.7) and (7.8) that

$$|R_{4,\varepsilon}^{A,B,C}| + |R_{5,\varepsilon}^{A,B,C}| \leq \frac{M}{\sqrt{n}} \sum_{D,E,F \in \mathcal{I}} \left\{ \frac{1}{2} \mathbb{E} \left[|Z_D Z_E Z_F| \mathcal{G}_{5,h}(1 + \|\mathfrak{X}_{1,\theta_4}^{(n)}\|_F) \right] \right. \\ \left. + |\sigma_{D,E}| \mathbb{E} \left[|Z_F| \mathcal{G}_{5,h}(1 + \|\mathfrak{X}_{1,\theta_5}^{(n)}\|_F) \right] \right\}.$$

We shall use the following explicit upper bound on M . Let $\mathfrak{G} \in \mathbb{R}^{\nu \times d}$ have iid standard normal entries. Since $\mathfrak{N}' \stackrel{\text{law}}{=} \mathfrak{G} \Sigma^{1/2}$, we have $\|\mathfrak{N}'\|_F \leq \|\Sigma^{1/2}\|_2 \|\mathfrak{G}\|_F = \|\Sigma\|_2^{1/2} \|\mathfrak{G}\|_F$. Also, for every $Q \in \mathcal{I}$ and every $p \in \{1, 3, 5\}$,

$$|(\mathfrak{N}')_Q| (1 + \|\mathfrak{N}'\|_F)^p \leq \|\mathfrak{N}'\|_F (1 + \|\mathfrak{N}'\|_F)^p \leq (1 + \|\mathfrak{N}'\|_F)^6 \leq 2^5 \{1 + \|\mathfrak{N}'\|_F^6\}.$$

Since $\|\mathfrak{G}\|_F^2 \sim \chi_{\nu d}^2$, it follows that $\mathbb{E}[\|\mathfrak{G}\|_F^6] = \nu d(\nu d + 2)(\nu d + 4)$, and thus

$$M \leq 8\pi \|\Sigma^{-1}\|_\infty \{1 + \|\Sigma\|_2^3 \nu d(\nu d + 2)(\nu d + 4)\} \equiv K.$$

Combining (7.5) and (7.6), we have the following bound:

$$|\mathbb{E}[\mathcal{A}^{\text{OU}}_{g_{h,\varepsilon}}(\mathfrak{X}^{(n)})]| \leq |R_{1,\varepsilon}| + |R_{2,\varepsilon}| + |R_{3,\varepsilon}| + \frac{1}{2\sqrt{n}} \sum_{A,B,C \in \mathcal{I}} |\mathbb{E}[Z_A Z_B Z_C]| (|R_{4,\varepsilon}^{A,B,C}| + |R_{5,\varepsilon}^{A,B,C}|).$$

We now find bounds on $|R_{1,\varepsilon}|$, $|R_{2,\varepsilon}|$, $|R_{3,\varepsilon}|$, $|R_{4,\varepsilon}^{A,B,C}|$, and $|R_{5,\varepsilon}^{A,B,C}|$ that do not involve $\theta_1, \dots, \theta_5$ or moments of $\mathfrak{X}_{-1}^{(n)}$. For $q \in \{0, \dots, 8\}$, set

$$\mu_q := \mathbb{E}[\|\mathfrak{Z}_1\|_F^q], \quad \Lambda_q := \begin{cases} 1, & q = 0, \\ \mu_2^{1/2}, & q = 1, \\ 2^q(q-1)^{q/2} \mu_q, & 2 \leq q \leq 8. \end{cases}$$

We claim that, for $0 \leq q \leq 8$,

$$\mathbb{E}[\|\mathfrak{X}_{-1}^{(n)}\|_F^q] \leq \Lambda_q. \quad (7.9)$$

The case $q = 0$ is immediate. For $q = 1$, by the Cauchy-Schwarz inequality and the independence and centering of $\mathfrak{Z}_2, \dots, \mathfrak{Z}_n$,

$$\mathbb{E}[\|\mathfrak{X}_{-1}^{(n)}\|_F] \leq \left(\mathbb{E}[\|\mathfrak{X}_{-1}^{(n)}\|_F^2] \right)^{1/2} = \left(\frac{1}{n} \sum_{i=2}^n \mathbb{E}[\|\mathfrak{Z}_i\|_F^2] \right)^{1/2} = \left(\frac{n-1}{n} \mu_2 \right)^{1/2} \leq \mu_2^{1/2}.$$

Now let $2 \leq q \leq 8$. By symmetrization and Khintchine's inequality, for independent Rademacher random variables $\varepsilon_2, \dots, \varepsilon_n$, independent of $\mathfrak{Z}_2, \dots, \mathfrak{Z}_n$, we have

$$\begin{aligned} \mathbb{E}[\|\mathfrak{X}_{-1}^{(n)}\|_F^q] &= \frac{1}{n^{q/2}} \mathbb{E} \left[\left\| \sum_{i=2}^n \mathfrak{Z}_i \right\|_F^q \right] \leq \frac{2^q}{n^{q/2}} \mathbb{E} \left[\left\| \sum_{i=2}^n \varepsilon_i \mathfrak{Z}_i \right\|_F^q \right] \\ &\leq \frac{2^q(q-1)^{q/2}}{n^{q/2}} \mathbb{E} \left[\left(\sum_{i=2}^n \|\mathfrak{Z}_i\|_F^2 \right)^{q/2} \right] \leq 2^q(q-1)^{q/2} \mu_q, \end{aligned}$$

where we applied Jensen's inequality to obtain the last bound. This proves (7.9).

For $r \in \mathbb{N}$ and $q \in \mathbb{N}_0$ with $r + q \leq 8$, define

$$\mathcal{L}_{q,r} := \begin{cases} \mu_r, & q = 0, \\ 2^{q-1} \{\Lambda_q \mu_r + \mu_{r+q}\}, & q \geq 1, \end{cases} \quad \mathcal{L}_{q,r}^+ := \begin{cases} \mu_r, & q = 0, \\ 2^{q-1} \{2^{q-1}(1 + \Lambda_q) \mu_r + \mu_{r+q}\}, & q \geq 1. \end{cases}$$

Indeed, since $\mathfrak{X}_{1,\theta}^{(n)} = \mathfrak{X}_{-1}^{(n)} + \theta n^{-1/2} \mathfrak{Z}_1$, the independence of $\mathfrak{X}_{-1}^{(n)}$ and $\mathfrak{Z}_1$ gives, for $q \geq 1$ and $A_1, \dots, A_r \in \mathcal{I}$,

$$\begin{aligned} \mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) \|\mathfrak{X}_{1,\theta}^{(n)}\|_F^q \right] &\leq 2^{q-1} \mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) \left\{ \|\mathfrak{X}_{-1}^{(n)}\|_F^q + \frac{\|\mathfrak{Z}_1\|_F^q}{n^{q/2}} \right\} \right] \\ &\leq 2^{q-1} \left\{ \Lambda_q \mathbb{E} \left[\prod_{\ell=1}^r |Z_{A_\ell}| \right] + \mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) \|\mathfrak{Z}_1\|_F^q \right] \right\} \\ &\leq 2^{q-1} \{\Lambda_q \mu_r + \mu_{r+q}\}. \end{aligned}$$

For $q = 0$, we simply use $\mathbb{E}[\prod_{\ell=1}^r |Z_{A_\ell}|] \leq \mu_r$. Thus, for $\theta \in [0, 1]$,

$$\mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) \|\mathfrak{X}_{1,\theta}^{(n)}\|_F^q \right] \leq \mathcal{L}_{q,r}, \quad r + q \leq 8. \quad (7.10)$$

Similarly, since $1 + \|\mathfrak{X}_{1,\theta}^{(n)}\|_F \leq 1 + \|\mathfrak{X}_{-1}^{(n)}\|_F + \|\mathfrak{Z}_1\|_F/\sqrt{n}$, we have, for $q \geq 1$,

$$\begin{aligned} \mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) (1 + \|\mathfrak{X}_{1,\theta}^{(n)}\|_F)^q \right] &\leq 2^{q-1} \mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) \left\{ (1 + \|\mathfrak{X}_{-1}^{(n)}\|_F)^q + \frac{\|\mathfrak{Z}_1\|_F^q}{n^{q/2}} \right\} \right] \\ &\leq 2^{q-1} \left\{ 2^{q-1}(1 + \Lambda_q) \mathbb{E} \left[\prod_{\ell=1}^r |Z_{A_\ell}| \right] + \mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) \|\mathfrak{Z}_1\|_F^q \right] \right\} \\ &\leq 2^{q-1} \{2^{q-1}(1 + \Lambda_q) \mu_r + \mu_{r+q}\}. \end{aligned}$$

Consequently, for $\theta \in [0, 1]$,

$$\mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) (1 + \|\mathfrak{X}_{1,\theta}^{(n)}\|_F)^q \right] \leq \mathcal{L}_{q,r}^+, \quad q \geq 0, \quad r + q \leq 8. \quad (7.11)$$

For $r \in \{1, 2, 3, 4\}$, define

$$\mathcal{U}_{4,h}^{[r]} := 12\mathcal{M}_2(f_h)\mathcal{L}_{0,r} + 48\mathcal{M}_3(f_h)\mathcal{L}_{2,r} + 16\mathcal{M}_4(f_h)\mathcal{L}_{4,r},$$

whenever the right-hand side is defined, and, for $r \in \{1, 3\}$, define

$$\mathcal{U}_{5,h}^{[r]} := 120\mathcal{M}_3(f_h)\mathcal{L}_{1,r}^+ + 160\mathcal{M}_4(f_h)\mathcal{L}_{3,r}^+ + 32\mathcal{M}_5(f_h)\mathcal{L}_{5,r}^+.$$

By (7.3),

$$\begin{aligned} \mathcal{G}_{4,h}(x) &= 12\mathcal{M}_2(f_h) + 48x^2\mathcal{M}_3(f_h) + 16x^4\mathcal{M}_4(f_h), \\ \mathcal{G}_{5,h}(x) &= 120x\mathcal{M}_3(f_h) + 160x^3\mathcal{M}_4(f_h) + 32x^5\mathcal{M}_5(f_h). \end{aligned}$$

Therefore, the bounds (7.10) and (7.11) imply that, for $\theta \in [0, 1]$,

$$\mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) \mathcal{G}_{4,h}(\|\mathfrak{X}_{1,\theta}^{(n)}\|_F) \right] \leq \mathcal{U}_{4,h}^{[r]}, \quad (7.12)$$

and

$$\mathbb{E} \left[\left(\prod_{\ell=1}^r |Z_{A_\ell}| \right) \mathcal{G}_{5,h}(1 + \|\mathfrak{X}_{1,\theta}^{(n)}\|_F) \right] \leq \mathcal{U}_{5,h}^{[r]}. \quad (7.13)$$

Set $\omega_8 := 1 \vee \max_{A \in \mathcal{I}} \mathbb{E}[|Z_A|^8]$. We have

$$\mu_8 = \mathbb{E}[\|\mathfrak{Z}_1\|_F^8] = \mathbb{E} \left[\left(\sum_{A \in \mathcal{I}} Z_A^2 \right)^4 \right] \leq (\nu d)^3 \sum_{A \in \mathcal{I}} \mathbb{E}[|Z_A|^8] \leq (\nu d)^4 \omega_8.$$

Hence, for $0 \leq q \leq 8$, we find that $\mu_q \leq \mu_8^{q/8} \leq (\nu d)^{q/2} \omega_8$. More generally, if $\rho \in \{0, 1, 2\}$ and $\rho + q_1 + \dots + q_s \leq 8$, then

$$\mu_2^{\rho/2} \prod_{j=1}^s \mu_{q_j} \leq (\nu d)^{(\rho + q_1 + \dots + q_s)/2} \omega_8. \quad (7.14)$$

Using the definitions of Λ_q , $\mathcal{L}_{q,r}$ and $\mathcal{L}_{q,r}^+$, the bound (7.14) yields

$$\begin{aligned} \mathcal{L}_{0,r} &\leq (\nu d)^{r/2} \omega_8, & \mathcal{L}_{2,r} &\leq 10(\nu d)^{(r+2)/2} \omega_8, & \mathcal{L}_{4,r} &\leq 1160(\nu d)^{(r+4)/2} \omega_8, \\ \mathcal{L}_{1,r}^+ &\leq 3(\nu d)^{(r+1)/2} \omega_8, & \mathcal{L}_{3,r}^+ &\leq (20 + 256\sqrt{2})(\nu d)^{(r+3)/2} \omega_8, & \mathcal{L}_{5,r}^+ &\leq 262416(\nu d)^{(r+5)/2} \omega_8. \end{aligned}$$

Similarly,

$$\begin{aligned} \mu_2 \mathcal{L}_{0,2} &\leq (\nu d)^2 \omega_8, & \mu_2 \mathcal{L}_{2,2} &\leq 10(\nu d)^3 \omega_8, & \mu_2 \mathcal{L}_{4,2} &\leq 1160(\nu d)^4 \omega_8, \\ \mu_2 \mathcal{L}_{1,1}^+ &\leq 3(\nu d)^2 \omega_8, & \mu_2 \mathcal{L}_{3,1}^+ &\leq (20 + 256\sqrt{2})(\nu d)^3 \omega_8, & \mu_2 \mathcal{L}_{5,1}^+ &\leq 262416(\nu d)^4 \omega_8. \end{aligned}$$

Consequently,

$$\begin{aligned} \mathcal{U}_{4,h}^{[4]} &\leq \omega_8 \{ 12(\nu d)^2 \mathcal{M}_2(f_h) + 480(\nu d)^3 \mathcal{M}_3(f_h) + 18560(\nu d)^4 \mathcal{M}_4(f_h) \}, \\ \mu_2 \mathcal{U}_{4,h}^{[2]} &\leq \omega_8 \{ 12(\nu d)^2 \mathcal{M}_2(f_h) + 480(\nu d)^3 \mathcal{M}_3(f_h) + 18560(\nu d)^4 \mathcal{M}_4(f_h) \}, \\ \mathcal{U}_{4,h}^{[1]} &\leq \omega_8 \{ 12(\nu d)^{1/2} \mathcal{M}_2(f_h) + 480(\nu d)^{3/2} \mathcal{M}_3(f_h) + 18560(\nu d)^{5/2} \mathcal{M}_4(f_h) \}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{U}_{5,h}^{[3]} &\leq \omega_8 \left\{ 360(\nu d)^2 \mathcal{M}_3(f_h) + (3200 + 40960\sqrt{2})(\nu d)^3 \mathcal{M}_4(f_h) + 8397312(\nu d)^4 \mathcal{M}_5(f_h) \right\}, \\ \mu_2 \mathcal{U}_{5,h}^{[1]} &\leq \omega_8 \left\{ 360(\nu d)^2 \mathcal{M}_3(f_h) + (3200 + 40960\sqrt{2})(\nu d)^3 \mathcal{M}_4(f_h) + 8397312(\nu d)^4 \mathcal{M}_5(f_h) \right\}. \end{aligned}$$

Moreover, $|\sigma_{A,B}| = |\mathbb{E}[Z_A Z_B]| \leq \mathbb{E}[|Z_A Z_B|] \leq \mu_2$, so that

$$\sum_{A,B \in \mathcal{I}} |\sigma_{A,B}| \leq (\nu d)^2 \mu_2.$$

Using (7.12) in the bounds for $R_{1,\varepsilon}, R_{2,\varepsilon}, R_{3,\varepsilon}$, and using (7.13) in the bound for $R_{4,\varepsilon}^{A,B,C} + R_{5,\varepsilon}^{A,B,C}$, gives

$$|R_{1,\varepsilon}| + |R_{2,\varepsilon}| + |R_{3,\varepsilon}| \leq \frac{1}{n} \left\{ \frac{(\nu d)^4}{6} \mathcal{U}_{4,h}^{[4]} + \frac{(\nu d)^4 \mu_2}{2} \mathcal{U}_{4,h}^{[2]} + \frac{\nu d}{2} \left(\sum_{A,B,C \in \mathcal{I}} |\mathbb{E}[Z_A Z_B Z_C]| \right) \mathcal{U}_{4,h}^{[1]} \right\},$$

and, for every fixed $A, B, C \in \mathcal{I}$,

$$|R_{4,\varepsilon}^{A,B,C}| + |R_{5,\varepsilon}^{A,B,C}| \leq \frac{K}{\sqrt{n}} \left\{ \frac{(\nu d)^3}{2} \mathcal{U}_{5,h}^{[3]} + (\nu d)^3 \mu_2 \mathcal{U}_{5,h}^{[1]} \right\}.$$

It follows from (7.5) and (7.6) that

$$\begin{aligned} |\mathbb{E}[\mathcal{A}^{\text{OU}}_{g_{h,\varepsilon}}(\mathfrak{X}^{(n)})]| &\leq \frac{1}{n} \left\{ \frac{(\nu d)^4}{6} \mathcal{U}_{4,h}^{[4]} + \frac{(\nu d)^4 \mu_2}{2} \mathcal{U}_{4,h}^{[2]} + \frac{\nu d}{2} \left(\sum_{A,B,C \in \mathcal{I}} |\mathbb{E}[Z_A Z_B Z_C]| \right) \mathcal{U}_{4,h}^{[1]} \right\} \\ &\quad + \frac{K}{2n} \left(\sum_{A,B,C \in \mathcal{I}} |\mathbb{E}[Z_A Z_B Z_C]| \right) \left\{ \frac{(\nu d)^3}{2} \mathcal{U}_{5,h}^{[3]} + (\nu d)^3 \mu_2 \mathcal{U}_{5,h}^{[1]} \right\}. \end{aligned}$$

Since $\nu d \geq 1$ and $K \leq 1 + K$, the preceding display yields

$$|\mathbb{E}[\mathcal{A}^{\text{OU}}_{g_{h,\varepsilon}}(\mathfrak{X}^{(n)})]| \leq \frac{(\nu d)^8 (1 + K) \omega_8}{n} \{ \beta_2 \mathcal{M}_2(f_h) + \beta_3 \mathcal{M}_3(f_h) + \beta_4 \mathcal{M}_4(f_h) + \beta_5 \mathcal{M}_5(f_h) \},$$

where

$$\beta_j := u_j + v_j \sum_{A,B,C \in \mathcal{I}} |\mathbb{E}[Z_A Z_B Z_C]|, \quad j \in \{2, 3, 4, 5\},$$

with

$$(u_2, u_3, u_4, u_5) := \left(8, 320, \frac{37120}{3}, 0 \right), \quad (v_2, v_3, v_4, v_5) := \left(6, 510, 11680 + 30720\sqrt{2}, 6297984 \right).$$

Finally, we can use (3.22) in Theorem 3.6 to obtain $\mathcal{M}_m(f_h) \leq \frac{1}{2m} \mathcal{M}_m^{\mathcal{D},\Sigma}(h)$ for $m \in \{2, 3, 4, 5\}$. Thus, uniformly in $\varepsilon \in (0, 1]$,

$$\begin{aligned} |\mathbb{E}[\mathcal{A}^{\text{OU}}_{g_{h,\varepsilon}}(\mathfrak{X}^{(n)})]| &\leq \frac{(\nu d)^8 (1 + K) \omega_8}{n} \left\{ \alpha_2 \mathcal{M}_2^{\mathcal{D},\Sigma}(h) + \alpha_3 \mathcal{M}_3^{\mathcal{D},\Sigma}(h) + \alpha_4 \mathcal{M}_4^{\mathcal{D},\Sigma}(h) + \alpha_5 \mathcal{M}_5^{\mathcal{D},\Sigma}(h) \right\}, \end{aligned}$$

where in obtaining $\alpha_2, \dots, \alpha_5$ the values of $r_2, \dots, r_5$ and $s_2, \dots, s_5$ are obtained by rounding numbers up to the nearest integer. Since h is identified with its bounded continuous extension to $\mathcal{S}_+^d$ and $S_\varepsilon(\mathfrak{X}^{(n)}) = \mathfrak{W}_\nu^{(n)} + \varepsilon I_d \rightarrow \mathfrak{W}_\nu^{(n)}$ almost surely, dominated convergence gives

$$\lim_{\varepsilon \downarrow 0} \mathbb{E}[h(S_\varepsilon(\mathfrak{X}^{(n)}))] = \mathbb{E}[h(\mathfrak{W}_\nu^{(n)})].$$

Letting $\varepsilon \downarrow 0$ in (7.1) and using the preceding uniform bound gives the conclusion.

7.3. Proof of Proposition 4.3

Let $\alpha_\bullet := \sum_{j=1}^N \alpha_j$. By assumption, the α_j are positive integers, so we may take independent random vectors $\{\mathbf{X}_{j,r} : j \in [N], r \in [\alpha_j]\}$ such that $\mathbf{X}_{j,r} \sim \mathcal{N}_d(\mathbf{0}_d, \Sigma_j)$. By the standard Gaussian representation of the Wishart distribution, we may write $\mathfrak{G}_j = \sum_{r=1}^{\alpha_j} \mathbf{X}_{j,r} \mathbf{X}_{j,r}^\top$ for $j \in [N]$, and therefore

$$\mathfrak{T} = \sum_{j=1}^N \sum_{r=1}^{\alpha_j} \mathbf{X}_{j,r} \mathbf{X}_{j,r}^\top. \quad (7.15)$$

Since $\alpha_\bullet \geq d + 1$, the Gaussian vectors appearing in (7.15) span $\mathbb{R}^d$ almost surely, and hence $\mathfrak{T} \in \mathcal{S}_{++}^d$ almost surely. Let f_h denote the solution of the Stein equation associated with the target law $\mathfrak{W}_\nu \sim \mathcal{W}_d(\nu, \tilde{\Sigma}_\nu)$. Since $\nu > 3d - 3$, Theorem 3.6 applies to this target distribution. For each (j, r) , define the leave-one-vector-out matrix

$$\mathfrak{T}_{j,r} := \mathfrak{T} - \mathbf{X}_{j,r} \mathbf{X}_{j,r}^\top.$$

We assumed $\alpha_\bullet \geq d + 1$, so the matrix $\mathfrak{T}_{j,r}$ is a sum of at least d independent full-covariance Gaussian rank-one matrices. These Gaussian vectors span $\mathbb{R}^d$ almost surely, so $\mathfrak{T}_{j,r} \in \mathcal{S}_{++}^d$ almost surely. Consequently, for every $\mathbf{x} \in \mathbb{R}^d$, the matrix $\mathfrak{T}_{j,r} + \mathbf{x} \mathbf{x}^\top$ belongs to $\mathcal{S}_{++}^d$ almost surely, and all derivatives of f_h appearing below are evaluated inside their domain.

