



# **Adaptive inference with random ellipsoids through Conformal Conditional Linear Expectation**

Iain Henderson, Adrien Mazoyer, Fabrice Gamboa

## **► To cite this version:**

Iain Henderson, Adrien Mazoyer, Fabrice Gamboa. Adaptive inference with random ellipsoids through Conformal Conditional Linear Expectation. 2025. <hal-04489842v3>

**HAL Id: hal-04489842**

**<https://hal.science/hal-04489842v3>**

Preprint submitted on 26 Jun 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Adaptive inference with random ellipsoids through Conformal Conditional Linear Expectation

Iain Henderson<sup>\*1</sup>, Adrien Mazoyer<sup>2</sup>, and Fabrice Gamboa<sup>2</sup>

<sup>1</sup>Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, Toulouse, 31055, France

<sup>2</sup>Institut de Mathématiques de Toulouse, Université de Toulouse, CNRS, UPS, 31062, Toulouse, France

## Abstract

We propose two new conformity scores for conformal prediction, in a general multivariate regression framework. The underlying score functions are based on a covariance analysis of the residuals and the input points. We give theoretical guarantees on the prediction sets, which consist in explicit ellipsoids. We study the asymptotic properties of the ellipsoids, and show that their volume is reduced compared to that of classic balls, under ellipticity assumptions. Finally, we illustrate the effectiveness of all our results on an in-depth numerical study, including heavy-tailed as well as non-elliptical distributions.

**Keywords** Conformal prediction, distribution free inference, adaptive confidence, confidence ellipsoid

## 1 Introduction

Uncertainty quantification aims to provide mathematical techniques to quantify uncertainties in computational and real worlds. One very common problem in uncertainty quantification is to provide a confidence set for a given prediction method. When the prediction method is based on a statistical model that is not too complex, classical statistical tools can lead to a confidence set. In black box models, as those developed in machine learning or in complicated regression models the inference process to build such confidence sets is not so straightforward. One now classical way to build such confidence sets stands on a non-parametric learning of the fluctuations of the predictors. This technique called *conformal inference* allows the construction of a confidence set for a given trained predictor by only observing its behaviour on a calibration sample.

Conformal inference has a long story beginning early in the forty's by the pioneer works in reliability of Wilks ([51, 52]). Conformal prediction has then been subsequently developed in the early 2000s by a research group around Vapnik and Vovk (see e.g. [49]). With the spectacular progress and the massive implementation of black-box models in neural machine learning, there has been an impressive revival of conformal prediction for at least five or seven years. In a nutshell, conformal prediction is a general concept that includes many different means of providing confidence of pointwise predictions produced using a statistical model or machine learning, without any knowledge on the predictor and under mild assumptions. Indeed, the construction of a prediction set having a high probability of containing a quantity of interest is a general challenging problem in the uncertainty quantification paradigm. More formally, let  $\mathcal{X}$  and  $\mathcal{Y}$  be some state spaces. Let further  $(X, Y) \in \mathcal{X} \times \mathcal{Y}$  be a random vector. Given a sample  $(X_i, Y_i)_{i=1, \dots, n}$  and a predictor  $\hat{f} : \mathcal{X} \rightarrow \mathcal{Y}$  (independent of the sample), for  $X_{n+1} \in \mathcal{X}$  (a new draw of  $X$ ), we aim to build a random measurable region  $C_\alpha^n(X_{n+1}) \subseteq \mathcal{Y}$  as *tight as possible*, such that for a given  $\alpha$  (close to 0 and set by the user), we have  $\mathbb{P}(Y_{n+1} \in C_\alpha^n(X_{n+1})) \geq 1 - \alpha$ . In some ideal cases, the known statistical properties or characteristics of the predictor  $\hat{f}$  can be sole used to construct  $C_\alpha^n(\cdot)$ . However, this usually requires model assumptions on the data, that are sometimes too restrictive or unrealistic. Even worse, it may happen that none of the properties of  $\hat{f}$  are useful for building  $C_\alpha^n(\cdot)$ . The aim of Conformal Inference (CI) is to construct such regions, with a very appealing advantage: CI requires only mild assumptions on the sample distribution and on the predictor. Actually, the only hypothesis needed is an exchangeability one, that will be discussed later (see the beginning of Section 2). Roughly speaking, the intuition behind CI is that a relevant

---

<sup>\*</sup>iain.henderson@isae-supaero.fr

prediction for the unknown output  $Y_{n+1}$  should be any  $y \in \mathcal{Y}$  such that  $(X_{n+1}, y, \hat{f}(X_{n+1}))$  seems to be *conform* to the sample  $(X_i, Y_i, \hat{f}(X_i))_{i=1, \dots, n}$ . The main ingredient here is to define and to work with an appropriate score of conformity. Of course, the appropriateness of a given score greatly depends on the context. In particular, the shape and accuracy of the prediction set is directly related both to the performance of the predictor  $\hat{f}$  and to the conformity score.

Since its introduction in the early 2000s (see e.g. [50]), conformal prediction has been developed in many different contexts as for example classification [3, 12], regression [7, 29, 43], functional data [13], outlier detection [5, 30], neural networks [41], etc. For more insights and examples, we refer to the recent reviews [2, 17, 24, 28]. In many works dealing with CI, the size and shape of  $C_\alpha^n(X_{n+1})$  only mildly depend on  $X_{n+1}$ . This can make sense if for example the covariance of the residual  $Y - \hat{f}(X)$  is homogeneous. Dealing with heterogeneous data can be approached with different objectives in mind. The goal of conformal inference would be to obtain a conditional coverage property for  $C_\alpha^n(X_{n+1})$ , that is  $\mathbb{P}(Y_{n+1} \in C_\alpha^n(X_{n+1}) | X_{n+1}) \geq 1 - \alpha$ . However, it is theoretically impossible to ensure the conditional coverage with a finite sample [48]. A less demanding property could be then to be able to capture some geometric features of the conditional distribution  $\mathbb{P}(Y|X)$ , typically heteroscedasticity in data. The resulting prediction set should then adapt in terms of shape and size with respect to the input. For example, [28] proposed to use some normalizing statistic in the score, based on the previous sample points. In this last work, the trick consists in normalizing the residuals, for example with respect to their median. Notice furthermore that many of the works in CI consider only the case where  $\mathcal{Y} \subseteq \mathbb{R}$ , leading thus to prediction intervals. The direct extension to  $\mathcal{Y} \subseteq \mathbb{R}^\ell$  then consists in using a norm of the residual vector  $R := Y - \hat{f}(X)$ , leading to prediction balls centered on  $\hat{f}(X)$ . However, our wish is that the geometry of the prediction set adapts with the input, which cannot be achieved by considering balls. The multidimensional framework is still widely studied, with some ongoing works considering responses taking values in separable metric spaces [32], exploring approaches based on transport tools [45, 54], or formulation in terms of minimum-volume coverage for arbitrary norm-based sets [6].

In this work, we propose and fully investigate CI with a score built on the empirical covariance of  $(X, Y - \hat{f}(X))$ , where *both*  $X$  and  $Y$  are multidimensional. The covariance estimate and score include the prediction point  $X_{n+1}$  and the potential residual  $y - \hat{f}(X_{n+1})$ . We show that our procedure amounts to estimating the conditional linear expectation of the residual  $Y - \hat{f}(X)$  given the input  $X$ , formulated in a CI setting (see Remark in Section 3.3). We will thus call our method the conformal conditional linear expectation (CCLE). The main advantage of the associated region  $C_\alpha^n$  is that it is well tailored to multivariate outputs  $Y \in \mathbb{R}^\ell$ , while remaining easy to compute. Moreover, it leads to adaptive prediction sets in terms of size and shape. It is important to recall that despite the conditional aspect of the score proposed here, its main motivation is not to improve the conditional coverage. Rather, our primary goals are (i) the adaptivity of the prediction set with reference to the distribution of  $(X, R)$  and (ii) decreasing the volume of the prediction set. In Section 4, we discuss the necessary adjustments to recover the conditional coverage in the asymptotic regime.

The following theorem summarises and gives in a nutshell our main results (Theorem 3.1, Theorem 3.6, Proposition 4.7, Proposition 4.9 and Proposition 4.10).

**Theorem 1.1.** *Let  $((X_1, Y_1), \dots, (X_{n+1}, Y_{n+1}))$  be exchangeable pairs of random vectors and  $\alpha \in (1/(n+1), 1)$ . Then,*

- *There exists explicit  $\mathcal{A}_n, \rho_{n,\alpha}$  and  $\tilde{Y}_{n+1}$ , only depending on  $((X_1, Y_1), \dots, (X_n, Y_n))$ ,  $X_{n+1}$  and  $\alpha$ , such that the  $\ell$ -dimensional ellipsoid*

$$E_\alpha^n = \{y \in \mathbb{R}^\ell : (y - \tilde{Y}_{n+1})^\top \mathcal{A}_n^{-1} (y - \tilde{Y}_{n+1}) \leq \rho_{n,\alpha}\} \quad (1)$$

*satisfies  $\mathbb{P}(Y_{n+1} \in E_\alpha^n) \geq 1 - \alpha$ .*

- *Set  $R_1 = Y_1 - \hat{f}(X_1)$  and assume that  $(X_1, R_1)$  follows an elliptical distribution. Then, as  $n \rightarrow +\infty$  and under mild assumptions on the distribution of  $(X_1, R_1)$ , the volume  $\text{Vol}(E_\alpha^n)$  converges in distribution towards a random variable  $\text{Vol}(E_\alpha^\infty)$  satisfying*

$$\text{Vol}(E_\alpha^\infty) \leq \text{Vol}(B_\alpha^\infty) \text{ a.s..} \quad (2)$$

*Here,  $\text{Vol}(B_\alpha^\infty)$  is the (deterministic) limit of the volume of the ball obtained using the standard score  $\|Y - \hat{f}(X)\|$  (see Proposition 4.6 for its expression).*

In other words, we show that the prediction set obtained using CCLE is an ellipsoid  $E_\alpha^n \subset \mathbb{R}^\ell$ . Furthermore, the volume of this prediction set is asymptotically smaller than that of the classical conformal prediction ball, under

mild ellipticity assumptions (Section 4.3). Elliptical distributions are those whose level sets are elliptical, such as the Gaussian ones. As we will see later, we are actually able to identify two different  $\rho_{n,\alpha}$  satisfying Theorem 1.1, each with different properties concerning the associated ellipsoid. The first one, corresponding to the joint Mahalanobis score (Section 3.2), leads to ellipsoid  $E_\alpha^n$  such that  $\mathbb{P}(E_\alpha^n = \emptyset) > 0$  and  $\mathbb{P}(E_\alpha^n = \mathbb{R}^\ell) = 0$ ; whereas the second, corresponding to adjusted Mahalanobis score (Section 3.5), leads to ellipsoid  $F_\alpha^n$  such that  $\mathbb{P}(F_\alpha^n = \emptyset) = 0$  and  $\mathbb{P}(F_\alpha^n = \mathbb{R}^\ell) > 0$  (this probability tending to 0 as  $n$  tends to  $\infty$ ). Notably, the second ellipsoid reduces to the conditional prediction set with a minimal volume, in the case of Gaussian data, when  $n$  tends to  $\infty$  (Propositions 4.4 and 4.7).

Let us note that to calculate the conditional linear expectation, we will use a ridge estimator for the covariance. This is motivated by three facts. First, it stabilizes the numerical procedure [26], Section 4.4. Secondly, it simplifies several proofs (see e.g. Proposition 3.4). Last but not least, our method will be carried out in an infinite dimensions setting (see [19]) where finite size ellipsoids are rather well-understood (see e.g. Propositions 6.1 to 6.3 in [16]).

Before presenting the outline of our paper, let us further discuss the use of ellipsoidal confidence regions in CI. In [23] and [36] the authors construct local ellipsoidal confidence regions in a standard multivariate CI setting. While our method effectively relies on the conditional *linear* covariance (see Section 3.3), the local ellipsoids described in [36] correspond to a convex combination of empirical estimates of the global covariance of the residual  $R$  and the *true* conditional covariance  $\text{Cov}(R|X = x)$  (even though those local ellipsoids are not presented as estimates of  $\text{Cov}(R|X = x)$  in [36]). The latter are estimated by only selecting the observed residuals  $r_i$  (realisation of  $R_i$ ) such that the corresponding  $x_i$  (realisation of  $X_i$ ) are the  $k$  closest to the observed realisation  $x_{n+1}$  of  $X_{n+1}$ . A drawback of this nearest neighbour algorithm is that a sufficient amount of observed inputs  $x_i$  close to  $x_{n+1}$  is required to obtain meaningful estimates of  $\text{Cov}(R_{n+1}|X_{n+1} = x_{n+1})$ . The authors of [36] then show the practical efficiency of their CI method, but do not further study its theoretical properties. CCLE can also be seen as an intermediate solution between two paradigms: relying on the global (unconditional) covariance matrix  $\text{Cov}(R)$ , or on the local (fully conditional) covariance matrix  $\text{Cov}(R|X = x)$ . CCLE is then an alternative solution to the convex combination proposed in [36]. The recent paper [53] also proposes a CI on the residuals leading also to a confidence ellipsoid. The statistical frame therein is a single realisation of a time series. Here a stationarity assumption is replacing the exchangeability one. The method stands on the estimated empirical covariance of the residuals, estimated from the past. Notice that the method does not encompass any conditional information  $R|X$ . The authors show the practical advantage of their method when compared to copula-based CI [35], and establish conditional coverage properties, in the limit of the large observational horizon (this would correspond to the large sample size in our framework). In the case of both [36] and [53], the authors do not study the theoretical properties of the size (volume) of the corresponding confidence regions. We believe that this feature is important in practical applications, along with the conditional coverage. In our work, we study the volumetric properties of  $E_\alpha^n$  in Section 4.3. Finally, in both [36] and [53], the ellipsoids are centered at the predictor  $\hat{f}(X_{n+1})$ , while our method naturally corrects  $\hat{f}(X_{n+1})$  with the empirical conditional linear expectation of  $R$  given  $X$  (Section 3.3 and Proposition 4.5). This feature is interesting as  $\hat{f}$  can be biased, which may happen e.g. when  $\hat{f}$  is a black-box neural network.

The paper is organized as follows. We recall some preliminaries on split conformal prediction for regression in Section 2. The heart of our work is the Section 3 where we discuss the construction of the conformal score and give some initial results on the induced conformal prediction. This is followed by the Section 4 where we study some asymptotic properties of the conformal ellipsoid, notably its volume under ellipticity assumptions. We illustrate and demonstrate the practical effectiveness of our results on simulated data in Section 5. All the technical proofs and technical lemmas are postponed to Appendices B and C.

## 2 Background for Split Conformal Inference

Let  $(U_i)_{i=1,\dots,n+1}$  be  $(n+1)$  exchangeable vectors of random variables, with  $U_i = (X_i, Y_i) \in \mathcal{U} = \mathcal{X} \times \mathcal{Y}$ . This means that the joint distribution of  $(U_i)_{i=1,\dots,n+1}$  is invariant by any permutation of the symmetric group of order  $(n+1)$ . For example, an i.i.d. sample satisfies this assumption. In our framework we assume that  $(U_i)_{i=1,\dots,n}$ , called the calibration set (or calibration sample), has been observed and that we get  $X_{n+1}$ . We assume further that we have at hand a predictor  $\hat{f}(X_{n+1})$  of  $Y_{n+1}$ . In this paper, for the sake of simplicity we assume that this predictor is deterministic. For example, it had been built using a sample independent of  $(U_i)_{i=1,\dots,n+1}$  and so we are working conditionally to the training sample (this is the main assumption of *split* CI, in contrast with *full* CI where  $\hat{f}$  is retrained for each new sample point). The aim is to provide a prediction set for  $Y_{n+1}$ . More precisely, as discussed in the Introduction, our objective is to construct a *conservatively* valid prediction set for  $Y_{n+1}$  with a given level of confidence. Hence, we aim to build a set function  $C_\alpha^n$  defined on  $(\mathcal{X} \times \mathcal{Y})^n \times \mathcal{X}$  that will return a

measurable subset of  $\mathcal{Y}$  for  $Y_{n+1}$  such that

$$\mathbb{P}(Y_{n+1} \in C_\alpha^n((X_i, Y_i)_{i=1\dots n}; X_{n+1})) \geq 1 - \alpha. \quad (3)$$

Notice that in the above probability we have integrated over  $(X_i, Y_i)_{i=1\dots n+1}$ . The construction of  $C_\alpha^n$  stands on the notion of conformity. To begin with, we define a nonconformity score. It consists in a map  $S$  from  $\mathcal{U}^n \times \mathcal{U}$  to  $\mathbb{R}$ , symmetric in its first  $n$  arguments. The empirical distribution built from  $(S(U_1, \dots, U_n; U_j))_{j=1, \dots, n}$  will be the main tool to quantify how conformal to  $U_1, \dots, U_n$ ,  $u \in \mathcal{U}$  is. Archetypal scores in the usual regression framework with  $\mathcal{U} = \mathbb{R}^k \times \mathbb{R}^\ell$  are related to the magnitude of the empirical residual, the most popular example being

$$S((X_1, Y_1), \dots, (X_n, Y_n), (x, y)) = \|y - \hat{f}(x)\|, \quad (4)$$

where  $\|\cdot\|$  is the Euclidean norm of  $\mathbb{R}^\ell$ . Notice that in this last example the score function only depends on  $u = (x, y)$  but not on the calibration sample  $(U_1, \dots, U_n)$ . This will not be the case in the framework that we are developing in this paper (see the next section).

Let  $\alpha \in (1/(n+1), 1)$ . For the sake of conciseness, we will use the notation  $S(x, y) := S((X_1, Y_1), \dots, (X_n, Y_n); (x, y))$ . For  $i = 1, \dots, n+1$ , we will also write  $S_i := S(X_i, Y_i)$ . Following for example [2], the conformalized prediction set for  $Y_{n+1}$  when observing  $X_{n+1}$  is then

$$C_\alpha^n(X_{n+1}) \equiv C_\alpha^n((X_1, Y_1) \dots (X_n, Y_n); X_{n+1}) = \{y \in \mathcal{Y} : S(X_{n+1}, y) \leq S_{(n_\alpha)}\}. \quad (5)$$

Here,  $n_\alpha := \lceil (1 - \alpha)(n + 1) \rceil$  and  $S_{(1)} \leq \dots \leq S_{(n)}$  denotes the order statistics associated to  $(S_i)_{i=1\dots n}$ . By construction, the above set  $C_\alpha^n(X_{n+1})$  satisfies

$$\mathbb{P}(Y_{n+1} \in C_\alpha^n(X_{n+1})) \geq 1 - \alpha. \quad (6)$$

The key property ensuring that equation (6) holds is that  $(S_i)_{i=1, \dots, n+1}$  is exchangeable. Moreover, if the nonconformity scores  $(S_i)_{i=1\dots n}$  have a continuous joint distribution, then we have the following upper bound,

$$\mathbb{P}(Y_{n+1} \in C_\alpha^n(X_{n+1})) \leq 1 - \alpha + \frac{1}{n+1}. \quad (7)$$

Returning to the regression framework and considering the residual score (4), the prediction set for  $Y_{n+1}$  given the covariable  $X_{n+1}$  is

$$C_\alpha^n(X_{n+1}) = B\left(\hat{f}(X_{n+1}), S_{(n_\alpha)}\right),$$

i.e. the ball with center  $\hat{f}(X_{n+1})$  and radius  $S_{(n_\alpha)}$ .

### 3 The Mahalanobis score and main results

#### 3.1 Notations

We will assume that  $\mathcal{X} = \mathbb{R}^k$  and  $\mathcal{Y} = \mathbb{R}^\ell$ , ( $\ell, k \in \mathbb{N}$ ). For any positive integer  $n_1$  and  $n_2$ ,  $\mathcal{M}_{n_1, n_2}$  denotes the set of all  $n_1 \times n_2$  real matrices. For,  $i = 1 \dots, n$ ,  $e_i$  denotes the  $i^{\text{th}}$  element of the canonical basis of  $\mathbb{R}^n$ . Set  $p = k + \ell$ , for  $v \in \mathbb{R}^p \setminus \{0\}$ ,  $\pi_v^\perp$  denotes the orthogonal projector onto  $\text{Span}(v)^\perp$ . In others words,  $\pi_v^\perp = \mathbf{I}_p - \|v\|^{-2} v v^\top$ .  $\mathbf{1}$  denotes the vector compounded by ones on all its components (its dimension will be implicit from the context). For any  $\alpha \in (0, 1)$ , we set  $n_\alpha := \lceil (1 - \alpha)(n + 1) \rceil$ . All non-column matrices are denoted with bold letters. For any block matrix  $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$ , we will denote the Schur complement (whenever it exists) by  $\mathbf{M}/\mathbf{A} := \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ . The abbreviation SLLN stands for the strong law of large numbers. Given  $A \in \mathcal{M}_{n_1, n_2}$  ( $n_1, n_2 \geq 1$ ), its Frobenius norm is denoted as  $\|A\|_F^2 := \text{Tr}(AA^\top)$ . Given  $A, B \in \mathcal{M}_{n, n}$  two symmetric matrices,  $A \preceq B$  means that  $B - A$  is nonnegative definite. Let  $\xi$  be real random variable and  $\alpha \in (0, 1)$ ,  $q_{1-\alpha}(\xi)$  denotes its quantile of order  $1 - \alpha$ , i.e.  $q_{1-\alpha}(\xi) = \inf\{r \in \mathbb{R} : \mathbb{P}(\xi > r) \leq \alpha\}$ . Recall also that we have a predictor  $\hat{f}$  independent of the calibration sample and that we denote the residual by  $R := Y - \hat{Y}$ ,  $\hat{Y} = \hat{f}(X)$ . For a square integrable random vector  $\xi$ ,  $\mu_\xi$  and  $\Sigma_\xi$  denote its mean vector and covariance matrix, respectively. We abbreviate ‘‘almost surely’’ as ‘‘a.s.’’.  $\xrightarrow{\mathbb{P}}$  stands for convergence in probability,  $\stackrel{\text{d}}{=}$  and  $\stackrel{\text{d}}{\rightarrow}$  stand for equality and convergence in distribution, respectively. To finish, we represent all finite-dimensional vectors as column matrices.

### 3.2 Joint Mahalanobis score

As discussed in the previous section, a useful score used in CI is the (squared) norm of the residual. In this paper, we will consider a natural generalization of the previous score by considering some statistical evaluation of  $(R - \mu_R)^\top \Sigma_R^{-1} (R - \mu_R)$  (assuming that everything here is well defined). Before explaining how we will estimate and handle the previous quantity, let it first bulk up by including  $X$  in the score. For this purpose, let  $V := (X^\top \ R^\top)^\top \in \mathbb{R}^p$  and set

$$\mathcal{S} = (V - \mu_V)^\top \Sigma_V^{-1} (V - \mu_V). \quad (8)$$

As before, we assume here that everything is well defined. The intuition behind this score is to work with standardized quantities. One may recognize in (8) the so-called Mahalanobis quadratic form, which in turn defines the Mahalanobis metric. This metric is classical in multivariate analysis and widely used for example for classification tasks (see [33]). The idea of including the covariate  $X$  in the score is inspired by the CI framework for time series, where  $X$  can contain the  $k$  previous residuals (see e.g. [31], Section 4.2, or the end of this subsection) : the joint distribution  $(X^\top \ R^\top)^\top$  is relevant for predicting the future  $\ell$  residuals (i.e.  $R$ ).

In the context of CI, the position and covariance parameters of  $V$  are of course unknown or even undefined, and it is natural to replace them with their empirical estimators. Nevertheless, in order to preserve exchangeability and so the finite sample theoretical guarantees provided by the classical CI procedure, one has to process such empirical estimators in an exchangeable fashion. Of course, another way to proceed would be to estimate these parameters separately, similarly to the pre-trained predictor, leading to a *double split* procedure. The first option is more challenging and this is our approach here. To begin with, we now introduce precisely our working score. We first set

$$V_i = \begin{pmatrix} X_i \\ R_i \end{pmatrix} \quad (i = 1, \dots, n) \text{ and, for } z \in \mathbb{R}^\ell, \quad V_{n+1}(z) = \begin{pmatrix} X_{n+1} \\ z \end{pmatrix}. \quad (9)$$

Here, the variable  $z = y - \hat{f}(X_{n+1})$  (resp.  $y$ ) is used to locate the *most likely* potential residuals (resp. predictions) in the CI machinery. We store the observation vectors  $V_i$ ,  $i = 1, \dots, n$  and the guess  $V_{n+1}(z)$  in a matrix  $\mathbf{V}(z) \in \mathcal{M}_{n+1,p}$  whose first  $n$  row are the  $V_i^\top$ ,  $i = 1, \dots, n$  and the last one is  $V_{n+1}^\top(z)$ . Following equation (9), we decompose  $\mathbf{V}(z)$  as  $\mathbf{V}(z) = (\mathbf{X} \ \mathbf{R}(z))$  with  $\mathbf{X} \in \mathcal{M}_{n+1,k}$  and  $\mathbf{R}(z) \in \mathcal{M}_{n+1,\ell}$ . The empirically centered counterpart of  $\mathbf{V}(z)$  is  $\mathbf{W}(z) = \pi_{\mathbb{I}}^\perp \mathbf{V}(z)$ . We further compute the ridge empirical covariance matrix  $\hat{\Sigma}_\lambda(z) \in \mathcal{M}_{p,p}$ ,

$$\hat{\Sigma}_\lambda(z) = \frac{1}{n} \mathbf{W}(z)^\top \mathbf{W}(z) + \lambda \mathbf{I}. \quad (10)$$

Here,  $\lambda > 0$  is the ridge parameter. With all these supplementary notations, we are now able to define the score  $S_i(z)$  ( $1 \leq i \leq n+1$ ) as the square of the empirical Mahalanobis norm of  $W_i(z)$  (the  $i^{\text{th}}$  row of  $\mathbf{W}(z)$  written in column). Namely, we have

$$S_i(z) = \|\hat{\Sigma}_\lambda^{-1/2}(z) W_i(z)\|^2 = W_i(z)^\top \hat{\Sigma}_\lambda^{-1}(z) W_i(z). \quad (11)$$

It is readily seen that  $S_i(z)$  is the  $i^{\text{th}}$  diagonal element of the following matrix,

$$\begin{aligned} \mathbf{S}(z) &= \mathbf{W}(z) \left( \frac{1}{n} \mathbf{W}(z)^\top \mathbf{W}(z) + \lambda \mathbf{I} \right)^{-1} \mathbf{W}(z)^\top \\ &= n \mathbf{W}(z) \left( \mathbf{W}(z)^\top \mathbf{W}(z) + n \lambda \mathbf{I} \right)^{-1} \mathbf{W}(z)^\top. \end{aligned} \quad (12)$$

Remarkably, and contrarily to the norm residual score, the first  $n$  scores  $(S_1(z), \dots, S_n(z))$  also depend on  $z$ . This property induces a number of technical difficulties which are dealt with in Section 3.4. To alleviate the notations we will sometimes omit the dependence in  $z$  in the matrix valued functions.