By the Stein equation (3.17) and the expression of the Wishart extended generator in (3.2), with (α, Σ) replaced by $(\nu, \tilde{\Sigma}_\nu)$, we have

$$\begin{aligned} \mathbb{E}[h(\mathfrak{T})] - \mathbb{E}[h(\mathfrak{W}_\nu)] &= 2 \mathbb{E}[\text{tr}\{(\nu \tilde{\Sigma}_\nu - \mathfrak{T}) \nabla f_h(\mathfrak{T})\}] + 4 \mathbb{E}[\text{tr}\{\mathfrak{T} \nabla \tilde{\Sigma}_\nu \nabla f_h(\mathfrak{T})\}] \\ &= 2 \mathbb{E}[\text{tr}\{(\bar{\Sigma} - \mathfrak{T}) \nabla f_h(\mathfrak{T})\}] + 4 \mathbb{E}[\text{tr}\{\mathfrak{T} \nabla \tilde{\Sigma}_\nu \nabla f_h(\mathfrak{T})\}], \end{aligned} \quad (7.16)$$

since $\nu \tilde{\Sigma}_\nu = \bar{\Sigma}$. Fix (j, r) , define

$$F_{j,r}(\mathbf{x}) := \nabla f_h(\mathfrak{T}_{j,r} + \mathbf{x} \mathbf{x}^\top) \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^d,$$

and introduce the Σ_j -divergence

$$\text{div}_{\Sigma_j} F_{j,r}(\mathbf{x}) := \text{tr}\{\Sigma_j D F_{j,r}(\mathbf{x})\} = \sum_{a,b=1}^d (\Sigma_j)_{ab} \mathbf{e}_a^\top D F_{j,r}(\mathbf{x}) [\mathbf{e}_b].$$

Also, for $\mathbf{x} \in \mathbb{R}^d$ and $a \in [d]$, let

$$H_{\mathbf{x}}^{(a)} := \mathbf{x} \mathbf{e}_a^\top + \mathbf{e}_a \mathbf{x}^\top \in \mathcal{S}^d.$$

Writing $S_{\mathbf{x}} := \mathfrak{T}_{j,r} + \mathbf{x} \mathbf{x}^\top$, the chain rule gives

$$D F_{j,r}(\mathbf{x}) [\mathbf{e}_b] = D(\nabla f_h)(S_{\mathbf{x}}) [H_{\mathbf{x}}^{(b)}] \mathbf{x} + \nabla f_h(S_{\mathbf{x}}) \mathbf{e}_b.$$

Therefore,

$$\text{div}_{\Sigma_j} F_{j,r}(\mathbf{x}) = \text{tr}\{\Sigma_j \nabla f_h(S_{\mathbf{x}})\} + \sum_{a,b=1}^d (\Sigma_j)_{ab} \mathbf{e}_a^\top D(\nabla f_h)(S_{\mathbf{x}}) [H_{\mathbf{x}}^{(b)}] \mathbf{x}$$

$$\begin{aligned}
&= \text{tr}\{\Sigma_j \nabla f_h(S_{\mathbf{x}})\} + \frac{1}{2} \sum_{a,b=1}^d (\Sigma_j)_{ab} D^2 f_h(S_{\mathbf{x}})[H_{\mathbf{x}}^{(b)}, H_{\mathbf{x}}^{(a)}] \\
&= \text{tr}\{\Sigma_j \nabla f_h(S_{\mathbf{x}})\} + \frac{1}{2} \sum_{a,b=1}^d (\Sigma_j)_{ab} \sum_{i,\ell=1}^d x_i x_\ell D^2 f_h(S_{\mathbf{x}})[\mathbf{e}_i \mathbf{e}_b^\top + \mathbf{e}_b \mathbf{e}_i^\top, \mathbf{e}_\ell \mathbf{e}_a^\top + \mathbf{e}_a \mathbf{e}_\ell^\top] \\
&= \text{tr}\{\Sigma_j \nabla f_h(S_{\mathbf{x}})\} + 2 \sum_{i,\ell=1}^d x_i x_\ell \sum_{a,b=1}^d \nabla_{ib}(\Sigma_j)_{ba} \nabla_{a\ell} f_h(S_{\mathbf{x}}) \\
&= \text{tr}\{\Sigma_j \nabla f_h(S_{\mathbf{x}})\} + 2 \mathbf{x}^\top \nabla \Sigma_j \nabla f_h(S_{\mathbf{x}}) \mathbf{x}.
\end{aligned} \tag{7.17}$$

Moreover, by the assumption $\mathcal{M}_1^{\mathcal{D}, \tilde{\Sigma}_\nu}(h) + \mathcal{M}_2^{\mathcal{D}, \tilde{\Sigma}_\nu}(h) < \infty$ and Theorem 3.6 (ii), applied with $m = 1, 2$ and target scale matrix $\tilde{\Sigma}_\nu$, f_h has bounded first and second derivatives. Thus, conditionally on $\mathfrak{T}_{j,r}$, the map $F_{j,r}$ is C^1 with at most linear growth and its derivative has at most quadratic growth. Gaussian integration by parts below is therefore justified. Since $\mathbf{X}_{j,r}$ is centered Gaussian with covariance matrix Σ_j and is independent of $\mathfrak{T}_{j,r}$, Gaussian integration by parts yields

$$\mathbb{E}[\mathbf{X}_{j,r}^\top F_{j,r}(\mathbf{X}_{j,r})] = \mathbb{E}[\text{div}_{\Sigma_j} F_{j,r}(\mathbf{X}_{j,r})].$$

Using (7.17) and the identity $\mathfrak{T} = \mathfrak{T}_{j,r} + \mathbf{X}_{j,r} \mathbf{X}_{j,r}^\top$, we deduce

$$\mathbb{E}[\mathbf{X}_{j,r}^\top \nabla f_h(\mathfrak{T}) \mathbf{X}_{j,r}] = \mathbb{E}[\text{tr}\{\Sigma_j \nabla f_h(\mathfrak{T})\}] + 2 \mathbb{E}[\mathbf{X}_{j,r}^\top \nabla \Sigma_j \nabla f_h(\mathfrak{T}) \mathbf{X}_{j,r}]. \tag{7.18}$$

Summing (7.18) over $j \in [N]$ and $r \in [\alpha_j]$, we get from (7.15) and the definition $\bar{\Sigma} = \sum_{j=1}^N \alpha_j \Sigma_j$ that

$$\begin{aligned}
\mathbb{E}[\text{tr}\{\mathfrak{T} \nabla f_h(\mathfrak{T})\}] &= \sum_{j=1}^N \sum_{r=1}^{\alpha_j} \mathbb{E}[\mathbf{X}_{j,r}^\top \nabla f_h(\mathfrak{T}) \mathbf{X}_{j,r}] \\
&= \mathbb{E}[\text{tr}\{\bar{\Sigma} \nabla f_h(\mathfrak{T})\}] + 2 \sum_{j=1}^N \sum_{r=1}^{\alpha_j} \mathbb{E}[\mathbf{X}_{j,r}^\top \nabla \Sigma_j \nabla f_h(\mathfrak{T}) \mathbf{X}_{j,r}].
\end{aligned}$$

Substituting this identity into (7.16) and using again (7.15), we find

$$\begin{aligned}
|\mathbb{E}[h(\mathfrak{T})] - \mathbb{E}[h(\mathfrak{W}_\nu)]| &= 4 \left| \sum_{j=1}^N \sum_{r=1}^{\alpha_j} \mathbb{E}[\mathbf{X}_{j,r}^\top \nabla \tilde{\Sigma}_\nu \nabla f_h(\mathfrak{T}) \mathbf{X}_{j,r}] - \sum_{j=1}^N \sum_{r=1}^{\alpha_j} \mathbb{E}[\mathbf{X}_{j,r}^\top \nabla \Sigma_j \nabla f_h(\mathfrak{T}) \mathbf{X}_{j,r}] \right| \\
&\leq 4 \sum_{j=1}^N \sum_{r=1}^{\alpha_j} \mathbb{E}[|\mathbf{X}_{j,r}^\top \nabla (\tilde{\Sigma}_\nu - \Sigma_j) \nabla f_h(\mathfrak{T}) \mathbf{X}_{j,r}|],
\end{aligned} \tag{7.19}$$

where in obtaining the inequality we applied the triangle inequality. It remains to bound the quadratic form in (7.19). Let $B \in \mathcal{S}^d$. From the calculation in (7.17), with Σ_j replaced by B , and the definition of $\mathcal{M}_2(f_h)$ in (3.20), we have

$$|\mathbf{x}^\top \nabla B \nabla f_h(S_{\mathbf{x}}) \mathbf{x}| = \frac{1}{4} \left| \sum_{a,b=1}^d B_{ab} D^2 f_h(S_{\mathbf{x}})[H_{\mathbf{x}}^{(b)}, H_{\mathbf{x}}^{(a)}] \right| \leq \frac{\mathcal{M}_2(f_h)}{4} \sum_{a,b=1}^d |B_{ab}| \|H_{\mathbf{x}}^{(b)}\|_F \|H_{\mathbf{x}}^{(a)}\|_F.$$

By the Cauchy-Schwarz inequality,

$$\sum_{a,b=1}^d |B_{ab}| \|H_{\mathbf{x}}^{(b)}\|_F \|H_{\mathbf{x}}^{(a)}\|_F \leq \|B\|_F \left(\sum_{a,b=1}^d \|H_{\mathbf{x}}^{(b)}\|_F^2 \|H_{\mathbf{x}}^{(a)}\|_F^2 \right)^{1/2} = \|B\|_F \sum_{a=1}^d \|H_{\mathbf{x}}^{(a)}\|_F^2.$$

Now, $\|H_{\mathbf{x}}^{(a)}\|_F^2 = \|\mathbf{x} \mathbf{e}_a^\top + \mathbf{e}_a \mathbf{x}^\top\|_F^2 = 2\|\mathbf{x}\|_2^2 + 2x_a^2$, so that $\sum_{a=1}^d \|H_{\mathbf{x}}^{(a)}\|_F^2 = 2(d+1)\|\mathbf{x}\|_2^2$. Therefore,

$$|\mathbf{x}^\top \nabla B \nabla f_h(S_{\mathbf{x}}) \mathbf{x}| \leq \frac{d+1}{2} \mathcal{M}_2(f_h) \|B\|_F \|\mathbf{x}\|_2^2. \quad (7.20)$$

Applying (7.20) with $B = \tilde{\Sigma}_\nu - \Sigma_j$ and $\mathbf{x} = \mathbf{X}_{j,r}$, and taking expectations, yields

$$\mathbb{E}[|\mathbf{X}_{j,r}^\top \nabla(\tilde{\Sigma}_\nu - \Sigma_j) \nabla f_h(\mathfrak{T}) \mathbf{X}_{j,r}|] \leq \frac{d+1}{2} \|\tilde{\Sigma}_\nu - \Sigma_j\|_F \mathcal{M}_2(f_h) \text{tr}(\Sigma_j).$$

Combining this with (7.19), we obtain

$$|\mathbb{E}[h(\mathfrak{T})] - \mathbb{E}[h(\mathfrak{W}_\nu)]| \leq 2(d+1) \mathcal{M}_2(f_h) \sum_{j=1}^N \alpha_j \text{tr}(\Sigma_j) \|\Sigma_j - \tilde{\Sigma}_\nu\|_F.$$

Finally, by Theorem 3.6 (ii) with $m = 2$ and target scale matrix $\tilde{\Sigma}_\nu$, $\mathcal{M}_2(f_h) \leq \frac{1}{4} \mathcal{M}_2^{\mathcal{D}, \tilde{\Sigma}_\nu}(h)$. Substituting this into the previous display gives (4.5). This completes the proof.

7.4. Proof of Proposition 4.4

Fix $g \in C_b^2(\mathcal{S}_{++}^d)$ satisfying $m_g \leq g(X) \leq M_g$ for all $X \in \mathcal{S}_{++}^d$. For $t > 0$, let $\Sigma_t := (1 - e^{-2t})\Sigma$ and $p_t(W, Y) := f_{\alpha, \Sigma_t, e^{-2t}\Sigma_t^{-1}W}^{\mathcal{W}}(Y)$, so that, by the distributional relation (3.4) and Eq. (3.6) of Proposition 3.2,

$$(\mathcal{P}_t^{\mathcal{W}} h)(W) = \int_{\mathcal{S}_{++}^d} h(Y) p_t(W, Y) dY.$$

Also set

$$f_\xi(X) := \frac{\xi(dX)}{dX} = f_{\alpha, \Sigma, 0_{d \times d}}^{\mathcal{W}}(X), \quad X \in \mathcal{S}_{++}^d.$$

By (2.2), for every $W, Y \in \mathcal{S}_{++}^d$,

$$\begin{aligned} p_t(W, Y) f_\xi(W) &= \frac{|W|^{\alpha/2 - (d+1)/2} |Y|^{\alpha/2 - (d+1)/2}}{|2\Sigma|^{\alpha/2} |2\Sigma_t|^{\alpha/2} \Gamma_d(\alpha/2)^2} \\ &\quad \times \text{etr} \left\{ -\frac{1}{2} \Sigma_t^{-1} Y - \frac{e^{-2t}}{2} \Sigma_t^{-1} W - \frac{1}{2} \Sigma^{-1} W \right\} {}_0F_1 \left(\frac{\alpha}{2}; \frac{e^{-2t}}{4} \Sigma_t^{-1} W \Sigma_t^{-1} Y \right). \end{aligned}$$

Since $\Sigma_t^{-1} = (1 - e^{-2t})^{-1} \Sigma^{-1}$, we have $(e^{-2t}/2) \Sigma_t^{-1} + (1/2) \Sigma^{-1} = (1/2) \Sigma_t^{-1}$. Moreover, we have ${}_0F_1(\nu; AB) = {}_0F_1(\nu; BA)$ for square matrices A and B of the same size. Hence

$$p_t(W, Y) f_\xi(W) = p_t(Y, W) f_\xi(Y), \quad W, Y \in \mathcal{S}_{++}^d.$$

It follows that the Wishart semigroup is symmetric in $L^2(\xi)$. Indeed, for bounded measurable functions $\varphi, \psi : \mathcal{S}_{++}^d \rightarrow \mathbb{R}$,

$$\begin{aligned} \int_{\mathcal{S}_{++}^d} (\mathcal{P}_t^{\mathcal{W}} \varphi)(W) \psi(W) \xi(dW) &= \int_{\mathcal{S}_{++}^d} \int_{\mathcal{S}_{++}^d} \psi(W) \varphi(Y) p_t(W, Y) dY \xi(dW) \\ &= \int_{\mathcal{S}_{++}^d} \int_{\mathcal{S}_{++}^d} \psi(W) \varphi(Y) p_t(Y, W) dW \xi(dY) \\ &= \int_{\mathcal{S}_{++}^d} \varphi(Y) (\mathcal{P}_t^{\mathcal{W}} \psi)(Y) \xi(dY). \end{aligned}$$

Therefore, since $\mu_t = \mu \mathcal{P}_t^{\mathcal{W}}$ and $\mu(dW) = g(W) \xi(dW)$,

$$\int_{\mathcal{S}_{++}^d} \varphi(Y) \mu_t(dY) = \int_{\mathcal{S}_{++}^d} (\mathcal{P}_t^{\mathcal{W}} \varphi)(W) g(W) \xi(dW) = \int_{\mathcal{S}_{++}^d} \varphi(Y) (\mathcal{P}_t^{\mathcal{W}} g)(Y) \xi(dY).$$

This proves the first claim of the proposition, namely

$$g_t = \mathcal{P}_t^{\mathcal{W}} g. \quad (7.21)$$

For $h \in C_b^2(\mathcal{S}_{++}^d)$, the Kolmogorov backward equation for the Wishart semigroup gives

$$\partial_t \mathcal{P}_t^{\mathcal{W}} h = \mathcal{A}^{\mathcal{W}} \mathcal{P}_t^{\mathcal{W}} h, \quad t > 0. \quad (7.22)$$

Applying (7.22) with $h = g$ and using (7.21), we obtain

$$\partial_t g_t = \mathcal{A}^{\mathcal{W}} g_t. \quad (7.23)$$

Next, let $\varphi, \psi \in C_b^2(\mathcal{S}_{++}^d)$. By Proposition 3.2, the measure ξ is invariant for $(\mathcal{P}_t^{\mathcal{W}})_{t \geq 0}$, and therefore

$$\int_{\mathcal{S}_{++}^d} (\mathcal{P}_t^{\mathcal{W}} \varphi)(X) \xi(dX) = \int_{\mathcal{S}_{++}^d} \varphi(X) \xi(dX), \quad t \geq 0.$$

Using this invariance and Dynkin's formula gives $\int_{\mathcal{S}_{++}^d} \mathcal{A}^{\mathcal{W}} \varphi(X) \xi(dX) = 0$. Combining this with (4.6) and the symmetry of $\mathcal{P}_t^{\mathcal{W}}$, hence of $\mathcal{A}^{\mathcal{W}}$, we obtain

$$\begin{aligned} 0 &= \int_{\mathcal{S}_{++}^d} \mathcal{A}^{\mathcal{W}}(\varphi\psi)(X) \xi(dX) \\ &= \int_{\mathcal{S}_{++}^d} \varphi(X) \mathcal{A}^{\mathcal{W}} \psi(X) \xi(dX) + \int_{\mathcal{S}_{++}^d} \psi(X) \mathcal{A}^{\mathcal{W}} \varphi(X) \xi(dX) \\ &\quad + 8 \int_{\mathcal{S}_{++}^d} \text{tr}\{X \nabla \varphi(X) \Sigma \nabla \psi(X)\} \xi(dX) \\ &= 2 \int_{\mathcal{S}_{++}^d} \psi(X) \mathcal{A}^{\mathcal{W}} \varphi(X) \xi(dX) + 8 \int_{\mathcal{S}_{++}^d} \text{tr}\{X \nabla \varphi(X) \Sigma \nabla \psi(X)\} \xi(dX). \end{aligned}$$

Thus,

$$\int_{\mathcal{S}_{++}^d} \psi(X) \mathcal{A}^{\mathcal{W}} \varphi(X) \xi(dX) = -4 \int_{\mathcal{S}_{++}^d} \text{tr}\{X \nabla \varphi(X) \Sigma \nabla \psi(X)\} \xi(dX). \quad (7.24)$$

The boundedness assumption now gives the required domination. Since $\mathcal{P}_t^{\mathcal{W}}$ is Markov, (7.21) implies

$$m_g \leq g_t(X) \leq M_g, \quad X \in \mathcal{S}_{++}^d, \quad t \geq 0.$$

In particular, $\log(g_t)$ is bounded uniformly in X and t . Moreover, by the regularity of the Wishart semigroup on $C_b^2(\mathcal{S}_{++}^d)$, for every compact interval $I \subset (0, \infty)$, the functions g_t , ∇g_t and the second-order derivatives of g_t are bounded uniformly over $t \in I$. Hence, from the generator formula (3.2), there exists a constant $C_I \in (0, \infty)$ such that

$$|\mathcal{A}^{\mathcal{W}} g_t(X)| \leq C_I(1 + \text{tr}(X)), \quad X \in \mathcal{S}_{++}^d, \quad t \in I. \quad (7.25)$$

Since $\xi = \mathcal{W}_d(\alpha, \Sigma)$ has finite first moment, the right-hand side of (7.25) is integrable with respect to ξ . Thus dominated convergence justifies the entropy differentiation below. The lower bound on g_t also implies that $\log(g_t) \in C_b^2(\mathcal{S}_{++}^d)$ whenever $g_t \in C_b^2(\mathcal{S}_{++}^d)$, so the integration by parts identity (7.24) applies with $\varphi = g_t$ and $\psi = \log(g_t)$.

We can now differentiate the relative entropy. Using (7.23) and the domination above (this is the role of Assumption (A) in Remark 4.3), we have

$$\begin{aligned} \frac{d}{dt} \text{Ent}(\mu_t \parallel \xi) &= \frac{d}{dt} \int_{\mathcal{S}_{++}^d} g_t(X) \log(g_t(X)) \xi(dX) \\ &= \int_{\mathcal{S}_{++}^d} (1 + \log(g_t(X))) \partial_t g_t(X) \xi(dX) \\ &= \int_{\mathcal{S}_{++}^d} (1 + \log(g_t(X))) \mathcal{A}^{\mathcal{W}} g_t(X) \xi(dX). \end{aligned}$$

Since $\int_{\mathcal{S}_{++}^d} \mathcal{A}^{\mathcal{W}} g_t(X) \xi(dX) = 0$ (this is the first part of Assumption (B) in Remark 4.3), this reduces to

$$\frac{d}{dt} \text{Ent}(\mu_t \parallel \xi) = \int_{\mathcal{S}_{++}^d} \log(g_t(X)) \mathcal{A}^{\mathcal{W}} g_t(X) \xi(dX).$$

Applying (7.24) with $\varphi = g_t$ and $\psi = \log(g_t)$ (this is the second part of Assumption (B) in Remark 4.3) gives

$$\begin{aligned} \frac{d}{dt} \text{Ent}(\mu_t \parallel \xi) &= -4 \int_{\mathcal{S}_{++}^d} \text{tr}\{X \nabla g_t(X) \Sigma \nabla \log(g_t)(X)\} \xi(dX) \\ &= -4 \int_{\mathcal{S}_{++}^d} \text{tr}\{X \nabla \log(g_t)(X) \Sigma \nabla \log(g_t)(X)\} g_t(X) \xi(dX) \\ &= -J_{\mathcal{W}, \Sigma}(\mu_t \mid \xi), \end{aligned}$$

which proves (4.7). Then, the identity (4.8) follows from the chain rule and the change of variable $\tau = e^{-2t}$. Since $J_{\mathcal{W}, \Sigma}(\mu_t \mid \xi) \geq 0$, the monotonicity of $t \mapsto \text{Ent}(\mu_t \parallel \xi)$ is immediate.

7.5. Proof of Proposition 4.5

We work under the same assumptions as in Proposition 4.4, so fix $g \in C_b^2(\mathcal{S}_{++}^d)$ satisfying $m_g \leq g(X) \leq M_g$ for all $X \in \mathcal{S}_{++}^d$. By (7.21), for every $X \in \mathcal{S}_{++}^d$,

$$g_t(X) = (\mathcal{P}_t^{\mathcal{W}} g)(X) = \mathbb{E}[g(\mathfrak{W}_t) \mid \mathfrak{W}_0 = X].$$

Stochastic continuity gives $\mathfrak{W}_t \mid \{\mathfrak{W}_0 = X\} \rightarrow X$ in probability as $t \downarrow 0$. Since g is bounded and continuous, $g_t(X) \rightarrow g(X)$ as $t \downarrow 0$. Moreover, $0 < g_t(X) \leq \|g\|_\infty$ for $X \in \mathcal{S}_{++}^d$ and $t \geq 0$. Therefore, if $C_g := \sup_{0 \leq u \leq \|g\|_\infty} |u \log(u)| < \infty$, with the convention $0 \log(0) = 0$, then

$$|g_t(X) \log(g_t(X))| \leq C_g, \quad X \in \mathcal{S}_{++}^d, \quad t \geq 0.$$

Since ξ is a probability measure, dominated convergence yields

$$\text{Ent}(\mu_t \parallel \xi) \rightarrow \text{Ent}(\mu \parallel \xi), \quad t \downarrow 0. \quad (7.26)$$

Next, for fixed $X, Y \in \mathcal{S}_{++}^d$, the explicit transition density satisfies $p_t(X, Y) \rightarrow f_\xi(Y)$ as $t \rightarrow \infty$, by the convergence in distribution result in Eq. (3.5) of Proposition 3.2. Since $p_t(X, \cdot)$ and f_ξ are probability densities, we have $\int_{\mathcal{S}_{++}^d} |p_t(X, Y) - f_\xi(Y)| dY \rightarrow 0$ as $t \rightarrow \infty$, by Scheffé's theorem (Scheffé, 1947, p. 435–436). Hence, as $t \rightarrow \infty$,

$$\begin{aligned} \left| g_t(X) - \int_{\mathcal{S}_{++}^d} g(Y) \xi(dY) \right| &= \left| \int_{\mathcal{S}_{++}^d} g(Y) \{p_t(X, Y) - f_\xi(Y)\} dY \right| \\ &\leq \|g\|_\infty \int_{\mathcal{S}_{++}^d} |p_t(X, Y) - f_\xi(Y)| dY \rightarrow 0. \end{aligned}$$

Since $\int_{\mathcal{S}_{++}^d} g(Y) \xi(dY) = 1$, we have $g_t(X) \rightarrow 1$ as $t \rightarrow \infty$. Using the same domination by C_g , we obtain

$$\text{Ent}(\mu_t \parallel \xi) = \int_{\mathcal{S}_{++}^d} g_t(X) \log(g_t(X)) \xi(dX) \rightarrow 0, \quad t \rightarrow \infty. \quad (7.27)$$

Integrating (4.7) (under the assumptions of Proposition 4.4) from ε to T , with $0 < \varepsilon < T$, gives

$$\text{Ent}(\mu_\varepsilon \parallel \xi) - \text{Ent}(\mu_T \parallel \xi) = \int_\varepsilon^T J_{\mathcal{W}, \Sigma}(\mu_t \mid \xi) dt.$$

Letting $\varepsilon \downarrow 0$ and $T \rightarrow \infty$, and using (7.26), (7.27) and monotone convergence for the nonnegative integrand, proves the first identity in (4.9). The second identity in (4.9) follows from the change of variable $\tau = e^{-2t}$.

To prove (4.10), note that the Lebesgue density of ξ is

$$f_\xi(X) = \frac{|X|^{\alpha/2 - (d+1)/2} \text{etr}(-\Sigma^{-1}X/2)}{|2\Sigma|^{\alpha/2} \Gamma_d(\alpha/2)}, \quad X \in \mathcal{S}_{++}^d.$$

Using the definition of differential entropy, we obtain

$$\begin{aligned} \text{Ent}(\mu \parallel \xi) &= \int_{\mathcal{S}_{++}^d} f_{\mathfrak{X}}(X) \log \left(\frac{f_{\mathfrak{X}}(X)}{f_\xi(X)} \right) dX \\ &= -H(\mathfrak{X}) - \mathbb{E}[\log(f_\xi(\mathfrak{X}))] \\ &= -H(\mathfrak{X}) + \frac{1}{2} \text{tr}\{\Sigma^{-1} \mathbb{E}[\mathfrak{X}]\} - \left(\frac{\alpha - d - 1}{2} \right) \mathbb{E}[\log(|\mathfrak{X}|)] + \frac{\alpha}{2} \log(|2\Sigma|) + \log(\Gamma_d(\alpha/2)) \\ &= H(\mathfrak{W}_\infty) - H(\mathfrak{X}) + \frac{1}{2} \text{tr}\{\Sigma^{-1}(\mathbb{E}[\mathfrak{X}] - \mathbb{E}[\mathfrak{W}_\infty])\} \\ &\quad - \left(\frac{\alpha - d - 1}{2} \right) (\mathbb{E}[\log(|\mathfrak{X}|)] - \mathbb{E}[\log(|\mathfrak{W}_\infty|)]). \end{aligned}$$

Under the stated moment constraints, the last two terms vanish, so $\text{Ent}(\mu \parallel \xi) = H(\mathfrak{W}_\infty) - H(\mathfrak{X})$. Combining this with (4.9) yields (4.10). This completes the proof.

7.6. Proof of Proposition 4.6

The proof of our logarithmic Sobolev inequality for the Wishart measure uses Bartlett's decomposition of the Wishart law, tensorization of the one-dimensional Gaussian and gamma logarithmic Sobolev inequalities, and a change-of-variables computation identifying the resulting Dirichlet form with the Wishart Fisher information.

It is enough to prove the inequality first for smooth densities g which are bounded above and bounded away from zero. Indeed, if $J_{\mathcal{W}, \Sigma}(\mu \parallel \xi) < \infty$, standard truncation and mollification on $\mathcal{S}_{++}^d$ give smooth densities g_n bounded above and bounded away from zero such that $g_n \rightarrow g$ in $L^1(\xi)$ and

$$\limsup_{n \rightarrow \infty} J_{\mathcal{W}, \Sigma}(g_n \xi \parallel \xi) \leq J_{\mathcal{W}, \Sigma}(\mu \parallel \xi).$$

Applying the smooth inequality to $g_n \xi$ and using lower semicontinuity of relative entropy and of the closed Dirichlet form gives the stated inequality. We therefore assume throughout the proof below that g is smooth, bounded above and bounded away from zero.

Let γ_i denote the gamma law with shape parameter $(\alpha - i + 1)/2$ and rate parameter $1/2$, for $i \in [d]$, and let ϕ denote the standard Gaussian law on $\mathbb{R}$. Define the product measure

$$\beta_\alpha := \left(\bigotimes_{i=1}^d \gamma_i \right) \otimes \left(\bigotimes_{1 \leq j < i \leq d} \phi \right).$$

Since $\alpha \in [d, \infty)$, each gamma shape parameter $(\alpha - i + 1)/2$ is at least $1/2$, so Proposition 2 of Arras and Swan (2017) applies to every diagonal coordinate.

For $(\mathbf{y}, \mathbf{z}) \in (0, \infty)^d \times \mathbb{R}^{d(d-1)/2}$, let $L(\mathbf{y}, \mathbf{z})$ be the lower triangular matrix defined by

$$L_{ii}(\mathbf{y}, \mathbf{z}) = \sqrt{y_i}, \quad L_{ij}(\mathbf{y}, \mathbf{z}) = z_{ij}, \quad 1 \leq j < i \leq d,$$

and set

$$W(\mathbf{y}, \mathbf{z}) := \Sigma^{1/2} L(\mathbf{y}, \mathbf{z}) L(\mathbf{y}, \mathbf{z})^\top \Sigma^{1/2}, \quad G(\mathbf{y}, \mathbf{z}) := g(W(\mathbf{y}, \mathbf{z})).$$

By Bartlett's decomposition of the Wishart law (see, e.g., Muirhead (1982, Theorem 3.2.14) or Gupta and Nagar (2000, Theorem 3.3.4)), if $(\mathbf{Y}, \mathbf{Z}) \sim \beta_\alpha$, then $W(\mathbf{Y}, \mathbf{Z}) \sim \xi$. Hence

$$\int_{(0, \infty)^d \times \mathbb{R}^{d(d-1)/2}} G(\mathbf{y}, \mathbf{z}) \beta_\alpha(d\mathbf{y}, d\mathbf{z}) = \int_{\mathcal{S}_{++}^d} g(X) \xi(dX) = \int_{\mathcal{S}_{++}^d} \mu(dX) = 1$$

and

$$\begin{aligned} \text{Ent}(\mu \parallel \xi) &= \int_{\mathcal{S}_{++}^d} g(X) \log(g(X)) \xi(dX) \\ &= \int_{(0, \infty)^d \times \mathbb{R}^{d(d-1)/2}} G(\mathbf{y}, \mathbf{z}) \log(G(\mathbf{y}, \mathbf{z})) \beta_\alpha(d\mathbf{y}, d\mathbf{z}). \end{aligned}$$

We next rewrite Proposition 2 of Arras and Swan (2017) in the functional form needed below. Fix $i \in [d]$, and let f_{γ_i} denote the Lebesgue density of γ_i . If ν is a probability measure on $(0, \infty)$ with Lebesgue density f_ν , set

$$q(y) := \frac{f_\nu(y)}{f_{\gamma_i}(y)}, \quad y \in (0, \infty).$$

Then $\nu(dy) = q(y) \gamma_i(dy)$ and $\int_0^\infty q(y) \gamma_i(dy) = 1$. The standardized gamma Fisher information appearing in [Arras and Swan \(2017\)](#) can be written as

$$\int_0^\infty y \frac{(q'(y))^2}{q(y)} \gamma_i(dy),$$

because

$$\partial_y \log(f_\nu(y)) - \partial_y \log(f_{\gamma_i}(y)) = \partial_y \log(q(y)) = \frac{q'(y)}{q(y)}.$$

Since γ_i has rate parameter $1/2$, Proposition 2 of [Arras and Swan \(2017\)](#) gives

$$\int_0^\infty q(y) \log(q(y)) \gamma_i(dy) \leq 2 \int_0^\infty y \frac{(q'(y))^2}{q(y)} \gamma_i(dy).$$

Now let $u : (0, \infty) \rightarrow (0, \infty)$ be smooth and set

$$m_u := \int_0^\infty u(y) \gamma_i(dy), \quad q_u(y) := \frac{u(y)}{m_u}.$$

Applying the previous inequality to q_u gives

$$\int_0^\infty q_u(y) \log(q_u(y)) \gamma_i(dy) \leq 2 \int_0^\infty y \frac{(q'_u(y))^2}{q_u(y)} \gamma_i(dy).$$

Moreover,

$$\int_0^\infty q_u(y) \log(q_u(y)) \gamma_i(dy) = \frac{1}{m_u} \int_0^\infty u(y) \log(u(y)) \gamma_i(dy) - \log(m_u),$$

and

$$\int_0^\infty y \frac{(q'_u(y))^2}{q_u(y)} \gamma_i(dy) = \frac{1}{m_u} \int_0^\infty y \frac{(u'(y))^2}{u(y)} \gamma_i(dy).$$

Multiplying by m_u yields, for each $i \in [d]$,

$$\int_0^\infty u(y) \log(u(y)) \gamma_i(dy) - \left(\int_0^\infty u(y) \gamma_i(dy) \right) \log \left(\int_0^\infty u(y) \gamma_i(dy) \right) \leq 2 \int_0^\infty y \frac{(u'(y))^2}{u(y)} \gamma_i(dy).$$

Likewise, for every smooth $v : \mathbb{R} \rightarrow (0, \infty)$, the classical Gaussian logarithmic Sobolev inequality (see, e.g., [Boucheron et al., 2013](#), Section 5.3, p. 125) gives

$$\int_{\mathbb{R}} v(z) \log(v(z)) \phi(dz) - \left(\int_{\mathbb{R}} v(z) \phi(dz) \right) \log \left(\int_{\mathbb{R}} v(z) \phi(dz) \right) \leq \frac{1}{2} \int_{\mathbb{R}} \frac{(v'(z))^2}{v(z)} \phi(dz).$$

Tensorizing these one-dimensional inequalities over the independent Bartlett coordinates yields

$$\begin{aligned} \text{Ent}(\mu \parallel \xi) &\leq 2 \sum_{i=1}^d \int_{(0, \infty)^d \times \mathbb{R}^{d(d-1)/2}} y_i \frac{(\partial_{y_i} G(\mathbf{y}, \mathbf{z}))^2}{G(\mathbf{y}, \mathbf{z})} \beta_\alpha(d\mathbf{y}, d\mathbf{z}) \\ &\quad + \frac{1}{2} \sum_{1 \leq j < i \leq d} \int_{(0, \infty)^d \times \mathbb{R}^{d(d-1)/2}} \frac{(\partial_{z_{ij}} G(\mathbf{y}, \mathbf{z}))^2}{G(\mathbf{y}, \mathbf{z})} \beta_\alpha(d\mathbf{y}, d\mathbf{z}). \end{aligned} \tag{7.28}$$

Fix $(\mathbf{y}, \mathbf{z})$ and abbreviate $L := L(\mathbf{y}, \mathbf{z})$ and $W := W(\mathbf{y}, \mathbf{z})$. Also write

$$B(\mathbf{y}, \mathbf{z}) := \Sigma^{1/2} \nabla g(W(\mathbf{y}, \mathbf{z})) \Sigma^{1/2} L(\mathbf{y}, \mathbf{z}).$$

For $E_{ij} := \mathbf{e}_i \mathbf{e}_j^\top$ and $1 \leq j < i \leq d$, we have

$$\begin{aligned} \partial_{z_{ij}} W(\mathbf{y}, \mathbf{z}) &= \Sigma^{1/2} (E_{ij} L^\top + L E_{ji}) \Sigma^{1/2}, \\ \partial_{y_i} W(\mathbf{y}, \mathbf{z}) &= \frac{1}{2\sqrt{y_i}} \Sigma^{1/2} (E_{ii} L^\top + L E_{ii}) \Sigma^{1/2}. \end{aligned}$$

Since $\nabla g(W)$ is symmetric, the chain rule gives

$$\begin{aligned} \partial_{z_{ij}} G(\mathbf{y}, \mathbf{z}) &= \text{tr}\{\nabla g(W) \partial_{z_{ij}} W(\mathbf{y}, \mathbf{z})\} = 2B_{ij}(\mathbf{y}, \mathbf{z}), \\ \partial_{y_i} G(\mathbf{y}, \mathbf{z}) &= \text{tr}\{\nabla g(W) \partial_{y_i} W(\mathbf{y}, \mathbf{z})\} = \frac{1}{\sqrt{y_i}} B_{ii}(\mathbf{y}, \mathbf{z}). \end{aligned}$$

Substituting these identities into (7.28), we obtain

$$\begin{aligned} \text{Ent}(\mu \parallel \xi) &\leq 2 \int_{(0, \infty)^d \times \mathbb{R}^{d(d-1)/2}} \frac{\sum_{i=1}^d B_{ii}(\mathbf{y}, \mathbf{z})^2 + \sum_{1 \leq j < i \leq d} B_{ij}(\mathbf{y}, \mathbf{z})^2}{g(W(\mathbf{y}, \mathbf{z}))} \beta_\alpha(d\mathbf{y}, d\mathbf{z}) \\ &\leq 2 \int_{(0, \infty)^d \times \mathbb{R}^{d(d-1)/2}} \frac{\|B(\mathbf{y}, \mathbf{z})\|_F^2}{g(W(\mathbf{y}, \mathbf{z}))} \beta_\alpha(d\mathbf{y}, d\mathbf{z}). \end{aligned}$$

Moreover,

$$\begin{aligned} \|B(\mathbf{y}, \mathbf{z})\|_F^2 &= \text{tr}\{L^\top \Sigma^{1/2} \nabla g(W) \Sigma \nabla g(W) \Sigma^{1/2} L\} \\ &= \text{tr}\{\Sigma^{1/2} L L^\top \Sigma^{1/2} \nabla g(W) \Sigma \nabla g(W)\} \\ &= \text{tr}\{W \nabla g(W) \Sigma \nabla g(W)\}. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Ent}(\mu \parallel \xi) &\leq 2 \int_{(0, \infty)^d \times \mathbb{R}^{d(d-1)/2}} \frac{\text{tr}\{W(\mathbf{y}, \mathbf{z}) \nabla g(W(\mathbf{y}, \mathbf{z})) \Sigma \nabla g(W(\mathbf{y}, \mathbf{z}))\}}{g(W(\mathbf{y}, \mathbf{z}))} \beta_\alpha(d\mathbf{y}, d\mathbf{z}) \\ &= 2 \int_{\mathcal{S}_{++}^d} \frac{\text{tr}\{X \nabla g(X) \Sigma \nabla g(X)\}}{g(X)} \xi(dX) = \frac{1}{2} J_{\mathcal{W}, \Sigma}(\mu \parallel \xi), \end{aligned}$$

which is exactly (4.11). This completes the proof.

7.7. Proof of Proposition 4.7

We have $\nabla f_U(S) = U$ and $\text{tr}\{S \nabla \widehat{\Sigma} \nabla f_U(S)\} = 0$ for $S \in \mathcal{S}_{++}^d$. Therefore, substituting (4.13) into (4.12) yields

$$0 = 2 \text{tr} \left\{ \left(\alpha \widehat{\Sigma} - \frac{1}{n} \sum_{k=1}^n \mathfrak{W}^{(k)} \right) U \right\}.$$

Since U ranges over $\mathcal{S}^d$, trace duality implies $\widehat{\Sigma} = \frac{1}{\alpha n} \sum_{k=1}^n \mathfrak{W}^{(k)}$, as claimed.

7.8. Proof of Proposition 4.8

Similarly to the proof of Proposition 4.7, applying the linear probes in (4.13) to (4.14) and using trace duality gives $\widehat{M} = \frac{1}{n} \sum_{\ell=1}^n \mathfrak{W}^{(\ell)}$. We first consider the quadratic probe $q(S) = \text{tr}(S^2)$ from (4.15). We have

$$\nabla q(S) = 2S, \quad \nabla_{k\ell} q(S) = 2S_{k\ell}, \quad \nabla_{ij} \nabla_{k\ell} q(S) = \delta_{ik} \delta_{j\ell} + \delta_{il} \delta_{jk}.$$

Therefore,

$$(\nabla \widehat{M} \nabla q(S))_{i\ell} = \sum_{j,k=1}^d \widehat{M}_{jk} \nabla_{ij} \nabla_{k\ell} q(S) = \widehat{M}_{\ell i} + \delta_{i\ell} \text{tr}(\widehat{M}) = \widehat{M}_{i\ell} + \delta_{i\ell} \text{tr}(\widehat{M}),$$

where we used the symmetry of $\widehat{M}$. Hence,

$$\text{tr}\{S \nabla \widehat{M} \nabla q(S)\} = \text{tr}(S \widehat{M}) + \text{tr}(S) \text{tr}(\widehat{M}), \quad S \in \mathcal{S}_{++}^d.$$

Substituting q into (4.14), and using the identities $\overline{\text{tr}(\mathfrak{W} \widehat{M})} = \text{tr}(\widehat{M}^2)$ and $\overline{\text{tr}(\mathfrak{W})} = \text{tr}(\widehat{M})$, we obtain

$$\begin{aligned} 0 &= \overline{\mathcal{A}_{\widehat{M}, \widehat{\alpha}_{\text{quad}}}^{\mathfrak{W}} q(\mathfrak{W})} = 4 \overline{\text{tr}(\widehat{M} \mathfrak{W})} - 4 \overline{\text{tr}(\mathfrak{W}^2)} + \frac{4}{\widehat{\alpha}_{\text{quad}}} \overline{\text{tr}(\mathfrak{W} \widehat{M}) + \text{tr}(\mathfrak{W}) \text{tr}(\widehat{M})} \\ &= 4 \text{tr}(\widehat{M}^2) - 4 \overline{\text{tr}(\mathfrak{W}^2)} + \frac{4}{\widehat{\alpha}_{\text{quad}}} [\text{tr}(\widehat{M}^2) + \{\text{tr}(\widehat{M})\}^2]. \end{aligned}$$

Solving for $\widehat{\alpha}_{\text{quad}}$ and using $M = \alpha \Sigma$ yields the estimators $\widehat{\alpha}_{\text{quad}}$ and $\widehat{\Sigma}_{\text{quad}}$ defined in (4.16).

Next, we consider the logarithmic probe $\ell(S) = \text{tr}\{S \log(S) - S\}$ from (4.15). Let $u(x) = x \log(x) - x$, so that $\ell(S) = \text{tr}\{u(S)\}$ and $u'(x) = \log(x)$. By the differential calculus of spectral functions (Bhatia, 2007, Theorem 5.3.1), for every $E \in \mathcal{S}^d$,

$$D\ell(S)[E] = \left. \frac{d}{dt} \text{tr}\{u(S + tE)\} \right|_{t=0} = \text{tr}\{u'(S)E\} = \text{tr}\{\log(S)E\}.$$

On the other hand, since the gradient is defined with respect to the trace inner product on $\mathcal{S}^d$, $D\ell(S)[E] = \text{tr}\{\nabla \ell(S)E\}$ for $E \in \mathcal{S}^d$. Therefore, $\text{tr}\{(\nabla \ell(S) - \log(S))E\} = 0$ for all $E \in \mathcal{S}^d$. By trace duality, we deduce

$$\nabla \ell(S) = \log(S), \quad S \in \mathcal{S}_{++}^d.$$

It remains to compute the diffusion part of the Wishart Stein operator. It is enough to work in a basis in which S is diagonal because all the objects involved are orthogonally equivariant. Let $S = \text{diag}(\lambda_1, \dots, \lambda_d)$. At such an S , the differential of the matrix logarithm gives

$$\nabla_{ij}(\log(S))_{a\ell} = \frac{1}{2} \log^{[1]}(\lambda_a, \lambda_\ell) (\delta_{ai} \delta_{\ell j} + \delta_{aj} \delta_{\ell i}).$$

Therefore, for any $M \in \mathcal{S}^d$,

$$\begin{aligned} (\nabla M \nabla \ell(S))_{i\ell} &= \sum_{j,a=1}^d M_{ja} \nabla_{ij} \nabla_{a\ell} \ell(S) = \sum_{j,a=1}^d M_{ja} \nabla_{ij}(\log(S))_{a\ell} \\ &= \frac{1}{2} M_{\ell i} \log^{[1]}(\lambda_i, \lambda_\ell) + \frac{1}{2} \delta_{i\ell} \sum_{j=1}^d M_{jj} \log^{[1]}(\lambda_j, \lambda_i). \end{aligned}$$

Taking the trace after multiplication by S yields, in the basis in which $S = \text{diag}(\lambda_1, \dots, \lambda_d)$,

$$\text{tr}\{S\nabla M\nabla\ell(S)\} = \frac{1}{2}\text{tr}(M) + \frac{1}{2}\sum_{i,j=1}^d \lambda_i M_{jj} \log^{[1]}(\lambda_j, \lambda_i).$$

Returning to a general $S = H\Lambda H^\top$, the same diagonal computation applies with M replaced by its representation $M^{(S)} = H^\top M H$ in the eigenbasis of S . Hence

$$\text{tr}\{S\nabla M\nabla\ell(S)\} = \frac{1}{2}\text{tr}(M) + \frac{1}{2}\sum_{i,j=1}^d \lambda_i M_{jj}^{(S)} \log^{[1]}(\lambda_j, \lambda_i) = \mathcal{J}_S(M).$$

The expression is independent of the chosen spectral decomposition of S .

Now set

$$D_{n,\log} := \overline{\text{tr}(\mathfrak{W}\log(\mathfrak{W}))} - \text{tr}\{\widehat{M}\log(\mathfrak{W})\}, \quad J_{n,\log} := \overline{\mathcal{J}_{\mathfrak{W}}(\widehat{M})}.$$

Substituting ℓ into (4.14), and using the notation above, we obtain

$$0 = \overline{\mathcal{A}_{M, \widehat{\alpha}_{\log}}^{\mathfrak{W}} \ell(\mathfrak{W})} = 2 \text{tr}\{(\widehat{M} - \mathfrak{W})\log(\mathfrak{W})\} + \frac{4}{\widehat{\alpha}_{\log}} \overline{\mathcal{J}_{\mathfrak{W}}(\widehat{M})} = -2D_{n,\log} + \frac{4}{\widehat{\alpha}_{\log}} J_{n,\log}.$$

Solving for $\widehat{\alpha}_{\log}$ and using $M = \alpha\Sigma$ yields the estimators $\widehat{\alpha}_{\log}$ and $\widehat{\Sigma}_{\log}$ defined in (4.17).

7.9. Proof of Proposition 4.9

Define the population moment vector $\boldsymbol{\theta} \in \mathbb{R}^M$ by $\boldsymbol{\theta} := (\theta_1, \dots, \theta_M)^\top$, with $\theta_m := \mathbb{E}[y_{n,m}]$ for all $m \in \{1, \dots, M\}$. By the calculation above with the linear probe f_{U_m} , we have

$$\mathbb{E}[\text{tr}(\mathfrak{W}^{(1)}U_m)] = \alpha \text{tr}(\Sigma U_m),$$

and therefore, using the assumption $\Sigma = \Sigma(\boldsymbol{\beta}^*) = \sum_{j=1}^p \beta_j^* B_j$,

$$\theta_m = \frac{1}{\alpha} \mathbb{E}[\text{tr}(\mathfrak{W}^{(1)}U_m)] = \text{tr}(\Sigma U_m) = \sum_{j=1}^p \beta_j^* \text{tr}(B_j U_m) = (C\boldsymbol{\beta}^*)_m.$$

Thus, at the population level, we have the linear system $C\boldsymbol{\beta}^* = \boldsymbol{\theta}$.

Now consider the sample vector $\mathbf{y}_n$. For each $m \in \{1, \dots, M\}$,

$$y_{n,m} = \frac{1}{n} \sum_{k=1}^n Z_{k,m}, \quad Z_{k,m} := \frac{1}{\alpha} \text{tr}(\mathfrak{W}^{(k)}U_m).$$

Since $|Z_{k,m}| \leq \frac{1}{\alpha} \|U_m\|_F \|\mathfrak{W}^{(k)}\|_F \leq \frac{1}{\alpha} \|U_m\|_F \text{tr}(\mathfrak{W}^{(k)})$ and $\mathbb{E}[\text{tr}(\mathfrak{W}^{(1)})] = \alpha \text{tr}(\Sigma) < \infty$, the random variables $Z_{k,m}$ are integrable. Therefore, by the strong law of large numbers, $\mathbf{y}_n \rightarrow \boldsymbol{\theta}$ almost surely (a.s.). Hence, together with $C\boldsymbol{\beta}^* = \boldsymbol{\theta}$, we deduce

$$\widehat{\boldsymbol{\beta}}_n = (C^\top C)^{-1} C^\top \mathbf{y}_n \rightarrow (C^\top C)^{-1} C^\top \boldsymbol{\theta} = \boldsymbol{\beta}^*, \quad \text{a.s.}$$

Since $\boldsymbol{\beta} \mapsto \Sigma(\boldsymbol{\beta})$ is linear, it follows that

$$\widehat{\Sigma}_n = \Sigma(\widehat{\boldsymbol{\beta}}_n) \rightarrow \Sigma(\boldsymbol{\beta}^*) = \Sigma, \quad \text{a.s.}$$

Finally, $\widehat{\boldsymbol{\beta}}_n$ minimizes $\|C\boldsymbol{\beta} - \mathbf{y}_n\|_2^2$, so $\widehat{\Sigma}_n$ solves the projected Stein moment equations (4.18) in the least-squares sense. This concludes the proof.

Reproducibility

The R code that generated the tables, the figures and the simulation study results is available online in the GitHub repository of [Bailly and Ouimet \(2026\)](#).

Funding

Robert Gaunt is funded by EPSRC grant EP/Y008650/1. Frédéric Ouimet is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through Discovery Grant RGPIN-2026-04471 and Discovery Launch Supplement DGECR-2026-00449. Gabriel Bailly and Rainer von Sachs gratefully acknowledge support from the Fonds spéciaux de recherche (FSR) of UCLouvain.

References

- A. Anastasiou, A. Barp, F.-X. Briol, B. Ebner, R. E. Gaunt, F. Ghaderinezhad, J. Gorham, A. Gretton, C. Ley, Q. Liu, L. Mackey, C. J. Oates, G. Reinert, and Y. Swan. Stein’s method meets computational statistics: a review of some recent developments. *Statist. Sci.*, 38(1): 120–139, 2023. [doi:10.1214/22-STS863](#).
- B. Arras and C. Houdré. On Stein’s method for multivariate self-decomposable laws. *Electron. J. Probab.*, 24:Paper No. 128, 63, 2019a. [doi:10.1214/19-EJP378](#).
- B. Arras and C. Houdré. *On Stein’s method for infinitely divisible laws with finite first moment*. SpringerBriefs in Probability and Mathematical Statistics. Springer, Cham, 2019b. ISBN 978-3-030-15016-7; 978-3-030-15017-4. [doi:10.1007/978-3-030-15017-4](#).
- B. Arras and Y. Swan. A stroll along the gamma. *Stochastic Process. Appl.*, 127(11):3661–3688, 2017. [doi:10.1016/j.spa.2017.03.012](#).
- R. A. Askey and A. B. Olde Daalhuis. Generalized hypergeometric functions and Meijer G -function. In Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark, editors, *NIST Handbook of Mathematical Functions*, chapter 16, pages 403–418. Cambridge University Press, New York, 2010. URL <https://dlmf.nist.gov/16>. Online DLMF version: Release 1.2.6 of 2026-03-15.
- G. Bailly and F. Ouimet. SteinWishart, 2026. The GitHub repository is publicly available online at <https://github.com/FredericOuimetMcGill/SteinWishart>.
- D. Bakry. Remarques sur les semigroupes de Jacobi. In *Hommage à P. A. Meyer et J. Neveu*, number 236 in Astérisque, pages 23–39. Société mathématique de France, Paris, 1996. URL https://www.numdam.org/item/AST_1996__236__23_0/.
- D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de Probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer-Verlag, Berlin, 1985. [doi:10.1007/BFb0075847](#). URL https://www.numdam.org/item/SPS_1985__19__177_0/.
- D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften*. Springer, Cham, 2014. ISBN 978-3-319-00226-2; 978-3-319-00227-9. [doi:10.1007/978-3-319-00227-9](#).
- A. D. Barbour. Stein’s method for diffusion approximations. *Probab. Theory Related Fields*, 84 (3):297–322, 1990. [doi:10.1007/BF01197887](#).
- R. Bhatia. *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007. ISBN 978-0-691-12918-1; 0-691-12918-5.