**First intuition on the confidence regions associated to  $S(z)$**  Before continuing with the main results, let us consider the Gaussian case  $(X^\top \ R^\top)^\top \sim \mathcal{N}(0, \Sigma)$ . Anticipating Section 4, we replace the ridge empirical covariance with  $\Sigma$  and set  $\lambda = 0$ . Then, setting  $T(z) = z - \Sigma_{21} \Sigma_{11}^{-1} X$  with  $z = y - \hat{f}(X)$  and using the block inversion lemma (equation (90)), our score amounts to

$$S(z) = \begin{pmatrix} X \\ z \end{pmatrix}^\top \Sigma^{-1} \begin{pmatrix} X \\ z \end{pmatrix} = X^\top (\Sigma^{11})^{-1} X + T(z)^\top (\Sigma / \Sigma^{11})^{-1} T(z). \quad (13)$$



Assume also that the quantile of order  $1 - \alpha$  of  $S(R)$  (denoted by  $q_{1-\alpha}^\mathcal{E}$ ) is known, so that we can replace the empirical quantile usually computed in CI by its theoretical value  $q_{1-\alpha}^\mathcal{E}$ . The confidence region becomes

$$\mathcal{E}_\alpha^\infty = \{z : X^\top (\Sigma^{11})^{-1} X + T(z)^\top (\Sigma/\Sigma^{11})^{-1} T(z) \leq q_{1-\alpha}^\mathcal{E}\}. \quad (14)$$

Under our Gaussianity assumption,  $q_{1-\alpha}^\mathcal{E} = q_{1-\alpha}(\chi^2(p))$ . The ellipsoid  $\mathcal{E}_\alpha^\infty$  has two important features : it is centered on the conditional expectation  $\mathbb{E}[R|X] = \Sigma_{21}\Sigma_{11}^{-1}X$  and its matrix is equal to the conditional covariance  $\text{Cov}(R|X) = \Sigma/\Sigma^{11}$ . However, we can see in this simple case that  $\mathcal{E}_\alpha^\infty = \emptyset \iff X^\top (\Sigma^{11})^{-1} X \geq q_{1-\alpha}^\mathcal{E}$ . This observation suggests the introduction of  $S'(z)$ , an adjusted version of the score  $S(z)$ , defined as

$$S'(z) = S(z) - X^\top (\Sigma^{11})^{-1} X = T(z)^\top (\Sigma/\Sigma^{11})^{-1} T(z). \quad (15)$$

The score  $S'$  is studied in Section 3.5. This new score leads to a second confidence region

$$\mathcal{F}_\alpha^\infty = \{z : T(z)^\top (\Sigma/\Sigma^{11})^{-1} T(z) \leq q_{1-\alpha}^\mathcal{F}\}, \quad (16)$$

where  $q_{1-\alpha}^\mathcal{F} = q_{1-\alpha}(\chi^2(\ell))$  in the Gaussian case. Contrarily to  $\mathcal{E}_\alpha^\infty$ , this region is never empty as  $\Sigma_{21}\Sigma_{11}^{-1}X \in \mathcal{F}_\alpha^\infty$ . In return, its finite sample counterpart  $\mathcal{F}_\alpha^n$  can equal to  $\mathbb{R}^\ell$  (Section 3.5). Observe finally that when working with the score matrix (12), it is not obvious anymore that the prediction region built with the score matrix (12) is an ellipsoid, although we expect to obtain a prediction set similar to  $\mathcal{E}_\alpha^\infty$ .

**Computation of the score** Let us comment the effective computation of equation (12) before stating the upcoming Theorem 3.1. The evaluation of  $\mathbf{S}(z)$  in (12) is not directly tractable: for each candidate  $z$ , both a matrix inversion and matrix products have to be performed. Fortunately, it turns out that explicit inversion tricks can be used to overcome this first issue. First, notice that the matrix  $\mathbf{V}(z)$  is a linear function of  $z$ . More explicitly, we may write  $\mathbf{V}(z) = \mathbf{V}(0) + \mathbf{A}(z)$ , where  $\mathbf{A}(z)$  is the rank one matrix given by

$$\mathbf{A}(z) = e_{n+1} \left( \sum_{s=1}^{\ell} z_s e_{k+s} \right)^\top = e_{n+1} (\mathbf{L}z)^\top, \quad (17)$$

where  $\mathbf{L} \in \mathcal{M}_{p,\ell}$  is the matrix such that  $\mathbf{L}^\top = (\mathbf{0}_{\ell,k} \mid \mathbf{I}_\ell)$ . Thus,

$$\mathbf{W}(z) = \pi_\perp^\perp \mathbf{V}(0) + \pi_\perp^\perp e_{n+1} (\mathbf{L}z)^\top = \mathbf{W}(0) + v(\mathbf{L}z)^\top. \quad (18)$$

Here,

$$v = \pi_\perp^\perp e_{n+1} = e_{n+1} - \frac{1}{n+1} \mathbb{1} = \frac{1}{n+1} (-1 - 1 \dots - 1 \ n)^\top, \quad \text{with} \quad \|v\|^2 = \frac{n}{n+1}. \quad (19)$$

Equation (18) shows that  $\mathbf{W}(z)$  is a rank one perturbation of  $\mathbf{W}(0)$ . This crucial property is exploited in Lemma B.2, to prove our main Theorem 3.1.

*Remark* (Time series framework). In the context of time series [31, 44],  $U_i$  takes the form  $U_i = (U_i^t)_{t \in \mathbb{N}}$ . Given a multi-horizon predictor  $(\hat{U}_{t+1}^i, \dots, \hat{U}_{t+\ell}^i) = \hat{f}(U_{t-k+1}^i, \dots, U_t^i)$  one then computes the time series of the residuals,  $R_t^i = U_t^i - \hat{U}_t^i$ . At a given time  $t$ , we aim at localizing with high probability the next  $\ell$  residuals,  $Y = (R_{t+1}, \dots, R_{t+\ell})$  using the  $k$  previous ones,  $X = (R_{t-k+1}, \dots, R_t)$ . It may also happen that the time series and so the residuals are both multidimensional ( $R_s \in \mathbb{R}^d$ ). This is e.g. the case for time series describing the evolution of a position in  $\mathbb{R}^3$  ( $d = 3$ ). In this case,  $X$  is a vector in  $\mathbb{R}^{kd}$  and  $Y$  as a vector in  $\mathbb{R}^{\ell d}$ . The vector corresponding to the  $i^{\text{th}}$  time series takes the form

$$V_i = (X_i^\top \ Y_i^\top)^\top \in \mathbb{R}^{dk+d\ell}. \quad (20)$$

This time series setting falls in the framework of this paper and will be developed extensively in [18].

The most direct interpretation of the score vector  $\text{diag}(\mathbf{S}(z)) = (S_1(z), \dots, S_{n+1}(z))$  is the empirical Mahalanobis distance. This metric is e.g. also the central tool in linear discriminant analysis (LDA) [39], Chapter 9-10. This is closely related to the Karhunen-Loève decomposition, which consists in finding an orthonormal basis  $(\xi_i \otimes \phi_i)_{i \in \mathbb{N}} \subset L^2(\mathbb{P}) \otimes L^2(\mathbb{R}^d)$  on which a random field is naturally decomposed. It is worth to notice that other interpretations in terms of leverage score and shape theory can also be made. The interested reader will find some insights in Appendix A.1.

### 3.3 The first ellipsoid $\mathcal{E}_\alpha^n$ : main theorem

We first need to introduce some notations. While the score  $\mathbf{S}(z)$  is built on a procedure which is exchangeable in the  $(n+1)$ -sample, we will in fact need quantities that are obtained only using the  $n$  sample  $(V_1, \dots, V_n)$ . For this, we introduce the matrices  $\mathbf{B}_n \in \mathcal{M}_{n,p}$ ,  $\hat{\Sigma}_{n,\lambda} \in \mathcal{M}_{p,p}$  and  $\mathbf{P}_{n,\lambda} \in \mathcal{M}_{n,n}$  such that

$$(\mathbf{B}_n)_{ij} := (V_i)_j - \frac{1}{n} \sum_{k=1}^n (V_k)_j, \quad \hat{\Sigma}_{n,\lambda} := \frac{1}{n} \mathbf{B}_n^\top \mathbf{B}_n + \lambda \mathbf{I}_p = \begin{pmatrix} \hat{\Sigma}_{n,\lambda}^{11} & \hat{\Sigma}_{n,\lambda}^{12} \\ \hat{\Sigma}_{n,\lambda}^{21} & \hat{\Sigma}_{n,\lambda}^{22} \end{pmatrix}, \quad (21)$$

$$\mathbf{P}_{n,\lambda} := \mathbf{B}_n (\mathbf{B}_n^\top \mathbf{B}_n + n\lambda \mathbf{I}_p)^{-1} \mathbf{B}_n^\top = \frac{1}{n} \mathbf{B}_n \hat{\Sigma}_{n,\lambda}^{-1} \mathbf{B}_n^\top. \quad (22)$$

Above,  $\hat{\Sigma}_{n,\lambda}^{11} \in \mathcal{M}_{k,k}$ ,  $\hat{\Sigma}_{n,\lambda}^{22} \in \mathcal{M}_{\ell,\ell}$ ,  $\hat{\Sigma}_{n,\lambda}^{12} \in \mathcal{M}_{k,\ell}$ , and  $\hat{\Sigma}_{n,\lambda}^{21} = (\hat{\Sigma}_{n,\lambda}^{12})^\top \in \mathcal{M}_{\ell,k}$ .  $\mathbf{B}_n$  contains the data of the  $n$ -sample after being centered with their empirical mean, and  $\hat{\Sigma}_{n,\lambda}$  is the corresponding empirical ridge covariance matrix. The matrix  $\mathbf{P}_{n,\lambda}$  is a regularized orthogonal projector, that is,  $\mathbf{P}_{n,0}$  is an orthogonal projector,  $\mathbf{P}_{n,\lambda}^\top = \mathbf{P}_{n,\lambda}$  and  $0 \preceq \mathbf{P}_{n,\lambda} \preceq \mathbf{P}_{n,0}$  for  $\lambda > 0$ . Denoting  $p_{i,n} = (\mathbf{P}_{n,\lambda})_{ii}$  ( $p_{i,n} \in [0, 1]$ ) we set

$$q_{n,\alpha} := np_{(n_\alpha)}, \quad (23)$$

the  $n_\alpha^{\text{th}}$  order statistic of the  $n$ -tuple  $(np_{1,n}, \dots, np_{n,n})$ . Note that  $p_{(n_\alpha)}$  is the empirical quantile of order  $(1 - \alpha)(n+1)/n$  for the  $n$ -tuple  $(p_{1,n}, \dots, p_{n,n})$ , and not  $1 - \alpha$  (apply e.g. [47], p. 305). Contrarily to standard  $n$ -samples, the  $p_{i,n}$  are not independent. For example, they are constrained by the following deterministic inequality:  $\sum_{i=1}^n p_{i,n} = \text{Tr}(\mathbf{P}_{n,\lambda}) \leq \text{Tr}(\mathbf{P}_{n,0}) = \text{rank}(\mathbf{P}_{n,0}) \leq p$ .

We can now state our first main result.

**Theorem 3.1.** *Let  $\alpha \in (1/(n+1), 1)$ . let  $((X_1, Y_1), \dots, (X_{n+1}, Y_{n+1})) \in (\mathcal{X} \times \mathcal{Y})^{n+1}$  exchangeable pairs of random vectors. If  $q_{n,\alpha} \geq n-1$ , we set  $\mathcal{E}_\alpha^n := \mathbb{R}^\ell$ . If  $q_{n,\alpha} < n-1$ , we define  $\mathcal{E}_\alpha^n$  as the following ellipsoid,*

$$\mathcal{E}_\alpha^n = \{z \in \mathbb{R}^\ell : (z - Z_0^n)^\top \mathcal{A}_n^{-1} (z - Z_0^n) \leq \rho_{n,\alpha}\}, \quad (24)$$

where, setting  $\bar{X}_n := n^{-1} \sum_{i=1}^n X_i$ ,  $\bar{R}_n := n^{-1} \sum_{i=1}^n R_i$  and  $X_{n+1}^c = X_{n+1} - \bar{X}_n$ ,

$$\mathcal{A}_n = \hat{\Sigma}_{n,\lambda} / \hat{\Sigma}_{n,\lambda}^{11}, \quad (25)$$

$$Z_0^n = \hat{\Sigma}_n^{21} (\hat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c + \bar{R}_n, \quad (26)$$

$$\rho_{n,\alpha} = \frac{q_{n,\alpha} + 1}{1 - (q_{n,\alpha} + 1)/n} - 1 - (X_{n+1}^c)^\top (\hat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c. \quad (27)$$

Then the set  $\mathcal{E}_\alpha^n$  satisfies

$$\mathbb{P}(Y_{n+1} - \hat{Y}_{n+1} \in \mathcal{E}_\alpha^n) \geq 1 - \alpha. \quad (28)$$

The ellipsoid in Theorem 1.1 is recovered by  $E_\alpha^n = \{\hat{Y}_{n+1}\} + \mathcal{E}_\alpha^n$ , in the sense of the usual Minkowski set addition.

An important feature of Lemma B.3 is that  $\mathcal{A}_n$  does not depend on  $V_{n+1}$ . Likewise,  $\rho_{n,\alpha}$  and  $Z_0^n$  depend on  $V_{n+1}$  only through  $X_{n+1}$ . Interestingly, when the distribution of  $V_1$  is elliptical (see Section 4.3) and  $V_1$  admits second order moment, the matrix  $\mathcal{A}_n$  corresponds to an empirical estimate of the conditional covariance matrix  $\text{Cov}(R_1|X_1)$ , up to a random positive multiplicative constant (see [39], Theorem 1.5.4). In fact, one can see from equation (24) that only  $\rho_{n,\alpha} \mathcal{A}_n$  is uniquely defined. This allows us to interpret  $\mathcal{A}_n$  as an empirical conditional covariance matrix for such distributions, up to a rescaling of  $\rho_{n,\alpha}$  with the said multiplicative constant.

*Remark* (Empirical linear conditional expectation). Observe that  $Z_0^n = \hat{e}_n + \hat{\mathbf{A}}_n X_{n+1}$ , where

$$\hat{\mathbf{A}}_n = \hat{\Sigma}_n^{21} (\hat{\Sigma}_{n,\lambda}^{11})^{-1}, \quad \hat{e}_n = \bar{R}_n - \hat{\mathbf{A}}_n \bar{X}_n. \quad (29)$$

Interestingly, it is easily shown that  $(\hat{e}_n, \hat{\mathbf{A}}_n)$  is an M-estimator obtained by minimizing

$$T_n(e, \mathbf{A}) = \frac{1}{n} \sum_{i=1}^n \|R_i - e - \mathbf{A} X_i\|^2 + \lambda \|\mathbf{A}\|_F^2, \quad (30)$$



which simply corresponds to ridge multivariate linear regression. The continuous counterpart of  $T_n$  is  $T(e, \mathbf{A}) = \mathbb{E}[\|R - e - \mathbf{A}X\|^2] + \lambda\|\mathbf{A}\|_F^2$ , which is minimal for  $\mathbf{A}_* = \Sigma^{21}(\Sigma_\lambda^{11})^{-1}$  and  $e_* = \mathbb{E}[R_1] - \mathbf{A}_*\mathbb{E}[X_1]$ . The pair  $(e_*, \mathbf{A}_*)$  yields the best affine approximation of  $R$  in terms of  $X$ , in the sense of  $L^2(\mathbb{P})$ , and the resulting affine map is known as the linear conditional expectation ([26], e.g. Theorem 4.14). In comparison, the (full) conditional expectation is obtained by minimizing the map  $T(f) = \mathbb{E}[\|R - f(X)\|^2]$  over all measurable maps  $f$ . Finally,  $\mathcal{A}_n$  and  $(\hat{e}_n, \hat{\mathbf{A}}_n)$  are linked together by  $\min T_n = T_n(\hat{e}_n, \hat{\mathbf{A}}_n) = \text{Tr}(\mathcal{A}_n) - \ell\lambda$ .

In the previous theorem, the introduced objects are not exchangeable in terms of the  $(n+1)$ -sample. Nevertheless, we will show in next section how the ellipsoid  $\mathcal{E}_\alpha^n$  can be exchangeably rewritten.

### 3.4 From the score matrix $\mathbf{S}(z)$ to the ellipsoid $\mathcal{E}_\alpha^n$

An important first step consists in rewriting the score matrix  $\mathbf{S}(z)$ , using the fact that  $\mathbf{W}(z)$  can be written as a rank-one perturbation (see eq. (18)). Applying then Lemma B.2 to  $\mathbf{S}(z)$ , we obtain that  $\mathbf{S}(z)$  is of the form

$$\mathbf{S}(z) = n\mathbf{C}_n - n \frac{b^n(z)b^n(z)^\top}{1 + d_n(z)}. \quad (31)$$

Detailed expressions of  $\mathbf{C}_n$ ,  $b^n(z)$  and  $d_n(z)$  are given in Lemma B.4. Notably, the vector  $b^n(z)$  is affine in  $z$  and  $d^n(z)$  is a positive quadratic form in  $z$ . Recall that we are interested in the diagonal terms  $(\mathbf{S}(z))_{ii}$ , hereinafter denoted by  $S_i(z)$  for short. Equation (31) implies that the evaluation of the full score vector  $(S_1(z), \dots, S_{n+1}(z))$  requires only one matrix inversion ( $\mathbf{D}_\mu^{-1}$ , see Lemma B.2), instead of one inversion per  $z$  candidate if using formula (12). The second step consists in applying Lemma B.3, which shows that  $\mathbf{C}_n$  in equation (31) is given by

$$\mathbf{C}_n = \begin{pmatrix} \mathbf{P}_{n,\lambda} & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 0 \end{pmatrix} + ww^\top \in \mathcal{M}_{n+1,n+1}, \quad w := \frac{v}{\|v\|}. \quad (32)$$

Here,  $\mathbf{P}_{n,\lambda}$  is given in equation (22). In particular, for  $i \in \{1, \dots, n\}$ ,  $(\mathbf{C}_n)_{ii} = p_{i,n} + 1/n(n+1)$ . In fact,  $\mathbf{C}_n$  does not depend on  $V_{n+1}$ , while  $d_n(z)$  and  $b^n(z)$  only depend on  $V_{n+1}$  through  $X_{n+1}$ .

Note that up until now, the expression for the score  $\mathbf{S}(z)$  is exact. However, a limitation remains: the scores  $S_i(z)$  of the first  $n$  examples also depend on  $z = y - \hat{Y}^{n+1}$ . In particular, the conformal region  $\mathcal{C}_\alpha^n$  for our score is  $\mathcal{C}_\alpha^n = \{\hat{Y}_{n+1}\} + \mathcal{C}_\alpha^n$ , where

$$\mathcal{C}_\alpha^n = \{z \in \mathbb{R}^\ell : S_{n+1}(z) \leq S_{(n_\alpha)}(z)\}. \quad (33)$$

Here,  $S_{(n_\alpha)}(z)$  is the  $n_\alpha^{\text{th}}$  order statistic of the  $n$ -tuple  $(S_1(z), \dots, S_n(z))$ . As such, the computation of the region  $\mathcal{C}_\alpha^n$  still requires that the scoring rule be tested for each  $y \in \mathcal{Y}_{\text{trial}}$ . To alleviate this second computational difficulty, a simple approximation of the confidence region is given by the next lemma.

**Lemma 3.2** (Conservative approximation of  $\mathcal{C}_\alpha^n$ ). *Let  $\alpha \in (1/(n+1), 1)$ . Introduce the set  $\tilde{\mathcal{C}}_\alpha^n$ , defined as*

$$\tilde{\mathcal{C}}_\alpha^n := \left\{ z \in \mathbb{R}^\ell : S_{n+1}(z) \leq q_{n,\alpha} + \frac{1}{n+1} \right\}. \quad (34)$$

*Then  $\mathcal{C}_\alpha^n \subset \tilde{\mathcal{C}}_\alpha^n$ ; as a result,  $\mathbb{P}(Y_{n+1} - \hat{Y}_{n+1} \in \tilde{\mathcal{C}}_\alpha^n) \geq 1 - \alpha$ .*

Before commenting Lemma 3.2 above, we can already state our main result, which completes the proof of Theorem 3.1.

**Proposition 3.3.** *Under the assumptions of Theorem 3.1,  $\tilde{\mathcal{C}}_\alpha^n = \mathcal{E}_\alpha^n$ .*

The approximation given in equation (34) corresponds to discarding the contribution  $b^n(z)b^n(z)^\top$  in  $S_i(z)$  for  $i \neq n+1$  in equation (31), observing that  $S_i(z) \leq n(\mathbf{C}_n)_{ii}$  uniformly in  $z$ . This approximation is further justified by Proposition 3.4 below, which states that under additional moment assumptions, the surplus volume converges to 0 in probability. Note however that currently, we are not able to prove an upper bound similar to equation (7) for  $\tilde{\mathcal{C}}_\alpha^n$ .

**Proposition 3.4.** *Assume that the  $V_i$  are iid, that  $\lambda > 0$  and that  $\mathbb{E}[\|V_1\|^{4q}] < +\infty$  for some  $q > 1$ . Then for all compact set  $K \subset \mathbb{R}^\ell$ ,*

$$\text{Vol}((\mathcal{E}_\alpha^n \setminus \mathcal{C}_\alpha^n) \cap K) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (35)$$

The interested reader will find in Appendix A.2 discussions on the settings of the above Proposition, in particular the case  $\lambda = 0$  and the finite  $4q$ -th moment assumption.

We conclude this section with the following lemma, which provides a sufficient condition on  $n$  and  $p = k + \ell$  so that full space ellipsoids ( $\mathcal{E}_\alpha^n = \mathbb{R}^\ell$ ) never occur.

**Lemma 3.5** (Sufficient condition for bounded ellipsoids). *For all  $\lambda \geq 0$ , we have  $q_{n,\alpha} < n - 1$  almost surely as soon as  $n > p + 1$  and*

$$\alpha \geq \frac{p+1}{n+1}. \quad (36)$$

A sharper bound, where  $\lambda$  and  $\max \text{Spec}(\widehat{\Sigma}_{n,\lambda})$  appear, can be obtained by writing an SVD of  $\mathbf{P}_{n,\lambda}$ . Equation (36) is to be compared with the standard requirement in conformal inference that  $\alpha > 1/(n+1)$ . Note also that  $n+1$  and  $p$  are the number of rows and columns of  $\mathbf{V}$ , respectively. In practice, we observe that  $q_{n,\alpha} < n - 1$  may hold even if  $p + 1 \geq n$ , the proof of Lemma 3.5 being a worst-case analysis.

**Metrics of the ellipsoid** An important metric related to the shape of an ellipsoid is its principal eccentricity. Given an ellipsoid  $\mathcal{E} = \{z \in \mathbb{R}^\ell : (z - z_0)\mathcal{A}^{-1}(z - z_0) \leq 1\}$ , this corresponds to

$$e = \sqrt{1 - \lambda_m/\lambda_M} \in [0, 1), \quad (37)$$

where  $\lambda_m$  is the smallest eigenvalue of  $\mathcal{A}$  and  $\lambda_M$  is the largest one. Roughly speaking, this metric measures how “different” is a given ellipsoid from a ball, the latter corresponding to  $e = 0$ . Likewise, if  $e$  is close to 1 then  $\mathcal{E}$  is accordingly flat. The volume of  $\mathcal{E}$  above is

$$\text{Vol}(\mathcal{E}) = v_\ell \sqrt{\det(\mathcal{A})}, \quad \text{where} \quad v_\ell = \pi^{\ell/2}/\Gamma(\ell/2 + 1). \quad (38)$$

Here,  $v_\ell$  corresponds to the volume of the  $\ell$ -dimensional unit ball. In our case, the determinant of  $\mathcal{A}_n$  can be further computed as (see e.g. [20], Section 0.8.5)

$$\det(\mathcal{A}_n) = \det(\widehat{\Sigma}_{n,\lambda}/\widehat{\Sigma}_{n,\lambda}^{11}) = \det(\widehat{\Sigma}_{n,\lambda})/\det(\widehat{\Sigma}_{n,\lambda}^{11}). \quad (39)$$

**Case where  $\mathcal{E}_\alpha^n$  is empty** Empty confidence regions  $\mathcal{E}_\alpha^n$  correspond to  $\rho_{n,\alpha} < 0$ . This event has a non-null probability, as e.g. visible from equation (55) (still,  $\mathbb{P}(\mathcal{E}_\alpha^n = \emptyset) \leq \alpha$  by construction). To understand the event  $\{\mathcal{E}_\alpha^n = \emptyset\}$ , observe that  $\mathcal{E}_\alpha^n$  can be written as  $\mathcal{E}_\alpha^n = \mathcal{V}(X_{n+1}) \cap \mathcal{E}$ , where  $\mathcal{V}(X_{n+1})$  is an  $\ell$ -dimensional affine subspace of  $\mathbb{R}^{k+\ell}$  and  $\mathcal{E}$  is a  $(k + \ell)$ -dimensional ellipsoid. They are given by, for some explicit  $\beta_n > 0$ ,

$$\mathcal{V}(X_{n+1}) := \{v = (x^\top r^\top)^\top \in \mathbb{R}^{k+\ell} : x = X_{n+1}\},$$

$$\mathcal{E} := \{v \in \mathbb{R}^{k+\ell} : (v - \bar{V}_n)^\top (\widehat{\Sigma}_{n,\lambda})^{-1} (v - \bar{V}_n) \leq \beta_n\}, \quad \bar{V}_n = \frac{1}{n} \sum_{i=1}^n V_i.$$

(to obtain  $\mathcal{E}$ , combine equations (125) and (126)). In this setting,  $\mathcal{E}_\alpha^n$  is empty precisely when  $\mathcal{V}(X_{n+1}) \cap \mathcal{E} = \emptyset$ . This is interpreted as  $X_{n+1}$  being so abnormal (w.r.t. the score  $S(z)$ ) that no value of  $z$  can make the pair  $(X_{n+1}^\top z^\top)^\top$  conform to the samples  $(V_1, \dots, V_n)$ . While this is unexpected from a UQ perspective, this is natural in the context of outlier detection :  $X_{n+1}$  is so abnormal that  $(X_{n+1}^\top z^\top)^\top$  is always an outlier, whatever the value of  $z$ .