- S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013. ISBN 978-0-19-953525-5. doi:10.1093/acprof:oso/9780199535255.001.0001.
- M.-F. Bru. Wishart processes. *J. Theoret. Probab.*, 4(4):725–751, 1991. doi:10.1007/BF01259552.
- S. Chatterjee, J. Fulman, and A. Röllin. Exponential approximation by Stein’s method and spectral graph theory. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 8:197–223, 2011. URL <https://alea.impa.br/articles/v8/08-11.pdf>.
- L. H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probab.*, 3(3):534–545, 1975. doi:10.1214/aop/1176996359.
- L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal Approximation by Stein’s Method*. Probability and its Applications (New York). Springer, Heidelberg, 2011. ISBN 978-3-642-15006-7. doi:10.1007/978-3-642-15007-4.
- P. Chen, I. Nourdin, L. Xu, and X. Yang. Multivariate stable approximation by Stein’s method. *J. Theoret. Probab.*, 37(1):446–488, 2024. doi:10.1007/s10959-023-01244-x.
- C. Cuchiero, D. Filipović, E. Mayerhofer, and J. Teichmann. Affine processes on positive semidefinite matrices. *Ann. Appl. Probab.*, 21(2):397–463, 2011. doi:10.1214/10-AAP710.
- R. Dhoyer and C. A. Tudor. Limit behavior in high-dimensional regime for the Wishart tensors in Wiener chaos. *J. Theoret. Probab.*, 37(2):1445–1468, 2024. doi:10.1007/s10959-024-01328-2.
- P. Diaconis and S. Zabel. Closed form summation for classical distributions: Variations on a theme of de Moivre. *Statist. Sci.*, 6(3):284–302, 1991. doi:10.1214/ss/1177011699.
- B. Ebner, A. Fischer, R. E. Gaunt, B. Picker, and Y. Swan. Stein’s method of moments. *Scand. J. Stat.*, 52(4):1594–1624, 2025. doi:10.1111/sjos.70003.
- R. Eden and J. Víquez. Nourdin–Peccati analysis on Wiener and Wiener–Poisson space for general distributions. *Stochastic Process. Appl.*, 125(1):182–216, 2015. doi:10.1016/j.spa.2014.09.001.
- R. E. Gaunt. *Rates of convergence of Variance-Gamma approximations via Stein’s method*. PhD thesis, University of Oxford, 2013. The Queen’s College.
- R. E. Gaunt. Variance-Gamma approximation via Stein’s method. *Electron. J. Probab.*, 19(38): 33 pp., 2014. doi:10.1214/EJP.v19-3020.
- R. E. Gaunt and G. Reinert. Bounds for the chi-square approximation of Friedman’s statistic by Stein’s method. *Bernoulli*, 29(3):2008–2034, 2023. doi:10.3150/22-BEJ1530.
- R. E. Gaunt, A. M. Pickett, and G. Reinert. Chi-square approximation by Stein’s method with application to Pearson’s statistic. *Ann. Appl. Probab.*, 27(2):720–756, 2017. doi:10.1214/16-AAP1213.
- R. E. Gaunt, F. Ouimet, and D. Richards. Stein’s method for the matrix normal distribution. *ArXiv preprint*, 2026. doi:10.48550/arXiv.2601.11422.
- C. Genest, A. MacKay, and F. Ouimet. On noncentral Wishart mixtures of noncentral Wisharts and their use for testing random effects in factorial design models. *J. Math. Anal. Appl.*, 554(1):Paper No. 129897, 11 pp., 2026. doi:10.1016/j.jmaa.2025.129897.
- F. Götze. On the rate of convergence in the multivariate CLT. *Ann. Probab.*, 19(2):724–739, 1991. doi:10.1214/aop/1176990448.
- L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975. doi:10.2307/2373688.
- A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*, volume 104 of *Monographs and Surveys in Pure and Applied Mathematics*. Chapman & Hall/CRC, Boca Raton, FL, first

- edition, 2000. ISBN 1-58488-046-5. [doi:10.1201/9780203749289](https://doi.org/10.1201/9780203749289).
- L. R. Haff. An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, 9(4):531–544, 1979. [doi:10.1016/0047-259X\(79\)90056-3](https://doi.org/10.1016/0047-259X(79)90056-3).
- C. S. Herz. Bessel functions of matrix argument. *Ann. of Math. (2)*, 61(3):474–523, 1955. [doi:10.2307/1969810](https://doi.org/10.2307/1969810).
- C. G. Khatri. On certain distribution problems based on positive definite quadratic functions in normal vectors. *Ann. Math. Statist.*, 37(2):468–479, 1966. [doi:10.1214/aoms/1177699530](https://doi.org/10.1214/aoms/1177699530).
- C. G. Khatri. Series representations of distributions of quadratic form in the normal vectors and generalised variance. *J. Multivariate Anal.*, 1(2):199–214, 1971. [doi:10.1016/0047-259X\(71\)90011-X](https://doi.org/10.1016/0047-259X(71)90011-X).
- C. G. Khatri. Multivariate generalization of t' -statistic based on the mean square successive difference. *Commun. Statist. Theory Methods*, 18(5):1983–1992, 1989. [doi:10.1080/03610928908830015](https://doi.org/10.1080/03610928908830015).
- S. G. Krantz. *Function Theory of Several Complex Variables*. AMS Chelsea Publishing, Providence, RI, second edition, 2001. ISBN 0-8218-2724-3. [doi:10.1090/chel/340](https://doi.org/10.1090/chel/340). Reprint of the 1992 edition.
- M. Ledoux, I. Nourdin, and G. Peccati. Stein’s method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.*, 25(1):256–306, 2015. [doi:10.1007/s00039-015-0312-0](https://doi.org/10.1007/s00039-015-0312-0).
- H. M. Luk. *Stein’s Method for the Gamma Distribution and Related Statistical Applications*. PhD thesis, University of Southern California, 1994.
- G. Mijoule, M. Raič, G. Reinert, and Y. Swan. Stein’s density method for multivariate continuous distributions. *Electron. J. Probab.*, 28:Paper No. 59, 40, 2023. [doi:10.1214/22-EJP883](https://doi.org/10.1214/22-EJP883).
- D. Mikulincer. A CLT in Stein’s distance for generalized Wishart matrices and higher-order tensors. *Int. Math. Res. Not. IMRN*, 2022(10):7839–7872, 2022. [doi:10.1093/imrn/rnaa336](https://doi.org/10.1093/imrn/rnaa336).
- C. Mortarino. A decomposition for a stochastic matrix with an application to MANOVA. *J. Multivariate Anal.*, 92(1):134–144, 2005. [doi:10.1016/S0047-259X\(03\)00135-0](https://doi.org/10.1016/S0047-259X(03)00135-0).
- R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1982. ISBN 0-471-09442-0. [doi:10.1002/9780470316559](https://doi.org/10.1002/9780470316559).
- D. N. Naik and S. S. Rao. Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *J. Appl. Stat.*, 28(1):91–105, 2001. [doi:10.1080/02664760120011626](https://doi.org/10.1080/02664760120011626).
- I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus: From Stein’s Method to Universality*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012. ISBN 978-1-107-01777-1. [doi:10.1017/CBO9781139084659](https://doi.org/10.1017/CBO9781139084659).
- I. Nourdin and G. Zheng. Asymptotic behavior of large Gaussian correlated Wishart matrices. *J. Theoret. Probab.*, 35(4):2239–2268, 2022. [doi:10.1007/s10959-021-01133-1](https://doi.org/10.1007/s10959-021-01133-1).
- E. A. Peköz and A. Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.*, 39(2):587–608, 2011. [doi:10.1214/10-AOP559](https://doi.org/10.1214/10-AOP559).
- O. Pfaffel. Wishart processes. *ArXiv preprint*, 2012. [doi:10.48550/arXiv.1201.3256](https://doi.org/10.48550/arXiv.1201.3256).
- M. Raič. A multivariate CLT for decomposable random vectors with finite second moments. *J. Theoret. Probab.*, 17(3):573–603, 2004. [doi:10.1023/B:JOTP.0000040290.44087.68](https://doi.org/10.1023/B:JOTP.0000040290.44087.68).
- D. St. P. Richards. Functions of matrix argument. In F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors, *NIST Handbook of Mathematical Functions*, chapter 35, pages 767–774. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-14063-8. URL

<https://dlmf.nist.gov/35>.

- N. Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011. doi:10.1214/11-PS182.
- F. E. Satterthwaite. Synthesis of variance. *Psychometrika*, 6(5):309–316, 1941. doi:10.1007/BF02288586.
- F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bull.*, 2(6):110–114, 1946. doi:10.2307/3002019.
- H. Scheffé. A useful convergence theorem for probability distributions. *Ann. Math. Statist.*, 18(3):434–438, 1947. doi:10.1214/aoms/1177730390.
- M. Singull and T. Koski. On the distribution of matrix quadratic forms. *Comm. Statist. Theory Methods*, 41(18):3403–3415, 2012. doi:10.1080/03610926.2011.563009.
- C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. II: Probability Theory*, pages 583–602. University of California Press, Berkeley, CA, 1972.
- C. Stein. *Approximate Computation of Expectations*, volume 7 of *IMS Lecture Notes Monogr. Ser.* Institute of Mathematical Statistics, Hayward, CA, 1986. ISBN 0-940600-08-0. doi:10.1214/lnms/1215466568.
- W. Y. Tan. On probability distributions from mixtures of multivariate densities. *South African Statist. J.*, 14(1):47–59, 1980. URL https://hdl.handle.net/10520/AJA0038271X_604.
- W. Y. Tan and R. P. Gupta. On approximating a linear combination of central Wishart matrices with positive coefficients. *Comm. Statist. Theory Methods*, 12(22):2589–2600, 1983. doi:10.1080/03610928308828625.
- L. Xu. Approximation of stable law in Wasserstein-1 distance by Stein's method. *Ann. Appl. Probab.*, 29(1):458–504, 2019. doi:10.1214/18-AAP1424.

Supplementary material

S.1. Additional details for the multivariate Satterthwaite approximation application

This section details the feasible choices of degrees of freedom for the quantitative multivariate Satterthwaite approximation presented in Section 4.2, and explains how the accompanying numerical comparisons are computed. The first part of the section derives the five effective degrees-of-freedom criteria used in the comparisons. The second part gives an exact density representation for a sum of independent Wishart random matrices with different scale parameters. This representation is needed only for the numerical approximation of total variation distances in the two-dimensional experiment. The final part describes that numerical approximation and then reports the comparison of the proposed choices of ν .

Recall that

$$\mathfrak{T} = \sum_{j=1}^N \mathfrak{G}_j, \quad \mathfrak{G}_j \sim \mathcal{W}_d(\alpha_j, \Sigma_j), \quad \bar{\Sigma} = \mathbb{E}[\mathfrak{T}] = \sum_{j=1}^N \alpha_j \Sigma_j,$$

and that, for a proposed Wishart approximation $\mathfrak{W}_\nu \sim \mathcal{W}_d(\nu, \tilde{\Sigma}_\nu)$, the first moment matching condition fixes

$$\tilde{\Sigma}_\nu = \nu^{-1} \bar{\Sigma}.$$

Thus the only remaining scalar parameter is ν , which plays the role of an effective number of degrees of freedom. As shown in (4.4), exact matching of the second moment would require

$$\frac{2}{\nu} \mathcal{V}(\bar{\Sigma}) = 2 \sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j), \quad (9.29)$$

which is generally an overdetermined matrix equation for the single scalar ν . The first four criteria in the next section are scalar reductions of this second moment matching condition, while the fifth is obtained by minimizing the geometric part of the Stein bound.

S.1.1. Effective degrees of freedom criteria

Tan and Gupta (1983) proposed to match the *generalized* variances, that is, the determinants of the covariance matrices in (9.29). This gives

$$\nu_{\text{TG}} := \left(\frac{|\mathcal{V}(\bar{\Sigma})|}{|\sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j)|} \right)^{2/(d(d+1))}.$$

Alternatively, Khatri (1989) proposed to match the *total* variances, that is, the traces of the covariance matrices in (9.29). This gives

$$\nu_{\text{K}} := \frac{\text{tr}\{\mathcal{V}(\bar{\Sigma})\}}{\sum_{j=1}^N \alpha_j \text{tr}\{\mathcal{V}(\Sigma_j)\}}.$$

We also consider a least-squares choice of approximate degrees of freedom, namely

$$\nu_{\text{LS}} \in \arg \min_{\nu > 0} \left\| 2 \sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j) - \frac{2}{\nu} \mathcal{V}(\bar{\Sigma}) \right\|_F^2.$$

Equivalently, writing $\beta = \nu^{-1}$, the inverse degrees of freedom ν_{LS}^{-1} is the optimal coefficient for the least-squares projection of $\sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j)$ onto the ray spanned by $\mathcal{V}(\bar{\Sigma})$, viz. $\beta \mapsto \beta \mathcal{V}(\bar{\Sigma})$. Since $\langle \sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j), \mathcal{V}(\bar{\Sigma}) \rangle_F > 0$, the minimizer is unique and equals

$$\nu_{\text{LS}} = \frac{\|\mathcal{V}(\bar{\Sigma})\|_F^2}{\langle \sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j), \mathcal{V}(\bar{\Sigma}) \rangle_F}.$$

Let $p := d(d+1)/2$ and define the upper absolute sum $\Delta(W) := \sum_{1 \leq a \leq b \leq p} |W_{ab}|$ for $W \in \mathcal{S}^p$. Matching this scalar summary of the two covariance matrices in (9.29) gives

$$\nu_{\text{US}} := \frac{\Delta\{\mathcal{V}(\bar{\Sigma})\}}{\Delta\{\sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j)\}}.$$

Finally, one may choose the degrees of freedom by minimizing the geometric part of the upper bound in Proposition 4.3. Since the target shape parameter in the Wishart Stein bound must satisfy $\nu > 3d - 3$, we fix $\nu_0 > 3d - 3$ and define

$$\nu_{\text{UB}} \in \arg \min_{\nu \geq \nu_0} G(\nu), \quad G(\nu) := \sum_{j=1}^N \alpha_j \text{tr}(\Sigma_j) \|\Sigma_j - \nu^{-1} \bar{\Sigma}\|_F.$$

This objective $G(\nu)$ is convex in ν^{-1} , since each summand is the Frobenius norm of an affine function of ν^{-1} . The minimizer is unique whenever at least one Σ_j is not proportional to $\bar{\Sigma}$; otherwise, the set of minimizers may be an interval, and any minimizer may be selected.

Proposition S.1 (Order of the effective degrees of freedom). *Let $d \in \mathbb{N}$ be fixed and let $N = N_n$ depend on some positive integer parameter n (e.g., a sample size). Assume that $\alpha_1, \dots, \alpha_N > 0$ and that there exist constants $0 < c < C < \infty$, independent of n and j , such that*

$$cn \leq \sum_{j=1}^N \alpha_j \leq Cn, \quad \frac{c}{n} I_d \preceq \Sigma_j \preceq \frac{C}{n} I_d, \quad j \in [N],$$

where $\preceq$ denotes the Loewner order. Let $\nu_0 > 3d - 3$ be fixed, and let ν_{UB} be any minimizer of G over $[\nu_0, \infty)$. Then, as $n \rightarrow \infty$,

$$\nu_{\text{TG}} \asymp n, \quad \nu_{\text{K}} \asymp n, \quad \nu_{\text{LS}} \asymp n, \quad \nu_{\text{US}} \asymp n, \quad \nu_{\text{UB}} \asymp n.$$

In particular, each of the above choices satisfies $\nu = \mathcal{O}(n)$.

Remark S.1. Recall the Stein bound from Proposition 4.3:

$$|\mathbb{E}[h(\mathfrak{T})] - \mathbb{E}[h(\mathfrak{W}_\nu)]| \leq \frac{d+1}{2} \mathcal{M}_2^{\mathcal{D}, \tilde{\Sigma}_\nu}(h) \sum_{j=1}^N \alpha_j \text{tr}(\Sigma_j) \|\Sigma_j - \tilde{\Sigma}_\nu\|_F. \quad (9.30)$$

Assume that d is fixed and that $\|\Sigma_j\|_F = \mathcal{O}(n^{-1})$ uniformly in j , as $n \rightarrow \infty$. Since Σ_j is positive definite, this also implies $\text{tr}(\Sigma_j) = \mathcal{O}(n^{-1})$ uniformly in j . If $\sum_{j=1}^N \alpha_j = \mathcal{O}(n)$ and $\nu \asymp n$, then $\|\bar{\Sigma}\|_F = \mathcal{O}(1)$ and $\|\tilde{\Sigma}_\nu\|_F = \mathcal{O}(n^{-1})$. Hence,

$$\|\Sigma_j - \tilde{\Sigma}_\nu\|_F = \mathcal{O}(n^{-1})$$

uniformly in j . Therefore, the geometric factor (i.e., the sum) in the upper bound (9.30) satisfies

$$\sum_{j=1}^N \alpha_j \text{tr}(\Sigma_j) \|\Sigma_j - \tilde{\Sigma}_\nu\|_F = \mathcal{O}\left(n^{-2} \sum_{j=1}^N \alpha_j\right) = \mathcal{O}(n^{-1}).$$

Consequently, if $\mathcal{M}_2^{\mathcal{D}, \tilde{\Sigma}_\nu}(h) = \mathcal{O}(1)$ along the sequence, or if the test functions are normalized so that $\mathcal{M}_2^{\mathcal{D}, \tilde{\Sigma}_\nu}(h) = 1$, then the full upper bound is of order $\mathcal{O}(n^{-1})$. This applies, in particular, when N is fixed and $\alpha_j = \mathcal{O}(n)$ for all j , and when $N \asymp n$ and $\alpha_j = \mathcal{O}(1)$ uniformly in j .

Proof of Proposition S.1. Define

$$C_\Sigma := \sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j), \quad V_{\bar{\Sigma}} := \mathcal{V}(\bar{\Sigma}),$$

and let $p = d(d+1)/2$. The assumptions imply

$$c^2 I_d \preceq \bar{\Sigma} \preceq C^2 I_d,$$

and hence $\bar{\Sigma} \asymp I_d$ under the Loewner order. Since $W \mapsto \mathcal{V}(W)$ is homogeneous of degree two and is uniformly positive definite on compact subsets of $\mathcal{S}_{++}^d$ in vecp-coordinates (indeed, if the spectrum of W lies in a compact subinterval of $(0, \infty)$, then so does that of $\mathcal{V}(W) = B_d^\top (W \otimes W) B_d$ by the Rayleigh quotient, since the eigenvalues of $W \otimes W$ are pairwise products of those of W and B_d has full column rank), we also have, under the Loewner order,

$$V_{\bar{\Sigma}} \asymp I_p.$$

Moreover, since $\Sigma_j \asymp n^{-1} I_d$ uniformly in j , we have $\mathcal{V}(\Sigma_j) \asymp n^{-2} I_p$ uniformly in j . Therefore

$$C_\Sigma = \sum_{j=1}^N \alpha_j \mathcal{V}(\Sigma_j) \asymp n^{-1} I_p.$$

For the TG criterion $\nu_{\text{TG}} = (|V_{\bar{\Sigma}}|/|C_\Sigma|)^{1/p}$, $|V_{\bar{\Sigma}}| \asymp 1$ and $|C_\Sigma| \asymp n^{-p}$, so we get $\nu_{\text{TG}} \asymp n$. For Khatri's criterion, $\text{tr}(V_{\bar{\Sigma}}) \asymp 1$ and $\text{tr}(C_\Sigma) \asymp n^{-1}$, so $\nu_{\text{K}} \asymp n$. For the least-squares criterion, $\|V_{\bar{\Sigma}}\|_F^2 \asymp 1$ and $\langle C_\Sigma, V_{\bar{\Sigma}} \rangle_F \asymp n^{-1}$, so $\nu_{\text{LS}} \asymp n$. Finally, $\Delta(V_{\bar{\Sigma}}) \asymp 1$ and $\Delta(C_\Sigma) \asymp n^{-1}$, so $\nu_{\text{US}} \asymp n$.

It remains to consider ν_{UB} . Let

$$A_j := n \Sigma_j, \quad \beta := \nu^{-1}, \quad t := n \beta.$$

Then $A_j \asymp I_d$ uniformly in j , and minimizing $G(\nu)$ over $\nu \geq \nu_0$ is equivalent, up to the irrelevant multiplicative factor n^{-2} , to minimizing

$$H_n(t) := \sum_{j=1}^N \alpha_j \text{tr}(A_j) \|A_j - t \bar{\Sigma}\|_F, \quad 0 < t \leq \frac{n}{\nu_0}.$$

The matrices A_j and $\bar{\Sigma}$ are uniformly bounded and uniformly positive definite. Hence there exist constants $a, b > 0$, independent of n and j , such that

$$a I_d \preceq A_j \preceq b I_d, \quad a I_d \preceq \bar{\Sigma} \preceq b I_d.$$

Since A_j and $\bar{\Sigma}$ are uniformly positive definite, $\langle A_j, \bar{\Sigma} \rangle_F = \text{tr}(\bar{\Sigma}^{1/2} A_j \bar{\Sigma}^{1/2})$ is bounded away from zero uniformly in j , while $\|\bar{\Sigma}\|_F^2$ is uniformly bounded. Hence, choosing $t_0 > 0$ sufficiently small, uniformly in j and for all $0 \leq t \leq 2t_0$,

$$t\|\bar{\Sigma}\|_F^2 - \langle A_j, \bar{\Sigma} \rangle_F < 0.$$

Moreover, $A_j - t\bar{\Sigma} \neq 0$ on this interval. Thus, for all $0 \leq t \leq 2t_0$ and all $j \in [N]$,

$$\frac{d}{dt}\|A_j - t\bar{\Sigma}\|_F = \frac{t\|\bar{\Sigma}\|_F^2 - \langle A_j, \bar{\Sigma} \rangle_F}{\|A_j - t\bar{\Sigma}\|_F} < 0.$$

Since the weights $\alpha_j \text{tr}(A_j)$ are positive, H_n is strictly decreasing on $[0, 2t_0]$. Hence no minimizer belongs to $(0, t_0]$.

Similarly, choose $T < \infty$ sufficiently large. Then, uniformly in j and for all $t \geq T$,

$$t\|\bar{\Sigma}\|_F^2 - \langle A_j, \bar{\Sigma} \rangle_F > 0,$$

and $A_j - t\bar{\Sigma} \neq 0$. Therefore H_n is strictly increasing on $[T, \infty)$, and no minimizer belongs to (T, ∞) . For all sufficiently large n , the interval $(0, n/\nu_0]$ contains $[t_0, T]$. Hence every minimizer t_n^* of H_n satisfies

$$t_0 \leq t_n^* \leq T.$$

Since $\nu_{\text{UB}} = n/t_n^*$, it follows that

$$\frac{n}{T} \leq \nu_{\text{UB}} \leq \frac{n}{t_0},$$

and therefore $\nu_{\text{UB}} \asymp n$. □

S.1.2. Density of a sum of independent Wisharts with different shape and scale parameters

The next two sections provide the density and integration tools used for the total variation comparison in Section S.1.4. Unlike the Stein bound, the total variation distance requires evaluating the density of $\mathfrak{T}$ itself. When the scale matrices Σ_j are not all equal, the sum $\mathfrak{T}$ is not Wishart in general, and its density is not given by a single Wishart density. The following formula rewrites the convolution of the Wishart densities in matrix Dirichlet coordinates. It expresses the density $f_{\mathfrak{T}}$ of $\mathfrak{T}$ as a reference Wishart density multiplied by a correction factor $\Psi(X)$, and this is the formula inserted into the numerical total variation estimators in Section S.1.3.

Proposition S.2. *Let $N \geq 2$, and let $\alpha_j > d - 1$ and $\Sigma_j \in \mathcal{S}_{++}^d$ for all $j \in [N]$. Let $\mathfrak{G}_1, \dots, \mathfrak{G}_N$ be independent random matrices such that $\mathfrak{G}_j \sim \mathcal{W}_d(\alpha_j, \Sigma_j)$ for all $j \in [N]$. Then $\mathfrak{T} = \sum_{j=1}^N \mathfrak{G}_j$ has density*

$$f_{\mathfrak{T}}(X) = \frac{|\Sigma_N|^{\alpha_{\bullet}/2}}{\prod_{j=1}^N |\Sigma_j|^{\alpha_j/2}} f_{\alpha_{\bullet}, \Sigma_N}(X) \Psi(X), \quad X \in \mathcal{S}_{++}^d,$$

where $\alpha_{\bullet} := \sum_{j=1}^N \alpha_j$ and

$$\Psi(X) := \mathbb{E} \left[\text{etr} \left\{ -\frac{1}{2} \sum_{j=1}^{N-1} X^{1/2} (\Sigma_j^{-1} - \Sigma_N^{-1}) X^{1/2} \mathfrak{U}_j \right\} \right],$$

with $(\mathfrak{U}_1, \dots, \mathfrak{U}_{N-1}) \sim \text{MatrixDirichlet}_{d \times d}(\alpha_1, \dots, \alpha_N)$ ([Gupta and Nagar, 2000](#), Definition 2.6.1).

Proof of Proposition S.2. Let f_j denote the density of $\mathcal{W}_d(\alpha_j, \Sigma_j)$. For $X \in \mathcal{S}_{++}^d$,

$$f_{\mathfrak{T}}(X) = \int_{\substack{G_1 \succ 0, \dots, G_{N-1} \succ 0 \\ X - \sum_{j=1}^{N-1} G_j \succ 0}} \left\{ \prod_{j=1}^{N-1} f_j(G_j) \right\} f_N \left(X - \sum_{j=1}^{N-1} G_j \right) dG_1 \cdots dG_{N-1}.$$

Substituting the Wishart densities gives

$$\begin{aligned} f_{\mathfrak{T}}(X) &= C \int_{\substack{G_1 \succ 0, \dots, G_{N-1} \succ 0 \\ X - \sum_{j=1}^{N-1} G_j \succ 0}} \left\{ \prod_{j=1}^{N-1} |G_j|^{(\alpha_j - d - 1)/2} \right\} \left| X - \sum_{j=1}^{N-1} G_j \right|^{(\alpha_N - d - 1)/2} \\ &\quad \times \text{etr} \left(-\frac{1}{2} \left\{ \sum_{j=1}^{N-1} \Sigma_j^{-1} G_j + \Sigma_N^{-1} \left(X - \sum_{j=1}^{N-1} G_j \right) \right\} \right) dG_1 \cdots dG_{N-1}, \end{aligned}$$

where

$$C := \left(2^{\alpha_{\bullet} d/2} \prod_{j=1}^N \{ |\Sigma_j|^{\alpha_j/2} \Gamma_d(\alpha_j/2) \} \right)^{-1}.$$

Make the change of variables $G_j = X^{1/2} U_j X^{1/2}$ for $j = 1, \dots, N-1$. The integration domain becomes

$$\Delta_{d, N-1} := \left\{ (U_1, \dots, U_{N-1}) : U_j \succ 0, j = 1, \dots, N-1, I_d - \sum_{j=1}^{N-1} U_j \succ 0 \right\},$$

and on calculating the Jacobian, we obtain $dG_1 \cdots dG_{N-1} = |X|^{(N-1)(d+1)/2} dU_1 \cdots dU_{N-1}$. Hence, using the cyclic symmetry of the trace,

$$\begin{aligned} f_{\mathfrak{T}}(X) &= C |X|^{(\alpha_{\bullet} - d - 1)/2} \text{etr} \left(-\frac{1}{2} \Sigma_N^{-1} X \right) \\ &\quad \times \int_{\Delta_{d, N-1}} \left\{ \prod_{j=1}^{N-1} |U_j|^{(\alpha_j - d - 1)/2} \right\} \left| I_d - \sum_{j=1}^{N-1} U_j \right|^{(\alpha_N - d - 1)/2} \\ &\quad \times \text{etr} \left\{ -\frac{1}{2} \sum_{j=1}^{N-1} X^{1/2} (\Sigma_j^{-1} - \Sigma_N^{-1}) X^{1/2} U_j \right\} dU_1 \cdots dU_{N-1}. \end{aligned}$$

The density of $(\mathfrak{U}_1, \dots, \mathfrak{U}_{N-1}) \sim \text{MatrixDirichlet}_{d \times d}(\alpha_1, \dots, \alpha_N)$ is

$$f_{(\mathfrak{U}_1, \dots, \mathfrak{U}_{N-1})}(U_1, \dots, U_{N-1}) = \frac{\Gamma_d(\alpha_{\bullet}/2)}{\prod_{j=1}^N \Gamma_d(\alpha_j/2)} \left\{ \prod_{j=1}^{N-1} |U_j|^{(\alpha_j - d - 1)/2} \right\} \left| I_d - \sum_{j=1}^{N-1} U_j \right|^{(\alpha_N - d - 1)/2}.$$

It follows that

$$f_{\mathfrak{T}}(X) = \left(2^{\alpha_{\bullet} d/2} \left\{ \prod_{j=1}^N |\Sigma_j|^{\alpha_j/2} \right\} \Gamma_d(\alpha_{\bullet}/2) \right)^{-1} |X|^{(\alpha_{\bullet} - d - 1)/2} \text{etr} \left(-\frac{1}{2} \Sigma_N^{-1} X \right) \Psi(X).$$

Since

$$f_{\alpha_{\bullet}, \Sigma_N}(X) = \left(2^{\alpha_{\bullet} d/2} |\Sigma_N|^{\alpha_{\bullet}/2} \Gamma_d(\alpha_{\bullet}/2) \right)^{-1} |X|^{(\alpha_{\bullet} - d - 1)/2} \text{etr} \left(-\frac{1}{2} \Sigma_N^{-1} X \right),$$

the claimed expression follows. $\square$

S.1.3. Numerical approximation of the total variation distance

We now describe the numerical approximation of the total variation distance used to compare the feasible choices of ν in Section S.1.4. The purpose of this section is computational: Proposition S.2 reduces the evaluation of $f_{\mathfrak{T}}(X)$ to the evaluation of the scalar factor $\Psi(X)$, and the total variation estimators below repeatedly use this pointwise density evaluation. In the implementation used for the $d = 2$, $N = 3$ experiment reported in Section S.1.4, $\Psi(X)$ is evaluated deterministically by applying Weyl's integration formula to the matrix Dirichlet integral, rather than by simulating from the matrix Dirichlet distribution. This avoids introducing an additional inner Monte Carlo error at every value of X .

Let $d = 2$ and set

$$\beta_j := \frac{\alpha_j - 3}{2}, \quad j = 1, 2, 3, \quad \beta_{23} := \beta_2 + \beta_3 + \frac{3}{2}.$$

For fixed $X \in \mathcal{S}_{++}^2$, define

$$A_j(X) := X^{1/2}(\Sigma_j^{-1} - \Sigma_3^{-1})X^{1/2}, \quad j = 1, 2.$$

In the matrix Dirichlet integral, we first use the change of variables

$$U_2 = (I_2 - U_1)^{1/2}V(I_2 - U_1)^{1/2}, \quad 0 \prec V \prec I_2.$$

The Jacobian contributes the factor $|I_2 - U_1|^{3/2}$, so that

$$\begin{aligned} \Psi(X) &= c_\alpha \int_{0 \prec U_1 \prec I_2} \int_{0 \prec V \prec I_2} |U_1|^{\beta_1} |I_2 - U_1|^{\beta_{23}} |V|^{\beta_2} |I_2 - V|^{\beta_3} \\ &\quad \times \operatorname{etr} \left\{ -\frac{1}{2} A_1(X) U_1 - \frac{1}{2} A_2(X) (I_2 - U_1)^{1/2} V (I_2 - U_1)^{1/2} \right\} dV dU_1, \end{aligned}$$

where

$$c_\alpha := \frac{\Gamma_2(\alpha_\bullet/2)}{\Gamma_2(\alpha_1/2)\Gamma_2(\alpha_2/2)\Gamma_2(\alpha_3/2)}.$$

For a symmetric 2×2 matrix M with $0 \prec M \prec I_2$, we write the ordered eigendecomposition

$$M = R_\theta \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} R_\theta^\top, \quad 0 < \lambda_2 < \lambda_1 < 1, \quad 0 \leq \theta < \pi,$$

where

$$R_\theta := \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

Weyl's formula gives

$$dM = (\lambda_1 - \lambda_2) d\lambda_1 d\lambda_2 d\theta.$$

We apply this formula once to U_1 and once to V . For U_1 , write $\lambda_1 = r$ and $\lambda_2 = rq$, with $r, q \in (0, 1)$. For V , write $\mu_1 = s$ and $\mu_2 = st$, with $s, t \in (0, 1)$. Define

$$U(r, q, \theta) := R_\theta \begin{pmatrix} r & 0 \\ 0 & rq \end{pmatrix} R_\theta^\top, \quad V(s, t, \phi) := R_\phi \begin{pmatrix} s & 0 \\ 0 & st \end{pmatrix} R_\phi^\top,$$

and

$$B(r, q, \theta) := \{I_2 - U(r, q, \theta)\}^{1/2}.$$

Then the integral becomes

$$\begin{aligned} \Psi(X) = c_\alpha \int_{[0,1]^4} \int_{[0,\pi]^2} \text{etr} \left\{ -\frac{1}{2} A_1(X) U(r, q, \theta) - \frac{1}{2} A_2(X) B(r, q, \theta) V(s, t, \phi) B(r, q, \theta) \right\} \\ \times r^{2\beta_1+2} q^{\beta_1} (1-q)(1-r)^{\beta_{23}} (1-rq)^{\beta_{23}} s^{2\beta_2+2} t^{\beta_2} (1-t)(1-s)^{\beta_3} (1-st)^{\beta_3} d\phi d\theta dt ds dq dr. \end{aligned}$$

This is the formula implemented numerically for the density evaluations in the two-dimensional total variation calculation. The variables r, q, s, t are integrated using Gauss-Jacobi quadrature rules adapted to the endpoint weights, and the angular variables θ, ϕ are integrated using Gauss-Legendre quadrature on $[0, \pi]$. The cross factors $(1-rq)^{\beta_{23}}$ and $(1-st)^{\beta_3}$, together with the exponential factor, are evaluated at the quadrature nodes. The implementation evaluates the sum on the log scale when needed, and checks the accuracy of the quadrature by setting the exponential factor equal to one, in which case the integral should be close to one after multiplication by c_α .

The resulting approximation of $\Psi(X)$ is inserted into the density formula from Proposition S.2 with $N = 3$. Thus, in the numerical calculations, $f_{\mathfrak{T}}(X)$ is evaluated as

$$\widehat{f}_{\mathfrak{T}}(X) = \frac{|\Sigma_3|^{\alpha_\bullet/2}}{\prod_{j=1}^3 |\Sigma_j|^{\alpha_j/2}} f_{\alpha_\bullet, \Sigma_3}(X) \widehat{\Psi}(X),$$

where $\widehat{\Psi}(X)$ denotes the Weyl quadrature approximation defined by the preceding integral.

We compare the law of $\mathfrak{T}$ with that of $\mathfrak{W}_\nu \sim \mathcal{W}_2(\nu, \widetilde{\Sigma}_\nu)$, where $\widetilde{\Sigma}_\nu = \nu^{-1} \overline{\Sigma}$ and ν is chosen by one of the five effective degrees-of-freedom criteria. The total variation distance is

$$d_{\text{TV}}(\mathfrak{T}, \mathfrak{W}_\nu) = \frac{1}{2} \int_{\mathcal{S}_{++}} |f_{\mathfrak{T}}(X) - f_{\nu, \widetilde{\Sigma}_\nu}(X)| dX,$$

which we can approximate using a Bartlett unit-cube estimator, motivated by the identity

$$d_{\text{TV}}(\mathfrak{T}, \mathfrak{W}_\nu) = \frac{1}{2} \mathbb{E}_{\mathfrak{W}_\nu} \left[\left| \frac{f_{\mathfrak{T}}(\mathfrak{W}_\nu)}{f_{\nu, \widetilde{\Sigma}_\nu}(\mathfrak{W}_\nu)} - 1 \right| \right] = \frac{1}{2} \mathbb{E}_{\mathfrak{W}_\nu} \left[|\exp\{\ell_\nu(\mathfrak{W}_\nu)\} - 1| \right],$$

where

$$\ell_\nu(X) = \log f_{\mathfrak{T}}(X) - \log f_{\nu, \widetilde{\Sigma}_\nu}(X)$$

is the log-likelihood ratio. The expectation with respect to $\mathfrak{W}_\nu$ is then mapped to an integral over the unit cube by means of the Bartlett representation of Wishart random matrices (see e.g. [Muirhead, 1982](#), Section 3.2.4). Let $L_\nu L_\nu^\top = \widetilde{\Sigma}_\nu$ and $u = (u_1, u_2, u_3) \in (0, 1)^3$. Define

$$C_\nu(u) := \begin{pmatrix} \sqrt{F_{\chi_\nu^2}^{-1}(u_1)} & 0 \\ \Phi^{-1}(u_3) & \sqrt{F_{\chi_{\nu-1}^2}^{-1}(u_2)} \end{pmatrix},$$

where $F_{\chi_\nu^2}^{-1}$ is the quantile function of the chi-square distribution with ν degrees of freedom and Φ^{-1} is the standard normal quantile function. We then set

$$X_\nu(u) := L_\nu C_\nu(u) C_\nu(u)^\top L_\nu^\top.$$

If $\mathbf{v} = (v_1, v_2, v_3)$ is a random vector uniformly distributed on $(0, 1)^3$, then

$$F_{\chi_\nu^2}^{-1}(v_1) \sim \chi_\nu^2, \quad F_{\chi_{\nu-1}^2}^{-1}(v_2) \sim \chi_{\nu-1}^2, \quad \Phi^{-1}(v_3) \sim \mathcal{N}(0, 1).$$

Hence $C_\nu(\mathbf{v})$ is precisely the lower-triangular Bartlett factor for a two-dimensional Wishart random matrix with ν degrees of freedom. Consequently,

$$X_\nu(\mathbf{v}) = L_\nu C_\nu(\mathbf{v}) C_\nu(\mathbf{v})^\top L_\nu^\top \sim \mathcal{W}_2(\nu, \tilde{\Sigma}_\nu).$$

Therefore, for any integrable h on $\mathcal{S}_{++}^2$,

$$\mathbb{E}_{\mathfrak{W}_\nu}[h(\mathfrak{W}_\nu)] = \int_{\mathcal{S}_{++}^2} h(X) f_{\nu, \tilde{\Sigma}_\nu}(X) dX = \int_{(0,1)^3} h(X_\nu(\mathbf{v})) d\mathbf{v}. \quad (9.31)$$

Indeed, if $Y = L_\nu^{-1} X L_\nu^{-\top}$, then the inverse Bartlett coordinates are obtained from the Cholesky factor of Y :

$$u_1 = F_{\chi_\nu^2}(Y_{11}), \quad u_2 = F_{\chi_{\nu-1}^2}\left(Y_{22} - \frac{Y_{12}^2}{Y_{11}}\right), \quad u_3 = \Phi\left(\frac{Y_{12}}{\sqrt{Y_{11}}}\right),$$

and (9.31) follows from observing that the Jacobian of this transformation is precisely the Wishart density. Thus, we can rewrite the total variation distance as

$$d_{\text{TV}}(\mathfrak{T}, \mathfrak{W}_\nu) = \frac{1}{2} \int_{(0,1)^3} |\exp\{\ell_\nu(X_\nu(\mathbf{v}))\} - 1| d\mathbf{v}.$$

In practice, we will use the Weyl quadrature approximation of the log-likelihood ratio,

$$\begin{aligned} \widehat{\ell}_\nu(X) &:= \log \widehat{f}_{\mathfrak{T}}(X) - \log f_{\nu, \tilde{\Sigma}_\nu}(X) \\ &= \log \widehat{\Psi}(X) + (\nu - \alpha_\bullet) \log 2 + \log \Gamma_2(\nu/2) - \log \Gamma_2(\alpha_\bullet/2) + \frac{\nu}{2} \log |\tilde{\Sigma}_\nu| \\ &\quad - \frac{1}{2} \sum_{j=1}^3 \alpha_j \log |\Sigma_j| + \frac{\alpha_\bullet - \nu}{2} \log |X| - \frac{1}{2} \text{tr}\{(\Sigma_3^{-1} - \tilde{\Sigma}_\nu^{-1})X\}. \end{aligned}$$

Thus, for Sobol points $u_1, \dots, u_M \in (0, 1)^3$, the Bartlett unit-cube approximation to the total variation distance is

$$\widehat{\text{TV}}(\nu) := \frac{1}{2M} \sum_{m=1}^M \left| \exp\{\widehat{\ell}_\nu(X_\nu(u_m))\} - 1 \right|. \quad (9.32)$$

In the actual numerical implementation, the unit-cube coordinates are clipped by a very small numerical tolerance before applying the quantile functions. The computations over the points u_m are independent and are therefore parallelized by splitting the point set into batches.

S.1.4. Numerical comparisons

The purpose of the numerical study is to determine whether the geometric part of the Stein bound gives a reliable ordering of the scalar Satterthwaite reductions from the point of view of total variation distance. We therefore do not use the numerical experiment only to compare

the values of the different effective degrees of freedom: rather, for each configuration of the summands, we compare two orderings on the set

$$\mathcal{M} = \{\text{TG}, \text{K}, \text{LS}, \text{US}\}.$$

The first ordering is induced by the estimated total variation distance, while the second ordering is induced by the geometric part of the bound. The upper-bound choice ν_{UB} is not included in this ranking comparison: although it is useful as a diagnostic reference, it is obtained by directly minimizing the geometric term and is not a scalar reduction of the covariance-matching problem in the same sense as TG, K, LS and US. This distinction is also consistent with the one-dimensional situation, where the minimizer of the geometric upper-bound criterion may have a worse total variation performance than the classical Satterthwaite degree of freedom.

Comparisons setup. Throughout the experiment, we consider independent random matrices $\mathfrak{G}_j \sim \mathcal{W}_2(\alpha_j, \Sigma_j)$, $j = 1, 2, 3$, and their sum $\mathfrak{T} = \mathfrak{G}_1 + \mathfrak{G}_2 + \mathfrak{G}_3$. For each $k \in \mathcal{M}$, we write ν_k for the corresponding scalar reduction and

$$\mathfrak{W}_k \sim \mathcal{W}_2(\nu_k, \bar{\Sigma}/\nu_k), \quad \bar{\Sigma} = \sum_{j=1}^3 \alpha_j \Sigma_j.$$

The geometric quantity associated with k is denoted by

$$B(\nu_k) = \sum_{j=1}^3 \alpha_j \text{tr}(\Sigma_j) \left\| \Sigma_j - \nu_k^{-1} \bar{\Sigma} \right\|_F,$$

whereas the empirical total variation estimate is denoted by $\widehat{\text{TV}}(\nu_k)$, as defined in (9.32). All reported computations use the Bartlett-QMC estimator with 2500 Sobol points, and 8 quadrature nodes per coordinate for the Weyl quadrature. Throughout, we will denote

$$k_{\text{TV}} \in \arg \min_{k \in \mathcal{M}} \widehat{\text{TV}}(\nu_k), \quad k_B \in \arg \min_{k \in \mathcal{M}} B(\nu_k).$$

Description of the scenarios. The configurations are organized into three ranking regimes. The first is the diagonal common-eigenspace regime. Here $\Sigma_i = \Lambda_i$, where

$$\Lambda_i = \begin{pmatrix} \lambda_{i,1} & 0 \\ 0 & \lambda_{i,2} \end{pmatrix}, \quad \lambda_{i,1}, \lambda_{i,2} > 0.$$

The ratios $\lambda_{i,1}/\lambda_{i,2}$ are not all equal, so the scale matrices are not merely proportional. This regime is the cleanest non-trivial test case, since the matrices commute and the discrepancy is driven by eigenvalue heterogeneity rather than by competing eigenspaces. The eighteen base configurations are listed in Table S.1. This diagonal framework will be presented alongside the second regime: that is, the common-rotation regime where $\Sigma_i = R_\theta \Lambda_i R_\theta^\top$ with the same rotation matrix R_θ for all i . This is still a common-eigenspace regime, but it is no longer diagonal in the canonical coordinates unless $R_\theta = I_2$. Each of the eighteen base configurations is combined with the following five common rotation settings:

$$\theta \in \{0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ\}.$$

The third regime is the heterogeneous-rotation regime, where $\Sigma_i = R_{\theta_i} \Lambda_i R_{\theta_i}^\top$ and the rotations R_{θ_i} are not all equal. This is the genuine non-commuting stress test. The rotation profiles used here are angle vectors $(\theta_1, \theta_2, \theta_3)$. Each of the eighteen base configurations is combined with the following five heterogeneous rotation profiles:

$$(0^\circ, 10^\circ, -10^\circ), \quad (0^\circ, 20^\circ, -20^\circ), \quad (-30^\circ, 0^\circ, 30^\circ), \quad (0^\circ, 35^\circ, 70^\circ), \quad (-45^\circ, 15^\circ, 55^\circ).$$

Profile	Scenario	$(\alpha_1, \alpha_2, \alpha_3)$	Λ_1	Λ_2	Λ_3
mild crossing	01	(22, 22, 22)	(1.300, 0.900)	(1.050, 1.050)	(0.850, 1.300)
	02	(14, 22, 30)	(1.300, 0.900)	(1.050, 1.050)	(0.850, 1.300)
	03	(30, 22, 14)	(1.300, 0.900)	(1.050, 1.050)	(0.850, 1.300)
same-axis anisotropy	04	(22, 22, 22)	(1.500, 0.850)	(1.300, 0.950)	(1.100, 1.050)
	05	(14, 22, 30)	(1.500, 0.850)	(1.300, 0.950)	(1.100, 1.050)
	06	(30, 22, 14)	(1.500, 0.850)	(1.300, 0.950)	(1.100, 1.050)
opposite-axis anisotropy	07	(22, 22, 22)	(1.600, 0.750)	(1.050, 1.050)	(0.750, 1.600)
	08	(14, 22, 30)	(1.600, 0.750)	(1.050, 1.050)	(0.750, 1.600)
	09	(30, 22, 14)	(1.600, 0.750)	(1.050, 1.050)	(0.750, 1.600)
one eccentric summand	10	(22, 22, 22)	(1.750, 0.700)	(0.950, 1.200)	(0.950, 1.200)
	11	(14, 22, 30)	(1.750, 0.700)	(0.950, 1.200)	(0.950, 1.200)
	12	(30, 22, 14)	(1.750, 0.700)	(0.950, 1.200)	(0.950, 1.200)
approx. equal determinants	13	(22, 22, 22)	(1.600, 0.625)	(1.250, 0.800)	(0.900, 1.111)
	14	(14, 22, 30)	(1.600, 0.625)	(1.250, 0.800)	(0.900, 1.111)
	15	(30, 22, 14)	(1.600, 0.625)	(1.250, 0.800)	(0.900, 1.111)
mostly scalar scale contrast	16	(22, 22, 22)	(1.450, 1.250)	(0.950, 1.050)	(0.750, 0.950)
	17	(14, 22, 30)	(1.450, 1.250)	(0.950, 1.050)	(0.750, 0.950)
	18	(30, 22, 14)	(1.450, 1.250)	(0.950, 1.050)	(0.750, 0.950)

Table S.1: The eighteen base scenarios used in the diagonal common-eigenspace regime. The pair in column Λ_i denotes $(\lambda_{i,1}, \lambda_{i,2})$.

Agreement and rank correlation. We first report, for each regime, the agreement proportion: that is, the proportion of scenarios where $k_{TV} = k_B$. In the diagonal regime, 94% of the scenarios agree on the best choice, and the criteria are practically indistinguishable when agreement does not hold. In the common-rotation regime, this proportion decreases to 58%; in the heterogeneous-rotation regime, it is 70%. We also report Kendall’s rank correlation τ for each scenario, and display the numerical values in Figure S.1.

We observe that the ordering of a majority of scenarios, namely 66%, is perfectly described by the geometric criterion: in these configurations, the rank correlation is exactly 1. These configurations cover every scenario where $k_{TV} = k_B$. On the other hand, a substantial number of exact minimizers change after a common rotation, as indicated by the left panel of Figure S.1 covering the second regime. This behavior is consistent with the coordinate sensitivity of some scalar reductions, especially those involving entrywise or unscaled half-vectorized summaries,

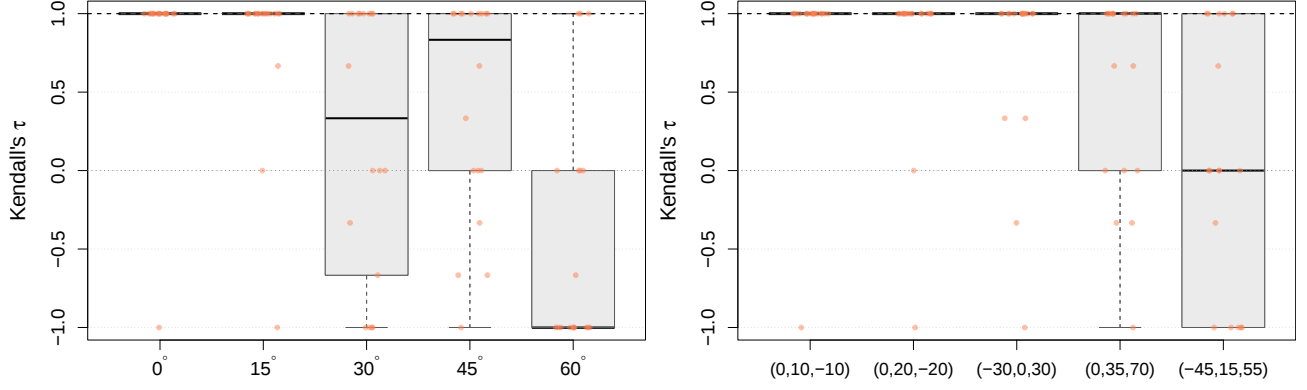


Figure S.1: Dispersion of Kendall's rank correlation, in the considered regimes.

that is, the LS and US criteria. Finally, it appears that the third regime is less sensitive to this phenomenon.

Regret analysis. We also report the relative total variation loss of choosing the geometric choice k_B over k_{TV} ,

$$\text{Loss}(k_B) := \frac{\widehat{\text{TV}}(\nu_{k_B})}{\widehat{\text{TV}}(\nu_{k_{TV}})} - 1.$$

This quantity is equal to 0 when the geometric criterion agrees with the minimizer of the total variation. Values close to 0 indicate that, even if the exact minimizer differs, the discrepancy is practically negligible in total variation. The numerical values are displayed in Figure S.2; most losses are below 5%, although a few outlying configurations have substantially larger losses.

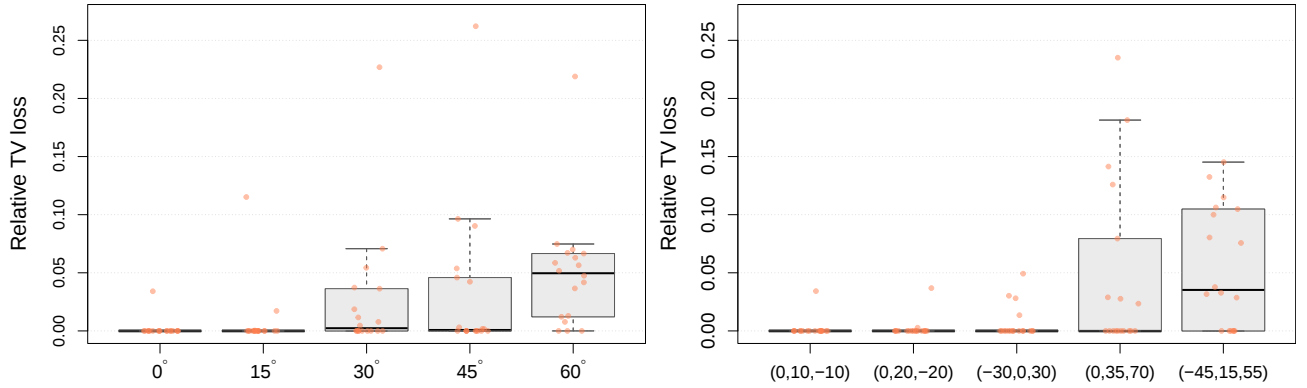


Figure S.2: Regret of choosing the minimizer of the geometric criterion, in the considered regimes.

S.2. Additional details for Stein's method-of-moments application

S.2.1. Numerical comparison of the log-MOM, quad-MOM, and MLE estimators

This section details the numerical experiment mentioned in Remark 4.7 and reported in Tables S.2 and 4.1. In this experiment, the values of α_0 and Σ_0 were selected randomly before the Monte Carlo replications. More precisely, ten values of α_0 were generated independently from the

uniform distribution on $[3, 30]$. Ten positive definite matrices Σ_0 were generated independently as follows. With $d = 3$, eigenvalues were drawn according to

$$\lambda_1, \lambda_2, \lambda_3 \stackrel{\text{iid}}{\sim} \text{LogUnif}(0.35, 3.50),$$

and an orthogonal matrix Q was generated by applying the QR construction to a matrix with independent standard normal entries. We then set

$$\Sigma_0 = Q \text{diag}(\lambda_1, \dots, \lambda_d) Q^\top.$$

The generated values of α_0 were crossed with the generated values of Σ_0 , giving one hundred configurations in total. For each configuration, we simulated 500 independent samples from $\mathcal{W}_3(\alpha_0, \Sigma_0)$, with sample sizes $n \in \{10, 10^2, 10^3, 10^4\}$. Within each replication, the samples were nested across n . The corresponding values of α_0 and Σ_0 are displayed in the first column of Table S.2, rounded to one decimal place for α_0 and two decimal places for the entries of Σ_0 .

For each replication, sample size, and parameter configuration, we computed three pairs of estimators: $(\hat{\alpha}_{\log}, \hat{\Sigma}_{\log})$, $(\hat{\alpha}_{\text{quad}}, \hat{\Sigma}_{\text{quad}})$, and $(\hat{\alpha}_{\text{MLE}}, \hat{\Sigma}_{\text{MLE}})$. The first two are Stein's method-of-moments estimators in Proposition 4.8, while the last one is the maximum likelihood estimator (MLE). More precisely, writing

$$\widehat{M} = \frac{1}{n} \sum_{k=1}^n \mathfrak{M}^{(k)},$$

the profile likelihood maximizer in the scale parameter, for fixed α , is $\widehat{\Sigma}(\alpha) = \widehat{M}/\alpha$. The MLE of the shape parameter is therefore obtained by solving the scalar equation

$$\overline{\log |\mathfrak{M}|} - \log |\widehat{M}| + d \log(\alpha) - d \log(2) - \psi_d(\alpha/2) = 0, \quad \alpha > d - 1,$$

where

$$\psi_d(a) := \sum_{j=1}^d \psi \left(a + \frac{1-j}{2} \right).$$

The scale matrix is then set to $\widehat{\Sigma}_{\text{MLE}} = \widehat{M}/\widehat{\alpha}_{\text{MLE}}$.

The denominators in (4.16) and (4.17) were checked against numerical zero. The log-MOM and quad-MOM estimates were marked invalid if the relevant denominator was numerically zero, if the estimated shape parameter did not satisfy $\hat{\alpha} > d - 1$, or if the scale estimate failed the numerical positive-definiteness check. The MLE was marked invalid if the profile likelihood equation could not be solved reliably or if the resulting estimate failed the same parameter-space checks. Table S.2 reports, for each configuration and sample size, the median and interquartile range over the valid Monte Carlo replications of the relative Frobenius error $\|\widehat{\Sigma} - \Sigma_0\|_F / \|\Sigma_0\|_F$. Within each configuration and sample size, the smallest median error and the smallest interquartile range among the three estimators are shown in bold. Table 4.1 gives the corresponding counts, by sample size, of how often each estimator attains the smallest median error or the smallest interquartile range. The MLE is included as an efficiency benchmark. The logarithmic and quadratic Stein's method-of-moments estimators are simpler to compute, since they are given explicitly by (4.16) and (4.17), and do not require solving the likelihood equation. This

experiment is intended to check whether the improvement of the logarithmic Stein's method-of-moments estimator over the quadratic, classical moment estimator persists across randomly generated parameter configurations.