Empty confidence regions can be problematic from an applied perspective. Moreover, empty  $\mathcal{E}_\alpha^n$  are not devoid of information, since  $\mathcal{A}_n$  and  $Z_0^n$  are still defined when  $\rho_{n,\alpha} < 0$ . To capitalize on  $\mathcal{A}_n$  and  $Z_0^n$  as well as avoid empty confidence regions, we can replace  $\rho_{n,\alpha}$  with  $\rho_{n,\alpha}^\varepsilon = \max(\rho_{n,\alpha}, \varepsilon^2)$ , where  $\varepsilon > 0$  is a length associated to a minimum user set volume  $V_{\min}$ . The ellipsoid is then given by

$$(z - Z_0^n)^\top (\rho_{n,\alpha}^\varepsilon \mathcal{A}_n)^{-1} (z - Z_0^n) \leq 1. \quad (40)$$

$\varepsilon$  is found by solving  $V_{\min} = v_\ell \sqrt{\det(\rho_{n,\alpha}^\varepsilon \mathcal{A}_n)} = v_\ell \varepsilon^\ell \sqrt{\det(\mathcal{A}_n)}$ , i.e.

$$\varepsilon = \left( V_{\min} / (v_\ell \det(\mathcal{A}_n)^{1/2}) \right)^{1/\ell}. \quad (41)$$

By construction, this procedure leads to overcoverage, but if  $V_{\min}$  is chosen small enough (e.g. if  $V_{\min} \ll \mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^n)]$ ), it will not affect the empirical average volume of the confidence ellipsoid  $\mathcal{E}_{\alpha,\varepsilon}^n$ , since the cutoff  $\rho_\varepsilon = \max(\rho, \varepsilon^2)$  is only activated when the volume of  $\mathcal{E}_\alpha^n$  is very small or null.

A more mathematically sound way of avoiding empty regions is found by using the score  $S'(z)$  introduced below.

### 3.5 The adjusted Mahalanobis score and the second ellipsoid $\mathcal{F}_\alpha^n$

Following equation (15), we can correct the score matrix (12) to prevent empty confidence regions, by subtracting the  $\mathbf{X}$  part of  $\mathbf{S}(z)$  to  $\mathbf{S}(z)$ . Explicitly, define  $\mathbf{F} := \pi_1^\perp \mathbf{X} \in \mathcal{M}_{n+1,k}$ , where the  $i^{\text{th}}$  row of  $\mathbf{X}$  is  $X_i^\top$ , and introduce

$$\mathbf{S}'(z) := \mathbf{S}(z) - \mathbf{S}_X, \quad \mathbf{S}_X = \mathbf{F} \left( \frac{1}{n} \mathbf{F}^\top \mathbf{F} + \lambda \mathbf{I}_k \right)^{-1} \mathbf{F}^\top. \quad (42)$$

As with  $\mathbf{S}(z)$ , the associated score for  $V_i$  is  $S'_i(z) = \mathbf{S}'(z)_{ii}$  and the associated conformal set is  $\{z : S'_{n+1}(z) \leq S'_{(n_\alpha)}(z)\}$ , where  $S'_{(n_\alpha)}(z)$  is an order statistic built from  $(S'_1(z), \dots, S'_n(z))$ . Because  $\mathbf{X}$  depends on  $X_{n+1}$  through a rank-one perturbation, namely  $\mathbf{X} = (X_1 \dots X_n \ 0_{k,1})^\top + e_{n+1} X_{n+1}^\top$ , we can use Lemmas B.2 and B.3 to rewrite  $\mathbf{S}_X$  as

$$\frac{1}{n} \mathbf{S}_X = \begin{pmatrix} \mathbf{P}_{n,\lambda}^{XX} & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 0 \end{pmatrix} + ww^\top - \frac{b_X^n (b_X^n)^\top}{1 + d_n^X}, \quad (43)$$

for some matrix  $\mathbf{P}_{n,\lambda}^{XX} \in \mathcal{M}_{n,n}$ , some vector  $b_X^n \in \mathcal{M}_{n+1,1}$  and some  $d_n^X \in \mathbb{R}^+$  (see Lemma B.5 for expressions). Notably,

$$d_n^X = \frac{1}{n+1} (X_{n+1}^c)^\top (\boldsymbol{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c, \quad X_{n+1}^c = X_{n+1} - \bar{X}_n.$$

The matrix  $\mathbf{S}'(z)$  is finally written as

$$\mathbf{S}'(z) = n \begin{pmatrix} \mathbf{P}_{n,\lambda} - \mathbf{P}_{n,\lambda}^{XX} & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 0 \end{pmatrix} - n \frac{b^n(z) b^n(z)^\top}{1 + d_n(z)} + n \frac{b_X^n (b_X^n)^\top}{1 + d_n^X}.$$

As in Lemma 3.2, we can discard the  $b^n(z)$  term in  $S'_i(z)$ , for  $i \neq n+1$ , leading to a second ellipsoidal set  $\mathcal{F}_\alpha^n$ . To describe this set, we finally need to introduce

$$p'_{i,n} := (\mathbf{P}_{n,\lambda} - \mathbf{P}_{n,\lambda}^{XX})_{ii} + \frac{(b_X^n)_i^2}{1 + d_n^X}, \quad \text{and} \quad q'_{n,\alpha} := n p'_{(n_\alpha)}, \quad (44)$$

where  $p'_{(n_\alpha)}$  is the order statistic of order  $n_\alpha$  of  $(p'_{1,n}, \dots, p'_{n,n})$ . As in Lemma 3.2,  $p'_{i,n}$  corresponds to  $\mathbf{S}'(z)_{ii}$  after discarding the  $b^n(z)_i$  term. The analog of Theorem 3.1 for the score  $S'(z)$  is the following, which describes a second ellipsoid  $\mathcal{F}_\alpha^n$ .

**Theorem 3.6.** *Take the assumptions of Theorem 3.1, and define*

$$t_n := \left( \frac{n+1}{n} \right) (1 + d_n^X). \quad (45)$$

*If  $t_n q'_{n,\alpha} \geq n$ , we set  $\mathcal{F}_\alpha^n = \mathbb{R}^\ell$ . Else we set*

$$\mathcal{F}_\alpha^n := \{z \in \mathbb{R}^\ell : (z - Z_0^n)^\top \mathcal{A}_n^{-1} (z - Z_0^n) \leq \rho'_{n,\alpha}\}, \quad (46)$$

*where  $Z_0^n$  and  $\mathcal{A}_n$  are given in Theorem 3.1, and where*

$$\rho'_{n,\alpha} := t_n^2 \frac{q'_{n,\alpha}}{1 - t_n q'_{n,\alpha}/n}. \quad (47)$$

*Then*

$$\mathbb{P}(Y_{n+1} - \hat{Y}_{n+1} \in \mathcal{F}_\alpha^n) \geq 1 - \alpha. \quad (48)$$

As for  $E_\alpha^n$  and  $\mathcal{E}_\alpha^n$ , the ellipsoid for  $Y_{n+1}$  is recovered by  $F_\alpha^n = \{\hat{Y}_{n+1}\} + \mathcal{F}_\alpha^n$ . The only difference between  $\mathcal{E}_\alpha^n$  and  $\mathcal{F}_\alpha^n$  is the parameter  $\rho'_{n,\alpha}$ . The next proposition shows that  $\mathcal{F}_\alpha^n$  is never empty. However, there is a chance that  $\mathcal{F}_\alpha^n = \mathbb{R}^\ell$ : there is no equivalent of Proposition 3.4 for  $\mathcal{F}_\alpha^n$ , except when  $n$  tends to  $\infty$ .

**Proposition 3.7.** *Under the assumptions of Theorem 3.1,  $\rho'_{n,\alpha} \geq 0$ . In particular,  $Z_0^n \in \mathcal{F}_\alpha^n$  and  $\mathcal{F}_\alpha^n$  is never empty. Finally, under the assumptions of Proposition 4.1,*

$$\mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell) \xrightarrow{n \rightarrow \infty} 0.$$

*Remark* (Behaviour of  $\mathcal{E}_\alpha^n$  and  $\mathcal{F}_\alpha^n$  w.r.t. large  $d_n^X$ ). Observe that a large  $d_n^X$  means that  $X_{n+1}$  is far away from  $\bar{X}_n$ , in the data metric  $\Sigma_{n,\lambda}$ . Interestingly,  $\mathcal{E}_\alpha^n$  and  $\mathcal{F}_\alpha^n$  have opposite behaviours with reference to large  $d_n^X$ . From equation (27), large values of  $d_n^X$  correspond to small  $\mathcal{E}_\alpha^n$  ( $\mathcal{E}_\alpha^n = \emptyset$  in the worst case, see Proposition 3.1), while they correspond to large  $\mathcal{F}_\alpha^n$  ( $\mathcal{F}_\alpha^n = \mathbb{R}^\ell$  in the worst case, see Proposition 3.6). In the first case, this is interpreted as  $X_{n+1}$  being so abnormal that no value of  $z$  can make the pair  $(X_{n+1}^\top z^\top)^\top$  conform to the samples  $(V_1, \dots, V_n)$ . In the second case, this is interpreted as the whole calibration sample  $(V_1, \dots, V_n)$  being uninformative with reference to the current sample  $(X_{n+1}^\top R_{n+1}^\top)^\top$ , when the new input  $X_{n+1}$  is too abnormal when compared to  $(X_1, \dots, X_n)$ .

## 4 Asymptotic analysis

### 4.1 The first asymptotic ellipsoid $\mathcal{E}_\alpha^\infty$

In the limit where  $n \rightarrow \infty$ , we are able to state the following result.

**Proposition 4.1.** *Assume that  $V_1$  has a well-defined covariance matrix  $\Sigma$ , and that the vectors  $V_i$ ,  $i \in \mathbb{N}$  are iid. Denote  $V_c := V_1 - \mathbb{E}[V_1]$ , with the decomposition  $V_c = (X_c^\top R_c^\top)^\top$ ,  $X_c \in \mathbb{R}^k$ ,  $R_c \in \mathbb{R}^\ell$ . Assume also that  $\lambda > 0$  and that the quantile function of  $\|(\Sigma + \lambda \mathbf{I}_p)^{-1/2} V_c\|^2$  is continuous on a neighbourhood of  $1 - \alpha$ . Write  $\Sigma + \lambda \mathbf{I}_p$  in blockwise fashion, according to the decomposition  $p = k + \ell$ :*

$$\Sigma_\lambda := \Sigma + \lambda \mathbf{I}_p = \begin{pmatrix} \Sigma_\lambda^{11} & \Sigma_\lambda^{12} \\ \Sigma_\lambda^{21} & \Sigma_\lambda^{22} \end{pmatrix}. \quad (49)$$

( $\Sigma_\lambda^{11} \in \mathcal{M}_{k,k}$ ,  $\Sigma_\lambda^{22} \in \mathcal{M}_{\ell,\ell}$ ,  $\Sigma_\lambda^{12} \in \mathcal{M}_{k,\ell}$ .) Denote  $q_{1-\alpha}^\mathcal{E} := q_{1-\alpha}(V_c^\top \Sigma_\lambda^{-1} V_c)$ . Then, as  $n \rightarrow \infty$ ,

1. (Asymptotic ellipsoid)

$$q_{n,\alpha} \xrightarrow[n \rightarrow \infty]{a.s.} q_{1-\alpha}^\mathcal{E}, \quad (50)$$

$$\mathcal{A}_n \xrightarrow[n \rightarrow \infty]{a.s.} \mathcal{A}_\infty := \Sigma_\lambda / \Sigma_\lambda^{11}, \quad (51)$$

$$\rho_{n,\alpha} \xrightarrow[n \rightarrow \infty]{d} \rho_{\infty,\alpha} := q_{1-\alpha}^\mathcal{E} - X_c^\top (\Sigma_\lambda^{11})^{-1} X_c, \quad (52)$$

$$Z_0^n \xrightarrow[n \rightarrow \infty]{d} Z_0^\infty := \Sigma_\lambda^{21} (\Sigma_\lambda^{11})^{-1} X_c + \mathbb{E}[R_1]. \quad (53)$$

2. (Asymptotic volume and probability of empty regions)

$$\text{Vol}(\mathcal{E}_\alpha^n) \xrightarrow[n \rightarrow \infty]{d} v_\ell \sqrt{\det(\Sigma_\lambda / \Sigma_\lambda^{11})} (q_{1-\alpha}^\mathcal{E} - X_c^\top (\Sigma_\lambda^{11})^{-1} X_c)_+^{\ell/2}, \quad (54)$$

$$\mathbb{P}(\mathcal{E}_\alpha^n = \emptyset) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(X_c^\top (\Sigma_\lambda^{11})^{-1} X_c > q_{1-\alpha}^\mathcal{E}) \leq \alpha. \quad (55)$$

Finally, if  $\Sigma$  is invertible, then all the results above remain true when  $\lambda = 0$ .

In equation (54), we underline that  $(x)_+^s = \max(0, x)^s \neq \max(0, x^s)$ ; in particular,  $(x)_+^s = 0$  if  $x \leq 0$ . We now define the random ellipsoid  $\mathcal{E}_\alpha^\infty$  as

$$\mathcal{E}_\alpha^\infty := \{z \in \mathbb{R}^\ell : (z - Z_0^\infty)^\top \mathcal{A}_\infty^{-1} (z - Z_0^\infty) \leq \rho_{\infty,\alpha}\}. \quad (56)$$

Observe that  $\mathcal{E}_\alpha^\infty$  is a confidence region of level  $1 - \alpha$  for  $R_1$ : from equation (90) and the expressions of  $\mathcal{A}_\infty$ ,  $Z_0^\infty$  and  $\rho_{\infty,\alpha}$ , we can show that (see equations (171) and (172))

$$V_c^\top \Sigma_\lambda^{-1} V_c \leq q_{1-\alpha}^\mathcal{E} \iff (R_c - Z_0^\infty)^\top \mathcal{A}_\infty^{-1} (R_c - Z_0^\infty) \leq \rho_{\infty,\alpha}, \quad (57)$$

and the left-hand side has probability  $1 - \alpha$  from the definition of  $q_{1-\alpha}^\mathcal{E}$ . Next, the distribution of the volume  $\text{Vol}(\mathcal{E}_\alpha^\infty)$  volume is equal to that of the right-hand side of (54). In this paper, we do not tackle the question of random set convergence, i.e. in what sense does the convergence  $\mathcal{E}_\alpha^n \rightarrow \mathcal{E}_\alpha^\infty$  hold. This study would require the use of random set theory [38], a perspective that we leave for future work.

Note that the limit volume  $\text{Vol}(\mathcal{E}_\alpha^\infty)$  is bounded almost surely, hence all its moments are finite (equation (75)). However, there is no guarantee that the moments of  $\text{Vol}(\mathcal{E}_\alpha^n)$  converge toward those of  $\text{Vol}(\mathcal{E}_\alpha^\infty)$ . Such a property

requires additional uniform integrability properties on  $\text{Vol}(\mathcal{E}_\alpha^n)$ , which have to be checked on a case-by-case basis. For example, this property holds in the Gaussian case (Proposition 4.2). The main obstacle in the proof of Proposition 4.1 is that of the convergence of the empirical quantile  $q_{n,\alpha}$  to  $q_{1-\alpha}^\mathcal{E}$ , as the diagonal elements of  $\mathbf{C}_n$  are not independent. Our proof of this result relies on the study of empirical characteristic functions and an almost sure application of Lévy's theorem (Lemma B.6). Note that the asymptotic matrix  $\mathbf{A}_\infty$  is the covariance matrix of  $\pi_X^\perp R$ , where  $\pi_X^\perp R$  is the orthogonal projection of  $R = (R_1, \dots, R_\ell)$  onto  $[\text{Span}(X_1, \dots, X_k)]^\perp$  in  $L^2(\mathbb{P})$ .

It is worth noticing that the numerical experiments highlight that the asymptotic ellipsoid (as well as its volume) may still exist when the data are heavy tailed. A starting point to study that case is to note that, according to equation (24), the region  $\mathcal{E}_\alpha^n$  is mostly determined by the product  $\rho_{n,\alpha} \mathbf{A}_n$ . Then scalings differing from the one used in Proposition 4.1 (i.e.  $(\rho_{n,\alpha}/u_n)(u_n \mathbf{A}_n)$  with  $u_n \neq 1$ ) may be introduced, in the hope that  $u_n \mathbf{A}_n$  and  $\rho_{n,\alpha}/u_n$  may converge in some sense. The interested reader will find first insights in Appendix A.3 for a further study of heavy-tailed data.

*Remark* (Conditional coverage). Although our method is built on a form of conditional procedure, it cannot fulfill the conditional coverage described in the introduction, even when  $n \rightarrow \infty$ . Indeed, if  $X_1 = x$ , then  $\mathcal{E}_\alpha^\infty$  is empty as soon as  $q_{1-\alpha}^\mathcal{E} < (x - \mathbb{E}[X_1])^\top (\Sigma^{11})^{-1} (x - \mathbb{E}[X_1])$  (Proposition 4.1). In particular, for all such  $x$ ,

$$\mathbb{P}(R_1 \in \mathcal{E}_\alpha^\infty | X_1 = x) = \mathbb{P}(R_1 \in \emptyset | X_1 = x) = 0 \neq 1 - \alpha. \quad (58)$$

To recover conditional coverage, the asymptotic quantile  $q_{1-\alpha}^\mathcal{E} = q_{1-\alpha}(V_c^\top \Sigma^{-1} V_c)$  should be replaced with  $q_{1-\alpha}^\mathcal{E}(x) := q_{1-\alpha}(V_c^\top \Sigma^{-1} V_c | X_1 = x)$  in equation (52). Indeed, let us define

$$\rho_{\infty,\alpha}(x) := q_{1-\alpha}^\mathcal{E}(x) - (x - \mathbb{E}[X_1])^\top (\Sigma^{11})^{-1} (x - \mathbb{E}[X_1]), \quad (59)$$

and define  $\mathcal{E}_\alpha^\infty(x)$  to be the set  $\mathcal{E}_\alpha^\infty$  where  $\rho_{\infty,\alpha}$  has been replaced with  $\rho_{\infty,\alpha}(x)$ . Then, from equation (172), we recover conditional coverage in the limit where  $n \rightarrow \infty$ :

$$\begin{aligned} \mathbb{P}(R_1 \in \mathcal{E}_\alpha^\infty(x) | X = x) &= \mathbb{P}((R_c - Z_0^\infty)^\top \mathbf{A}_\infty^{-1} (R_c - Z_0^\infty) \leq \rho_{\infty,\alpha}(x) | X_1 = x) \\ &= \mathbb{P}(V_c^\top \Sigma_\lambda^{-1} V_c \leq q_{1-\alpha}^\mathcal{E}(x) | X_1 = x) = 1 - \alpha. \end{aligned} \quad (60)$$

This solution also solves the problem of empty confidence regions described in Section 3. Of course, the main difficulty is that of the estimation of the conditional quantile  $q_{1-\alpha}(x)$ . This can be done e.g. using conformalized quantile regression [43], although the application of this method has to be further studied in our case.

**Gaussian data** In this paragraph, we assume that  $V_1$  is a Gaussian random vector. In this case,  $Z_0^\infty = \mathbb{E}[R_1 | X_1]$  and  $\mathbf{A}_\infty = \text{Cov}(R_1 | X_1)$ . In fact, the limit ellipsoid  $\mathcal{E}_\alpha^\infty$  is an  $\ell$ -dimensional section of the  $k + \ell$ -dimensional confidence ellipsoid obtained under Gaussianity assumptions over the random vector  $(X_1, R_1)$  (this is also true under general ellipticity assumptions, see equation (71) for a definition). CCLE can thus be understood as a “non asymptotic conformal generalisation” of the regions provided by standard Gaussian ellipsoids. In the Gaussian case, we can further describe the volume of  $\mathcal{E}_\alpha^\infty$  when  $\lambda = 0$ , notably its moments.

**Proposition 4.2.** *Assume that the vectors  $V_i$  are iid, the vector  $V_1$  is Gaussian,  $\lambda = 0$  and  $\min \text{Spec}(\Sigma) > 0$ . Denoting  $F_{\chi^2(m)}$  the CDF of the  $\chi^2(m)$  distribution, we have  $q_{1-\alpha}^\mathcal{E} = F_{\chi^2(k+\ell)}^{-1}(1 - \alpha)$ . Next, denoting  $B(x, y)$  the Euler Beta function, we set  $C_{k,\ell,q} := 2^{-k/2} v_\ell^q B(k/2, q\ell/2 + 1) / \Gamma(k/2)$ . Then for all  $q > 0$ ,*

$$\mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^n)^q] \xrightarrow{n \rightarrow \infty} \mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^\infty)^q] = C_{k,\ell,q} \det(\Sigma / \Sigma^{11})^{q/2} (q_{1-\alpha}^\mathcal{E})^{(k+q\ell)/2} {}_1F_1\left(\frac{k}{2}, \frac{k+q\ell}{2} + 1, -\frac{q_{1-\alpha}^\mathcal{E}}{2}\right), \quad (61)$$

$$\mathbb{P}(\mathcal{E}_\alpha^n = \emptyset) \xrightarrow{n \rightarrow \infty} \mathbb{P}(\mathcal{E}_\alpha^\infty = \emptyset) = 1 - F_{\chi^2(k)}\left(F_{\chi^2(k+\ell)}^{-1}(1 - \alpha)\right), \quad (62)$$

where  ${}_1F_1$  is the Kummer confluent hypergeometric function of the first kind. In particular,

$$\mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^n)] \xrightarrow{n \rightarrow \infty} \frac{2^{-k/2} \pi^{\ell/2}}{\Gamma(p/2 + 1)} \det(\Sigma / \Sigma^{11})^{1/2} (q_{1-\alpha}^\mathcal{E})^{p/2} {}_1F_1\left(\frac{k}{2}, \frac{p}{2} + 1, -\frac{q_{1-\alpha}^\mathcal{E}}{2}\right). \quad (63)$$

Formula (61) is empirically verified in Table 1, and Formula (62) is illustrated in Figure 2b. One may wish to understand the properties of  $\mathcal{E}_\alpha^n$  for finite sample size  $n$ . The interested reader will find in Appendix A.4 some discussions on that point, where we identify the distributions of  $\mathbf{A}_n$  and  $d_n^X$ , and approximations of those of  $\mathbf{P}_{n,0}$  and  $p_{i,n}$ . However we did not manage to identify that of  $\rho_{n,\alpha}$ .

## 4.2 The second asymptotic ellipsoid $\mathcal{F}_\alpha^\infty$

Under the assumptions of Proposition 4.1, the center  $Z_0^n$  and “matrix”  $\mathcal{A}_n$  describing the ellipsoid  $\mathcal{F}_\alpha^n$  obviously still converge to the limits given in equations (51) and (53), as they are the same as those of  $\mathcal{E}_\alpha^n$ . The limit of  $\rho'_{n,\alpha}$  is given in the following proposition.

**Proposition 4.3.** *Take the assumptions of Proposition 4.1, and set*

$$q_{1-\alpha}^{\mathcal{F}} := q_{1-\alpha}(T_c^\top \mathcal{A}_\infty^{-1} T_c), \quad \text{where} \quad T_c = R_c - \Sigma^{21}(\Sigma_\lambda^{11})^{-1} X_c. \quad (64)$$

If  $\mathbb{E}[\|X_1\|^{4q}] < +\infty$  for some  $q > 1$ , then

$$q'_{n,\alpha} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} q_{1-\alpha}^{\mathcal{F}}, \quad \rho'_{n,\alpha} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} q_{1-\alpha}^{\mathcal{F}}, \quad \text{and} \quad \text{Vol}(\mathcal{F}_\alpha^n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} v_\ell \det(\Sigma_\lambda / \Sigma_\lambda^{11})^{1/2} (q_{1-\alpha}^{\mathcal{F}})^{\ell/2}. \quad (65)$$

We denote by  $\mathcal{F}_\alpha^\infty$  the ellipsoid with center  $Z_0^\infty$ , matrix  $\mathcal{A}_\infty$  and “squared radius”  $q_{1-\alpha}^{\mathcal{F}}$ . While the convergence of  $q'_{n,\alpha}$  is weaker than that of  $q_{n,\alpha}$ , the convergence of  $\rho'_{n,\alpha}$  is as strong as that of  $\rho_{n,\alpha}$  (because the limit of  $\rho'_{n,\alpha}$  is constant, convergence in probability is equivalent to convergence in distribution).

Observe also that contrarily to Proposition 3.7, we additionally require that  $\mathbb{E}[\|X_1\|^{4q}] < +\infty$ . While this assumption is not necessarily tight, it suggests that the convergence of  $\mathcal{F}_\alpha^n$  towards  $\mathcal{F}_\alpha^\infty$  may be slower than that of  $\mathcal{E}_\alpha^n$  towards  $\mathcal{E}_\alpha^\infty$ , or that it should hold in a weaker sense. This is consistent with the numerical experiments on Cauchy data (Section 5.2).

The remark on the conditional coverage of  $\mathcal{E}_\alpha^\infty$  in the previous section also holds for  $\mathcal{F}_\alpha^\infty$  if replacing  $q_{1-\alpha}^{\mathcal{F}}$  with  $q_{1-\alpha}^{\mathcal{F}}(x) := q_{1-\alpha}(T_c^\top \mathcal{A}_\infty^{-1} T_c | X_1 = x)$ . In fact, it is equivalent to estimate  $q_{1-\alpha}^{\mathcal{F}}(x)$  or  $q_{1-\alpha}^{\mathcal{E}}(x)$ , as from equation (172),

$$q_{1-\alpha}^{\mathcal{E}}(x) = q_{1-\alpha}^{\mathcal{F}}(x) + (x - \mathbb{E}[X_1])^\top (\Sigma^{11})^{-1} (x - \mathbb{E}[X_1]).$$

In particular,  $\rho_{\infty,\alpha}(x) = q_{1-\alpha}^{\mathcal{F}}(x)$ , and the conditional ellipsoids  $\mathcal{E}_\alpha^\infty(x)$  and  $\mathcal{F}_\alpha^\infty(x)$  are the same. As opposed to  $\mathcal{E}_\alpha^\infty$  though, we recover the conditional coverage for Gaussian data, when using  $\mathcal{F}_\alpha^\infty$ .