Table S.2: Monte Carlo comparison of the logarithmic Stein method-of-moments estimator, the quadratic moment estimator, and the maximum likelihood estimator. Entries are the median and interquartile range, over $B = 500$ replications, of $\|\hat{\Sigma} - \Sigma_0\|_F / \|\Sigma_0\|_F$. The first column gives the value of α_0 and the explicit matrix Σ_0 . Within each (α_0, Σ_0) configuration and sample size, the smallest median and the smallest interquartile range are shown in bold.

α_0, Σ_0	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
10.8, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.225	0.134	0.0697	0.0402	0.0215	0.0113	0.00681	0.00437
	quad-MOM	0.248	0.15	0.0757	0.0482	0.024	0.0136	0.00743	0.00515
	MLE	0.227	0.126	0.0675	0.0365	0.021	0.0112	0.00673	0.00407
10.8, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.222	0.117	0.0709	0.0373	0.0216	0.0118	0.00713	0.00415
	quad-MOM	0.238	0.133	0.0783	0.0456	0.0234	0.013	0.00746	0.00455
	MLE	0.217	0.109	0.07	0.0366	0.0212	0.0119	0.00704	0.00384
10.8, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.214	0.124	0.0713	0.0395	0.0215	0.0125	0.00694	0.0037
	quad-MOM	0.24	0.154	0.0778	0.0472	0.0228	0.0147	0.00772	0.00462
	MLE	0.211	0.109	0.0696	0.0369	0.021	0.0126	0.0068	0.00359
10.8, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.225	0.117	0.0734	0.0364	0.0228	0.0106	0.00711	0.00351
	quad-MOM	0.239	0.134	0.0755	0.0431	0.0243	0.0128	0.00748	0.00395
	MLE	0.225	0.108	0.0736	0.0355	0.0226	0.0105	0.00698	0.00348
10.8, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.223	0.109	0.0703	0.0366	0.0231	0.0109	0.00704	0.00339
	quad-MOM	0.234	0.126	0.0726	0.04	0.0239	0.0125	0.00737	0.00401
	MLE	0.224	0.117	0.0695	0.0379	0.0228	0.0105	0.0071	0.0036
10.8, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.225	0.145	0.0635	0.0479	0.0205	0.0139	0.00646	0.0044
	quad-MOM	0.265	0.186	0.0736	0.0684	0.026	0.0215	0.00776	0.006
	MLE	0.206	0.129	0.0607	0.0408	0.0195	0.0122	0.00631	0.00382
10.8, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.219	0.194	0.0603	0.0591	0.0193	0.0176	0.00627	0.00605
	quad-MOM	0.271	0.266	0.0854	0.0899	0.0273	0.0271	0.00854	0.00979
	MLE	0.193	0.163	0.0518	0.0458	0.0168	0.0143	0.0052	0.00465
10.8, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.218	0.181	0.0611	0.0559	0.0193	0.0155	0.00654	0.00439
	quad-MOM	0.277	0.236	0.0808	0.0791	0.0244	0.0258	0.0085	0.00744
	MLE	0.189	0.152	0.0564	0.0452	0.0183	0.0139	0.0058	0.00413
10.8, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.211	0.148	0.0702	0.045	0.0218	0.0132	0.00645	0.0043
	quad-MOM	0.24	0.176	0.08	0.0592	0.0251	0.0191	0.00761	0.00557
	MLE	0.2	0.127	0.0671	0.0414	0.0204	0.0121	0.00615	0.00356
10.8, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.221	0.123	0.0693	0.0405	0.0218	0.0124	0.00679	0.00391
	quad-MOM	0.239	0.153	0.0754	0.0498	0.0229	0.0148	0.0073	0.0049
	MLE	0.211	0.108	0.0682	0.0367	0.0214	0.0118	0.00671	0.00373
13.0, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.208	0.122	0.0627	0.0366	0.0202	0.0127	0.00627	0.00353
	quad-MOM	0.224	0.146	0.0676	0.0369	0.0212	0.0159	0.0067	0.00432
	MLE	0.204	0.125	0.0628	0.0365	0.0199	0.0113	0.00621	0.00331
13.0, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.22	0.115	0.0654	0.0366	0.0212	0.0116	0.00674	0.00386
	quad-MOM	0.234	0.14	0.0685	0.0375	0.0224	0.0132	0.00731	0.00481
	MLE	0.217	0.108	0.0643	0.0355	0.0208	0.012	0.00662	0.00358
13.0, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.203	0.124	0.066	0.0408	0.0206	0.0132	0.00653	0.00341
	quad-MOM	0.217	0.149	0.0691	0.0483	0.0227	0.0165	0.00684	0.00455
	MLE	0.202	0.122	0.0658	0.0362	0.0201	0.0123	0.00656	0.00349

Continued on next page

α_0, Σ_0	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
13.0, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.215	0.122	0.0648	0.0338	0.0207	0.0108	0.00666	0.00363
	quad-MOM	0.236	0.132	0.0673	0.0373	0.0213	0.012	0.00697	0.00387
	MLE	0.214	0.114	0.0658	0.0335	0.0207	0.0106	0.00671	0.00344
13.0, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.218	0.102	0.0669	0.0326	0.0217	0.0104	0.00683	0.00342
	quad-MOM	0.232	0.122	0.0698	0.0388	0.022	0.012	0.00715	0.00387
	MLE	0.214	0.0952	0.0672	0.0322	0.0214	0.0104	0.00681	0.00333
13.0, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.198	0.137	0.06	0.0452	0.02	0.0131	0.00637	0.00446
	quad-MOM	0.247	0.189	0.0695	0.0611	0.0235	0.0175	0.00796	0.00571
	MLE	0.19	0.118	0.0574	0.0412	0.018	0.0118	0.00575	0.00423
13.0, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.202	0.19	0.0629	0.061	0.021	0.0194	0.00562	0.00566
	quad-MOM	0.264	0.248	0.0877	0.0945	0.0293	0.0309	0.00796	0.00887
	MLE	0.173	0.154	0.0512	0.0446	0.0176	0.0151	0.00494	0.00445
13.0, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.204	0.17	0.0568	0.0501	0.0184	0.0155	0.00579	0.00515
	quad-MOM	0.251	0.233	0.0776	0.0745	0.0243	0.0226	0.00778	0.00759
	MLE	0.182	0.141	0.0559	0.0401	0.0164	0.0139	0.00531	0.00422
13.0, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.213	0.135	0.0675	0.0478	0.0205	0.0136	0.00638	0.00478
	quad-MOM	0.244	0.162	0.0776	0.0611	0.0224	0.0189	0.00711	0.00639
	MLE	0.204	0.134	0.0591	0.0418	0.0193	0.013	0.00589	0.00414
13.0, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.206	0.139	0.0669	0.0424	0.0204	0.0121	0.00644	0.00389
	quad-MOM	0.228	0.162	0.0745	0.0556	0.0234	0.0146	0.00696	0.0049
	MLE	0.196	0.121	0.0644	0.0412	0.0203	0.0122	0.00623	0.00402
13.5, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.212	0.114	0.0646	0.0365	0.0205	0.0125	0.00663	0.00371
	quad-MOM	0.217	0.131	0.0686	0.0415	0.0219	0.0156	0.00695	0.00467
	MLE	0.202	0.114	0.0639	0.0326	0.0198	0.0116	0.00641	0.00354
13.5, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.214	0.13	0.0637	0.0345	0.0208	0.0115	0.00652	0.00352
	quad-MOM	0.232	0.134	0.0689	0.0405	0.0221	0.0129	0.00698	0.00457
	MLE	0.207	0.125	0.0641	0.0333	0.0202	0.0116	0.00644	0.00336
13.5, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.197	0.113	0.0661	0.0371	0.0202	0.0113	0.00617	0.00319
	quad-MOM	0.22	0.148	0.0717	0.0483	0.0212	0.0136	0.00672	0.00414
	MLE	0.196	0.109	0.0646	0.0352	0.0198	0.0113	0.00618	0.00311
13.5, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.216	0.116	0.0639	0.0371	0.021	0.0114	0.00657	0.00343
	quad-MOM	0.224	0.125	0.0652	0.0387	0.0217	0.0131	0.00693	0.00419
	MLE	0.217	0.117	0.0646	0.0374	0.0209	0.0108	0.00653	0.00317
13.5, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.216	0.116	0.0672	0.0359	0.0203	0.0103	0.00672	0.00353
	quad-MOM	0.231	0.132	0.0707	0.0435	0.0204	0.0112	0.007	0.00371
	MLE	0.212	0.113	0.0684	0.0342	0.0203	0.0101	0.00669	0.00327
13.5, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.199	0.153	0.0601	0.0432	0.0194	0.0153	0.00594	0.00423
	quad-MOM	0.232	0.178	0.0751	0.0609	0.0225	0.0195	0.00729	0.00578
	MLE	0.187	0.131	0.057	0.0345	0.0186	0.013	0.00564	0.00329
13.5, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.207	0.192	0.0599	0.0583	0.019	0.0183	0.00638	0.0059
	quad-MOM	0.288	0.258	0.0894	0.0898	0.0261	0.0284	0.00895	0.00958
	MLE	0.173	0.16	0.0515	0.0422	0.0164	0.0149	0.00542	0.00473
13.5, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.192	0.17	0.0569	0.0497	0.0183	0.0163	0.00591	0.00557
	quad-MOM	0.246	0.251	0.0748	0.0722	0.023	0.0251	0.00878	0.00885
	MLE	0.175	0.146	0.0532	0.0412	0.0171	0.0139	0.00563	0.00418
13.5, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.199	0.128	0.0607	0.0426	0.0201	0.0142	0.00646	0.00486
	quad-MOM	0.233	0.17	0.0693	0.0559	0.0219	0.0175	0.00723	0.00602
	MLE	0.197	0.125	0.0591	0.0416	0.0194	0.0119	0.00635	0.00403
13.5, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.206	0.131	0.0633	0.0387	0.0204	0.0128	0.00644	0.00428
	quad-MOM	0.228	0.16	0.0681	0.0493	0.0223	0.0154	0.00728	0.00504
	MLE	0.201	0.112	0.0608	0.0361	0.0192	0.0124	0.00617	0.00391

Continued on next page

α_0, Σ_0	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
15.8, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.209	0.12	0.0581	0.0344	0.0197	0.0121	0.00625	0.00376
	quad-MOM	0.234	0.138	0.0638	0.0409	0.022	0.0142	0.00682	0.00419
	MLE	0.201	0.116	0.0577	0.0335	0.0194	0.011	0.00601	0.00354
15.8, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.207	0.131	0.0634	0.0361	0.0192	0.0117	0.00595	0.00361
	quad-MOM	0.224	0.132	0.0671	0.0425	0.0206	0.0133	0.00651	0.00437
	MLE	0.203	0.122	0.0623	0.0374	0.0193	0.0115	0.00605	0.00344
15.8, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.206	0.121	0.0615	0.0398	0.0198	0.0135	0.00621	0.00409
	quad-MOM	0.216	0.143	0.068	0.0484	0.0213	0.0154	0.00683	0.00467
	MLE	0.197	0.115	0.061	0.0366	0.0198	0.0123	0.00607	0.0038
15.8, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.203	0.128	0.0644	0.0349	0.0194	0.0106	0.00632	0.0037
	quad-MOM	0.21	0.144	0.0671	0.0429	0.0199	0.0118	0.00677	0.0043
	MLE	0.204	0.123	0.0641	0.0332	0.0195	0.0111	0.00637	0.00349
15.8, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.205	0.114	0.0624	0.0375	0.02	0.0122	0.00634	0.00342
	quad-MOM	0.22	0.131	0.0636	0.0415	0.0208	0.0142	0.00656	0.00359
	MLE	0.206	0.11	0.0626	0.0356	0.0197	0.0118	0.00641	0.00334
15.8, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.194	0.136	0.0584	0.0458	0.0192	0.0138	0.00588	0.00444
	quad-MOM	0.223	0.175	0.0678	0.0577	0.0221	0.0188	0.00698	0.00643
	MLE	0.179	0.126	0.0552	0.0387	0.0175	0.0125	0.00575	0.00377
15.8, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.209	0.198	0.0609	0.0597	0.0188	0.0189	0.00595	0.00624
	quad-MOM	0.277	0.246	0.08	0.0867	0.0273	0.0305	0.00822	0.00993
	MLE	0.184	0.164	0.0478	0.0453	0.0149	0.0141	0.00478	0.00438
15.8, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.196	0.162	0.0558	0.0529	0.019	0.0174	0.00601	0.00552
	quad-MOM	0.25	0.22	0.0724	0.0792	0.0239	0.0253	0.00792	0.00863
	MLE	0.166	0.143	0.0498	0.0445	0.0168	0.0143	0.00528	0.00449
15.8, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.199	0.132	0.0609	0.0436	0.0193	0.0145	0.00571	0.00423
	quad-MOM	0.226	0.161	0.0685	0.0543	0.0221	0.019	0.00675	0.00579
	MLE	0.192	0.117	0.0576	0.0368	0.0183	0.0129	0.00557	0.00399
15.8, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.198	0.134	0.063	0.0361	0.019	0.0122	0.00612	0.00402
	quad-MOM	0.221	0.151	0.0667	0.0398	0.0206	0.0147	0.00679	0.00502
	MLE	0.199	0.132	0.0597	0.0381	0.0191	0.0119	0.00624	0.00399
18.5, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.198	0.121	0.0612	0.0379	0.0187	0.0107	0.00587	0.00384
	quad-MOM	0.208	0.148	0.0631	0.0426	0.0197	0.0128	0.00638	0.00453
	MLE	0.197	0.12	0.0584	0.0383	0.0181	0.0113	0.00559	0.00374
18.5, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.193	0.13	0.059	0.0342	0.0196	0.01	0.00595	0.00347
	quad-MOM	0.209	0.137	0.0614	0.0429	0.0205	0.013	0.00626	0.00465
	MLE	0.188	0.127	0.0582	0.0317	0.0191	0.0102	0.00593	0.00358
18.5, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.192	0.12	0.0571	0.0332	0.0188	0.0119	0.00569	0.00402
	quad-MOM	0.218	0.144	0.0625	0.0402	0.0204	0.0142	0.00601	0.00443
	MLE	0.188	0.11	0.0565	0.0313	0.0181	0.0116	0.00571	0.00382
18.5, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.198	0.128	0.059	0.0364	0.0187	0.0116	0.00603	0.00352
	quad-MOM	0.2	0.14	0.0605	0.0386	0.0195	0.0123	0.00617	0.0041
	MLE	0.203	0.121	0.059	0.0367	0.0186	0.0111	0.00605	0.00338
18.5, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.194	0.112	0.0612	0.0325	0.0189	0.0106	0.00612	0.00375
	quad-MOM	0.205	0.131	0.0629	0.0397	0.0188	0.0112	0.00634	0.004
	MLE	0.195	0.113	0.0615	0.0337	0.0187	0.0103	0.00611	0.00365
18.5, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.193	0.151	0.0524	0.0431	0.0183	0.0153	0.00578	0.00428
	quad-MOM	0.23	0.196	0.0659	0.0616	0.0212	0.0211	0.00691	0.00579
	MLE	0.173	0.148	0.0502	0.0365	0.0176	0.0144	0.00546	0.00404
18.5, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.195	0.194	0.0552	0.0569	0.0185	0.0188	0.00582	0.00562
	quad-MOM	0.266	0.256	0.0746	0.0897	0.0245	0.0277	0.0077	0.00814
	MLE	0.16	0.155	0.0491	0.0484	0.015	0.0146	0.0049	0.00473

Continued on next page

α_0, Σ_0	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
18.5, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.19	0.163	0.0566	0.0504	0.0186	0.0167	0.00526	0.00544
	quad-MOM	0.256	0.234	0.0733	0.0725	0.0232	0.0264	0.00717	0.00728
	MLE	0.167	0.14	0.0466	0.0447	0.0157	0.0137	0.0048	0.00451
18.5, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.189	0.147	0.0571	0.0419	0.0181	0.0132	0.00603	0.0046
	quad-MOM	0.217	0.183	0.0639	0.0556	0.0203	0.0169	0.00702	0.00562
	MLE	0.18	0.134	0.0525	0.0386	0.0174	0.0107	0.0054	0.00417
18.5, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.201	0.136	0.0574	0.0341	0.019	0.0133	0.00551	0.00346
	quad-MOM	0.215	0.161	0.0631	0.0449	0.0217	0.0169	0.00636	0.00433
	MLE	0.196	0.124	0.0557	0.0344	0.0184	0.0124	0.00549	0.00324
19.2, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.195	0.129	0.058	0.0395	0.0185	0.0114	0.00625	0.00413
	quad-MOM	0.21	0.138	0.0656	0.0501	0.0197	0.0148	0.0071	0.00508
	MLE	0.185	0.118	0.056	0.0388	0.0174	0.011	0.00588	0.00405
19.2, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.198	0.134	0.0589	0.0402	0.0178	0.0103	0.00604	0.00395
	quad-MOM	0.207	0.149	0.0617	0.0435	0.0196	0.0121	0.00652	0.00466
	MLE	0.194	0.129	0.0573	0.0403	0.018	0.0103	0.00577	0.00382
19.2, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.194	0.134	0.0561	0.0355	0.0181	0.0122	0.00572	0.00364
	quad-MOM	0.205	0.138	0.0619	0.0443	0.0192	0.0148	0.00621	0.00412
	MLE	0.196	0.132	0.0541	0.033	0.0177	0.0109	0.00577	0.00341
19.2, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.199	0.118	0.0637	0.0399	0.0191	0.0116	0.00603	0.00398
	quad-MOM	0.211	0.131	0.0646	0.0439	0.0197	0.0134	0.00642	0.00422
	MLE	0.195	0.11	0.0632	0.042	0.019	0.0113	0.00586	0.00379
19.2, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.196	0.114	0.0606	0.0354	0.0191	0.0102	0.00579	0.00346
	quad-MOM	0.201	0.125	0.0603	0.0403	0.0206	0.0117	0.00616	0.00344
	MLE	0.195	0.114	0.0606	0.0363	0.0189	0.0104	0.00571	0.00336
19.2, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.194	0.154	0.0562	0.0436	0.0181	0.0145	0.00557	0.00451
	quad-MOM	0.231	0.186	0.0622	0.0543	0.0221	0.0194	0.00661	0.00614
	MLE	0.187	0.136	0.0538	0.0376	0.0164	0.0132	0.00526	0.00393
19.2, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.201	0.2	0.0567	0.0589	0.0172	0.0178	0.00598	0.00643
	quad-MOM	0.267	0.257	0.0784	0.0835	0.0254	0.0261	0.00815	0.00949
	MLE	0.166	0.159	0.0479	0.0479	0.015	0.0138	0.00476	0.0041
19.2, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.197	0.181	0.0536	0.0496	0.0177	0.0162	0.00539	0.00514
	quad-MOM	0.244	0.254	0.0697	0.0798	0.0239	0.0245	0.00722	0.00836
	MLE	0.183	0.153	0.0474	0.0407	0.0156	0.0137	0.00491	0.00413
19.2, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.192	0.144	0.0578	0.0458	0.0181	0.0139	0.00552	0.00436
	quad-MOM	0.223	0.187	0.066	0.0589	0.0208	0.0187	0.00646	0.00596
	MLE	0.181	0.141	0.0521	0.0398	0.0165	0.0128	0.00539	0.00392
19.2, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.183	0.135	0.0568	0.0352	0.0188	0.0128	0.00594	0.00417
	quad-MOM	0.207	0.145	0.063	0.046	0.0203	0.0163	0.00652	0.00467
	MLE	0.175	0.123	0.0557	0.0359	0.0176	0.0116	0.00556	0.00398
21.1, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.185	0.123	0.0581	0.0404	0.0183	0.0108	0.00556	0.00372
	quad-MOM	0.2	0.15	0.0645	0.0492	0.0184	0.0132	0.00601	0.00463
	MLE	0.184	0.119	0.058	0.0376	0.0176	0.0114	0.00566	0.00349
21.1, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.191	0.117	0.0589	0.0398	0.0176	0.0117	0.00552	0.00338
	quad-MOM	0.209	0.135	0.0633	0.0449	0.0189	0.013	0.00569	0.00396
	MLE	0.187	0.111	0.0585	0.0373	0.0172	0.0109	0.00551	0.00345
21.1, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.198	0.144	0.0599	0.0374	0.0177	0.0123	0.00581	0.00391
	quad-MOM	0.213	0.159	0.0622	0.0456	0.0184	0.0132	0.00622	0.00473
	MLE	0.198	0.128	0.0585	0.0365	0.0179	0.011	0.00571	0.00395
21.1, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.189	0.119	0.0566	0.0356	0.0185	0.0114	0.00564	0.00369
	quad-MOM	0.203	0.137	0.0611	0.0376	0.0197	0.0124	0.00612	0.00399
	MLE	0.188	0.114	0.0541	0.0349	0.0185	0.0118	0.00548	0.00355

Continued on next page

α_0, Σ_0	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
21.1, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.186	0.125	0.0585	0.035	0.0189	0.0111	0.00579	0.00377
	quad-MOM	0.194	0.133	0.0611	0.038	0.0195	0.0136	0.00601	0.00398
	MLE	0.187	0.118	0.0587	0.035	0.0188	0.0111	0.00576	0.00372
21.1, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.19	0.153	0.0534	0.0463	0.0168	0.014	0.0058	0.00463
	quad-MOM	0.216	0.192	0.0651	0.0561	0.021	0.0201	0.00647	0.0063
	MLE	0.169	0.142	0.0522	0.0406	0.0166	0.0131	0.00544	0.0044
21.1, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.185	0.201	0.0531	0.0568	0.0179	0.018	0.00539	0.00587
	quad-MOM	0.245	0.268	0.0793	0.0916	0.0242	0.0286	0.00758	0.00851
	MLE	0.155	0.146	0.0468	0.0441	0.0158	0.0152	0.00459	0.0046
21.1, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.207	0.185	0.0557	0.0563	0.0173	0.0163	0.00527	0.00499
	quad-MOM	0.242	0.228	0.0722	0.0744	0.0243	0.0239	0.00694	0.00786
	MLE	0.172	0.149	0.0489	0.0453	0.015	0.0125	0.00461	0.00394
21.1, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.189	0.141	0.0546	0.043	0.0178	0.0138	0.0055	0.00455
	quad-MOM	0.204	0.172	0.0619	0.0488	0.0205	0.0189	0.0065	0.0055
	MLE	0.183	0.134	0.0529	0.0354	0.0164	0.0115	0.0053	0.00368
21.1, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.183	0.13	0.0541	0.0365	0.0176	0.0109	0.00553	0.0043
	quad-MOM	0.21	0.158	0.0606	0.0442	0.0188	0.0142	0.00599	0.00474
	MLE	0.175	0.127	0.0533	0.0336	0.0165	0.0113	0.00556	0.00378
22.0, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.188	0.133	0.0533	0.0379	0.0174	0.0138	0.00545	0.00353
	quad-MOM	0.207	0.141	0.0569	0.0453	0.0187	0.0162	0.00598	0.00404
	MLE	0.189	0.107	0.0529	0.0337	0.0173	0.013	0.00538	0.00353
22.0, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.188	0.132	0.0583	0.041	0.0179	0.0128	0.00547	0.00332
	quad-MOM	0.194	0.146	0.0645	0.0435	0.0194	0.0147	0.00587	0.00397
	MLE	0.184	0.134	0.0568	0.0388	0.0176	0.0113	0.00539	0.00327
22.0, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.186	0.117	0.0583	0.0364	0.0178	0.0111	0.00536	0.00381
	quad-MOM	0.203	0.133	0.0621	0.0414	0.0191	0.0145	0.00575	0.00456
	MLE	0.181	0.111	0.0578	0.0387	0.0175	0.0105	0.00534	0.00354
22.0, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.192	0.129	0.0561	0.0398	0.0174	0.0109	0.00565	0.00394
	quad-MOM	0.196	0.14	0.0587	0.0379	0.0186	0.0115	0.00602	0.00441
	MLE	0.188	0.127	0.0548	0.0399	0.0176	0.0109	0.00557	0.00367
22.0, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.184	0.12	0.0588	0.0404	0.0182	0.0113	0.00559	0.00354
	quad-MOM	0.196	0.134	0.0612	0.0432	0.0182	0.0121	0.00569	0.00382
	MLE	0.184	0.121	0.0598	0.0396	0.0186	0.0111	0.00565	0.00345
22.0, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.186	0.146	0.0532	0.046	0.0167	0.0139	0.0057	0.00439
	quad-MOM	0.21	0.184	0.0626	0.0613	0.0207	0.0195	0.00687	0.00542
	MLE	0.179	0.126	0.0496	0.0394	0.0158	0.0122	0.00534	0.00368
22.0, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.189	0.201	0.0558	0.0566	0.0165	0.0199	0.00542	0.00521
	quad-MOM	0.243	0.269	0.0781	0.0861	0.0231	0.0302	0.00763	0.00872
	MLE	0.16	0.158	0.047	0.0446	0.0145	0.0145	0.00468	0.00443
22.0, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.189	0.162	0.0548	0.0513	0.0179	0.0162	0.0054	0.00535
	quad-MOM	0.245	0.232	0.072	0.0787	0.025	0.0235	0.00711	0.00784
	MLE	0.161	0.143	0.0482	0.0393	0.0155	0.0137	0.00482	0.00436
22.0, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.196	0.151	0.0571	0.0447	0.0181	0.0133	0.00551	0.00465
	quad-MOM	0.219	0.177	0.0657	0.0532	0.02	0.0178	0.0066	0.00547
	MLE	0.187	0.127	0.055	0.0411	0.0167	0.0112	0.00511	0.0038
22.0, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.185	0.136	0.0568	0.0435	0.0168	0.0124	0.00566	0.00387
	quad-MOM	0.201	0.153	0.0654	0.0509	0.0187	0.0144	0.00602	0.00491
	MLE	0.179	0.127	0.0557	0.0411	0.0163	0.0111	0.00543	0.0036
23.1, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.192	0.129	0.0569	0.0423	0.0178	0.0114	0.00559	0.00397
	quad-MOM	0.206	0.145	0.0615	0.0477	0.0191	0.0149	0.0062	0.00446
	MLE	0.183	0.122	0.056	0.0414	0.017	0.0105	0.00564	0.00392

Continued on next page

α_0, Σ_0	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
23.1, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.191	0.126	0.0569	0.036	0.0174	0.0111	0.00558	0.00364
	quad-MOM	0.198	0.143	0.0586	0.0425	0.0183	0.0116	0.00591	0.0043
	MLE	0.185	0.113	0.0549	0.0365	0.0173	0.0115	0.00564	0.00366
23.1, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.177	0.14	0.0575	0.0408	0.0174	0.0128	0.00561	0.00401
	quad-MOM	0.192	0.17	0.0641	0.0479	0.0192	0.0149	0.00633	0.00456
	MLE	0.18	0.135	0.0572	0.0399	0.0177	0.0118	0.00555	0.0039
23.1, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.198	0.122	0.0547	0.0362	0.018	0.013	0.00544	0.0036
	quad-MOM	0.205	0.142	0.0591	0.0421	0.0195	0.0139	0.0059	0.00428
	MLE	0.197	0.128	0.0547	0.0338	0.0179	0.0125	0.00543	0.0037
23.1, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.19	0.119	0.0584	0.0341	0.0179	0.011	0.00555	0.00313
	quad-MOM	0.2	0.134	0.0592	0.0349	0.0182	0.0117	0.00566	0.00354
	MLE	0.19	0.119	0.0572	0.0348	0.0182	0.011	0.00553	0.00329
23.1, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.188	0.146	0.0518	0.0445	0.0182	0.0157	0.00554	0.00434
	quad-MOM	0.213	0.176	0.0593	0.0588	0.0228	0.0198	0.00613	0.0057
	MLE	0.18	0.14	0.05	0.04	0.0166	0.0126	0.00536	0.00415
23.1, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.189	0.181	0.0541	0.0599	0.0174	0.0176	0.00535	0.00565
	quad-MOM	0.255	0.249	0.0762	0.084	0.0222	0.0263	0.00803	0.00977
	MLE	0.15	0.148	0.0473	0.0441	0.0144	0.0134	0.00442	0.00421
23.1, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.186	0.185	0.0568	0.0585	0.0172	0.0173	0.00536	0.00549
	quad-MOM	0.254	0.231	0.0746	0.0777	0.0227	0.026	0.00702	0.00819
	MLE	0.164	0.148	0.0524	0.0469	0.0149	0.0142	0.00493	0.00423
23.1, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.188	0.152	0.056	0.0476	0.0176	0.0142	0.00534	0.00422
	quad-MOM	0.214	0.179	0.0624	0.0578	0.0212	0.0179	0.00591	0.00587
	MLE	0.171	0.121	0.0511	0.0389	0.0167	0.0106	0.00519	0.00372
23.1, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.189	0.143	0.0554	0.0419	0.0184	0.0128	0.00549	0.00364
	quad-MOM	0.211	0.158	0.0615	0.0519	0.0203	0.0161	0.00603	0.00461
	MLE	0.185	0.132	0.0527	0.0372	0.0173	0.013	0.00526	0.00325
29.5, $\begin{pmatrix} 1.31 & -0.06 & 0.11 \\ -0.06 & 1.13 & 0.40 \\ 0.11 & 0.40 & 0.97 \end{pmatrix}$	log-MOM	0.177	0.137	0.0555	0.042	0.0162	0.0117	0.00528	0.00402
	quad-MOM	0.19	0.151	0.0565	0.0479	0.0175	0.013	0.00558	0.00487
	MLE	0.172	0.135	0.0533	0.0418	0.0158	0.0117	0.00514	0.0036
29.5, $\begin{pmatrix} 1.40 & -0.08 & 0.09 \\ -0.08 & 1.39 & 0.00 \\ 0.09 & 0.00 & 0.80 \end{pmatrix}$	log-MOM	0.184	0.126	0.0577	0.0397	0.0166	0.0117	0.00512	0.0037
	quad-MOM	0.196	0.147	0.0608	0.0454	0.0174	0.0134	0.00552	0.00419
	MLE	0.177	0.12	0.057	0.0399	0.0163	0.011	0.005	0.00351
29.5, $\begin{pmatrix} 0.41 & 0.07 & 0.03 \\ 0.07 & 0.61 & 0.11 \\ 0.03 & 0.11 & 0.43 \end{pmatrix}$	log-MOM	0.174	0.128	0.051	0.0377	0.016	0.0117	0.00505	0.00325
	quad-MOM	0.191	0.151	0.0556	0.0429	0.0176	0.013	0.00567	0.00391
	MLE	0.166	0.129	0.0497	0.0399	0.0157	0.0118	0.00502	0.00341
29.5, $\begin{pmatrix} 2.08 & -0.28 & 0.05 \\ -0.28 & 1.83 & 0.05 \\ 0.05 & 0.05 & 2.21 \end{pmatrix}$	log-MOM	0.177	0.143	0.0527	0.0344	0.0175	0.0117	0.00533	0.00391
	quad-MOM	0.183	0.16	0.054	0.0425	0.0179	0.0133	0.00566	0.00436
	MLE	0.179	0.138	0.0526	0.0357	0.0176	0.0117	0.00527	0.00391
29.5, $\begin{pmatrix} 2.18 & -0.01 & 0.09 \\ -0.01 & 2.10 & -0.02 \\ 0.09 & -0.02 & 2.19 \end{pmatrix}$	log-MOM	0.179	0.124	0.0527	0.0393	0.0172	0.012	0.00521	0.0035
	quad-MOM	0.192	0.136	0.0533	0.0407	0.0185	0.0131	0.00556	0.00348
	MLE	0.18	0.122	0.0529	0.0383	0.017	0.0114	0.00519	0.00349
29.5, $\begin{pmatrix} 0.65 & 0.21 & -0.26 \\ 0.21 & 1.61 & -0.14 \\ -0.26 & -0.14 & 0.61 \end{pmatrix}$	log-MOM	0.182	0.168	0.0502	0.0477	0.0169	0.0153	0.00542	0.00462
	quad-MOM	0.206	0.196	0.0614	0.0602	0.0198	0.0204	0.00623	0.0068
	MLE	0.165	0.145	0.0474	0.0406	0.0149	0.0128	0.00503	0.00444
29.5, $\begin{pmatrix} 0.62 & -0.46 & -0.43 \\ -0.46 & 1.63 & 0.94 \\ -0.43 & 0.94 & 1.16 \end{pmatrix}$	log-MOM	0.185	0.179	0.056	0.0616	0.0175	0.0185	0.00519	0.00587
	quad-MOM	0.243	0.249	0.0791	0.0907	0.0233	0.0281	0.00729	0.00868
	MLE	0.152	0.154	0.0457	0.0459	0.0137	0.0137	0.00398	0.00439
29.5, $\begin{pmatrix} 2.63 & -0.10 & 0.62 \\ -0.10 & 0.71 & -0.18 \\ 0.62 & -0.18 & 0.76 \end{pmatrix}$	log-MOM	0.182	0.172	0.0538	0.0533	0.0169	0.0166	0.00498	0.00526
	quad-MOM	0.228	0.222	0.0672	0.0779	0.0227	0.0227	0.00668	0.00717
	MLE	0.156	0.146	0.0465	0.0457	0.0148	0.0132	0.00461	0.00425

Continued on next page

α_0, Σ_0	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
29.5, $\begin{pmatrix} 1.24 & 0.45 & 0.17 \\ 0.45 & 1.50 & -0.47 \\ 0.17 & -0.47 & 0.74 \end{pmatrix}$	log-MOM	0.176	0.155	0.0561	0.0481	0.0164	0.0131	0.00529	0.00458
	quad-MOM	0.207	0.183	0.0632	0.0627	0.0186	0.0175	0.00589	0.0056
	MLE	0.161	0.141	0.0507	0.0401	0.0151	0.0128	0.00495	0.004
29.5, $\begin{pmatrix} 1.38 & 0.24 & 0.37 \\ 0.24 & 1.09 & -0.38 \\ 0.37 & -0.38 & 1.63 \end{pmatrix}$	log-MOM	0.18	0.134	0.0541	0.043	0.0159	0.0126	0.0052	0.00381
	quad-MOM	0.196	0.168	0.0587	0.05	0.018	0.0147	0.00576	0.00441
	MLE	0.171	0.13	0.0505	0.0364	0.0155	0.0123	0.00507	0.00354

S.2.2. Numerical comparison of the projected and naive Stein MOM estimators

This section details the numerical experiment mentioned in Remark 4.9. The results are reported in Tables S.3, 4.2, and S.4. In this experiment, the values of α_0 and Σ_0 were selected randomly before the Monte Carlo replications, with Σ_0 constrained to belong to the two-dimensional compound-symmetry subspace

$$\Sigma_0 = \Sigma(\beta^*) = \beta_1^* B_1 + \beta_2^* B_2,$$

where $B_1 = I_{10}$ and $B_2 = \mathbf{1}_{10}\mathbf{1}_{10}^\top - I_{10}$. Ten values of α_0 were generated independently from the uniform distribution on $[12, 80]$. Ten structured matrices Σ_0 were generated independently by drawing the compound-symmetry eigenvalues

$$\lambda_\perp \sim \text{LogUnif}(0.50, 3.00), \quad \lambda_\parallel \sim \text{LogUnif}(0.50, 15.00),$$

and then setting, with $d = 10$,

$$\beta_2^* = \frac{\lambda_\parallel - \lambda_\perp}{d}, \quad \beta_1^* = \lambda_\perp + \beta_2^*.$$

The generated values of α_0 were crossed with the generated values of β^* , giving one hundred structured configurations in total. For each structured configuration, we simulated 500 independent samples from $\mathcal{W}_{10}(\alpha_0, \Sigma_0)$, with sample sizes $n \in \{10, 10^2, 10^3, 10^4\}$. Within each replication, the samples were nested across n . The corresponding values of α_0 and β^* are displayed in the first columns of Tables S.3 and S.4, rounded to one decimal place for α_0 and two decimal places for the coordinates of β^* .

The probe matrices $\{U_m\}_{m=1}^M$ were chosen as an orthonormal basis of $\mathcal{S}^{10}$ for the Frobenius inner product, and the templates were the matrices B_1 and B_2 above. For each replication, sample size, and structured configuration, we computed the projected Stein's method-of-moments estimator $\tilde{\Sigma}_n$ from Proposition 4.9 and the naive (unstructured) Stein's method-of-moments estimator

$$\tilde{\Sigma}_n := \frac{1}{\alpha_0 n} \sum_{k=1}^n \mathfrak{W}^{(k)}.$$

Writing

$$\widehat{M} = \frac{1}{n} \sum_{k=1}^n \mathfrak{W}^{(k)},$$

the unstructured MLE is $\widehat{M}/\alpha_0$, because α_0 is known. It is therefore identical to $\tilde{\Sigma}_n$ and is not included as a separate competitor. The reported projected and naive scale estimates were marked

invalid if the corresponding scale estimate failed the numerical positive-definiteness check. No shape parameter is estimated in this experiment, since α_0 is known.

Table S.3 reports, for each structured configuration and sample size, the median and interquartile range over the valid Monte Carlo replications of the relative Frobenius error $\|\widehat{\Sigma} - \Sigma_0\|_F / \|\Sigma_0\|_F$. Within each structured configuration and sample size, the smaller median error and the smaller interquartile range among the two estimators are shown in bold. Table 4.2 gives the corresponding counts, by sample size, of how often each estimator attains the smaller median error or the smaller interquartile range. Table S.4 reports the median and interquartile range of the ratio $\|\widehat{\Sigma}_n - \Sigma_0\|_F / \|\widehat{\Sigma}_n - \Sigma_0\|_F$, corresponding to the entries labeled proj/naive. Values below one favor the projected Stein's method-of-moments estimator. This experiment is intended to check whether the gain from projecting the Stein moment equations onto the structured subspace persists across randomly generated parameter configurations, while comparing the projected estimator with the usual unstructured moment benchmark.

Table S.3: Monte Carlo comparison of the projected Stein method-of-moments estimator and the (unstructured) Stein method-of-moments estimator. Entries are the median and interquartile range, over $B = 500$ replications, of $\|\widehat{\Sigma} - \Sigma_0\|_F / \|\Sigma_0\|_F$. Within each structured configuration and sample size, the smallest median and the smallest interquartile range are shown in bold.

α_0, β^*	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
23.3,(0.71, 0.18)	proj-SMOM	0.0562	0.0623	0.0185	0.0199	0.00536	0.00589	0.00175	0.00194
	naive-SMOM	0.171	0.0365	0.0551	0.012	0.0171	0.00316	0.00543	0.00101
23.3,(0.74, 0.09)	proj-SMOM	0.0471	0.0409	0.0152	0.0137	0.00453	0.0041	0.00149	0.00136
	naive-SMOM	0.203	0.0293	0.0646	0.00964	0.0206	0.00297	0.00644	0.000905
23.3,(3.01, 1.15)	proj-SMOM	0.0599	0.0652	0.0192	0.0229	0.00566	0.007	0.00165	0.00207
	naive-SMOM	0.139	0.0371	0.045	0.0127	0.0141	0.00379	0.00437	0.00114
23.3,(1.01, 0.50)	proj-SMOM	0.0613	0.0681	0.0178	0.023	0.00584	0.0074	0.00205	0.00254
	naive-SMOM	0.118	0.0381	0.0371	0.0139	0.0118	0.00437	0.00381	0.00148
23.3,(1.35, 0.35)	proj-SMOM	0.0573	0.0585	0.0182	0.0185	0.00588	0.00636	0.00157	0.00176
	naive-SMOM	0.171	0.0372	0.0538	0.00982	0.0172	0.00371	0.00531	0.000993
23.3,(1.22, 0.07)	proj-SMOM	0.0422	0.0346	0.0131	0.0103	0.00411	0.0037	0.00134	0.00123
	naive-SMOM	0.212	0.0306	0.0676	0.00898	0.0213	0.00289	0.00675	0.000904
23.3,(1.44, 0.44)	proj-SMOM	0.0525	0.0632	0.0188	0.0207	0.00609	0.00684	0.0016	0.00189
	naive-SMOM	0.16	0.035	0.0512	0.0126	0.0163	0.00395	0.00503	0.00109
23.3,(1.24, 0.22)	proj-SMOM	0.0493	0.0469	0.016	0.016	0.00503	0.00512	0.00163	0.00162
	naive-SMOM	0.192	0.0306	0.0609	0.00957	0.0192	0.00318	0.00606	0.000964
23.3,(0.73, 0.04)	proj-SMOM	0.0405	0.0342	0.0128	0.0109	0.00413	0.0032	0.00126	0.000996
	naive-SMOM	0.213	0.0323	0.0684	0.00882	0.0212	0.00275	0.00674	0.000945
23.3,(2.74, -0.02)	proj-SMOM	0.034	0.0276	0.0102	0.00732	0.00314	0.00246	0.00109	0.000863
	naive-SMOM	0.216	0.0275	0.0684	0.00809	0.0215	0.00294	0.00681	0.000929
34.1,(0.71, 0.18)	proj-SMOM	0.0437	0.0499	0.0145	0.016	0.00435	0.00481	0.00123	0.00132
	naive-SMOM	0.141	0.028	0.0458	0.00969	0.0143	0.0028	0.00444	0.000832
34.1,(0.74, 0.09)	proj-SMOM	0.0384	0.0317	0.0116	0.0106	0.00379	0.00373	0.00114	0.000949
	naive-SMOM	0.168	0.0237	0.053	0.00828	0.0169	0.00249	0.00531	0.000729
34.1,(3.01, 1.15)	proj-SMOM	0.047	0.0551	0.0151	0.0194	0.00561	0.00554	0.00156	0.00192
	naive-SMOM	0.114	0.0311	0.0369	0.0102	0.0121	0.00335	0.00378	0.00108
34.1,(1.01, 0.50)	proj-SMOM	0.0485	0.059	0.0148	0.0177	0.00482	0.00572	0.00156	0.00179
	naive-SMOM	0.0964	0.0368	0.0299	0.0111	0.00974	0.00355	0.00309	0.00114

Continued on next page

α_0, β^*	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
34.1,(1.35, 0.35)	proj-SMOM	0.0465	0.0474	0.0142	0.016	0.00483	0.0049	0.00141	0.00166
	naive-SMOM	0.141	0.0322	0.0449	0.00906	0.0144	0.00293	0.00451	0.00086
34.1,(1.22, 0.07)	proj-SMOM	0.0349	0.0263	0.0114	0.00982	0.00345	0.00296	0.00111	0.000905
	naive-SMOM	0.175	0.0285	0.0555	0.00722	0.0175	0.00265	0.00556	0.000708
34.1,(1.44, 0.44)	proj-SMOM	0.0484	0.0556	0.0164	0.0166	0.0045	0.00514	0.00141	0.00157
	naive-SMOM	0.132	0.0279	0.0424	0.00869	0.0131	0.00339	0.00419	0.000932
34.1,(1.24, 0.22)	proj-SMOM	0.0429	0.0416	0.0132	0.014	0.00432	0.00404	0.0013	0.00122
	naive-SMOM	0.158	0.0272	0.0503	0.00803	0.0159	0.00243	0.00498	0.000805
34.1,(0.73, 0.04)	proj-SMOM	0.0346	0.0276	0.0107	0.00954	0.00338	0.00263	0.00105	0.000814
	naive-SMOM	0.177	0.023	0.0559	0.00744	0.0177	0.00218	0.00555	0.00073
34.1,(2.74, -0.02)	proj-SMOM	0.0272	0.0211	0.00912	0.00649	0.00275	0.0022	0.000859	0.000613
	naive-SMOM	0.178	0.0242	0.0563	0.00654	0.0178	0.00214	0.00563	0.000746
43.0,(0.71, 0.18)	proj-SMOM	0.0417	0.0454	0.0124	0.0136	0.00416	0.00436	0.00124	0.00144
	naive-SMOM	0.124	0.026	0.0395	0.00778	0.0126	0.00264	0.00401	0.000812
43.0,(0.74, 0.09)	proj-SMOM	0.0344	0.0298	0.0108	0.00937	0.00356	0.00299	0.00116	0.000983
	naive-SMOM	0.151	0.0244	0.0469	0.00658	0.0151	0.00189	0.00479	0.000617
43.0,(3.01, 1.15)	proj-SMOM	0.044	0.0539	0.0135	0.0171	0.00486	0.00584	0.00139	0.00156
	naive-SMOM	0.104	0.0302	0.0329	0.00888	0.0106	0.00328	0.00329	0.000929
43.0,(1.01, 0.50)	proj-SMOM	0.0451	0.0524	0.0134	0.0166	0.00449	0.00539	0.00132	0.00174
	naive-SMOM	0.0875	0.0303	0.0272	0.0103	0.00868	0.00312	0.00276	0.00103
43.0,(1.35, 0.35)	proj-SMOM	0.0424	0.0452	0.0124	0.0138	0.00394	0.0043	0.00129	0.00145
	naive-SMOM	0.127	0.0256	0.0399	0.00784	0.0127	0.00249	0.00395	0.00079
43.0,(1.22, 0.07)	proj-SMOM	0.0303	0.0248	0.00949	0.00817	0.003	0.0024	0.000979	0.000765
	naive-SMOM	0.154	0.0204	0.0498	0.00687	0.0158	0.00237	0.00492	0.000692
43.0,(1.44, 0.44)	proj-SMOM	0.0406	0.0528	0.0127	0.0142	0.00426	0.00459	0.00134	0.00148
	naive-SMOM	0.118	0.0269	0.0369	0.00832	0.0118	0.00256	0.00376	0.00086
43.0,(1.24, 0.22)	proj-SMOM	0.037	0.0326	0.0112	0.0115	0.00391	0.004	0.00121	0.00116
	naive-SMOM	0.14	0.0251	0.0448	0.00726	0.0141	0.00261	0.00447	0.000726
43.0,(0.73, 0.04)	proj-SMOM	0.0308	0.0224	0.00916	0.00766	0.0031	0.00257	0.000956	0.00073
	naive-SMOM	0.158	0.0207	0.0497	0.00674	0.0157	0.00218	0.00497	0.00067
43.0,(2.74, -0.02)	proj-SMOM	0.0253	0.0201	0.00806	0.00671	0.00245	0.00196	0.000744	0.000573
	naive-SMOM	0.159	0.0217	0.0504	0.00684	0.016	0.00215	0.00505	0.000636
46.6,(0.71, 0.18)	proj-SMOM	0.0389	0.0427	0.0127	0.0134	0.00381	0.00389	0.00125	0.00118
	naive-SMOM	0.122	0.0232	0.0385	0.00766	0.0121	0.00235	0.00379	0.00072
46.6,(0.74, 0.09)	proj-SMOM	0.0318	0.0294	0.01	0.00904	0.00323	0.00293	0.000982	0.00081
	naive-SMOM	0.143	0.0211	0.0461	0.00697	0.0145	0.00204	0.00454	0.000611
46.6,(3.01, 1.15)	proj-SMOM	0.0408	0.0523	0.0135	0.0157	0.00375	0.00444	0.00126	0.0014
	naive-SMOM	0.101	0.0272	0.032	0.00886	0.00983	0.00287	0.00308	0.000858
46.6,(1.01, 0.50)	proj-SMOM	0.0441	0.0488	0.0149	0.0184	0.00432	0.00578	0.0014	0.00187
	naive-SMOM	0.082	0.029	0.0269	0.0109	0.00839	0.00375	0.00266	0.00109
46.6,(1.35, 0.35)	proj-SMOM	0.0377	0.0418	0.0116	0.0124	0.00388	0.00421	0.00121	0.00126
	naive-SMOM	0.12	0.0236	0.0382	0.00725	0.0121	0.00249	0.00383	0.000732
46.6,(1.22, 0.07)	proj-SMOM	0.0306	0.0255	0.00886	0.00718	0.00286	0.0023	0.000926	0.000745
	naive-SMOM	0.151	0.0199	0.0472	0.00628	0.0151	0.00229	0.00477	0.000629
46.6,(1.44, 0.44)	proj-SMOM	0.0427	0.0436	0.0128	0.0148	0.00425	0.00481	0.00136	0.00149
	naive-SMOM	0.114	0.0244	0.0355	0.00845	0.0113	0.00292	0.00359	0.000837
46.6,(1.24, 0.22)	proj-SMOM	0.0373	0.0326	0.0113	0.0114	0.00373	0.0038	0.00108	0.00103
	naive-SMOM	0.137	0.0245	0.0432	0.00658	0.0137	0.00224	0.00428	0.000688
46.6,(0.73, 0.04)	proj-SMOM	0.029	0.0224	0.009	0.00783	0.00279	0.00224	0.000905	0.00073
	naive-SMOM	0.151	0.0199	0.0474	0.00615	0.015	0.00203	0.00477	0.000616

Continued on next page

α_0, β^*	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
46.6, (2.74, -0.02)	proj-SMOM	0.0223	0.0183	0.00807	0.00592	0.00238	0.00192	0.000786	0.000581
	naive-SMOM	0.154	0.0201	0.0484	0.00594	0.0153	0.00193	0.00485	0.000644
61.0, (0.71, 0.18)	proj-SMOM	0.0368	0.0404	0.0113	0.0112	0.0032	0.00358	0.00105	0.00117
	naive-SMOM	0.107	0.0225	0.0333	0.00643	0.0106	0.00215	0.00332	0.000669
61.0, (0.74, 0.09)	proj-SMOM	0.0288	0.0245	0.00979	0.0083	0.00294	0.00258	0.000946	0.000841
	naive-SMOM	0.125	0.0199	0.04	0.00574	0.0127	0.00182	0.00401	0.000528
61.0, (3.01, 1.15)	proj-SMOM	0.0387	0.0446	0.0113	0.0129	0.00371	0.00425	0.00116	0.00131
	naive-SMOM	0.088	0.0234	0.0278	0.00683	0.00894	0.00246	0.00276	0.000646
61.0, (1.01, 0.50)	proj-SMOM	0.0375	0.0471	0.0114	0.0133	0.00389	0.00467	0.00113	0.00137
	naive-SMOM	0.0722	0.0253	0.0227	0.008	0.00725	0.00286	0.00229	0.000836
61.0, (1.35, 0.35)	proj-SMOM	0.0348	0.0331	0.01	0.0118	0.00318	0.00365	0.00112	0.00119
	naive-SMOM	0.106	0.02	0.0338	0.00719	0.0106	0.00207	0.00336	0.000711
61.0, (1.22, 0.07)	proj-SMOM	0.0254	0.0208	0.00809	0.00672	0.0025	0.00215	0.000809	0.000656
	naive-SMOM	0.131	0.0181	0.0413	0.00561	0.0132	0.0017	0.00417	0.000548
61.0, (1.44, 0.44)	proj-SMOM	0.0352	0.0391	0.0112	0.013	0.00366	0.00425	0.00114	0.00127
	naive-SMOM	0.0978	0.022	0.0316	0.00737	0.0101	0.00228	0.00311	0.00067
61.0, (1.24, 0.22)	proj-SMOM	0.0303	0.0304	0.0104	0.0105	0.00312	0.00313	0.00106	0.00109
	naive-SMOM	0.118	0.0204	0.0374	0.0065	0.0119	0.00195	0.00377	0.000599
61.0, (0.73, 0.04)	proj-SMOM	0.0257	0.0208	0.00829	0.00619	0.0026	0.00172	0.000772	0.000661
	naive-SMOM	0.133	0.0191	0.0418	0.00539	0.0132	0.00173	0.00414	0.000586
61.0, (2.74, -0.02)	proj-SMOM	0.0213	0.0159	0.00625	0.00478	0.00198	0.00165	0.000594	0.000532
	naive-SMOM	0.133	0.0184	0.0424	0.00574	0.0133	0.00171	0.00423	0.000572
61.5, (0.71, 0.18)	proj-SMOM	0.035	0.034	0.0109	0.0121	0.00346	0.00372	0.00104	0.00115
	naive-SMOM	0.104	0.0191	0.0333	0.00654	0.0107	0.00204	0.0033	0.000685
61.5, (0.74, 0.09)	proj-SMOM	0.0292	0.0242	0.00937	0.00805	0.00306	0.00265	0.000923	0.000779
	naive-SMOM	0.125	0.0184	0.0401	0.00521	0.0126	0.00192	0.00396	0.000577
61.5, (3.01, 1.15)	proj-SMOM	0.0364	0.0422	0.0114	0.0136	0.00377	0.00387	0.00112	0.00139
	naive-SMOM	0.087	0.0223	0.0276	0.00775	0.00875	0.00228	0.00273	0.000748
61.5, (1.01, 0.50)	proj-SMOM	0.0393	0.0452	0.0116	0.0134	0.00351	0.00455	0.0012	0.00137
	naive-SMOM	0.0722	0.0264	0.0232	0.00815	0.00721	0.00257	0.0023	0.000819
61.5, (1.35, 0.35)	proj-SMOM	0.0344	0.0386	0.0106	0.0118	0.00342	0.00348	0.00113	0.00115
	naive-SMOM	0.105	0.0212	0.0332	0.00684	0.0104	0.00223	0.00333	0.000588
61.5, (1.22, 0.07)	proj-SMOM	0.0252	0.0202	0.00777	0.007	0.00255	0.00202	0.000815	0.000668
	naive-SMOM	0.13	0.0186	0.0412	0.00537	0.0131	0.00181	0.00412	0.000616
61.5, (1.44, 0.44)	proj-SMOM	0.0315	0.0403	0.0118	0.0125	0.00355	0.00408	0.00108	0.00125
	naive-SMOM	0.097	0.0228	0.0314	0.00723	0.00985	0.0023	0.0031	0.000733
61.5, (1.24, 0.22)	proj-SMOM	0.0303	0.0289	0.00928	0.0098	0.00327	0.00307	0.000922	0.000962
	naive-SMOM	0.118	0.0195	0.0371	0.00591	0.0118	0.00189	0.00372	0.000615
61.5, (0.73, 0.04)	proj-SMOM	0.0273	0.0193	0.00851	0.00643	0.00241	0.00203	0.000769	0.000641
	naive-SMOM	0.132	0.0167	0.0416	0.00584	0.0131	0.00191	0.00415	0.00052
61.5, (2.74, -0.02)	proj-SMOM	0.0206	0.0154	0.00663	0.00484	0.00213	0.00163	0.000662	0.000495
	naive-SMOM	0.133	0.0171	0.0421	0.0057	0.0134	0.00168	0.00421	0.000524
63.7, (0.71, 0.18)	proj-SMOM	0.0303	0.0355	0.0105	0.0101	0.00309	0.0039	0.00105	0.00111
	naive-SMOM	0.104	0.0212	0.033	0.0067	0.0103	0.00219	0.00324	0.00064
63.7, (0.74, 0.09)	proj-SMOM	0.0276	0.0236	0.00859	0.00804	0.00292	0.00217	0.000881	0.000814
	naive-SMOM	0.123	0.0174	0.0391	0.00601	0.0124	0.00178	0.00394	0.000556
63.7, (3.01, 1.15)	proj-SMOM	0.0372	0.0446	0.0106	0.012	0.00331	0.00398	0.00105	0.00125
	naive-SMOM	0.0876	0.0242	0.0267	0.00653	0.00839	0.00196	0.00271	0.000706
63.7, (1.01, 0.50)	proj-SMOM	0.0324	0.0449	0.0113	0.0144	0.00361	0.00449	0.0011	0.00139
	naive-SMOM	0.0695	0.0262	0.0223	0.00809	0.00703	0.00268	0.00223	0.000798