**Proposition 4.4.** *Assume that  $V_1 \sim \mathcal{N}(m, \Sigma)$ , that  $\min \text{Spec} \Sigma > 0$ , and set  $\lambda = 0$ . Then*

$$\forall x \in \mathbb{R}^k, \quad \mathbb{P}(Y_1 - \hat{Y}_1 \in \mathcal{F}_\alpha^\infty | X_1 = x) = \mathbb{P}(Y_1 - \hat{Y}_1 \in \mathcal{F}_\alpha^\infty) = 1 - \alpha.$$

This result is an immediate consequence of the fact that  $R_1 - \Sigma^{21}(\Sigma^{11})^{-1} X_1$  is independent from  $X_1$ , when  $V_1$  is Gaussian with  $\text{Cov}(V_1) = \Sigma$ . Anticipating Section 4.3, this property is unique to Gaussian distributions, among the family of elliptical distributions (see equation (71) for a definition). In the general elliptical case, the conditional distributions  $R_1 | X_1 = x$  and  $R_1 - \Sigma^{21}(\Sigma^{11})^{-1} X_1$  still have the same dispersion matrix  $\Sigma / \Sigma^{11}$ , but they differ in their generator. Coming back to the Gaussian case, numerical evaluations of formula (63) show that if  $V_1$  is Gaussian and  $\lambda = 0$ , then on average,  $\mathcal{F}_\alpha^\infty$  is better than  $\mathcal{E}_\alpha^\infty$  in terms of volume (see Figure 2c for a plot) :

$$\forall \alpha \in (0, 1), \quad \text{Vol}(\mathcal{F}_\alpha^\infty) \leq \mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^\infty)]. \quad (66)$$

It remains possible that  $\text{Vol}(\mathcal{E}_\alpha^\infty) \leq \text{Vol}(\mathcal{F}_\alpha^\infty)$ , which corresponds to the event  $\{X_c^\top (\Sigma^{11})^{-1} X_c \geq q_{1-\alpha}^{\mathcal{E}} - q_{1-\alpha}^{\mathcal{F}}\}$ .

**The corrected predictor** An interesting feature of the ellipsoids for  $Y_{n+1}$  is that they are not centered at the predictor  $\hat{Y}_{n+1}$ . Their common center is  $\tilde{Y}_{n+1} = Z_0^n + \hat{Y}_{n+1}$  which is effectively a correction of  $\hat{Y}_{n+1}$ . In fact, there is no guaranty that  $\hat{Y}_{n+1}$  lies in  $E_\alpha^n$  or  $F_\alpha^n$ , which is desirable as the predictor may e.g. be biased. On the contrary, the next proposition shows that under same moment assumptions as for Proposition 4.1, the corrected predictor  $\tilde{Y}_{n+1}$  is asymptotically unbiased.

**Proposition 4.5.** *If  $\mathbb{E}[\|Y_1\|] < +\infty$  then under the assumptions of Proposition 4.1,  $\tilde{Y}_{n+1}$  is asymptotically unbiased, i.e.*

$$\mathbb{E}[\tilde{Y}_{n+1}] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[Y_1]. \quad (67)$$

Moreover, introducing the matrix  $M_\lambda := \Sigma_\lambda / \Sigma_\lambda^{11} - \lambda(\mathbf{I}_\ell + \Sigma^{21}(\Sigma_\lambda^{11})^{-2} \Sigma^{12})$ , and assuming there exists  $q > 1$  such that  $\mathbb{E}[\|V_1\|^{4q}] < +\infty$ , we have the asymptotic covariance and mean squared error,

$$\text{Cov}(\tilde{Y}_{n+1} - Y_{n+1}) \xrightarrow[n \rightarrow \infty]{} M_\lambda, \quad \mathbb{E}[\|\tilde{Y}_{n+1} - Y_{n+1}\|_2^2] \xrightarrow[n \rightarrow \infty]{} \text{Tr}(M_\lambda). \quad (68)$$



### 4.3 Comparison with the norm residual score

In regression, the standard score is the norm of the residual,  $S'_i(y) = \|y - \hat{Y}_{n+1}\|$ , where  $\|\cdot\|$  is the Euclidean norm in  $\mathbb{R}^\ell$ . The corresponding confidence region for  $Y_{n+1}$  is a ball  $B_\alpha^n$  centered at  $\hat{Y}_{n+1}$ , with squared radius  $\beta_{n,\alpha}$  given by the  $n_\alpha^{\text{th}}$  order statistic of  $(\|R_1\|^2, \dots, \|R_n\|^2)$ . As for  $\mathcal{E}_\alpha^n$  and  $\mathcal{B}_\alpha^n$ , we introduce  $\mathcal{B}_\alpha^n$  the centered ball with squared radius  $\beta_{n,\alpha}$ , so that  $B_\alpha^n = \{\hat{Y}_{n+1}\} + \mathcal{B}_\alpha^n$ . We begin with describing the asymptotic behaviour of  $\mathcal{B}_\alpha^n$ .

**Proposition 4.6.** *Denote  $\beta_{n,\alpha}$  the  $n_\alpha^{\text{th}}$  order statistic of  $(\|R_1\|^2, \dots, \|R_n\|^2)$ , which are assumed iid, and assume that the quantile function of  $\|R_1\|^2$  is continuous on a neighbourhood of  $1 - \alpha$ . Then*

$$\beta_{n,\alpha} \xrightarrow{a.s.} q_{1-\alpha}(\|R_1\|^2). \quad (69)$$

As a result, the asymptotic volume of the associated ball is deterministic, and given by

$$\text{Vol}(\mathcal{B}_\alpha^n) \xrightarrow{a.s.} v_\ell q_{1-\alpha}(\|R_1\|^2)^{\ell/2}. \quad (70)$$

Equation (69) is quite intuitive and natural, though we did not manage to find it stated as such in standard textbooks. We thus provide a proof in the appendix<sup>1</sup>. Similarly to  $\mathcal{E}_\alpha^n$ , we define  $\mathcal{B}_\alpha^\infty$  to be the deterministic ball centered at 0 and with radius  $q_{1-\alpha}(\|R_1\|)$ . Note that, even in the case of Gaussian residuals,  $q_{1-\alpha}(\|R_1\|^2)$  cannot be expected to be further simplified, as  $\|R_1\|^2$  would follow a generalized chi-squared distribution.

**Elliptical distributions** We can now compare the volumes  $\text{Vol}(\mathcal{E}_\alpha^\infty)$  and  $\text{Vol}(\mathcal{B}_\alpha^\infty)$ . For this, our main assumption is that the vector  $V_1$  follows an absolutely continuous elliptical distribution. This means that there exists a nonnegative function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that the density  $f$  of  $V_1$  is of the form

$$f(v) \propto g((v - \mu)^\top \Sigma^{-1}(v - \mu)), \quad (71)$$

for some  $\mu \in \mathbb{R}^p$  and  $\Sigma \in \mathcal{M}_{p,p}$ ,  $\Sigma \succ 0$  ([39], Section 1.5). Above,  $\mu$ ,  $\Sigma$  and  $g$  are respectively the location parameter, the dispersion matrix and the generator of the distribution (71) ( $\Sigma$  is defined up to a multiplicative constant, which can be absorbed in  $g$ ). Elliptical distributions include multivariate Gaussian and Cauchy distributions. Under the assumptions of Proposition 4.1,  $\Sigma = \text{Cov}(V_1)$  up to a deterministic constant ([39], p. 34), although elliptical distributions do not necessarily have finite second order moments. As argued in Appendix A.3, this constant is irrelevant to define  $\mathcal{E}_\alpha^\infty$  (in particular its volume) in a unique way, so that in the rest of this section, we can assume that  $\Sigma = \text{Cov}(V_1)$ . This ellipticity assumption is equivalent to the existence of a random vector  $T = (T_1, \dots, T_p)^\top \in \mathbb{R}^p$  with an absolutely continuous spherical distribution, i.e. with a density of the form  $f(x) = g(\|x\|^2)$ , such that  $V_1 = \mu + \Sigma^{1/2}T$ .

We first consider the case where  $k = 0$  and  $\lambda = 0$ . In this case,  $\Sigma_\lambda = \Sigma = \Sigma^{22}$ , and  $\mathcal{E}_\alpha^\infty = \mathcal{F}_\alpha^\infty$ . Furthermore, denoting  $\lambda_1, \dots, \lambda_\ell$  the eigenvalues of  $\Sigma$ , we have  $\det(\Sigma) = \lambda_1 \dots \lambda_\ell$ . Under such assumptions,  $V_c^\top \Sigma^{-1} V_c = \sum_{i=1}^\ell T_i^2$  and  $\|R_1\|^2$  is equal to  $\sum_{i=1}^\ell \lambda_i (T_i + s_i)^2$  in distribution, for some  $s = (s_1, \dots, s_\ell)^\top \in \mathbb{R}^\ell$  (see the upcoming Lemma 4.8 for a more general result). In particular,

$$\begin{aligned} \left( \frac{\text{Vol}(\mathcal{E}_\alpha^\infty)}{\text{Vol}(\mathcal{B}_\alpha^\infty)} \right)^{2/\ell} &= \det(\Sigma)^{1/\ell} \frac{q_{1-\alpha}(\sum_{i=1}^\ell T_i^2)}{q_{1-\alpha}(\sum_{i=1}^\ell \lambda_i (T_i + s_i)^2)} \\ &= \frac{q_{1-\alpha}(\sum_{i=1}^\ell T_i^2)}{(\lambda_1 \dots \lambda_\ell)^{-1/\ell} q_{1-\alpha}(\sum_{i=1}^\ell \lambda_i (T_i + s_i)^2)} = \frac{q_{1-\alpha}(\sum_{i=1}^\ell T_i^2)}{q_{1-\alpha}(\sum_{i=1}^\ell \delta_i (T_i + s_i)^2)}, \end{aligned} \quad (72)$$

where  $\delta_i = \lambda_i / (\lambda_1 \dots \lambda_\ell)^{1/\ell} > 0$  verify  $\delta_1 \dots \delta_\ell = 1$ . The comparison of  $\text{Vol}(\mathcal{E}_\alpha^\infty)$  and  $\text{Vol}(\mathcal{B}_\alpha^\infty)$  is then settled by the next proposition.

**Proposition 4.7.** *Let  $t > 0$  and  $(T_1, \dots, T_\ell)$  be a random vector with an absolutely continuous spherical distribution, i.e. with a density of the form  $f(x) = g(\|x\|^2)$ ,  $x \in \mathbb{R}^\ell$ ; assume furthermore that  $g$  is non-increasing. For a fixed  $t$ , introduce  $F_t(s, \delta) := \mathbb{P}(\sum_{i=1}^\ell \delta_i (T_i + s_i)^2 \leq t)$ , where  $s = (s_1, \dots, s_\ell)^\top \in \mathbb{R}^\ell$  and  $\delta = (\delta_1, \dots, \delta_\ell)^\top \in \mathbb{R}_+^\ell$ . Then*

$$((0, \dots, 0), (1, \dots, 1)) \in \arg \max_{\substack{(s, \delta) \in \mathbb{R}^\ell \times \mathbb{R}_+^\ell \\ \prod_{i=1}^\ell \delta_i = 1}} F_t(s, \delta). \quad (73)$$

In particular, when  $\lambda = 0$ ,  $k = 0$  and under the ellipticity assumption (71),  $\text{Vol}(\mathcal{E}_\alpha^\infty) \leq \text{Vol}(\mathcal{B}_\alpha^\infty)$ .

<sup>1</sup>This proof is a copy of that of Théorème 8.9, p. 90 of the lecture notes [4] (in French).

The property that the generator  $g$  is nonincreasing is an unimodality assumption (see [10], Definition 3.1.1). Proposition 4.7 relies on the observation that centered balls correspond to level sets of spherical distributions. Note that no moment assumptions are required for equation (73) to hold.

We now consider the case where  $k \geq 1$ , still with  $\lambda = 0$ . We start with  $\mathcal{E}_\alpha^\infty$ . Introduce a random vector  $T = (T_1, \dots, T_p)^\top$  with a spherical distribution, such that  $V_1 = \mu + \Sigma^{1/2}T$ . To compare the volumes of  $\mathcal{E}_\alpha^\infty$  and  $\mathcal{B}_\alpha^\infty$ , we will need the following lemma which describes the distribution of  $\|R_1\|^2$  in terms of that of  $T$ .

**Lemma 4.8.** *Assume that  $T \in \mathbb{R}^p$  has a spherical distribution and  $V_1 = \mu + \Sigma^{1/2}T$  for some  $\mu \in \mathbb{R}^p$  and  $\Sigma \in \mathcal{M}_{p,p}$ ,  $\Sigma \succ 0$ . Denote  $(\lambda_1, \dots, \lambda_\ell)$  the eigenvalues of  $\Sigma^{22}$ . Then, there exists  $s \in \mathbb{R}^\ell$  such that we have the equality in distribution*

$$\|R_1\|^2 \stackrel{d}{=} \sum_{i=1}^{\ell} \lambda_i (T_i + s_i)^2. \quad (74)$$

Above, the vector  $s$  is given by  $s = \mathbf{D}^{-1/2} \mathbf{P}^\top \mu$ , where  $\Sigma^{22} = \mathbf{P} \mathbf{D} \mathbf{P}^\top$  is an eigendecomposition of  $\Sigma^{22}$ . To compare  $\text{Vol}(\mathcal{E}_\alpha^\infty)$  and  $\text{Vol}(\mathcal{B}_\alpha^\infty)$ , we will mimic the case  $k = 0$  (see equation (72)). To do so, we will require the bound

$$0 \leq (q_{1-\alpha}^\mathcal{E} - X_c^\top (\Sigma^{11})^{-1} X_c)_+ \leq q_{1-\alpha}^\mathcal{E} = q_{1-\alpha} \left( \sum_{i=1}^{k+\ell} T_i^2 \right) \quad (75)$$

as well as the ratio  $c_\alpha(k, \ell)$  defined as

$$c_\alpha(k, \ell) := \frac{q_{1-\alpha} \left( \sum_{i=1}^{k+\ell} T_i^2 \right)}{q_{1-\alpha} \left( \sum_{i=1}^{\ell} T_i^2 \right)} > 1. \quad (76)$$

From these two equations as well as Proposition 4.7 and Lemma 4.8, we are able to prove Proposition 4.9 below.

**Proposition 4.9.** *Denote the dependence of  $\mathcal{E}_\alpha^\infty$  on  $\lambda$  as  $\mathcal{E}_{\alpha,\lambda}^\infty$ . Under the assumptions of Proposition 4.1, if*

$$c_\alpha(k, \ell) \left( \frac{\det(\Sigma / \Sigma^{11})}{\det(\Sigma^{22})} \right)^{1/\ell} \leq 1, \quad (77)$$

*then  $\text{Vol}(\mathcal{E}_{\alpha,0}^\infty) \leq \text{Vol}(\mathcal{B}_\alpha^\infty)$ . Moreover, if the inequality in equation (77) is strict, then there exists  $\lambda_0 > 0$  such that for all  $\lambda \in [0, \lambda_0]$ ,  $\text{Vol}(\mathcal{E}_{\alpha,\lambda}^\infty) \leq \text{Vol}(\mathcal{B}_\alpha^\infty)$ . Finally,  $\lambda_0$  can be chosen as the unique solution to the algebraic equation*

$$\det(\Sigma_{\lambda_0}) = c_\alpha(k, \ell)^{-\ell} \det(\Sigma_{\lambda_0}^{11}) \det(\Sigma^{22}). \quad (78)$$

The condition (77) can be relaxed to having the first part of equation (202) smaller than one. The algebraic equation for finding  $\lambda_0$  is correspondingly changed as

$$\det(\Sigma_{\lambda_0}) = \left( \frac{q_{1-\alpha} \left( \sum_{i=1}^{k+\ell} T_i^2 \right)}{q_{1-\alpha} \left( \sum_{i=1}^{\ell} \delta_i (T_i + s_i)^2 \right)} \right)^{-\ell} \det(\Sigma_{\lambda_0}^{11}) \det(\Sigma^{22}). \quad (79)$$

The condition (77) enables to identify two distinctive features of the data, when aiming at optimizing the volume of  $\mathcal{E}_\alpha^\infty$ : the intrinsic distribution of the  $T_i$  and the values of  $k$  and  $\ell$  on the one hand, and the structure of  $\Sigma$  on the other hand. In particular it shows that in order to minimize the volume of  $\mathcal{E}_\alpha^\infty$ , we are faced with a tradeoff when using many explanatory variables. As  $k$  increases, the determinant of the Schur complement decreases (this can be seen from the block inversion lemma), but the ratio  $c_\alpha(k, \ell)$  increases at the same time. This tradeoff is illustrated in Section 5.1 (Table 2).

*Remark* (Optimal choice of explanatory variables). In view of this tradeoff, a natural question is that of the selection of the best variables within  $X_1$  to explain the response variable  $R_1$ , especially if  $k$  is large. For example, if  $R_1$  is independent of  $(X_1)_1$ , the first coordinate of  $X_1$ , then taking it into account will not decrease the determinant of the Schur complement, as (when  $\lambda = 0$ ) it is equal to the conditional covariance  $\text{Cov}(R_1|X_1)$ . It will, however, increase the ratio  $c_\alpha(k, \ell)$ , thus increasing the volume of  $\mathcal{E}_\alpha^\infty$ . Likewise, if  $(X_1)_1 = (X_1)_2$ , then only one of those random variables is required to explain  $R_1$ . In our context, this question boils down to that of the research of active subspaces, which is e.g. studied in [11]. The use of active subspace methods in our framework is left for future work.

To deal with the case  $k \geq 1$ , we used the bound (75), which may seem very conservative. This is not entirely true, in the following sense. Setting  $S = X_c^\top (\Sigma^{11})^{-1} X_c$ , the expected volume  $\mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^\infty)]$  is given by  $\mathbb{E}[(q_{1-\alpha}^\mathcal{E} - S)_+^{\ell/2}] = (q_{1-\alpha}^\mathcal{E})^{\ell/2} \mathbb{E}[(1 - S/q_{1-\alpha}^\mathcal{E})_+^{\ell/2}]$ , up to a deterministic constant. Assuming that  $q_{1-\alpha}^\mathcal{E} \rightarrow +\infty$  when  $\alpha \rightarrow 0^+$ , we then deduce from the bound  $0 \leq (1 - S/q)_+ \leq 1$  ( $q > 0$ ) and the dominated convergence theorem that

$$\mathbb{E}[(q_{1-\alpha}^\mathcal{E} - X_c^\top (\Sigma^{11})^{-1} X_c)_+^{\ell/2}] \underset{\alpha \rightarrow 0^+}{\sim} (q_{1-\alpha}^\mathcal{E})^{\ell/2}. \quad (80)$$

In particular, the right-hand side in equation (80) corresponds to the bound one would obtain using the conservative bound (75).

*Remark* (Conditional coverage for elliptical distributions). If the distribution of  $V_1$  is elliptical, it is possible to obtain the conditional distribution  $R_1|X_1$  from the observation of  $\beta \mapsto q_\beta(V_c^\top \Sigma^{-1} V_c)$ , as we can deduce the spherical density  $g$  appearing in equation (71) from this quantile map. For all  $x$ , using the definition of the conditional density and the knowledge (or estimates) of  $g, \mu$  and  $\Sigma$ , one can derive the conditional distribution  $R|X = x$ , and  $q_{1-\alpha}(V_c^\top \Sigma^{-1} V_c|X_1 = x)$  in particular, in integral/implicit form. As remarked in Section 4.1, this conditional quantile is the required quantity to recover conditional coverage using our methodology. The use of this observation in a practical setting is left for future work.

Contrarily to  $\mathcal{E}_\alpha^\infty$ , the case  $k \geq 1, \lambda = 0$  is easily dealt with for  $\mathcal{F}_\alpha^\infty$ .

**Proposition 4.10.** *Set  $\lambda = 0$ , and assume that  $V_1$  follows an elliptical distribution with a non-increasing generator  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ . Then we have, almost surely,*

$$\text{Vol}(\mathcal{F}_\alpha^\infty) \leq \text{Vol}(\mathcal{B}_\alpha^\infty). \quad (81)$$

The key to Proposition 4.10 is the observation that  $T_c^\top \mathcal{A}_\infty^{-1} T_c \stackrel{d}{=} \sum_{i=1}^\ell T_i^2$ . This fact can be used together with Lemma 4.8 and Proposition 4.7 to obtain the desired result. As opposed to Proposition 4.9, the dependence on  $k$  of the asymptotic volume of  $\mathcal{F}_\alpha^\infty$  is milder than that of  $\mathcal{E}_\alpha^\infty$ , when  $V_1$  is elliptical. In particular, minimizing  $\text{Vol}(\mathcal{F}_\alpha^\infty)$  as a function of  $k$  is equivalent to minimizing the determinant

$$\det(\Sigma/\Sigma^{11}) = \det(\Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12}), \quad (82)$$

which can be shown to be a decreasing function of  $k$  using the block-inversion lemma. Thus, one can only hope that this quantity becomes stationary for large  $k$ , and large values of  $k$  (a form of complexity) are not penalised, contrarily to  $\mathcal{E}_\alpha^\infty$ .

**Non elliptical distributions** In this paragraph, we assume that  $k = 0$  and  $\lambda = 0$ , so that  $\Sigma_\lambda = \Sigma = \Sigma^{22} = \text{Cov}(R_1)$  and  $\mathcal{E}_\alpha^\infty = \mathcal{F}_\alpha^\infty$ . Without ellipticity assumptions, the comparison of  $\text{Vol}(\mathcal{B}_\alpha^\infty)$  and  $\text{Vol}(\mathcal{E}_\alpha^\infty)$  can still be studied for small  $\alpha$  by comparing the tail behaviour of the distributions of  $\|\Sigma^{-1/2}(R_1 - \mathbb{E}[R_1])\|^2$  and  $\|R_1\|^2 / \det(\Sigma)^{1/\ell}$ . Indeed, if

$$\mathbb{P}(\|\Sigma^{-1/2}(R_1 - \mathbb{E}[R_1])\|^2 > t) \leq \mathbb{P}(\|R_1\|^2 / \det(\Sigma)^{1/\ell} > t) \quad \text{for large } t, \quad (83)$$

then we can show that  $\text{Vol}(\mathcal{E}_\alpha^\infty) \leq \text{Vol}(\mathcal{B}_\alpha^\infty)$  for  $\alpha$  small enough (e.g. adapt the proof of Proposition 4.11, equations (221) and (222)). In this context, the precise decay of the tail distribution of nonnegative random variables is studied in [37], when the decay is either exponential or polynomial. More precisely, [37] proves Tauberian theorems, which link this decay with properties of the Laplace transform of the said distribution. The full use of the results of [37] in our context is left for future work. For the time being we simply observe that, without ellipticity assumptions, it is unreasonable to expect that a scoring rule solely based on a covariance analysis of the residuals will always yield smaller confidence regions than those of the norm residual score. In that sense, the next proposition provides a counterexample for which our method underperforms when compared to the norm residual score. This proposition relies on [37], Theorem 1.2.

**Proposition 4.11.** *Let  $\lambda_1, \dots, \lambda_\ell > 0$  be such that there exists  $i \neq j$  such that  $\lambda_i \neq \lambda_j$ , and define  $\delta_i := \lambda_i / (\lambda_1 \cdots \lambda_\ell)^{1/\ell}$ . Then there exists a random vector  $T \in \mathbb{R}^\ell$  with  $\mathbb{E}[T] = 0$  and  $\text{Cov}(T) = \mathbf{I}_\ell$  such that for large  $t$ ,*

$$\mathbb{P}\left(\sum_{i=1}^\ell T_i^2 > t\right) > \mathbb{P}\left(\sum_{i=1}^\ell \delta_i T_i^2 > t\right). \quad (84)$$

In particular, if  $R_1 = \mathbf{D}T$  where  $\mathbf{D}$  is diagonal such that  $\mathbf{D}_{ii} = \lambda_i^{1/2}$ , then  $\mathbb{E}[R_1] = 0$ ,  $\Sigma = \text{Cov}(R_1) = \mathbf{D}^2$  and for large  $t$ ,

$$\mathbb{P}(\|\Sigma^{-1/2}R_1\|^2 > t) > \mathbb{P}(\|R_1\|^2 / \det(\Sigma)^{1/\ell} > t). \quad (85)$$

Thus, if  $k = 0$  and  $\lambda = 0$ , then for  $\alpha$  small enough,  $\text{Vol}(\mathcal{B}_\alpha^\infty) < \text{Vol}(\mathcal{E}_\alpha^\infty)$ .

We see that the inequality (85) is reversed when compared to (83). The vector  $T$  above is built using independent gamma distributions :  $T_i^2 \sim \gamma(\delta_i, \delta_i^{-1})$ . The proof of Proposition 4.11 relies on the fact that, despite  $T_i$  being centered with unit variance, the tails of  $T_i^2$  do not decay at the same rate (in particular, the  $T_i$  are not identically distributed). Note that in general, such a property cannot hold if the  $T_i$  are built from a one dimensional family of distributions, as fixing the variance would determine the decay rate. This is the reason why we used general gamma distributions  $\gamma(k, \theta)$ . In the case  $k \geq 1$ , the same counterexample where  $\ell$  is replaced by  $p$  works for  $\mathcal{F}_\alpha^\infty$ . Indeed, from the independence of the  $T_i$ ,  $\Sigma$  is diagonal, hence  $Z_0^\infty = 0$  a.s.,  $\Sigma/\Sigma^{11} = \Sigma^{22}$  and the conclusion of Proposition 4.11 applies. Counter-examples in expectation with  $\mathcal{E}_\alpha^\infty$  are easily built for  $\alpha \rightarrow 0^+$  using centered spherical distributions ( $\mu = 0$ ,  $\Sigma = \mathbf{I}_p$ ). Indeed, adapting equation (201) applied to such distributions and using equation (80), we have  $\mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^\infty)]/\text{Vol}(\mathcal{B}_\alpha^\infty) \sim_{\alpha \rightarrow 0^+} c_\alpha(k, \ell)^{\ell/2} > 1$ .

## 5 Numerical application

The code for generating both data and figures is given in <https://github.com/iain-pl-henderson/PyConfCov>. The workflow is as follow. We first train the predictor on  $n_{\text{split}}$  data points  $(X_i, Y_i)$ . Next, for each of the  $n_{\text{test}}$  test points, we generate  $n_{\text{calib}}$  calibration points and perform CCLE. By averaging over those  $n_{\text{test}}$  experiments, we obtain a realisation of the empirical average volume and coverage of the resulting ellipsoids, denoted by  $\hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{E}_\alpha^{n_{\text{calib}}})]$  and  $1 - \hat{\mathbb{E}}_{n_{\text{test}}}[\alpha_{\mathcal{E}_\alpha^{n_{\text{calib}}}}]$  (replace  $\mathcal{E}_\alpha^{n_{\text{calib}}}$  with  $\mathcal{F}_\alpha^{n_{\text{calib}}}$  if using  $\mathcal{F}_\alpha^{n_{\text{calib}}}$ ). We use the same notations for the balls of the norm residual score, replacing  $\mathcal{E}$  with  $\mathcal{B}$  in those notations. We perform this complete procedure (except for the training of the predictor)  $n_{\text{histo}}$  times, to obtain a histogram of the distributions of  $\hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{E}_\alpha^{n_{\text{calib}}})]$  and  $1 - \hat{\mathbb{E}}_{n_{\text{test}}}[\alpha_{\mathcal{E}_\alpha^{n_{\text{calib}}}}]$ . We denote the empirical means of those histograms as  $\hat{\mathbb{E}}_{n_{\text{histo}}} \hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{E}_\alpha^{n_{\text{calib}}})]$ , and so forth.

### 5.1 Gaussian data

We consider the case of iid observations of Gaussian random vectors  $U_1, \dots, U_{n+1}$ ,  $U_i \in \mathbb{R}^p$ , where  $k = 6$ ,  $\ell = 3$  and  $p = k + \ell = 9$ , following a  $\mathcal{N}(0, \Sigma)$  distribution. Here, we choose  $\Sigma$  such that  $\Sigma_{i,j} = k_{3/2}(i - j)$ , where  $k_\nu$  is a Matérn covariance function of order  $\nu$ , with variance  $\sigma^2 = 1$  and lengthscale  $L = 5$  (see equation (107)). Such a Gaussian vector is a sampling of a centered one dimensional Gaussian process  $(U_t)_{t>0}$  at integer grid points  $\{1, \dots, 9\}$ , where  $(U_t)_{t>0}$  has the said 3/2-Matérn covariance function. The considered problem is thus that of the simultaneous conformal inference for the  $\ell = 3$  next time steps, given the previous  $k = 6$  time steps. In the numerical experiments, we will represent the results  $(U_7, U_8, U_9)$  in  $(x, y, z)$  coordinates.

The 3/2-Matérn Gaussian process  $(U_t)_{t>0}$  is a continuous AR(2) process, meaning that it is a solution to a second order SDE [42], Appendix B.2. As a consequence, the knowledge of  $U_{t-dt}$  and  $U_t$  determines the value of  $U_{t+dt}$  up to a noise of variance of order  $O(dt)$ . If the lengthscale  $L$  is large, then the vector  $(U_1, \dots, U_9)$  can be understood as a sampling of the Gaussian vector  $(U_t)$  with a small sampling period  $T_{\text{sample}}$  compared to the characteristic length of the process ( $T_{\text{sample}} = 1 \ll L$ ). As such, using our covariance based score, we expect that for large lengthscales  $L$ ,  $U_7$  ( $x$  coordinate) will hold low uncertainty,  $U_8$  ( $y$  coordinate) medium uncertainty and  $U_9$  ( $z$  coordinate) will hold most of the uncertainty.

The predictor  $\hat{f}$  is obtained by ridge regression trained on previous split data, that is  $\hat{Y}_i = \hat{\beta}^\top X_i$  with  $\hat{\beta} = (X_s^\top X_s + \mu_0 I)^{-1} X_s^\top Y_s$ , where  $(X_s, Y_s) \in \mathcal{M}_{n_{\text{split}}, p}$  denotes the training data, stored row-wise. Following the split conformal inference framework, the matrix  $\hat{\beta}$  will be considered deterministic in the numerical experiments. All the results correspond to  $\alpha = 0.1$ .

**Comparison with the norm residual score** The histograms in Figure 2a correspond to the Gaussian data described above. They correspond to  $n_{\text{split}} = 5000$ ,  $n_{\text{calib}} = 200$ ,  $n_{\text{test}} = 100$  and  $n_{\text{histo}} = 2000$ . We see that  $\hat{\mathbb{E}}_{n_{\text{histo}}} \hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{E}_\alpha^{n_{\text{calib}}})] = 1.54$  and  $\hat{\mathbb{E}}_{n_{\text{histo}}} \hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{F}_\alpha^{n_{\text{calib}}})] = 0.895$  while we have  $\hat{\mathbb{E}}_{n_{\text{histo}}} \hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{B}_\alpha^{n_{\text{calib}}})] = 9.35$ , meaning that the volume of the confidence regions is, on average, divided by 6 using the score  $S(z)$  and divided by 10 using the score  $S'(z)$ . We also see that our method slightly overcovers (top left and middle histograms):

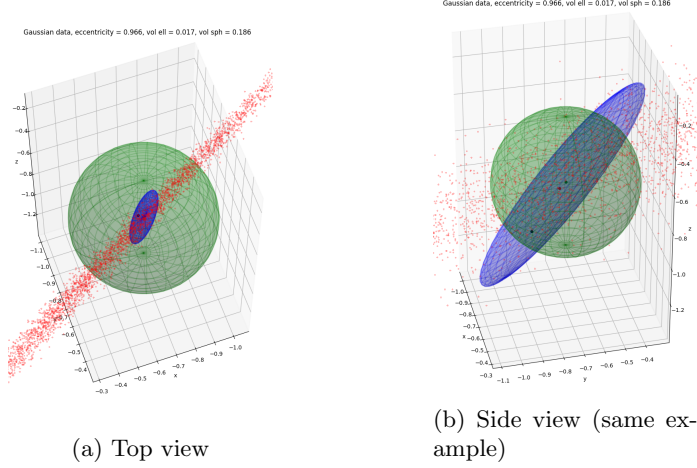


Figure 1: Sample example of  $\mathcal{E}_\alpha^n$ , Gaussian data. Black dot: real value position. Large red dot: center of ellipsoid. Green dot: predictor  $\hat{Y}_{n+1}$  (also the center of the sphere). Small red dots: other residuals ( $n = 2000$ ).

the average coverage for  $\mathcal{E}_\alpha^{n_{\text{calib}}}$  and  $\mathcal{F}_\alpha^{n_{\text{calib}}}$  is  $1 - \hat{\mathbb{E}}_{n_{\text{test}}}[\alpha_{\mathcal{E}_\alpha^{n_{\text{calib}}}}] \simeq 1 - \hat{\mathbb{E}}_{n_{\text{test}}}[\alpha_{\mathcal{F}_\alpha^{n_{\text{calib}}}}] \simeq 0.905$ , while it is  $1 - \hat{\mathbb{E}}_{n_{\text{test}}}[\alpha_{\mathcal{B}_\alpha^{n_{\text{calib}}}}] = 0.9004$  for the sphere, for a target of 0.9. This is expected from the construction of  $\mathcal{E}_\alpha^{n_{\text{calib}}}$  and  $\mathcal{F}_\alpha^{n_{\text{calib}}}$ , since  $n_{\text{calib}} = 200$  is far from the asymptotic regime (see Lemma 3.2 and Proposition 3.4). Still, concerning the volume of  $\mathcal{F}_\alpha^n$  (middle column of Figure 2a, bottom row), we can see that the convergence of  $\text{Vol}(\mathcal{F}_\alpha^n)$  to its limit  $\text{Vol}(\mathcal{F}_\alpha^\infty)$  is already quite advanced, from the width of the associated histogram.

A random realisation of  $\mathcal{E}_\alpha^n$  is given in Figure 1 (the corresponding  $\mathcal{F}_\alpha^n$  has the same appearance, the only difference with  $\mathcal{E}_\alpha^n$  being the squared radius  $\rho_{n,\alpha}$ ). This figure shows that the calibration residuals (red dots) are not distributed along the blue ellipsoid. This is also expected: the calibration residuals have a covariance  $\text{Cov}(R_1) = \Sigma^{22}$ , while the matrix associated to  $\mathcal{E}_\alpha^\infty$  is  $\text{Cov}(R_1|X_1) = \Sigma_\lambda^{22} - \Sigma^{21}(\Sigma_\lambda^{11})^{-1}\Sigma^{12}$ .

**Study of the moments  $\mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^n)^q]$**  Table 1 provides an empirical validation of formula (61). These results were obtained on a centered 3/2-Matérn Gaussian random vector with variance 1 and lengthscale  $L = 5$ . The sample sizes are  $n_{\text{split}} = 5000$ ,  $n_{\text{calib}} = 50000$ ,  $n_{\text{test}} = 40$ ,  $n_{\text{histo}} = 1000$ . The relative error corresponds to

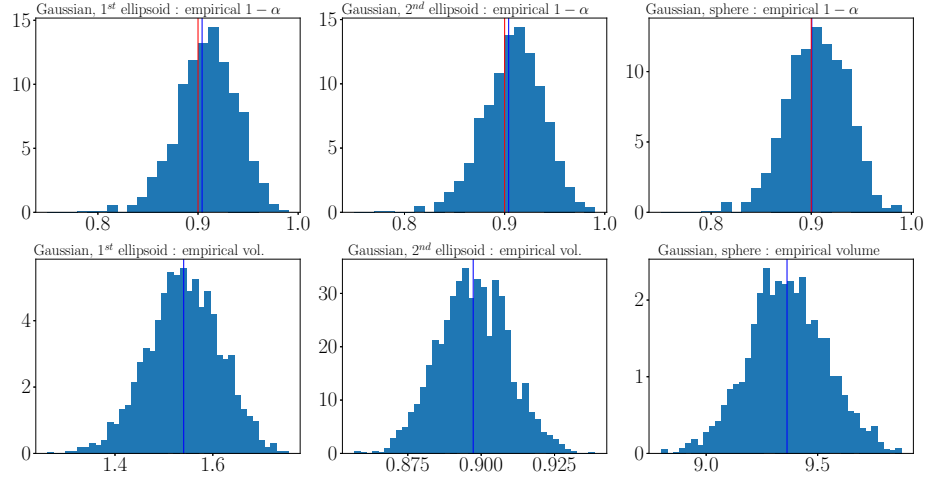
$$e_{\text{rel}} = \left| \mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^\infty)^q] - \hat{\mathbb{E}}_{n_{\text{histo}}} \hat{\mathbb{E}}_{n_{\text{test}}} [\text{Vol}(\mathcal{E}_\alpha^{n_{\text{calib}}})^q] \right| / \mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^\infty)^q]. \quad (86)$$

**Optimal number of input dimensions  $k$**  In this paragraph, we illustrate the tradeoff principle concerning  $\mathcal{E}_\alpha^n$ , as discussed following Proposition 4.9. we consider a variance  $\sigma^2 = 1$  and a lengthscale  $L = 5$ . Following [42], Section 4.2, Gaussian processes with covariance function  $k_{q-1/2}$  are AR( $q$ ) processes. As such, we expect the optimal value of  $k$  to be  $k_{\text{opt}} = q$ . Indeed, seeing the AR( $q$ ) relation as a linear recursive sequence of order  $q$ , one should at least use  $(U_6, \dots, U_{7-q})$  to predict  $U_7$  accurately, and the minimal reasonable value of  $k$  is thus equal to  $q$ . This fact is confirmed by Table 2, which provides empirical estimates of  $\mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^n)]$ . Do note that in numerical experiments, especially for  $q = 3$  and  $q = 4$ , it may happen that the optimal value identified with such empirical estimates be  $k_{\text{opt}} = q + 1$  and not  $q$ . Indeed, the corresponding values in the table below are 0.135 versus 0.142 ( $q = 3$ ) and 2.21e-02 versus 2.29e-02 ( $q = 4$ ). In fact it should be observed in this example that, after  $k_{\text{opt}}$  is reached, the average volume only slightly increases for  $k > k_{\text{opt}}$ . On the contrary, until  $k_{\text{opt}}$  is reached, the average volume is potentially much poorer.

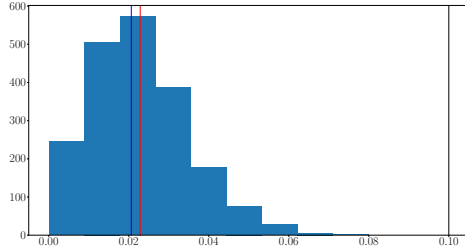
*Remark.* Even though CCLE is formulated in the standard regression framework, the presented example is essentially a time-series example. However, we do not address the typical issues encountered in conformal inference for

$q$	$\mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^\infty)^q]$	$\hat{\mathbb{E}}_{n_{\text{histo}}} \hat{\mathbb{E}}_{n_{\text{test}}} [\text{Vol}(\mathcal{E}_\alpha^{n_{\text{calib}}})^q]$	Rel. err. $e_{\text{rel}}$
1	1.4007	1.4039	2.3e-3
2	2.4183	2.4295	4.6e-3
3	4.5815	4.6096	6.1e-3

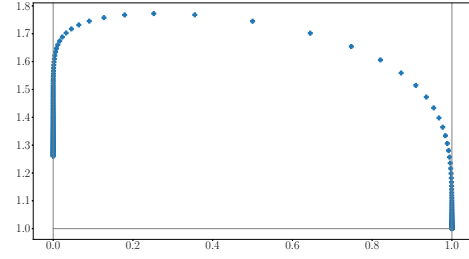
Table 1: Empirical validation of formula (61).



(a) Histograms of empirical volumes and coverage (1<sup>st</sup> row : 25 bins; 2<sup>nd</sup> row : 40 bins). Red lines :  $x = 1 - \alpha = 0.9$ . Blue lines : mean of each histogram. 1<sup>st</sup> row (left to right) : histograms of  $1 - \hat{\mathbb{E}}_{n_{\text{test}}}[\alpha_{\mathcal{E}_{\alpha}^{n_{\text{calib}}}}]$ ,  $1 - \hat{\mathbb{E}}_{n_{\text{test}}}[\alpha_{\mathcal{F}_{\alpha}^{n_{\text{calib}}}}]$ ,  $1 - \hat{\mathbb{E}}_{n_{\text{test}}}[\alpha_{\mathcal{B}_{\alpha}^{n_{\text{calib}}}}]$ . 2<sup>nd</sup> row (left to right) : histograms of  $\hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{E}_{\alpha}^{n_{\text{calib}}})]$ ,  $\hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{F}_{\alpha}^{n_{\text{calib}}})]$ ,  $\hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{B}_{\alpha}^{n_{\text{calib}}})]$ .



(b) Histogram of the proportion of empty  $\mathcal{E}_{\alpha}^{n_{\text{calib}}}$ . Blue line : mean of the histogram. red line : asymptotic value given by eq. (62). Black line :  $\alpha = 0.1$ .



(c) Plot of the ratio  $\mathbb{E}[\text{Vol}(\mathcal{E}_{\alpha}^{\infty})]/\text{Vol}(\mathcal{F}_{\alpha}^{\infty})$ , as a function of  $\alpha \in [10^{-15}, 1 - 10^{-15}]$ , for Gaussian data. This ratio is always greater than 1.

Figure 2: Numerical results, Gaussian data with  $\alpha = 0.1$ ,  $n_{\text{split}} = 5000$ ,  $n_{\text{calib}} = 200$ ,  $n_{\text{test}} = 500$  and  $n_{\text{histo}} = 1000$ .

time series, such as the balance between longitudinal and transversal coverage [31]. The analysis of our method in the time-series framework is the topic of an upcoming paper.

## 5.2 Cauchy data

Proposition 4.1 suggests that the confidence regions  $\mathcal{E}_{\alpha}^n$  may be expected to blow up in volume for heavy tailed distributions, as the limiting ellipsoid is expressed in terms of the covariance of the data, which is undefined in this case. On the contrary, since the limit sphere from the standard score  $S'_i = \|Y^i - \hat{Y}^i\|^2$  have a deterministic asymptotic volume, one could expect the standard score to beat our covariance based score in terms of volume. Here we consider a central multivariate Cauchy distribution,  $U_1 \sim C(0, \Sigma)$  [27], where  $\Sigma$  is the same as that of Section 5.1, i.e.  $\Sigma_{ij} = k_{3/2}(i - j)$ ,  $\sigma^2 = 1$ ,  $L = 5$  ( $\Sigma$  cannot be interpreted as a covariance matrix anymore). In particular,  $C(0, \Sigma)$  is an elliptical distribution. The numerical results are given in Figure 4, which corresponds to

Type \ $k$	0	1	2	3	4	5	6
$q = 1$	22.9	<b>15.2</b>	17.1	19.1	20.5	22.4	23.8
$q = 2$	4.96	1.32	<b>1.12</b>	1.23	1.33	1.43	1.53
$q = 3$	2.31	3.02e-01	1.49e-01	<b>1.35e-01</b>	1.42e-01	1.54e-01	1.63e-01
$q = 4$	1.61	1.25e-01	3.69e-02	2.37e-02	<b>2.21e-02</b>	2.29e-02	2.41e-02

Table 2: Empirical volume of the confidence ellipsoid  $\mathcal{E}_{\alpha}^n$ , for  $\alpha = 0.1$ , for Gaussian data.



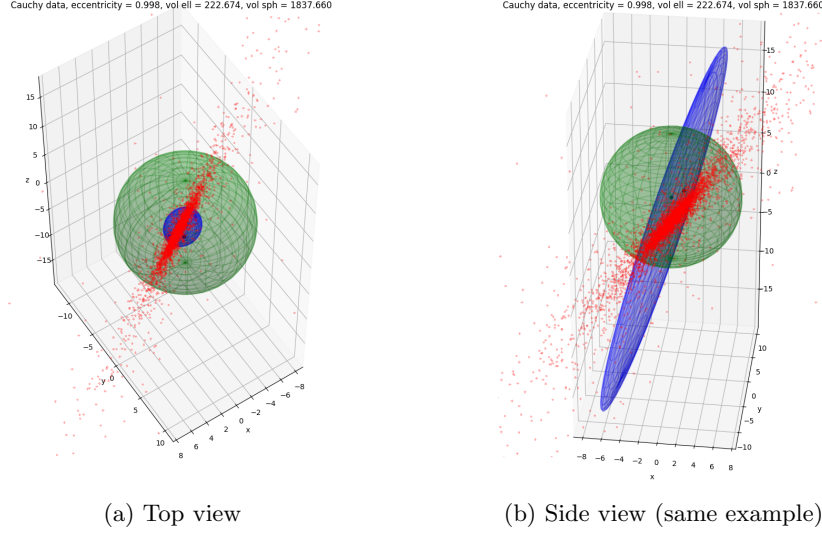


Figure 3: Sample example of  $\mathcal{E}_\alpha^n$ , Cauchy data with  $\rho = 0.8$ . Black dot: real value position. Large red dot: center of ellipsoid. Green dot: predictor  $\hat{Y}_{n+1}$  (also the center of the sphere). Small red dots: other residuals ( $n = 6000$ ).

$n_{\text{split}} = 3 \times 10^6$ ,  $n_{\text{calib}} = 10^4$ ,  $n_{\text{test}} = 800$  and  $n_{\text{histo}} = 10^3$  (we have used high values for  $n_{\text{split}}$  and  $n_{\text{calib}}$ , hoping that the corresponding numerical simulations may illustrate a form of convergence of our method for Cauchy data). For these parameters, we see that our method still exhibits an empirical average volume of  $\hat{\mathbb{E}}_{n_{\text{histo}}} \hat{\mathbb{E}}_{n_{\text{test}}} [\text{Vol}(\mathcal{E}_\alpha^{n_{\text{calib}}})] \simeq 790$ , whereas that of the residual score is  $\hat{\mathbb{E}}_{n_{\text{histo}}} \hat{\mathbb{E}}_{n_{\text{test}}} [\text{Vol}(\mathcal{B}_\alpha^{n_{\text{calib}}})] \simeq 1560$ . Interestingly, these volumes also seem to exhibit a form of asymptotic normality, suggesting that  $\mathcal{E}_\alpha^n$  is indeed robust to distributions with infinite variance.

It also seems that our method is able to recover the matrix  $\Sigma/\Sigma^{11}$  (which dictates the dispersion of  $R_1 - \Sigma^{21}(\Sigma^{21})^{-1}X_1$ ), as the shape and orientation of the ellipsoid for Cauchy data is similar to that of the ellipsoid for Gaussian data (Figures 1 and 3). We refer to Appendix A.3 for further comments on this behaviour. Of course, the volumes for the Cauchy distribution are several orders of magnitude larger than that of the Gaussian data, but this is expected as Cauchy distributions are heavy tailed.

Let us turn to  $\mathcal{F}_\alpha^n$ : because Cauchy data are heavy-tailed and Proposition 3.7 assumes a finite  $4q^{\text{th}}$  moment, we expect that arbitrarily large ellipsoids occur even in the regime of large  $n_{\text{calib}}$  (for example we obtained 9 full space ellipsoids out of  $200 \times 500 = 10^5$  tests, for  $n_{\text{calib}} = 1000$ ). For heavy-tailed distributions, histograms are not easily interpreted because of the large disparity of the observed values. Instead we present in Table 3 with empirical quantiles of the distribution of  $\hat{\mathbb{E}}_{n_{\text{test}}} [\text{Vol}(\mathcal{F}_\alpha^{n_{\text{calib}}})]$  compared to those of  $\hat{\mathbb{E}}_{n_{\text{test}}} [\text{Vol}(\mathcal{B}_\alpha^{n_{\text{calib}}})]$ , as a function of  $n_{\text{calib}}$ . This table shows that, from a quantile point of view, the use of  $\mathcal{F}_\alpha^n$  is increasingly interesting as  $n_{\text{calib}}$  increases, although empirical mean and variance still suffer from the heavy-tailed nature of Cauchy distributions. Note that  $\mathcal{E}_\alpha^n$  does not seem to suffer from this property, as shown in Figure 4 (and as could be conjectured from Proposition 3.1 and Lemma 3.5).

CI set	Quantile								Mean	Std. dev.	Nb. inf.
	$n_{\text{calib}}$	0.1	0.25	0.5	0.75	0.95	0.99	1			
$\mathcal{F}_\alpha^n$	500	<b>3.6e2</b>	<b>3.9e2</b>	<b>4.3e2</b>	2.4e4	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	36
$\mathcal{B}_\alpha^n$		9.3e3	9.5e3	9.7e3	<b>9.9e3</b>	<b>1e4</b>	<b>1e4</b>	<b>1.1e4</b>	<b>9.7e3</b>	<b>3.1e2</b>	<b>0</b>
$\mathcal{F}_\alpha^n$	$10^3$	<b>3.4e2</b>	<b>3.5e2</b>	<b>3.7e2</b>	<b>4.3e2</b>	3.3e5	$\infty$	$\infty$	$\infty$	$\infty$	9
$\mathcal{B}_\alpha^n$		8.84e3	8.94e3	8.61e3	9.07e3	<b>9.2e3</b>	<b>9.38e3</b>	<b>9.51e3</b>	<b>9.07e3</b>	<b>1.8e2</b>	<b>0</b>
$\mathcal{F}_\alpha^n$	$10^4$	<b>3.1e2</b>	<b>3.3e2</b>	<b>3.4e2</b>	<b>3.6e2</b>	<b>4e2</b>	<b>8.2e3</b>	9.2e8	1.8e6	4e7	<b>0</b>
$\mathcal{B}_\alpha^n$		8.54e3	8.58e3	8.61e3	8.65e3	8.69e3	8.73e3	<b>8.75e3</b>	<b>8.61e3</b>	<b>5.2e1</b>	<b>0</b>

Table 3: Cauchy data : Empirical quantiles of the distribution of  $\hat{\mathbb{E}}_{n_{\text{test}}} [\text{Vol}(\mathcal{F}_\alpha^{n_{\text{calib}}})]$  (empirical volume average over  $n_{\text{test}} = 200$ ), estimated from a sample of size  $n_{\text{histo}} = 500$ , as a function of  $n_{\text{calib}}$ . Last column : number of full space  $\mathcal{F}_\alpha^n$  over the  $n_{\text{test}} \times n_{\text{histo}} = 10^5$  test samples used to estimate the quantiles of  $\hat{\mathbb{E}}_{n_{\text{test}}} [\text{Vol}(\mathcal{F}_\alpha^{n_{\text{calib}}})]$ .

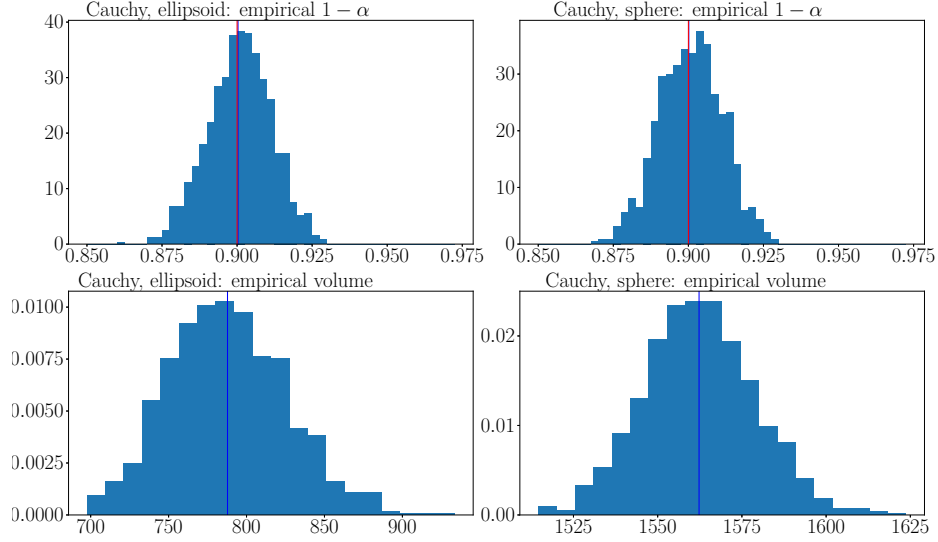


Figure 4: Empirical volumes and coverage, Cauchy data.  $n_{\text{split}} = 3 \times 10^6$ ,  $n_{\text{calib}} = 10^4$ ,  $n_{\text{test}} = 800$  and  $n_{\text{histo}} = 10^3$ . 50 bins were used. Red vertical lines correspond to  $x = 0.9$ . Blue vertical lines correspond to the mean of each histogram.