Continued on next page

α_0, β^*	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
63.7, (1.35, 0.35)	proj-SMOM	0.0366	0.0375	0.011	0.0121	0.00324	0.00377	0.00103	0.0011
	naive-SMOM	0.104	0.022	0.0323	0.0058	0.0103	0.00198	0.00332	0.000677
63.7, (1.22, 0.07)	proj-SMOM	0.0247	0.0213	0.00777	0.00706	0.00255	0.00211	0.000771	0.000586
	naive-SMOM	0.127	0.0173	0.0408	0.0054	0.0128	0.00161	0.00405	0.000568
63.7, (1.44, 0.44)	proj-SMOM	0.0353	0.0401	0.0114	0.0124	0.00357	0.0038	0.00113	0.00127
	naive-SMOM	0.0965	0.0217	0.0304	0.0068	0.00968	0.00195	0.00308	0.000625
63.7, (1.24, 0.22)	proj-SMOM	0.0304	0.0269	0.00897	0.0086	0.00326	0.003	0.000959	0.000972
	naive-SMOM	0.116	0.0184	0.0366	0.00602	0.0116	0.00189	0.00366	0.000577
63.7, (0.73, 0.04)	proj-SMOM	0.026	0.0195	0.00756	0.00668	0.00246	0.00183	0.00074	0.000565
	naive-SMOM	0.128	0.0166	0.041	0.00508	0.013	0.00165	0.00407	0.000503
63.7, (2.74, -0.02)	proj-SMOM	0.0211	0.0146	0.00639	0.00513	0.00207	0.00168	0.000631	0.000525
	naive-SMOM	0.13	0.0162	0.0411	0.00539	0.0131	0.0017	0.0041	0.000517
71.6, (0.71, 0.18)	proj-SMOM	0.0286	0.0322	0.0104	0.0102	0.00291	0.0033	0.00104	0.0011
	naive-SMOM	0.097	0.0187	0.031	0.00586	0.00972	0.00199	0.00311	0.000568
71.6, (0.74, 0.09)	proj-SMOM	0.0267	0.0257	0.00822	0.00765	0.00274	0.00232	0.000828	0.000785
	naive-SMOM	0.115	0.0179	0.0365	0.00554	0.0117	0.00162	0.00369	0.000585
71.6, (3.01, 1.15)	proj-SMOM	0.0325	0.0364	0.0105	0.0121	0.00351	0.0037	0.00105	0.00125
	naive-SMOM	0.0801	0.0219	0.0257	0.00744	0.00814	0.00235	0.00256	0.000668
71.6, (1.01, 0.50)	proj-SMOM	0.0312	0.0394	0.0116	0.0126	0.00344	0.00366	0.00107	0.00138
	naive-SMOM	0.066	0.0223	0.0211	0.00733	0.00666	0.00231	0.00211	0.000849
71.6, (1.35, 0.35)	proj-SMOM	0.0298	0.0319	0.0101	0.0119	0.00314	0.0033	0.000961	0.00102
	naive-SMOM	0.0972	0.0189	0.031	0.00645	0.0097	0.00195	0.0031	0.000647
71.6, (1.22, 0.07)	proj-SMOM	0.0239	0.0199	0.00759	0.00626	0.00231	0.00184	0.000793	0.000602
	naive-SMOM	0.121	0.0189	0.0384	0.00515	0.0121	0.00177	0.00387	0.000513
71.6, (1.44, 0.44)	proj-SMOM	0.0327	0.0359	0.00988	0.0118	0.00351	0.00358	0.00104	0.00112
	naive-SMOM	0.0913	0.0213	0.029	0.00647	0.00924	0.00199	0.00283	0.00063
71.6, (1.24, 0.22)	proj-SMOM	0.0274	0.0284	0.00916	0.00896	0.00291	0.00307	0.00096	0.000928
	naive-SMOM	0.11	0.0166	0.0344	0.00581	0.0109	0.00179	0.00347	0.000608
71.6, (0.73, 0.04)	proj-SMOM	0.0235	0.0191	0.00732	0.00562	0.00243	0.00201	0.000746	0.000646
	naive-SMOM	0.122	0.0173	0.0383	0.00462	0.0121	0.00167	0.00387	0.000525
71.6, (2.74, -0.02)	proj-SMOM	0.0187	0.0154	0.00606	0.00459	0.00191	0.00136	0.000586	0.00045
	naive-SMOM	0.124	0.0162	0.039	0.00417	0.0123	0.00159	0.00387	0.000499
72.3, (0.71, 0.18)	proj-SMOM	0.0289	0.0327	0.00976	0.0105	0.00302	0.00337	0.00101	0.00104
	naive-SMOM	0.0973	0.0186	0.0313	0.00628	0.00966	0.00204	0.0031	0.00062
72.3, (0.74, 0.09)	proj-SMOM	0.027	0.0215	0.00849	0.00733	0.0026	0.00222	0.000909	0.000732
	naive-SMOM	0.115	0.0171	0.0368	0.00549	0.0116	0.00164	0.00372	0.000534
72.3, (3.01, 1.15)	proj-SMOM	0.0356	0.0423	0.0106	0.0126	0.00307	0.00376	0.00103	0.00108
	naive-SMOM	0.0812	0.0214	0.0255	0.00686	0.00795	0.00236	0.00253	0.000643
72.3, (1.01, 0.50)	proj-SMOM	0.0333	0.0403	0.0107	0.0131	0.00334	0.0042	0.00103	0.00127
	naive-SMOM	0.0668	0.0272	0.0208	0.00786	0.00655	0.00249	0.00209	0.00075
72.3, (1.35, 0.35)	proj-SMOM	0.0303	0.0351	0.00981	0.0105	0.00299	0.00325	0.00101	0.00108
	naive-SMOM	0.0958	0.02	0.0307	0.00614	0.00972	0.00195	0.0031	0.000591
72.3, (1.22, 0.07)	proj-SMOM	0.0241	0.0169	0.00714	0.00596	0.00245	0.00191	0.000728	0.000576
	naive-SMOM	0.121	0.0167	0.0379	0.0053	0.012	0.00158	0.0038	0.000521
72.3, (1.44, 0.44)	proj-SMOM	0.0293	0.0358	0.0107	0.0124	0.00293	0.0036	0.000959	0.00113
	naive-SMOM	0.0899	0.0207	0.0292	0.00676	0.00903	0.00206	0.00289	0.000669
72.3, (1.24, 0.22)	proj-SMOM	0.0295	0.0297	0.00965	0.00862	0.00293	0.00274	0.000864	0.000962
	naive-SMOM	0.11	0.0172	0.0346	0.00587	0.011	0.00181	0.00342	0.000536
72.3, (0.73, 0.04)	proj-SMOM	0.0225	0.0178	0.00705	0.00583	0.00221	0.0019	0.000691	0.000589
	naive-SMOM	0.12	0.0168	0.0381	0.00507	0.0121	0.00163	0.00382	0.000487

Continued on next page

α_0, β^*	Estimator	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
72.3,(2.74, -0.02)	proj-SMOM	0.0177	0.0147	0.00607	0.00438	0.00198	0.00134	0.00059	0.000435
	naive-SMOM	0.123	0.0144	0.0385	0.00506	0.0123	0.00157	0.00385	0.000487
79.3,(0.71, 0.18)	proj-SMOM	0.0278	0.027	0.00896	0.00957	0.0033	0.00331	0.000963	0.00101
	naive-SMOM	0.0926	0.0173	0.029	0.00591	0.00932	0.00183	0.00293	0.000578
79.3,(0.74, 0.09)	proj-SMOM	0.0269	0.0224	0.00771	0.00644	0.00236	0.00211	0.000758	0.000627
	naive-SMOM	0.111	0.0154	0.035	0.00527	0.0109	0.00151	0.00352	0.000513
79.3,(3.01, 1.15)	proj-SMOM	0.0316	0.0379	0.00985	0.012	0.00307	0.00389	0.000973	0.00118
	naive-SMOM	0.0764	0.0203	0.0245	0.00699	0.00765	0.00217	0.00243	0.000658
79.3,(1.01, 0.50)	proj-SMOM	0.0366	0.043	0.0101	0.013	0.00331	0.00403	0.00101	0.00126
	naive-SMOM	0.0669	0.0272	0.0201	0.00709	0.00639	0.00234	0.00201	0.000779
79.3,(1.35, 0.35)	proj-SMOM	0.028	0.0332	0.00859	0.00936	0.00323	0.00343	0.00099	0.0011
	naive-SMOM	0.0937	0.019	0.0292	0.00543	0.00938	0.00202	0.00297	0.000628
79.3,(1.22, 0.07)	proj-SMOM	0.0227	0.018	0.00738	0.00598	0.00235	0.00181	0.000702	0.000552
	naive-SMOM	0.114	0.0141	0.0366	0.00463	0.0116	0.00151	0.00361	0.000522
79.3,(1.44, 0.44)	proj-SMOM	0.0298	0.035	0.00909	0.0115	0.00329	0.00372	0.00093	0.00097
	naive-SMOM	0.0862	0.0206	0.0276	0.0064	0.0087	0.00203	0.00273	0.00055
79.3,(1.24, 0.22)	proj-SMOM	0.0271	0.0274	0.0085	0.00828	0.00266	0.00261	0.000845	0.000876
	naive-SMOM	0.103	0.0166	0.0326	0.00566	0.0103	0.00177	0.00331	0.000564
79.3,(0.73, 0.04)	proj-SMOM	0.0215	0.0165	0.00655	0.00554	0.00233	0.00186	0.000709	0.000581
	naive-SMOM	0.115	0.0145	0.0366	0.00441	0.0116	0.00155	0.00364	0.000444
79.3,(2.74, -0.02)	proj-SMOM	0.0177	0.0141	0.00618	0.00456	0.00183	0.00131	0.000593	0.000465
	naive-SMOM	0.118	0.0144	0.0371	0.00503	0.0118	0.00148	0.00372	0.000469

Table S.4: Monte Carlo ratios of Frobenius errors for the projected Stein method-of-moments estimator. Entries are the median and interquartile range, over $B = 500$ replications, of $\|\hat{\Sigma}_{\text{proj}} - \Sigma_0\|_F / \|\hat{\Sigma}_{\text{comp}} - \Sigma_0\|_F$, where the competitor is the (unstructured) Stein method-of-moments estimator. Values below one favour the projected Stein estimator.

α_0, β^*	Ratio	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
23.3,(0.71, 0.18)	proj/naive	0.343	0.316	0.346	0.31	0.325	0.302	0.33	0.287
23.3,(0.74, 0.09)	proj/naive	0.239	0.197	0.236	0.208	0.22	0.196	0.241	0.199
23.3,(3.01, 1.15)	proj/naive	0.462	0.367	0.442	0.397	0.423	0.381	0.401	0.407
23.3,(1.01, 0.50)	proj/naive	0.551	0.409	0.515	0.432	0.523	0.424	0.556	0.462
23.3,(1.35, 0.35)	proj/naive	0.334	0.299	0.342	0.312	0.35	0.302	0.304	0.326
23.3,(1.22, 0.07)	proj/naive	0.202	0.152	0.195	0.153	0.195	0.171	0.2	0.171
23.3,(1.44, 0.44)	proj/naive	0.342	0.337	0.399	0.327	0.392	0.357	0.343	0.352
23.3,(1.24, 0.22)	proj/naive	0.271	0.221	0.27	0.245	0.269	0.241	0.275	0.248
23.3,(0.73, 0.04)	proj/naive	0.188	0.156	0.19	0.164	0.2	0.151	0.19	0.139
23.3,(2.74, -0.02)	proj/naive	0.156	0.132	0.149	0.104	0.151	0.112	0.16	0.126
34.1,(0.71, 0.18)	proj/naive	0.319	0.299	0.332	0.312	0.313	0.289	0.286	0.264
34.1,(0.74, 0.09)	proj/naive	0.23	0.19	0.225	0.197	0.231	0.207	0.219	0.183
34.1,(3.01, 1.15)	proj/naive	0.429	0.376	0.438	0.417	0.492	0.362	0.436	0.414
34.1,(1.01, 0.50)	proj/naive	0.527	0.415	0.519	0.427	0.531	0.423	0.536	0.421
34.1,(1.35, 0.35)	proj/naive	0.337	0.301	0.327	0.305	0.353	0.304	0.332	0.32
34.1,(1.22, 0.07)	proj/naive	0.206	0.155	0.209	0.168	0.2	0.172	0.202	0.163
34.1,(1.44, 0.44)	proj/naive	0.375	0.381	0.392	0.352	0.353	0.343	0.351	0.344
34.1,(1.24, 0.22)	proj/naive	0.285	0.243	0.271	0.258	0.277	0.246	0.266	0.224
34.1,(0.73, 0.04)	proj/naive	0.194	0.15	0.195	0.165	0.191	0.146	0.191	0.14

Continued on next page

α_0, β^*	Ratio	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
34.1,(2.74, -0.02)	proj/naive	0.155	0.114	0.16	0.12	0.153	0.119	0.157	0.107
43.0,(0.71, 0.18)	proj/naive	0.347	0.306	0.317	0.312	0.336	0.308	0.317	0.324
43.0,(0.74, 0.09)	proj/naive	0.234	0.194	0.234	0.183	0.234	0.19	0.249	0.21
43.0,(3.01, 1.15)	proj/naive	0.44	0.395	0.419	0.42	0.481	0.422	0.429	0.374
43.0,(1.01, 0.50)	proj/naive	0.533	0.438	0.514	0.46	0.529	0.449	0.502	0.457
43.0,(1.35, 0.35)	proj/naive	0.347	0.306	0.329	0.31	0.314	0.3	0.338	0.331
43.0,(1.22, 0.07)	proj/naive	0.202	0.152	0.196	0.153	0.19	0.152	0.204	0.156
43.0,(1.44, 0.44)	proj/naive	0.369	0.379	0.369	0.331	0.376	0.35	0.367	0.344
43.0,(1.24, 0.22)	proj/naive	0.274	0.22	0.255	0.247	0.281	0.262	0.273	0.236
43.0,(0.73, 0.04)	proj/naive	0.194	0.142	0.185	0.153	0.198	0.156	0.198	0.147
43.0,(2.74, -0.02)	proj/naive	0.159	0.126	0.159	0.13	0.154	0.123	0.152	0.113
46.6,(0.71, 0.18)	proj/naive	0.336	0.307	0.346	0.315	0.325	0.289	0.34	0.298
46.6,(0.74, 0.09)	proj/naive	0.226	0.201	0.221	0.182	0.233	0.201	0.22	0.174
46.6,(3.01, 1.15)	proj/naive	0.436	0.412	0.441	0.414	0.409	0.374	0.409	0.369
46.6,(1.01, 0.50)	proj/naive	0.556	0.421	0.589	0.452	0.545	0.463	0.534	0.456
46.6,(1.35, 0.35)	proj/naive	0.33	0.306	0.318	0.313	0.331	0.305	0.335	0.298
46.6,(1.22, 0.07)	proj/naive	0.205	0.163	0.191	0.149	0.191	0.15	0.194	0.152
46.6,(1.44, 0.44)	proj/naive	0.385	0.307	0.372	0.36	0.393	0.347	0.39	0.355
46.6,(1.24, 0.22)	proj/naive	0.279	0.248	0.273	0.249	0.286	0.256	0.26	0.23
46.6,(0.73, 0.04)	proj/naive	0.199	0.147	0.191	0.163	0.186	0.15	0.194	0.145
46.6,(2.74, -0.02)	proj/naive	0.148	0.119	0.167	0.117	0.156	0.122	0.161	0.121
61.0,(0.71, 0.18)	proj/naive	0.348	0.329	0.339	0.298	0.326	0.304	0.323	0.324
61.0,(0.74, 0.09)	proj/naive	0.233	0.193	0.252	0.189	0.234	0.198	0.237	0.207
61.0,(3.01, 1.15)	proj/naive	0.448	0.406	0.42	0.391	0.429	0.39	0.446	0.404
61.0,(1.01, 0.50)	proj/naive	0.55	0.467	0.53	0.426	0.555	0.444	0.505	0.436
61.0,(1.35, 0.35)	proj/naive	0.322	0.281	0.315	0.32	0.32	0.31	0.349	0.31
61.0,(1.22, 0.07)	proj/naive	0.193	0.165	0.195	0.16	0.193	0.158	0.196	0.157
61.0,(1.44, 0.44)	proj/naive	0.372	0.358	0.374	0.368	0.377	0.367	0.377	0.341
61.0,(1.24, 0.22)	proj/naive	0.26	0.229	0.286	0.261	0.276	0.236	0.289	0.26
61.0,(0.73, 0.04)	proj/naive	0.195	0.157	0.201	0.145	0.199	0.132	0.191	0.155
61.0,(2.74, -0.02)	proj/naive	0.159	0.117	0.152	0.11	0.154	0.119	0.142	0.127
61.5,(0.71, 0.18)	proj/naive	0.338	0.302	0.34	0.326	0.339	0.305	0.328	0.304
61.5,(0.74, 0.09)	proj/naive	0.236	0.188	0.235	0.19	0.247	0.201	0.242	0.188
61.5,(3.01, 1.15)	proj/naive	0.44	0.401	0.43	0.403	0.429	0.376	0.436	0.399
61.5,(1.01, 0.50)	proj/naive	0.551	0.434	0.537	0.418	0.514	0.446	0.536	0.443
61.5,(1.35, 0.35)	proj/naive	0.338	0.328	0.327	0.321	0.334	0.299	0.354	0.327
61.5,(1.22, 0.07)	proj/naive	0.193	0.147	0.193	0.16	0.198	0.158	0.2	0.157
61.5,(1.44, 0.44)	proj/naive	0.349	0.354	0.385	0.333	0.386	0.354	0.369	0.343
61.5,(1.24, 0.22)	proj/naive	0.259	0.239	0.252	0.242	0.292	0.239	0.252	0.237
61.5,(0.73, 0.04)	proj/naive	0.215	0.133	0.202	0.147	0.188	0.147	0.192	0.153
61.5,(2.74, -0.02)	proj/naive	0.157	0.119	0.158	0.115	0.16	0.12	0.163	0.116
63.7,(0.71, 0.18)	proj/naive	0.306	0.301	0.323	0.292	0.311	0.306	0.328	0.312
63.7,(0.74, 0.09)	proj/naive	0.224	0.189	0.225	0.195	0.235	0.172	0.231	0.194
63.7,(3.01, 1.15)	proj/naive	0.441	0.419	0.42	0.366	0.416	0.392	0.408	0.375
63.7,(1.01, 0.50)	proj/naive	0.497	0.489	0.523	0.47	0.55	0.456	0.511	0.42
63.7,(1.35, 0.35)	proj/naive	0.357	0.312	0.34	0.329	0.319	0.322	0.315	0.303
63.7,(1.22, 0.07)	proj/naive	0.196	0.16	0.195	0.167	0.198	0.168	0.193	0.139
63.7,(1.44, 0.44)	proj/naive	0.368	0.351	0.39	0.351	0.382	0.34	0.378	0.357
63.7,(1.24, 0.22)	proj/naive	0.26	0.231	0.25	0.221	0.28	0.253	0.262	0.247

Continued on next page

α_0, β^*	Ratio	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
63.7,(0.73, 0.04)	proj/naive	0.201	0.146	0.187	0.162	0.194	0.136	0.185	0.149
63.7,(2.74, -0.02)	proj/naive	0.166	0.112	0.155	0.119	0.161	0.13	0.155	0.118
71.6,(0.71, 0.18)	proj/naive	0.309	0.305	0.341	0.301	0.311	0.305	0.348	0.324
71.6,(0.74, 0.09)	proj/naive	0.24	0.207	0.231	0.212	0.232	0.2	0.23	0.203
71.6,(3.01, 1.15)	proj/naive	0.431	0.378	0.424	0.38	0.458	0.36	0.445	0.397
71.6,(1.01, 0.50)	proj/naive	0.498	0.43	0.564	0.418	0.529	0.43	0.534	0.451
71.6,(1.35, 0.35)	proj/naive	0.309	0.305	0.336	0.318	0.32	0.316	0.321	0.294
71.6,(1.22, 0.07)	proj/naive	0.195	0.153	0.199	0.159	0.193	0.15	0.212	0.151
71.6,(1.44, 0.44)	proj/naive	0.363	0.335	0.355	0.334	0.391	0.342	0.37	0.349
71.6,(1.24, 0.22)	proj/naive	0.256	0.24	0.266	0.246	0.28	0.245	0.279	0.25
71.6,(0.73, 0.04)	proj/naive	0.196	0.152	0.194	0.153	0.198	0.156	0.196	0.155
71.6,(2.74, -0.02)	proj/naive	0.148	0.127	0.158	0.113	0.154	0.114	0.153	0.117
72.3,(0.71, 0.18)	proj/naive	0.302	0.306	0.321	0.305	0.325	0.318	0.345	0.295
72.3,(0.74, 0.09)	proj/naive	0.236	0.174	0.238	0.186	0.222	0.188	0.247	0.193
72.3,(3.01, 1.15)	proj/naive	0.46	0.427	0.441	0.379	0.423	0.401	0.41	0.344
72.3,(1.01, 0.50)	proj/naive	0.534	0.444	0.526	0.454	0.539	0.457	0.522	0.44
72.3,(1.35, 0.35)	proj/naive	0.326	0.322	0.327	0.302	0.316	0.293	0.328	0.312
72.3,(1.22, 0.07)	proj/naive	0.202	0.138	0.187	0.156	0.207	0.154	0.19	0.154
72.3,(1.44, 0.44)	proj/naive	0.349	0.36	0.381	0.361	0.327	0.347	0.357	0.353
72.3,(1.24, 0.22)	proj/naive	0.275	0.244	0.29	0.231	0.274	0.233	0.265	0.271
72.3,(0.73, 0.04)	proj/naive	0.188	0.152	0.191	0.148	0.189	0.157	0.185	0.154
72.3,(2.74, -0.02)	proj/naive	0.145	0.114	0.157	0.108	0.16	0.119	0.154	0.115
79.3,(0.71, 0.18)	proj/naive	0.321	0.249	0.316	0.307	0.365	0.309	0.345	0.302
79.3,(0.74, 0.09)	proj/naive	0.252	0.202	0.228	0.179	0.218	0.189	0.226	0.186
79.3,(3.01, 1.15)	proj/naive	0.422	0.39	0.427	0.396	0.435	0.401	0.428	0.393
79.3,(1.01, 0.50)	proj/naive	0.563	0.449	0.551	0.422	0.533	0.452	0.539	0.428
79.3,(1.35, 0.35)	proj/naive	0.311	0.315	0.304	0.288	0.344	0.335	0.35	0.324
79.3,(1.22, 0.07)	proj/naive	0.197	0.156	0.203	0.151	0.204	0.159	0.197	0.145
79.3,(1.44, 0.44)	proj/naive	0.36	0.346	0.353	0.345	0.39	0.358	0.358	0.311
79.3,(1.24, 0.22)	proj/naive	0.27	0.239	0.267	0.225	0.267	0.227	0.265	0.243
79.3,(0.73, 0.04)	proj/naive	0.187	0.142	0.183	0.153	0.199	0.156	0.199	0.155
79.3,(2.74, -0.02)	proj/naive	0.149	0.117	0.161	0.117	0.157	0.115	0.16	0.123