### 5.3 Non elliptical data : the inverse Dirichlet distribution

In this section we assume that the distribution of  $(X_1^\top Y_1^\top)^\top$  is supported on the positive orthant  $\mathbb{R}_+^p$ , with a density of the form

$$f(v) \propto \frac{1}{(1 + |v|_1)^{b+|a|_1}} \prod_{i=1}^p v_i^{a_i-1}, \quad v \in \mathbb{R}_+^p. \quad (87)$$

In equation (87),  $b > 0$ ,  $(a_1, \dots, a_p)$  is a vector of parameters with  $a_i > 0$ , and  $|x|_1 = \sum_{i=1}^p |x_i|$ .  $V_1$  is said to follow the inverted Dirichlet distribution which we denote by  $V_1 \sim \text{IDirichlet}(a_1, \dots, a_p; b)$  ([40], Example 8.6). Depending on its parameters, this distribution is mildly to heavily non elliptical, as is visible from Figure 5. Observe also that  $b$  controls the decay of the distribution. We next use a linear regressor trained on  $n_{\text{split}} = 5000$  examples. Practical experiments show that our method is efficient when the distribution of  $V_1$  is approximately elliptical. This can be seen in Figure 6, which corresponds to  $k = 1$ ,  $\ell = 3$ ,  $b = 3$  and  $a = (1, 1, 10^{-1}, 10^{-2})$ . Its bottom histograms show that for this distribution, the volume of  $\mathcal{F}_\alpha^n$  is in average much than that of  $\mathcal{B}_\alpha^n$ . A sample corresponding to this histogram is given in Figure 5, subfigures (a) and (b). These figures show that the set  $\mathcal{F}_\alpha^n$  was able to capture part of the directionality associated to the parameters  $b$  and  $a$  above. We next consider an unfavourable case, where  $b = 1$  and  $a = 10^{-2} \times (1, 1, 1, 1)$ . A sample example is given in Figure 5, subfigures (c) and (d). This figure shows that the calibration residuals are, for the most part, oriented along one of the three  $x, y$  or  $z$  axes. on this example, compared to the ball  $\mathcal{B}_\alpha^n$ , the ellipsoid  $\mathcal{F}_\alpha^n$  seems to be overly sensitive to the observation of  $X_1$ . As for the Cauchy data (Table 3), unfavourable test cases are best studied with a table of empirical quantiles which we provide in Table 4. This table shows that, even for  $n_{\text{calib}} = 10^4$ , CCLE is severely outperformed by the norm-residual score (at least two order of magnitudes between corresponding quantiles, even for the empirical quantile of order 0.1).

## 6 Conclusion

In this article, we have introduced two covariance based scores, for conformal inference in multivariate regression. We have shown that the resulting confidence regions are conservatively and accurately approximated with an explicit ellipsoid. We have also studied their asymptotic properties, and compared them with the regions stemming from the standard conformal score function (norm of the residual) as used in the literature. Our first set of numerical experiments shows that our scores typically perform better than that of the standard norm score, in the sense that the corresponding volume is typically much lower than that of the spheres associated with the residual norm score, provided that the residuals  $R_i = Y_i - \hat{Y}_i$  are not uncorrelated with the input  $X_i$ . The next step will be to apply our methodology on time series, where the adaptivity of our score function may be expected to improve the performance

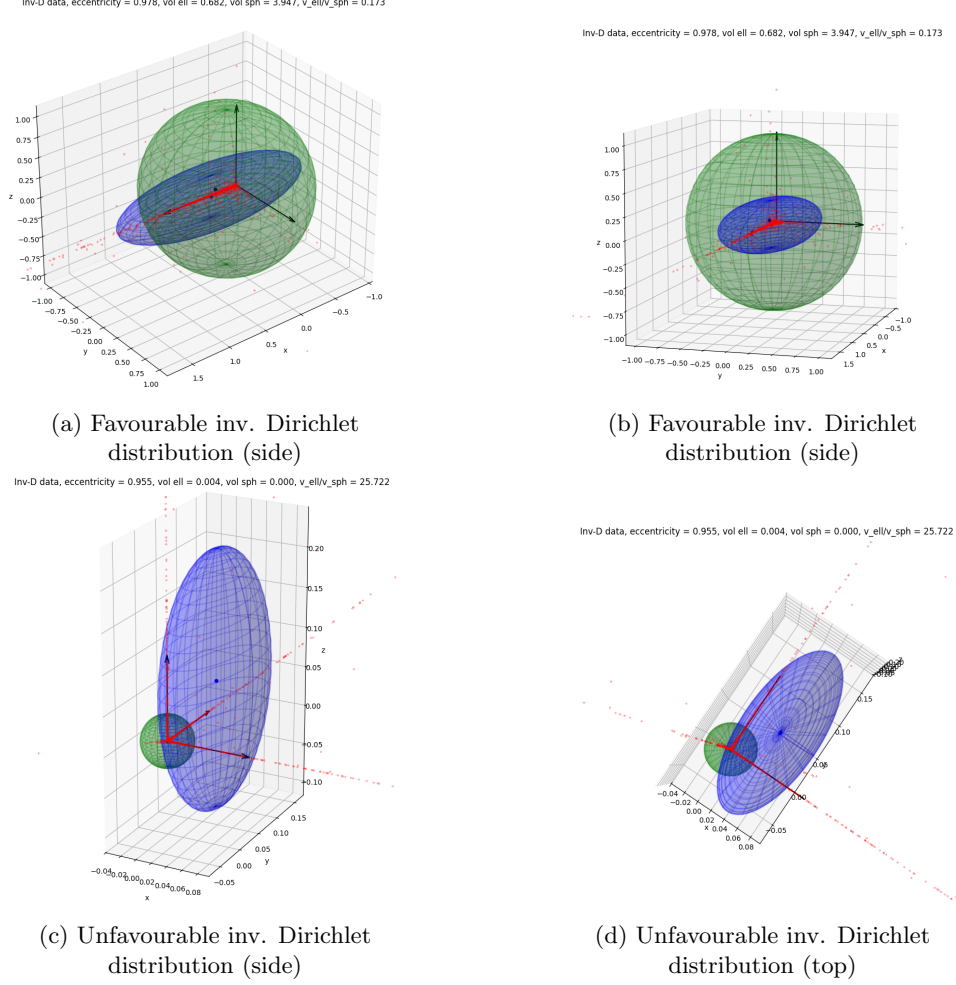


Figure 5: Two sample examples, inverse Dirichlet data. Black dot: real value position. Blue dot: center of ellipsoid. Green dot: predictor  $\hat{Y}_{n+1}$  (also the center of the sphere). Small red dots: other residuals (6000). Subfigures (a) and (b) : side views of an example simulated with a favourable inv. Dirichlet distribution (parameters  $b = 3$ ,  $a = (1, 1, 10^{-1}, 10^{-2})$ ). Subfigures (c) and (d) : side and top views of an example simulated with an unfavourable inv. Dirichlet distribution (parameters  $b = 1$ ,  $a = 10^{-2} \times (1, 1, 1, 1)$ ).

CI set	Quantile	0.1	0.25	0.5	0.75	0.95	0.99	1	Mean	Std. dev.	Nb. inf.
$\mathcal{F}_{\alpha}^n$	$n_{\text{calib}}$										
$\mathcal{B}_{\alpha}^n$	500	8.2e-3	1.1e-2	1.8e-2	4.2e-2	4.2	4.7e4	$\infty$	$\infty$	$\infty$	3
$\mathcal{F}_{\alpha}^n$	10 <sup>3</sup>	<b>5.4e-4</b>	<b>5.8e-4</b>	<b>6.3e-4</b>	<b>6.9e-4</b>	<b>7.7e-4</b>	<b>8.6e-4</b>	<b>9.2e-4</b>	<b>6.3e-4</b>	<b>7.7e-5</b>	<b>0</b>
$\mathcal{B}_{\alpha}^n$	10 <sup>3</sup>	1.2e-2	1.5e-2	2.3e-2	4.8e-2	8.2e-1	6.7e2	$\infty$	$\infty$	$\infty$	1
$\mathcal{F}_{\alpha}^n$	10 <sup>4</sup>	<b>2.3e-4</b>	<b>2.4e-4</b>	<b>2.6e-4</b>	<b>2.7e-4</b>	<b>2.9e-4</b>	<b>3.1e-4</b>	<b>3.5e-4</b>	<b>2.6e-4</b>	<b>2.3e-5</b>	<b>0</b>
$\mathcal{B}_{\alpha}^n$	10 <sup>4</sup>	4e-2	4.9e-2	6.8e-2	1.1e-1	4.9e-1	2.1	1e3	2.3	45.7	0
$\mathcal{F}_{\alpha}^n$		<b>3.3e-4</b>	<b>3.3e-4</b>	<b>3.4e-4</b>	<b>3.4e-4</b>	<b>3.5e-4</b>	<b>3.5e-4</b>	<b>3.6e-4</b>	<b>3.4e-4</b>	<b>7.1e-6</b>	<b>0</b>

Table 4: Inverse Dirichlet data with parameters  $b = 1$  and  $a = 10^{-2} \times (1, 1, 1, 1)$ : empirical quantiles of the distribution of  $\hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{F}_{\alpha}^{n_{\text{calib}}})]$  (empirical volume average over  $n_{\text{test}} = 200$ ), estimated from a sample of size  $n_{\text{histo}} = 500$ , as a function of  $n_{\text{calib}}$ . Last column : number of full space  $\mathcal{F}_{\alpha}^n$  over the  $n_{\text{test}} \times n_{\text{histo}} = 10^5$  test samples used to estimate the quantiles of  $\hat{\mathbb{E}}_{n_{\text{test}}}[\text{Vol}(\mathcal{F}_{\alpha}^{n_{\text{calib}}})]$ .

of CI for time series, especially concerning the longitudinal coverage (coverage within a single sample of the time series); this study will be accompanied by a Python package. We will soon extend our methodology to functional data [19]. Finally, a limitation of our scoring rule is that it may not be expected to be optimal for multimodal or non elliptical distributions of the residuals: the shape of  $\tilde{\mathcal{C}}_{\alpha}^n$  is bound to be an ellipsoid. As such, generalizations of our score may be sought to include moments of order 3, to take asymmetry into account, although the matrix

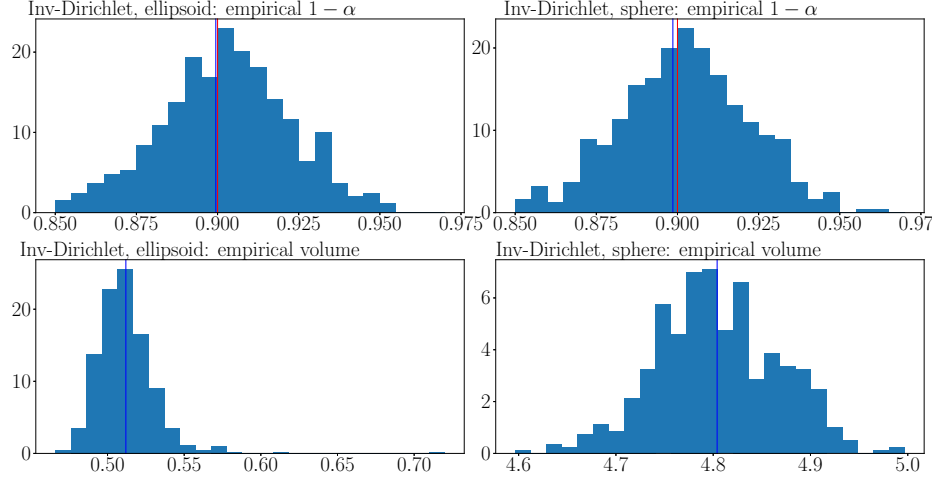


Figure 6: Empirical volumes and coverage, Inverse Dirichlet data with  $b = 3, a = (1, 1, 10^{-1}, 5.10^{-2})$ .  $n_{\text{split}} = 5000$ ,  $n_{\text{calib}} = 1000$ ,  $n_{\text{test}} = 200$  and  $n_{\text{histo}} = 500$ . 25 bins were used. Red vertical lines correspond to  $x = 0.9$ . Blue vertical lines correspond to the mean of each histogram.

algebra may become intractable.

## A Additional remarks

### A.1 Interpretations of the score vector $\text{diag}(\mathbf{S}(z))$

**Leverage score** Consider a multivariate linear model between the matrices  $\mathbf{R}$  and  $\mathbf{X}$ ,  $\mathbf{R} = \mathbf{X}\beta + \varepsilon$ . The ridge estimator in this linear regression model is  $\hat{\beta} = (\mathbf{X}^\top \mathbf{X} + \mu \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{R}$ . Let us denote  $\hat{\mathbf{R}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^\top \mathbf{X} + \mu \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{R} = \mathbf{M}\mathbf{R}$ . The leverage score of the  $i^{\text{th}}$  example is the diagonal coefficient  $M_{ii} \in [0, 1]$ . As  $M = \nabla_R \hat{R}$ , the leverage score can be interpreted as a self-influence score. A high leverage score  $M_{ii}$  means that the  $i^{\text{th}}$  example  $(X_i, R_i)$  is very influential in the behaviour of the linear model [8, 25]. The leverage score has been used, for example, for outliers detection [8]. This last task can also be tackled using conformal inference [5]. Note further that the leverage score does not depend on any underlying model nor on  $Y$ . Hence, no observational model is required for computing the leverage score. Our score is thus a form of leverage score, where we replaced the matrix  $\mathbf{M}$  by  $\mathbf{M}' = \mathbf{W}(\mathbf{W}^\top \mathbf{W} + \mu \mathbf{I}_p)^{-1} \mathbf{W}^\top$ . Hence, we jointly consider both the explanatory variables and the residual errors in the leverage score. Further, there is another difference with the classical leverage score in the centering step, as we consider  $\mathbf{W}$  instead of  $\mathbf{V}$ .

**Shape theory** In shape theory [15], iid random vectors are observed and stacked, say row-wise, in a matrix  $\mathbf{V}$ . In shape theory, one then typically studies the rectangular matrix  $\mathbf{V}$  according to its polar decomposition,  $\mathbf{V} = \mathbf{P}\mathbf{T}$ , where  $\mathbf{T}$  is the *modulus* and  $\mathbf{P}$  is the *orientation* ( $\mathbf{T} = (\mathbf{V}^\top \mathbf{V})^{1/2}$  and  $\mathbf{P} = \mathbf{V}(\mathbf{V}^\top \mathbf{V})^{-1/2}$ ). The matrices  $\mathbf{P}$  and  $\mathbf{T}$  are then studied separately ([14], [9], Section 1.3.3). In particular, the random orthogonal projector  $\mathbf{P}\mathbf{P}^\top$  is an orientation statistic [9], Section 1.3.3, encoding directional information of the data. In our case, when  $\lambda = 0$ , the connection between our score (12) and the orientation statistic  $\mathbf{P}\mathbf{P}^\top$  is the equality  $\mathbf{V}(\mathbf{V}^\top \mathbf{V})^{-1} \mathbf{V}^\top = \mathbf{P}\mathbf{P}^\top$ . As for the leverage score though, the data is usually not centered in shape theory: our score in equation (12) is, when  $\lambda = 0$ ,  $n^{-1} \mathbf{S} = \mathbf{W}(\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \neq \mathbf{V}(\mathbf{V}^\top \mathbf{V})^{-1} \mathbf{V}^\top$ . In our framework the centering step is useful to deal with potentially biased predictors  $\hat{f}$ , i.e. non centered residuals.

### A.2 Different assumptions for Proposition 3.4

**The case  $\lambda = 0$**  Our proof of Proposition 3.4 uses the fact that  $\lambda > 0$  (equation (137)): if  $\lambda = 0$ , we should require that  $\min \text{Spec}(\mathbf{\Sigma}) > 0$ . However, even if we can show that  $\min \text{Spec}(\hat{\mathbf{\Sigma}}_n)$  tends to  $\min \text{Spec}(\mathbf{\Sigma})$  almost surely (proof of Proposition 4.1), adapting the proof of Proposition 3.4 to  $\lambda = 0$  would require the control of the moments of  $\min \text{Spec}(\hat{\mathbf{\Sigma}}_n)^{-1}$ , which is a difficult task in itself. We leave the details of the corresponding proof to

the interested reader, e.g. using that almost surely,  $\min \text{Spec}(\widehat{\Sigma}_{n,0}) \rightarrow \min \text{Spec}(\Sigma)$ , from the SLLN on  $\widehat{\Sigma}_{n,0}$  and the continuity of the spectrum for symmetric matrices (see e.g. [20, Problem 1, p. 198]).

**Finite moment of  $V_1$**  The assumption that  $\mathbb{E}[\|V_1\|^{4q}] < +\infty$  (which is also used in several other places) enables us to use Jensen's inequality, to show that  $\sup_{z \in K} \max_{1 \leq i \leq n} n b_i^n(z)^2 \rightarrow 0$  in probability. Under the weaker assumption that  $\mathbb{E}[\|V_1\|^2] < +\infty$ , we can still show that the first  $n$  coordinates of  $b^n$  verify  $n \sum_{i=1}^n b_i^n(z)^2 = (1 + \|z\|^2) O_{\mathbb{P}}(1)$ , where the  $O_{\mathbb{P}}(1)$  term is uniform in  $z$  (e.g. adapt equation (164)). Unfortunately, this result alone is not strong enough to imply equation (35).

### A.3 Infinite variance case

From equation (24),  $\mathcal{E}_\alpha^\infty$  displays a scaling invariance, which we expect to have important implications for the analysis of our method in the case of heavy tailed data [46], Condition 1.1. For the moment, let us view  $\Sigma_\lambda$  and  $V_1$  as independent parameters. Observing that  $\mathcal{A}_\infty$ ,  $Z_0^\infty$  and  $\rho_{\infty,\alpha}$  are functions of  $\Sigma_\lambda$ , we denote  $\mathcal{E}_\alpha^\infty = \mathcal{E}_\alpha(\Sigma_\lambda, V_1)$ . We now prove that, as sets,  $\mathcal{E}_\alpha(\Sigma_\lambda, V_1) = \mathcal{E}_\alpha(\delta \Sigma_\lambda, V_1)$  for all  $\delta > 0$ . Indeed, assume that  $\Sigma_\lambda$  is replaced with  $\delta \Sigma_\lambda$  for some  $\delta > 0$ , then from equation (53)  $Z_0^\infty$  is left unchanged. Likewise,  $\mathcal{A}_\infty$  is changed to  $\delta \mathcal{A}_\infty$ , while

$$\begin{aligned} \frac{\rho_{\alpha,\infty}}{\delta} &= \frac{1}{\delta} q_{1-\alpha}(V_c^\top \Sigma_\lambda^{-1} V_c) - \frac{1}{\delta} X_c^\top (\Sigma_\lambda^{11})^{-1} X_c \\ &= q_{1-\alpha}(V_c^\top (\delta \Sigma_\lambda)^{-1} V_c) - X_c^\top (\delta \Sigma_\lambda^{11})^{-1} X_c, \end{aligned} \quad (88)$$

so that the product  $\rho_{\alpha,\infty} \mathcal{A}_\infty = (\delta^{-1} \rho_{\alpha,\infty})(\delta \mathcal{A}_\infty)$  is also unchanged when  $\Sigma_\lambda$  is replaced with  $\delta \Sigma_\lambda$ . Hence, following equation (57),  $\mathcal{E}_\alpha(\Sigma_\lambda, V_1) = \mathcal{E}_\alpha(\delta \Sigma_\lambda, V_1)$ . This scaling invariance is, in fact, a key property for building the Tyler dispersion matrix estimator [46] (see equation (71) and its comments for the dispersion matrix), which is known to be robust to infinite variance in the case of elliptical distributions. The same scaling arguments can also be made for  $\mathcal{F}_\alpha^\infty$ , but the convergence of  $\rho'_{n,\alpha}$  to  $q_{1-\alpha}^F$  is not at all clear (see e.g. the assumptions of Proposition 4.3), as can be seen from the numerical experiments with Cauchy data. The further study of our method for heavy-tailed data is left for future work.

### A.4 Non asymptotic analysis in the Gaussian case

One may wish to understand the properties of  $\mathcal{E}_\alpha^n$  for finite sample size  $n$ . For example, assuming that  $\lambda = 0$ , it is known that  $\mathcal{A}_n = \widehat{\Sigma}_n / \widehat{\Sigma}_n^{11}$  follows the Wishart distribution  $W_\ell(n-k-1, n^{-1} \Sigma / \Sigma^{11})$  [39], Theorem 3.2.10. We now turn to the quantities appearing in  $\rho_{n,\alpha}$ . Observing that  $\sqrt{n/(n+1)} X_{n+1}^c \sim \mathcal{N}(0, \Sigma^{11})$ , that  $n \widehat{\Sigma}_n^{11} \sim W_k(n-1, \Sigma^{11})$  and that both are independent [34], Corollary 3.3.3.2, we obtain that  $[(n-1)/(n+1)](X_{n+1}^c)^\top (\widehat{\Sigma}_n^{11})^{-1} X_{n+1}^c$  follows the Hotelling  $T^2$  distribution  $T^2(k, n-1)$  [34], Theorem 3.5.1. Let us now also assume that  $\mathbb{E}[V_1]$  is known. In this case, it is reasonable to replace  $n^{-1} \sum_{i=1}^n V_i$  with  $\mathbb{E}[V_1]$  in  $\mathbf{B}_n$  (equation (21)), so that  $\mathbf{B}_n$  follows the matrix normal distribution  $\mathbf{B}_n \sim \mathcal{N}(0, \mathbf{I}_n, \Sigma)$  [9], Section 1.5.3. In this case,  $\mathbf{P}_{n,0}$  is a random orthogonal projector whose distribution is uniform over the so-called “special manifold”  $\mathbf{P}_{p,n-p} \subset \mathcal{M}_{n,n}$  of orthogonal projectors of rank  $p$ , which is isomorphic to the Grassmanian manifold  $G_{p,n}$  [9], Theorem 2.4.9. Finally, building on [39], Exercise 3.15 p. 117, we can show that  $(\mathbf{P}_{n,0})_{ii}$  follows a Beta distribution  $B(p/2, (n-p)/2)$ . Note though that those results, while instructive, are not sufficient to describe the distribution of  $\rho_{n,\alpha}$ .

When  $\mathbb{E}[V_1]$  is unknown, we expect to obtain perturbations of the distributions above for  $\mathbf{P}_{n,0}$  and  $(\mathbf{P}_{n,0})_{ii}$ . Still, to our knowledge, there are no simple expressions available in this case. Indeed, the procedure for building  $\mathbf{B}_n$  correlates the columns of  $\mathbf{B}_n$  through equation (94), and  $\mathbf{B}_n \sim \mathcal{N}(0, \pi_{\mathbb{1}}^\perp, \Sigma)$  (here,  $\mathbb{1} \in \mathbb{R}^n$ ). As such, the distribution of  $\mathbf{P}_{n,0}$  does not seem to be available in closed form since neither  $\pi_{\mathbb{1}}^\perp$  nor  $\Sigma$  are equal to the identity matrix [9], Chapter 2.

## B Technical lemmas

The following lemma will be very useful in several proofs.

**Lemma B.1** (Block matrix inversion, [20], equation (7.7.5) p. 472). *If  $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{C} \end{pmatrix}$  is invertible and  $\mathbf{A}$  is invertible, then  $\mathbf{M}/\mathbf{A} = \mathbf{C} - \mathbf{B}^\top \mathbf{A}^{-1} \mathbf{B}$  is invertible and*

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1} \mathbf{B} (\mathbf{M}/\mathbf{A})^{-1} \mathbf{B}^\top \mathbf{A}^{-1} & -\mathbf{A}^{-1} \mathbf{B} (\mathbf{M}/\mathbf{A})^{-1} \\ -(\mathbf{M}/\mathbf{A})^{-1} \mathbf{B}^\top \mathbf{A}^{-1} & (\mathbf{M}/\mathbf{A})^{-1} \end{pmatrix}. \quad (89)$$

A central consequence of this lemma is that if  $\mathbf{M}$  is partitioned as above and  $(x^\top y^\top)^\top$  is a similarly partitioned vector, then setting  $t = y - \mathbf{A}^{-1}\mathbf{B}x$ ,

$$\begin{pmatrix} x \\ y \end{pmatrix}^\top \mathbf{M}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = x^\top \mathbf{A}^{-1}x + t^\top (\mathbf{M}/\mathbf{A})^{-1}t. \quad (90)$$

Next, the efficient computation of the score matrix is enabled by Lemma B.2 below, which provides an “explicit” representation of equation 12, in the case where  $W(z)$  is a rank one perturbation of a reference matrix.

**Lemma B.2.** *If  $\mu \geq 0$ ,  $\mathbf{M} \in \mathcal{M}_{n,p}$  is of the form  $\mathbf{M} = \mathbf{A} + wu^\top$  where  $w \in \mathcal{M}_{n,1}$ ,  $\|w\| = 1$  and  $u \in \mathcal{M}_{p,1}$ , then there exists  $\mathbf{C} \in \mathcal{M}_{n,n}$ ,  $b \in \mathcal{M}_{n,1}$  and  $d \in \mathbb{R}^+$  such that*

$$\mathbf{M}(\mathbf{M}^\top \mathbf{M} + \mu \mathbf{I})^{-1} \mathbf{M}^\top = \mathbf{C} - \frac{bb^\top}{1+d}. \quad (91)$$

$\mathbf{C}$ ,  $b$  and  $d$  are given by

$$\mathbf{C} = \mathbf{B}\mathbf{D}_\mu^{-1}\mathbf{B}^\top + ww^\top, \quad b = \mathbf{B}\mathbf{D}_\mu^{-1}r - w, \quad d = r^\top \mathbf{D}_\mu^{-1}r, \quad (92)$$

where

$$\mathbf{B} = (I - ww^\top)\mathbf{A} = \pi_w^\perp \mathbf{A}, \quad \mathbf{D}_\mu = \mathbf{B}^\top \mathbf{B} + \mu \mathbf{I}, \quad r = u + \mathbf{A}^\top w. \quad (93)$$

Above,  $\mathbf{B}_\mu^+ := \mathbf{B}\mathbf{D}_\mu^{-1}$  is a regularized pseudo-inverse of  $\mathbf{B}^\top$ : when  $\mu = 0$ ,  $\mathbf{B}_\mu^+ \mathbf{B}^\top$  is the orthogonal projector onto the range of  $\mathbf{B}$ . Note also that  $\mathbf{B}^\top w = \mathbf{A}^\top (I - ww^\top)w = 0$ , hence  $\mathbf{C} = \mathbf{B}_\mu^+ \mathbf{B}^\top + ww^\top$  is also a ( $\mu$ -regularized) orthogonal projector. Due to the fact that the score matrix is built upon centered data, the quantities in equation (91) can be further simplified, as stated in the following lemma.

**Lemma B.3.** *Assume that the matrix  $\mathbf{A}$  in Lemma B.2 lies in  $\mathcal{M}_{n+1,p}$ , that it is of the form  $\mathbf{A} = \pi_1^\perp \mathbf{V}$  for some matrix  $V \in \mathcal{M}_{n+1,p}$ , and that  $w$  in Lemma B.2 is  $w = v/\|v\|$ , where  $v$  is given in equation (19). Then  $\mathbf{B}$  is given by*

$$\mathbf{B}_{ij} = \mathbf{V}_{ij} - \frac{1}{n} \sum_{l=1}^n \mathbf{V}_{il}, \quad e_{n+1}^\top \mathbf{B} = 0 \quad (\text{null last row}). \quad (94)$$

The last equation implies that for  $\mathbf{C}$  and  $b$  given in Lemma 3.1,

$$(bb^\top)_{n+1,n+1} = b_{n+1}^2 = \mathbf{C}_{n+1,n+1} = w_{n+1}^2 = n/(n+1). \quad (95)$$

The key property for this lemma to hold is that  $\mathbb{1}^\top v = 0$ . Applying this lemma to our score, we can further describe the elements of the score matrix  $\mathbf{S}(z)$ .

**Lemma B.4.** *In the expression of the score matrix*

$$\mathbf{S}(z) = n\mathbf{C}_n - n \frac{b^n(z)b^n(z)^\top}{1+d_n(z)}, \quad (96)$$

the matrix  $\mathbf{C}_n$ , the vector  $b^n(z)$  and the scalar  $d_n(z)$  are given by

$$\mathbf{C}_n = \begin{pmatrix} \mathbf{P}_{n,\lambda} & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 0 \end{pmatrix} + ww^\top, \quad (97)$$

$$d_n(z) = \frac{1}{n} r_n(z)^\top \widehat{\Sigma}_{n,\lambda}^{-1} r_n(z), \quad b^n(z) = \frac{1}{n} \begin{pmatrix} \mathbf{B}_n \widehat{\Sigma}_{n,\lambda}^{-1} r_n(z) \\ 0 \end{pmatrix} - w. \quad (98)$$

Here,  $r_n(z)$  is given by

$$r_n(z) = \|v\| \left[ \begin{pmatrix} X_{n+1} \\ z \end{pmatrix} - \frac{1}{n} \sum_{i=1}^n V_i \right] \in \mathcal{M}_{p,1}. \quad (99)$$

Above,  $r_n(z)$  corresponds to  $r$  in Lemma B.2. Do note that above,  $v$  and  $w$  also depend on  $n$ . We can use the same arguments (lemmas B.2 and B.3) to obtain a similar expression for the matrix  $\mathbf{S}_X$  defined in equation (42).



**Lemma B.5.** We can write the matrix  $\mathbf{S}_X$  as

$$\mathbf{S}_X = n\mathbf{C}_n^X - n \frac{b_X^n (b_X^n)^\top}{1 + d_n^X}, \quad (100)$$

where the vector  $b_X^n$  and the scalar  $d_n^X$  are given by

$$d_n^X = \frac{1}{n+1} (X_{n+1}^c)^\top (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c, \quad b_X^n = \frac{\|v\|}{n} \begin{pmatrix} \mathbf{B}_n^X (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c \\ 0 \end{pmatrix} - w. \quad (101)$$

In  $b_X^n$ ,  $\mathbf{B}_n^X$  is the left  $n \times k$  block of  $\mathbf{B}_n$  and  $\mathbf{C}_n^X$  is given by

$$\mathbf{C}_n^X = \begin{pmatrix} \mathbf{P}_{n,\lambda}^{XX} & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 0 \end{pmatrix} + ww^\top, \quad \text{where} \quad \mathbf{P}_{n,\lambda}^{XX} = \mathbf{B}_n^X \left( (\mathbf{B}_n^X)^\top \mathbf{B}_n^X + n\lambda \mathbf{I}_k \right)^{-1} (\mathbf{B}_n^X)^\top.$$

Above, be wary that  $\mathbf{P}_{n,\lambda}^{XX}$  is not a sub-block of  $\mathbf{P}_{n,\lambda}$ . Our last lemma is pivotal to Section 4.

**Lemma B.6.** Take the assumptions of Proposition 4.1. Then we have that almost surely,

$$\forall t \in \mathbb{R}^p, \quad \frac{1}{n} \sum_{j=1}^n \exp \left( it^\top \widehat{\Sigma}_{n,\lambda}^{-1/2} (V_i - \bar{V}_n) \right) \xrightarrow[n \rightarrow \infty]{} \mathbb{E} \left[ \exp \left( it^\top \Sigma_\lambda^{-1/2} V_c \right) \right], \quad (102)$$

where  $\bar{V}_n = n^{-1} \sum_{i=1}^n V_i$  and  $V_c = V_1 - \mathbb{E}[V_1]$ . In particular, under the assumptions of Proposition 4.1,

$$q_{n,\alpha} \xrightarrow[n \rightarrow \infty]{a.s.} q_{1-\alpha}^\mathcal{E}. \quad (103)$$

Next, for all  $i \in \{1, \dots, n\}$ , set  $p_{i,n}'' := (\mathbf{P}_{n,\lambda} - \mathbf{P}_{n,\lambda}^{XX})_{ii}$  and define  $q_{n,\alpha}''$  as the order statistic of order  $n_\alpha$  of the  $n$ -tuple  $(np_{1,n}'', \dots, np_{n,n}'')$ . Then under the assumptions of Proposition 4.1,

$$q_{n,\alpha}'' \xrightarrow[n \rightarrow \infty]{a.s.} q_{1-\alpha}^\mathcal{F}. \quad (104)$$

Finally, if additionally  $\mathbb{E}[\|X_1\|^{4q}] < +\infty$  for some  $q > 1$  and taking the definition of  $q'_{n,\alpha}$  in equation (44), then

$$q'_{n,\alpha} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} q_{1-\alpha}^\mathcal{F}. \quad (105)$$

The main difficulty of this lemma is the fact that  $q_{n,\alpha}$  (resp.  $q'_{n,\alpha}$ ) is an order statistic built from  $(np_{1,n}, \dots, np_{n,n})$  (resp.  $(np'_{1,n}, \dots, np'_{n,n})$ ), which are identically distributed but not independent.

**Matérn covariance functions** We conclude this section with the expressions of the Matérn covariance functions that are used in Section 5. Setting  $H = |h|\sqrt{2\nu}/L$ ,  $k_\nu$  is given by [42], Section 4.2 p 85,

$$k_{\nu=1/2}(h) = \sigma^2 \exp(-H), \quad (106)$$

$$k_{\nu=3/2}(h) = \sigma^2 (1 + H) \exp(-H), \quad (107)$$

$$k_{\nu=5/2}(h) = \sigma^2 (1 + H + H^2/3) \exp(-H), \quad (108)$$

$$k_{\nu=7/2}(h) = \sigma^2 (1 + H + 2H^2/5 + H^3/15) \exp(-H). \quad (109)$$

## C Proofs

*Proof of Lemma B.2.* We have

$$\begin{aligned} \mathbf{M}^\top \mathbf{M} &= \mathbf{A}^\top \mathbf{A} + (uw^\top \mathbf{A} + \mathbf{A}^\top wu^\top) + uu^\top = \mathbf{A}^\top \mathbf{A} + (u + \mathbf{A}^\top w)(u + \mathbf{A}^\top w)^\top - \mathbf{A}^\top ww^\top \mathbf{A} \\ &= (\mathbf{A}^\top \mathbf{A} - \mathbf{A}^\top ww^\top \mathbf{A}) + rr^\top = \mathbf{B}^\top \mathbf{B} + rr^\top. \end{aligned} \quad (110)$$

With  $\mathbf{D}_\mu := \mathbf{B}^\top \mathbf{B} + \mu \mathbf{I}$ , the Sherman-Morrison formula yields

$$(\mathbf{M}^\top \mathbf{M} + \mu \mathbf{I})^{-1} = (\mathbf{B}^\top \mathbf{B} + \mu \mathbf{I} + rr^\top)^{-1} = \mathbf{D}_\mu^{-1} - \frac{\mathbf{D}_\mu^{-1} rr^\top \mathbf{D}_\mu^{-1}}{1 + r^\top \mathbf{D}_\mu^{-1} r} = \mathbf{D}_\mu^{-1} - \frac{\mathbf{D}_\mu^{-1} rr^\top \mathbf{D}_\mu^{-1}}{1 + d}. \quad (111)$$

Moreover,

$$\mathbf{M} = \mathbf{A} + wu^\top = (I - ww^\top) \mathbf{A} + ww^\top \mathbf{A} + wu^\top = \mathbf{B} + w(\mathbf{A}^\top w + u)^\top = \mathbf{B} + wr^\top. \quad (112)$$

Next, denoting  $\mathbf{B}_\mu^+ := \mathbf{B} \mathbf{D}_\mu^{-1}$ ,

$$\begin{aligned} \mathbf{M}(\mathbf{M}^\top \mathbf{M} + \mu \mathbf{I})^{-1} \mathbf{M}^\top &= \mathbf{M} \left( \mathbf{D}_\mu^{-1} - \frac{\mathbf{D}_\mu^{-1} rr^\top \mathbf{D}_\mu^{-1}}{1 + d} \right) (\mathbf{B}^\top + rw^\top) \\ &= \mathbf{M} \left( \mathbf{D}_\mu^{-1} \mathbf{B}^\top + \mathbf{D}_\mu^{-1} rw^\top - \frac{\mathbf{D}_\mu^{-1} rr^\top \mathbf{D}_\mu^{-1} \mathbf{B}^\top}{1 + d} - \frac{\mathbf{D}_\mu^{-1} r(r^\top \mathbf{D}_\mu^{-1} r)w^\top}{1 + d} \right) \\ &= \mathbf{M} \left( \mathbf{D}_\mu^{-1} \mathbf{B}^\top + \mathbf{D}_\mu^{-1} rw^\top - \frac{\mathbf{D}_\mu^{-1} rr^\top \mathbf{D}_\mu^{-1} \mathbf{B}^\top}{1 + d} - \frac{d}{1 + d} \mathbf{D}_\mu^{-1} rw^\top \right) \\ &= (\mathbf{B} + wr^\top) \left( \mathbf{D}_\mu^{-1} \mathbf{B}^\top + \frac{\mathbf{D}_\mu^{-1} rw^\top}{1 + d} - \frac{\mathbf{D}_\mu^{-1} rr^\top \mathbf{D}_\mu^{-1} \mathbf{B}^\top}{1 + d} \right) \\ &= \mathbf{B}_\mu^+ \mathbf{B}^\top + \frac{\mathbf{B}_\mu^+ rw^\top}{1 + d} - \frac{\mathbf{B}_\mu^+ rr^\top (\mathbf{B}_\mu^+)^\top}{1 + d} + wr^\top (\mathbf{B}_\mu^+)^\top \\ &\quad + \frac{d}{1 + d} ww^\top - \frac{d}{1 + d} wr^\top (\mathbf{B}_\mu^+)^\top \\ &= \mathbf{B}_\mu^+ \mathbf{B}^\top + \frac{\mathbf{B}_\mu^+ rw^\top}{1 + d} - \frac{\mathbf{B}_\mu^+ rr^\top (\mathbf{B}_\mu^+)^\top}{1 + d} + \frac{wr^\top (\mathbf{B}_\mu^+)^\top}{1 + d} + \frac{d}{1 + d} ww^\top \\ &= \mathbf{B}_\mu^+ \mathbf{B}^\top + \frac{d}{1 + d} ww^\top - \frac{(\mathbf{B}_\mu^+ r - w)(\mathbf{B}_\mu^+ r - w)^\top - ww^\top}{1 + d} \\ &= \mathbf{B}_\mu^+ \mathbf{B}^\top + ww^\top - \frac{(\mathbf{B}_\mu^+ r - w)(\mathbf{B}_\mu^+ r - w)^\top}{1 + d}, \end{aligned}$$

which concludes the proof.  $\square$

*Proof of Lemma B.3.* Observe first that  $\mathbf{B} = \pi_v^\perp \pi_1^\perp \mathbf{V}$ , with  $\pi_v^\perp \pi_1^\perp = \mathbf{I}_{n+1} - (n+1)^{-1} \mathbb{1} \mathbb{1}^\top - ww^\top$ , since  $w^\top \mathbb{1} = 0$ . Now, denoting  $\mathbb{1}_n \in \mathbb{R}^n$  the vector made up of ones (the difference with  $\mathbb{1}$  is that  $\mathbb{1} \in \mathbb{R}^{n+1}$ ), a simple computation shows that

$$ww^\top = \begin{pmatrix} \frac{1}{n(n+1)} \mathbb{1}_n \mathbb{1}_n^\top & -\frac{1}{n+1} \mathbb{1}_n \\ -\frac{1}{n+1} \mathbb{1}_n^\top & \frac{n}{n+1} \end{pmatrix} = \frac{1}{n+1} \begin{pmatrix} \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\top & -\mathbb{1}_n \\ -\mathbb{1}_n^\top & n \end{pmatrix}. \quad (113)$$

Therefore,

$$\begin{aligned} \mathbf{I}_{n+1} - \frac{1}{n+1} \mathbb{1} \mathbb{1}^\top - ww^\top &= \begin{pmatrix} \mathbf{I}_n & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 1 \end{pmatrix} - \frac{1}{n+1} \begin{pmatrix} \mathbb{1}_n \mathbb{1}_n^\top & \mathbb{1}_n \\ \mathbb{1}_n^\top & 1 \end{pmatrix} - \frac{1}{n+1} \begin{pmatrix} \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\top & -\mathbb{1}_n \\ -\mathbb{1}_n^\top & n \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\top & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 0 \end{pmatrix}. \end{aligned} \quad (114)$$

In particular, writing  $\mathbf{V}$  blockwise as  $\mathbf{V}^\top = (\mathbf{V}_0^\top \ \mathbf{V}_{n+1}^\top)$ , where  $\mathbf{V}_0 \in \mathcal{M}_{n,p}$  and  $\mathbf{V}_{n+1} \in \mathcal{M}_{1,p}$ , then

$$\mathbf{B} = \left( \mathbf{I}_{n+1} - \frac{1}{n+1} \mathbb{1} \mathbb{1}^\top - ww^\top \right) \mathbf{V} = \begin{pmatrix} (\mathbf{I}_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^\top) \mathbf{V}_0 \\ \mathbf{0}_{1,n+1} \end{pmatrix}, \quad (115)$$

which finishes the proof of equation (94). As a consequence, the last coordinate of  $\mathbf{D}_\mu^+ r$  is null:  $(\mathbf{D}_\mu^+ r)_{n+1} = (\mathbf{B} \mathbf{D}_\mu^{-1} r)_{n+1} = (e_{n+1}^\top \mathbf{B}) \mathbf{D}_\mu^{-1} r = 0$ . Thus,

$$(bb^\top)_{n+1,n+1} = \left( (\mathbf{B}_\mu^+ r - w)(\mathbf{B}_\mu^+ r - w)^\top \right)_{n+1,n+1} = \left( (\mathbf{B}_\mu^+ r - w)_{n+1} \right)^2 = w_{n+1}^2 = \frac{n}{n+1}. \quad (116)$$

Using the same reasoning,  $(\mathbf{B}_\mu^+ \mathbf{B}^\top)_{n+1,n+1} = 0$  and thus,  $\mathbf{C}_{n+1,n+1} = w_{n+1}^2 = n/(n+1)$ .  $\square$

*Proof of Lemma B.4.* Given  $z \in \mathbb{R}^\ell$ , we apply Lemma B.2 to  $n^{-1}\mathbf{S}(z)$  (equation (85)) by setting  $\mathbf{A} = \mathbf{W}(0)$ ,  $u = \|v\|\mathbf{L}z$ ,  $w = v/\|v\|$  and  $\mu = n\lambda$ . We thus express  $n^{-1}\mathbf{S}(z)$  in the form

$$n^{-1}\mathbf{S}(z) = \mathbf{C} - \frac{bb^\top}{1+d},$$

where  $\mathbf{C}, b$  and  $d$  are given in Lemma B.2. We next apply Lemma B.3, using that  $\mathbf{A} = \mathbf{W}(0) = \pi_1^\perp \mathbf{V}(0)$ . The matrix  $\mathbf{B}$  is thus given by

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_n \\ \mathbf{0}_{1,p} \end{pmatrix} \in \mathcal{M}_{n+1,p}, \quad (117)$$

where  $\mathbf{B}_n$  is given in equation (21). From equation (117) and Lemma B.2, we obtain the expression of  $\mathbf{C}_n$  (equation (32)). Next, from Lemma B.2,  $d_n(z) = r_n(z)^\top \mathbf{D}_{n\lambda}^{-1} r_n(z)$ , where  $r_n(z) = \|v\|\mathbf{L}z + \mathbf{W}(0)^\top w$ . We observe that  $\mathbf{D}_{n\lambda}^{-1} = n^{-1}\widehat{\Sigma}_{n,\lambda}^{-1}$ , which yields the expression of  $d_n(z)$ . The same observation and equation (117) yield the expression of  $b^n(z)$ . Finally, we simplify  $r_n(z)$ , as

$$\begin{aligned} r_n(z) &= \|v\|\mathbf{L}z + \mathbf{W}(0)^\top w = \|v\| \begin{pmatrix} 0 \\ z \end{pmatrix} + \|v\|^{-1} \mathbf{V}(0)^\top \pi_1^\perp v \\ &= \|v\| \begin{pmatrix} 0 \\ z \end{pmatrix} + \|v\|^{-1} \mathbf{V}(0)^\top v = \|v\| \begin{pmatrix} 0 \\ z \end{pmatrix} + \|v\|^{-1} \begin{pmatrix} X_{n+1} - (n+1)^{-1} \sum_{i=1}^{n+1} X_i \\ -(n+1)^{-1} \sum_{i=1}^n R_i \end{pmatrix} \\ &= \|v\| \left[ \begin{pmatrix} 0 \\ z \end{pmatrix} + \|v\|^{-2} \begin{pmatrix} X_{n+1} - (n+1)^{-1} \sum_{i=1}^{n+1} X_i \\ -(n+1)^{-1} \sum_{i=1}^n R_i \end{pmatrix} \right]. \end{aligned} \quad (118)$$

Equation (99) is finally obtained by further noticing that

$$\|v\|^{-2} \left( X_{n+1} - \frac{1}{n+1} \sum_{i=1}^{n+1} X_i \right) = \frac{n+1}{n} \left( \frac{n}{n+1} X_{n+1} - \frac{1}{n+1} \sum_{i=1}^n X_i \right) = X_{n+1} - \frac{1}{n} \sum_{i=1}^n X_i, \quad (119)$$

and performing a similar computation for  $(n+1)^{-1} \sum_{i=1}^n R_i$ .  $\square$

*Proof of Lemma B.5.* The expression for  $\mathbf{S}_X$  is obtained using that

$$\mathbf{X} = (X_1 \dots X_n 0_{k,1})^\top + e_{n+1} X_{n+1}^\top.$$

Hence we can obtain an expression for  $\mathbf{S}_X$  similar to equation (31) by performing the three substitutions

$$\mathbf{V} \leftrightarrow \mathbf{X}, \quad \mathbf{L} \leftrightarrow \mathbf{I}_k, \quad z \leftrightarrow X_{n+1}. \quad (120)$$

With the substitution (120), the vector  $r_n(z)$  in Lemma B.4 is replaced with

$$r_n = \|v\|(X_{n+1} - \bar{X}_n) = \|v\|X_{n+1}^c.$$

Next,  $\mathbf{B}_n$  is replaced with its  $n \times k$  left sub-block which we denote by  $\mathbf{B}_n^X$ , which verifies

$$(\mathbf{B}_n^X)_{ij} = (X_i)_j - \frac{1}{n} \sum_{k=1}^n (X_k)_j, \quad 1 \leq i \leq n, \quad 1 \leq j \leq k.$$

The matrix  $\widehat{\Sigma}_{n,\lambda}$  is next replaced with its upper  $k \times k$  sub-block  $\widehat{\Sigma}_{n,\lambda}^{11}$ , since

$$\widehat{\Sigma}_{n,\lambda}^{11} = \frac{1}{n} (\mathbf{B}_n^X)^\top \mathbf{B}_n^X + \lambda \mathbf{I}_k.$$

Next,  $d_n(z)$  is replaced with

$$d_n^X = \frac{1}{n} r_n^\top (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} r_n = \frac{\|v\|^2}{n} (X_{n+1}^c)^\top (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c = \frac{1}{n+1} (X_{n+1}^c)^\top (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c.$$

The vector  $b_n(z)$  is replaced with

$$b_X^n = \frac{1}{n} \left( \mathbf{B}_n^X (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} r_n \right) - w = \frac{\|v\|}{n} \left( \mathbf{B}_n^X (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c \right) - w.$$

Finally, the matrix  $\mathbf{P}_{n,\lambda}$  is replaced with

$$\mathbf{P}_{n,\lambda}^{XX} := \mathbf{B}_n^X \left( (\mathbf{B}_n^X)^\top \mathbf{B}_n^X + n\lambda \mathbf{I}_k \right)^{-1} (\mathbf{B}_n^X)^\top = \frac{1}{n} \mathbf{B}_n^X (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} (\mathbf{B}_n^X)^\top.$$

This finishes the proof.  $\square$

*Proof of Lemma 3.2.* First, observe from Lemma B.3 that

$$\frac{1}{n} S_{n+1}(z) = \frac{n}{n+1} \left( 1 - \frac{1}{1+d_n(z)} \right). \quad (121)$$

Now, recall that the standard conformal region  $\mathcal{C}_\alpha^n$  corresponding to our score is given by

$$\mathcal{C}_\alpha^n = \{z \in \mathbb{R}^\ell : S_{n+1}(z) \leq S_{(n_\alpha)}(z)\}, \quad (122)$$

where  $S_{(n_\alpha)}(z)$  is the order statistic of order  $n_\alpha$  of the  $n$ -tuple  $(S_1(z), \dots, S_n(z))$ . Next, denote  $C_{(n_\alpha)}$  the order statistic of the  $n$ -tuple  $((\mathbf{C}_n)_{11}, \dots, (\mathbf{C}_n)_{nn})$ . Observe that, by definition, at least  $100 \times (1-\alpha)^{\frac{n+1}{n}}\%$  values of this tuple are less than or equal to  $C_{(n_\alpha)}$ ; but for each of such values  $(\mathbf{C}_n)_{ii}$  and for all  $z$ , we have  $S_i(z) \leq n(\mathbf{C}_n)_{ii}$  (equation (31)). Hence, at least  $100 \times (1-\alpha)^{\frac{n+1}{n}}\%$  values of  $(S_1(z), \dots, S_n(z))$  are less or equal than  $nC_{(n_\alpha)}$ . Thus,  $S_{(n_\alpha)}(z) \leq nC_{(n_\alpha)}$ , and

$$\mathcal{C}_\alpha^n \subset \{z \in \mathbb{R}^\ell : S_{n+1}(z) \leq nC_{(n_\alpha)}\}. \quad (123)$$

To finish, recall that  $(\mathbf{C}_n)_{ii} = p_{i,n} + 1/n(n+1)$ : in particular,  $nC_{(n_\alpha)} = q_{n,\alpha} + 1/(n+1)$ . This shows that the set on the right hand side of equation (123) is  $\tilde{\mathcal{C}}_\alpha^n$ .  $\square$

*Proof of Proposition 3.3.* We start by rewriting the equation defining  $\tilde{\mathcal{C}}_\alpha^n$ . Starting from equations (121) and (123),

$$z \in \tilde{\mathcal{C}}_\alpha^n \iff -\left( \frac{n^2}{n+1} \right) \frac{1}{1+d_n(z)} \leq q_{n,\alpha} + \frac{1}{n+1} - \frac{n^2}{n+1} = q_{n,\alpha} - (n-1). \quad (124)$$

From the equation above, if  $q_{n,\alpha} \geq n-1$ , then  $\tilde{\mathcal{C}}_\alpha^n = \mathbb{R}^\ell$ . Assume now that  $q_{n,\alpha} < n-1$ , then

$$z \in \tilde{\mathcal{C}}_\alpha^n \iff 1 + d_n(z) \leq \left( \frac{n^2}{n+1} \right) \frac{1}{n-1-q_{n,\alpha}} \iff d_n(z) \leq \frac{q_{n,\alpha} + \frac{1}{n+1}}{n-1-q_{n,\alpha}}. \quad (125)$$

Now, recall from Lemma B.4 that  $d_n(z) = n^{-1} r_n(z)^\top \widehat{\Sigma}_{n,\lambda}^{-1} r_n(z)$ . Following equation (99) and the notation  $X_{n+1}^c := X_{n+1} - \bar{X}_n$ , we set  $z_c := z - \bar{R}_n$ , so that

$$r_n(z) = \|v\| \begin{pmatrix} X_{n+1}^c \\ z_c \end{pmatrix}. \quad (126)$$

Using that  $\|v\|^2 = n/(n+1)$  and equation (90),  $d_n(z)$  is rewritten as

$$(n+1)d_n(z) = \frac{n+1}{n} r_n(z)^\top \widehat{\Sigma}_{n,\lambda}^{-1} r_n(z) = \frac{n+1}{n} \|v\|^2 \begin{pmatrix} X_{n+1}^c \\ z_c \end{pmatrix}^\top \widehat{\Sigma}_{n,\lambda}^{-1} \begin{pmatrix} X_{n+1}^c \\ z_c \end{pmatrix} = \begin{pmatrix} X_{n+1}^c \\ z_c \end{pmatrix}^\top \widehat{\Sigma}_{n,\lambda}^{-1} \begin{pmatrix} X_{n+1}^c \\ z_c \end{pmatrix} \quad (127)$$

$$= (z_c - z_0)^\top \mathcal{A}_n^{-1} (z_c - z_0) + (X_{n+1}^c)^\top (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c, \quad (128)$$

with  $\mathcal{A}_n = \widehat{\Sigma}_{n,\lambda} / \widehat{\Sigma}_{n,\lambda}^{11}$  and  $z_0 = \widehat{\Sigma}_n^{21} (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c$ . Going back to equation (125), the final equation of  $\tilde{\mathcal{C}}_\alpha^n$  is obtained by setting  $Z_0^n := z_0 + \bar{R}_n$ , and writing

$$z \in \tilde{\mathcal{C}}_\alpha^n \iff (z - Z_0^n)^\top \mathcal{A}_n^{-1} (z - Z_0^n) \leq (n+1) \frac{q_{n,\alpha} + 1/(n+1)}{n-1-q_{n,\alpha}} - (X_{n+1}^c)^\top (\widehat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c.$$

The quantile term is simplified as

$$(n+1) \frac{q_{n,\alpha} + 1/(n+1)}{n-1-q_{n,\alpha}} = \frac{(n+1)q_{n,\alpha} + 1 + n - n}{n-1-q_{n,\alpha}} = \frac{n(q_{n,\alpha} + 1)}{n-1-q_{n,\alpha}} - 1 = \frac{q_{n,\alpha} + 1}{1 - (q_{n,\alpha} + 1)/n} - 1, \quad (129)$$

which finishes to show that  $\tilde{C}_\alpha^n = \mathcal{E}_\alpha^n$ .  $\square$

*Proof of Proposition 3.4.* The core idea for this result is the observation that for all  $z \in K$ ,

$$q_{n,\alpha} - \sup_{z \in K} \max_{1 \leq i \leq n} nb_i^n(z)^2 \leq S_{(n_\alpha)}(z) - 1/(n+1) \leq q_{n,\alpha}, \quad (130)$$

and the fact that under our moment assumptions,

$$b^n(K) := \sup_{z \in K} \max_{1 \leq i \leq n} nb_i^n(z)^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (131)$$

Let us prove the equation above. For this, denote  $Q_n(z) := (\mathbf{B}_n \hat{\Sigma}_{n,\lambda}^{-1} r_n(z))_1 = (V_1^c)^\top \hat{\Sigma}_{n,\lambda}^{-1} r_n(z)$ , where  $V_1^c = V_1 - n^{-1} \sum_{i=1}^n V_i$  and  $r_n(z)$  is given in equation (99). Next, following equation (98), write that for all  $z \in \mathbb{R}^\ell$ ,

$$\begin{aligned} b_1^n(z) &= \frac{Q_n(z)}{n} - w_1 = \frac{Q_n(z)}{n} - \frac{1}{\sqrt{n(n+1)}}, \\ (nb_1^n(z)^2)^q &= (\sqrt{n}b_1^n(z))^{2q} = \left( \frac{Q_n(z)}{\sqrt{n}} - \frac{1}{\sqrt{n+1}} \right)^{2q} \lesssim \frac{Q_n(z)^{2q}}{n^q} + \frac{1}{n^q}. \end{aligned} \quad (132)$$

Above, we used that  $|a+b|^{2q} \leq 2^{q-1}(|a|^{2q} + |b|^{2q})$ . Observe now that almost surely, because of the continuity of maps of the form  $z \mapsto \mathbf{A}z + b$ ,

$$b^n(K) = \sup_{z \in K} \max_{1 \leq i \leq n} nb_i^n(z)^2 = \sup_{z \in K \cap \mathbb{Q}^\ell} \max_{1 \leq i \leq n} nb_i^n(z)^2. \quad (133)$$

Hence,  $b^n(K)$  is a well-defined random variable (we apply the same reasoning to  $\sup_{z \in K} |Q_n(z)|$ ). Next (explanation below),

$$\begin{aligned} \mathbb{E}[b^n(K)]^q &= \mathbb{E} \left[ \sup_{z \in K} \max_{1 \leq i \leq n} nb_i^n(z)^2 \right]^q \leq \mathbb{E} \left[ \sup_{z \in K} \max_{1 \leq i \leq n} (nb_i^n(z)^2)^q \right] \leq \mathbb{E} \left[ \sup_{z \in K} \sum_{i=1}^n (nb_i^n(z)^2)^q \right] \\ &\leq \mathbb{E} \left[ \sum_{i=1}^n \sup_{z \in K} (nb_i^n(z)^2)^q \right] = n \mathbb{E} \left[ \sup_{z \in K} (nb_1^n(z)^2)^q \right] \\ &\lesssim \frac{1}{n^{q-1}} \mathbb{E} \left[ \sup_{z \in K} |Q_n(z)|^{2q} \right] + \frac{1}{n^{q-1}}. \end{aligned} \quad (134)$$

We used Jensen's inequality in the first inequality, and equation (132) in equation (134). We now prove that the assumption yields that  $\sup_n \mathbb{E}[\sup_{z \in K} |Q_n(z)|^{2q}] < +\infty$ . For this, use equation (99) to decompose  $r_n(z)$  as

$$r_n(z) = \|v\|(\mathbf{L}z + v_n), \quad v_n = \begin{pmatrix} X_{n+1} \\ 0 \end{pmatrix} - \frac{1}{n} \sum_{i=1}^n V_i. \quad (135)$$

(Be wary that  $v_n$  above has no link with  $v$ .) Next (explanation below),

$$|Q_n(z)|^2 = \left( (V_1^c)^\top \hat{\Sigma}_{n,\lambda}^{-1} r_n(z) \right)^2 \leq (V_1^c)^\top \hat{\Sigma}_{n,\lambda}^{-1} V_1^c \times r_n(z)^\top \hat{\Sigma}_{n,\lambda}^{-1} r_n(z) \quad (136)$$

$$\leq \lambda^{-2} \|V_1^c\|^2 \|r_n(z)\|^2 \lesssim \|V_1^c\|^2 (\|z\|^2 + \|v_n\|^2), \quad (137)$$

$$\sup_{z \in K} |Q_n(z)|^2 \lesssim \|V_1^c\|^2 \left( \sup_{z \in K} \|z\|^2 + \|v_n\|^2 \right). \quad (138)$$

Above, we applied the Cauchy-Schwarz inequality for the inner product  $(u, v) \mapsto u^\top \mathbf{A}v$  where  $\mathbf{A}$  is symmetric PSD in equation (136). In equation (137), we used that  $\hat{\Sigma}_{n,\lambda}^{-1} \preccurlyeq \lambda^{-1}I$ ,  $\|v\| \leq 1$  and  $\|r_n(z)\|^2 \leq 2(\|z\|^2 + \|v_n\|^2)$ ,

the latter inequality being easily deduced from equation (135). Applying the Cauchy-Schwarz inequality for the expectation and the fact that  $|a + b|^{2q} \lesssim |a|^{2q} + |b|^{2q}$ ,

$$\begin{aligned} \mathbb{E} \left[ \sup_{z \in K} |Q_n(z)|^{2q} \right] &\lesssim \mathbb{E}[\|V_1^c\|^{4q}]^{1/2} \mathbb{E} \left[ \left( \sup_{z \in K} \|z\|^2 + \|v_n\|^2 \right)^{2q} \right]^{1/2} \\ &\lesssim \mathbb{E}[\|V_1^c\|^{4q}]^{1/2} \left( \sup_{z \in K} \|z\|^{4q} + \mathbb{E}[\|v_n\|^{4q}] \right)^{1/2}. \end{aligned} \quad (139)$$

Similarly, from the triangle inequality and the convexity of  $(\cdot)^{4q}$ , the right-hand side of equation (139) is bounded as

$$\begin{aligned} \mathbb{E}[\|V_1^c\|^{4q}] &\lesssim \mathbb{E}[\|V_1\|^{4q}] + \mathbb{E} \left[ \left\| \frac{1}{n} \sum_{i=1}^n V_i \right\|^{4q} \right] \leq 2\mathbb{E}[\|V_1\|^{4q}], \\ \mathbb{E}[\|v_n\|^{4q}] &\lesssim \mathbb{E}[\|X_{n+1}\|^{4q}] + \mathbb{E} \left[ \left\| \frac{1}{n} \sum_{i=1}^n V_i \right\|^{4q} \right] \leq 2\mathbb{E}[\|V_1\|^{4q}]. \end{aligned}$$

Hence, from the assumption,  $\sup_n \mathbb{E} \left[ \sup_{z \in K} |Q_n(z)|^{2q} \right] < +\infty$  and from equation (134),  $b^n(K) \rightarrow 0$  in  $L^1(\mathbb{P})$ . Thus,

$$b^n(K) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (140)$$

We now prove equation (35). For this, we introduce the increasing function  $f_n$  and the nonnegative random variable  $T_n$ , such that  $\rho_{n,\alpha} = f_n(q_{n,\alpha}) - T_n$ , following equation (27). We also denote  $\bar{q}_{n,\alpha} := q_{n,\alpha} - b^n(K)$  ( $\bar{q}_{n,\alpha} \leq q_{n,\alpha}$ ) and  $\bar{\rho}_{n,\alpha} := f_n(\bar{q}_{n,\alpha}) - T_n$  ( $\bar{\rho}_{n,\alpha} \leq \rho_{n,\alpha}$ ). Finally we define  $\bar{\mathcal{E}}_\alpha^n$  as the following ellipsoid,

$$\bar{\mathcal{E}}_\alpha^n := \{z \in \mathbb{R}^\ell : (z - Z_0^n)^\top \mathcal{A}_n^{-1} (z - Z_0^n) \leq (\bar{\rho}_{n,\alpha})_+\}. \quad (141)$$

By copying the proof of Proposition 3.3,  $\bar{\mathcal{E}}_\alpha^n$  is also the set defined by  $\{z \in \mathbb{R}^\ell : S_{n+1}(z) \leq \bar{q}_{n,\alpha} + (n+1)^{-1}\}$ . For all  $z \in K$ , the inequality  $\bar{q}_{n,\alpha} + (n+1)^{-1} \leq S_{(n,\alpha)}(z) \leq q_{n,\alpha} + (n+1)^{-1}$  then yields that

$$(\bar{\mathcal{E}}_\alpha^n \cap K) \subset (\mathcal{C}_\alpha^n \cap K) \subset (\mathcal{E}_\alpha^n \cap K). \quad (142)$$

We now study the associated volumes. First, given  $s \geq 1$ , recall that from the mean value theorem,  $|x^s - y^s| \leq s \max(|x|, |y|)^{s-1} |x - y|$ ,  $x, y \geq 0$ , and observe that, because  $f_n$  is increasing and, for  $x, t \geq 0$ ,  $(x - t)_+ \leq x$ ,

$$\max((\rho_{n,\alpha})_+, (\bar{\rho}_{n,\alpha})_+)^{s-1} \leq \max(f_n(q_{n,\alpha}), f_n(\bar{q}_{n,\alpha}))^{s-1} = f_n(q_{n,\alpha})^{s-1}. \quad (143)$$

If  $\ell \geq 2$ , this fact and equation (142) together imply that

$$\begin{aligned} 0 \leq \text{Vol}((\mathcal{E}_\alpha^n \setminus \mathcal{C}_\alpha^n) \cap K) &\leq \text{Vol}((\mathcal{E}_\alpha^n \setminus \bar{\mathcal{E}}_\alpha^n) \cap K) \leq \text{Vol}(\mathcal{E}_\alpha^n \setminus \bar{\mathcal{E}}_\alpha^n) = \text{Vol}(\mathcal{E}_\alpha^n) - \text{Vol}(\bar{\mathcal{E}}_\alpha^n) \\ &\leq \det(\mathcal{A}_n)^{1/2} ((\rho_{n,\alpha})_+^{\ell/2} - (\bar{\rho}_{n,\alpha})_+^{\ell/2}) \leq \det(\mathcal{A}_n)^{1/2} \frac{\ell}{2} f_n(q_{n,\alpha})^{\ell/2-1} |(\rho_{n,\alpha})_+ - (\bar{\rho}_{n,\alpha})_+| \\ &\leq \det(\mathcal{A}_n)^{1/2} \frac{\ell}{2} f_n(q_{n,\alpha})^{\ell/2-1} |\rho_{n,\alpha} - \bar{\rho}_{n,\alpha}| \end{aligned} \quad (144)$$

$$\leq \det(\mathcal{A}_n)^{1/2} \frac{\ell}{2} f_n(q_{n,\alpha})^{\ell/2-1} |f_n(q_{n,\alpha}) - f_n(\bar{q}_{n,\alpha})|. \quad (145)$$

(We used that  $(\cdot)_+$  is 1-Lipschitz in equation (144).) But it is clear, from the definition of  $f_n$  as well as Lemma B.6, that  $f_n(q_{n,\alpha}) \rightarrow q_{1-\alpha}^\mathcal{E}$  in probability, while Lemma B.6 and equation (140) imply that  $f_n(\bar{q}_{n,\alpha}) \rightarrow q_{1-\alpha}^\mathcal{E}$  in probability. Combined with equation (51) and the continuity of  $\det(\cdot)$ , equation (145) implies the desired equation (35). If  $\ell = 1$ , the same proof can be adapted using this time that  $|x^{1/2} - y^{1/2}| \leq |x - y|^{1/2}$ ,  $x, y \geq 0$ .  $\square$

*Proof of Lemma 3.5.* From equation (32) and observing that  $0 \preccurlyeq \mathbf{P}_{n,\lambda} \preccurlyeq \mathbf{P}_{n,0}$ , we may first write that

$$\text{Tr}(\mathbf{P}_{n,\lambda}) \leq \text{Tr}(\mathbf{P}_{n,0}) \leq p. \quad (146)$$



Now, denote  $p_{(i)}$  the  $i^{\text{th}}$  order statistic of  $(p_{1,n}, \dots, p_{n,n})$ ,  $p_{(1)} \leq \dots \leq p_{(n)}$ . Then

$$\text{Tr}(\mathbf{P}_{n,\lambda}) = \sum_{i=1}^n p_{i,n} = \sum_{i=1}^n p_{(i)} \geq \sum_{i=n_\alpha}^n p_{(i)} \geq p_{(n_\alpha)}(n+1-n_\alpha) = \frac{q_{n,\alpha}}{n}(n+1-n_\alpha). \quad (147)$$

Moreover,

$$\begin{aligned} n+1-n_\alpha &= n+1 - \lceil (1-\alpha)(n+1) \rceil = n+1 + \lfloor -(1-\alpha)(n+1) \rfloor \\ &= \lfloor n+1 - (1-\alpha)(n+1) \rfloor = \lfloor \alpha(n+1) \rfloor. \end{aligned} \quad (148)$$

Thus, combining equations (146), (147) and (148),

$$q_{n,\alpha} \leq \frac{np}{n+1-n_\alpha} = \frac{np}{\lfloor \alpha(n+1) \rfloor}. \quad (149)$$

Hence, a sufficient condition for  $q_{n,\alpha} < n-1$  is  $np/\lfloor \alpha(n+1) \rfloor < n-1$ , that is,

$$\frac{np}{n-1} < \lfloor \alpha(n+1) \rfloor. \quad (150)$$

But this is equivalent to  $\lfloor np/(n-1) \rfloor + 1 \leq \lfloor \alpha(n+1) \rfloor$ , which is also equivalent to

$$\left\lfloor \frac{np}{n-1} \right\rfloor + 1 = p + \left\lfloor \frac{p}{n-1} \right\rfloor + 1 \leq \alpha(n+1). \quad (151)$$

Thus, if  $p < n-1$ , then  $\lfloor p/(n-1) \rfloor = 0$  and we obtain the sufficient condition  $\alpha \geq (p+1)/(n+1)$ .  $\square$

*Proof of Theorem 3.6.* The computations follow those of Proposition 3.3. Using Lemmas B.2 and B.3 on  $\mathbf{S}(z)$  and  $\mathbf{S}_X$ , we rewrite the score matrix as

$$\begin{aligned} \frac{1}{n} \mathbf{S}'(z) &= \begin{pmatrix} \mathbf{P}_{n,\lambda} & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 0 \end{pmatrix} + ww^\top - \frac{b^n(z)b^n(z)^\top}{1+d_n(z)} - \left( \begin{pmatrix} \mathbf{P}_{n,\lambda}^{XX} & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 0 \end{pmatrix} + ww^\top - \frac{b_X^n(b_X^n)^\top}{1+d_n^X} \right) \\ &= \begin{pmatrix} \mathbf{P}_{n,\lambda} - \mathbf{P}_{n,\lambda}^{XX} & \mathbf{0}_{n,1} \\ \mathbf{0}_{1,n} & 0 \end{pmatrix} - \frac{b^n(z)b^n(z)^\top}{1+d_n(z)} + \frac{b_X^n(b_X^n)^\top}{1+d_n^X}. \end{aligned} \quad (152)$$

In particular,

$$\forall i \in \{1, \dots, n\}, \quad S'_i(z) = n(\mathbf{P}_{n,\lambda} - \mathbf{P}_{n,\lambda}^{XX})_{ii} + n \frac{(b_X^n)_i^2}{1+d_n^X} - n \frac{b^n(z)_i^2}{1+d_n(z)}, \quad (153)$$

$$S'_{n+1}(z) = \frac{n^2}{n+1} \left( \frac{1}{1+d_n^X} - \frac{1}{1+d_n(z)} \right). \quad (154)$$

Equation (154) is a consequence of Lemma B.3 :  $b^n(z)_{n+1}^2 = (b_X^n)_{n+1}^2 = n/(n+1)$ . Now, since  $d_n(z) \geq 0$ , the following uniform bound in  $z$  holds for all  $i \in \{1, \dots, n\}$  :

$$S'_i(z) \leq n(\mathbf{P}_{n,\lambda} - \mathbf{P}_{n,\lambda}^{XX})_{ii} + n \frac{(b_X^n)_i^2}{1+d_n^X} = np'_{i,n}. \quad (155)$$

Using the same arguments as in Lemma 3.2, we have that  $S'_{(n_\alpha)}(z) \leq np'_{(n_\alpha)} = q'_{n,\alpha}$  (see equation (44) for  $q'_{n,\alpha}$ ), hence

$$\{z : S'_{n+1}(z) \leq S'_{(n_\alpha)}(z)\} \subset \{z : S'_{n+1}(z) \leq q'_{n,\alpha}\} =: \mathcal{F}_\alpha^n. \quad (156)$$

We now show that the set  $\mathcal{F}_\alpha^n$  above corresponds to that of Theorem 3.6. First, observe from equation (90) and the definitions of  $d_n(z)$  and  $d_n^X$  that

$$d_n(z) = \frac{1}{n+1} (z - Z_0^n)^\top \mathcal{A}_n^{-1} (z - Z_0^n) + d_n^X \geq 0.$$

Since  $d_n(z)$  is quadratic in  $z$  with  $\min_z d_n(z) = d_n(Z_0^n) = d_n^X$ , equation (154) implies that

$$\max_{z \in \mathbb{R}^\ell} S'_{n+1}(z) = \left( \frac{n^2}{n+1} \right) \frac{1}{1+d_n^X} = \frac{n}{t_n}.$$

Hence, from the definition of  $\mathcal{F}_\alpha^n$  (equation (156)),  $\mathcal{F}_\alpha^n = \mathbb{R}^\ell$  if and only if  $n/t_n \leq q'_{n,\alpha}$ , which settles the case where  $\mathcal{F}_\alpha^n = \mathbb{R}^\ell$ . Assume now that  $n > q'_{n,\alpha} t_n$ . Then,

$$\begin{aligned} z \in \mathcal{F}_\alpha^n &\iff \frac{n^2}{n+1} \left( \frac{1}{1+d_n^X} - \frac{1}{1+d_n(z)} \right) \leq q'_{n,\alpha} \iff \frac{1}{1+d_n(z)} \geq \frac{1}{1+d_n^X} - \frac{n+1}{n^2} q'_{n,\alpha} \\ &\iff 1+d_n(z) \leq \frac{1}{\frac{1}{1+d_n^X} - \frac{n+1}{n^2} q'_{n,\alpha}} = \frac{1+d_n^X}{1 - \frac{n+1}{n^2} (1+d_n^X) q'_{n,\alpha}} = \frac{1+d_n^X}{1 - t_n q'_{n,\alpha}/n} \\ &\iff (n+1)d_n(z) \leq (n+1) \left( \frac{1+d_n^X}{1 - t_n q'_{n,\alpha}/n} - 1 \right) \\ &\iff (z_c - z_0)^\top \mathcal{A}_n^{-1} (z_c - z_0) \leq (n+1) \left( \frac{1+d_n^X}{1 - t_n q'_{n,\alpha}/n} - 1 \right) - (n+1)d_n^X =: \rho'_{n,\alpha}. \end{aligned}$$

Moreover, the right-hand side  $\rho'_{n,\alpha}$  above can be rewritten as

$$\begin{aligned} \rho'_{n,\alpha} &= (n+1) \left( \frac{1+d_n^X}{1 - t_n q'_{n,\alpha}/n} - 1 \right) - (n+1)d_n^X = (n+1) \left( \frac{1+d_n^X}{1 - t_n q'_{n,\alpha}/n} - 1 - d_n^X \right) \\ &= (n+1)(1+d_n^X) \left( \frac{1}{1 - t_n q'_{n,\alpha}/n} - 1 \right) = (n+1)(1+d_n^X) \frac{t_n q'_{n,\alpha}/n}{1 - t_n q'_{n,\alpha}/n} = t_n^2 \frac{q'_{n,\alpha}}{1 - t_n q'_{n,\alpha}/n}, \end{aligned}$$

which finishes the proof.  $\square$

*Proof of Proposition 3.7.* We first prove that  $S'_i(z) \geq 0$  for all  $1 \leq i \leq n+1$ . We first introduce the matrices  $\mathbf{F}$  and  $\mathbf{G}(z)$  such that  $\mathbf{W}(z) = \pi_{\mathbb{I}}^\perp (\mathbf{X} \mathbf{R}(z)) = (\pi_{\mathbb{I}}^\perp \mathbf{X} \pi_{\mathbb{I}}^\perp \mathbf{R}(z)) =: (\mathbf{F} \mathbf{G}(z))$  and  $\widehat{\Sigma}_\lambda$  in equation (10) as

$$\widehat{\Sigma}_\lambda = \frac{1}{n} \mathbf{W}(z)^\top \mathbf{W}(z) + \lambda \mathbf{I}_p = \begin{pmatrix} \widehat{\Sigma}_\lambda^{11} & \widehat{\Sigma}_\lambda^{12} \\ \widehat{\Sigma}_\lambda^{21} & \widehat{\Sigma}_\lambda^{22} \end{pmatrix}, \quad \text{with} \quad \mathbf{S}(z) = (\mathbf{F} \mathbf{G}(z)) \widehat{\Sigma}_\lambda^{-1} \begin{pmatrix} \mathbf{F}^\top \\ \mathbf{G}(z) \end{pmatrix}.$$

Generalizing equation (90) to matrices (using the block inversion lemma B.1) and observing that  $\mathbf{S}_X = \mathbf{F}(\widehat{\Sigma}_\lambda^{11})^{-1} \mathbf{F}^\top$ ,

$$\begin{aligned} \mathbf{S}(z) &= \mathbf{F}(\widehat{\Sigma}_\lambda^{11})^{-1} \mathbf{F}^\top + (\mathbf{G}(z) - \widehat{\Sigma}^{21}(\widehat{\Sigma}_\lambda^{11})^{-1} \mathbf{F})(\widehat{\Sigma}_\lambda / \widehat{\Sigma}_\lambda^{11})^{-1} (\mathbf{G}(z) - (\widehat{\Sigma}_\lambda^{11})^{-1} \mathbf{F})^\top, \\ \mathbf{S}(z) - \mathbf{S}_X &= (\mathbf{G}(z) - \widehat{\Sigma}^{21}(\widehat{\Sigma}_\lambda^{11})^{-1} \mathbf{X})(\widehat{\Sigma}_\lambda / \widehat{\Sigma}_\lambda^{11})^{-1} (\mathbf{G}(z) - (\widehat{\Sigma}_\lambda^{11})^{-1} \mathbf{X})^\top \succcurlyeq 0. \end{aligned}$$

Hence, for all  $i \in \{1, \dots, n+1\}$ ,

$$S'_i(z) = (\mathbf{S}(z) - \mathbf{S}_X)_{ii} = e_i^\top (\mathbf{S}(z) - \mathbf{S}_X) e_i \geq 0.$$

In particular, for all  $i \in \{1, \dots, n\}$ ,

$$n p'_{i,n} \geq S'_i(z) \geq 0,$$

and as a consequence,  $q'_{n,\alpha} \geq 0$ . From the definition of  $\rho'_{n,\alpha}$ , we deduce that  $\rho'_{n,\alpha} \geq 0$ . Finally, since  $d_n(Z_0^n) = d_n^X$ , we have from equation (154) that

$$S'_{n+1}(Z_0^n) = 0 \leq q'_{n,\alpha},$$

and thus, from the definition of  $\mathcal{F}_\alpha^n$  (equation (156)),  $Z_0^n \in \mathcal{F}_\alpha^n$ . We now study the quantity  $\mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell)$ . We first write

$$\mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell) = \mathbb{P}\left( \frac{n}{n+1} \frac{1}{1+d_n^X} \leq \frac{q'_{n,\alpha}}{n} \right).$$

Set  $p''_{i,n} := (\mathbf{P}_{n,\lambda} - \mathbf{P}_{n,\lambda}^{XX})_{ii}$  for  $i \in \{1, \dots, n\}$ . Because  $d_n^X \geq 0$ , we have the bound

$$p'_{i,n} = (\mathbf{P}_{n,\lambda} - \mathbf{P}_{n,\lambda}^{XX})_{ii} + \frac{(b_X^n)_i^2}{1 + d_n^X} \leq p''_{i,n} + \sum_{j=1}^n (b_X^n)_j^2.$$

Hence,

$$q'_{n,\alpha} \leq q''_{n,\alpha} + n \sum_{j=1}^n (b_X^n)_j^2, \quad (157)$$

where  $q''_{n,\alpha}$  is the order statistic of order  $n_\alpha$  of the  $n$ -tuple  $(np''_{1,n}, \dots, np''_{n,n})$ . Using equation (157),

$$\mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell) \leq \mathbb{P}\left(\frac{n}{n+1} \frac{1}{1 + d_n^X} \leq \frac{q''_{n,\alpha}}{n} + \sum_{j=1}^n (b_X^n)_j^2\right) = \mathbb{P}\left(\frac{q''_{n,\alpha}}{n} + \sum_{j=1}^n (b_X^n)_j^2 + \left(1 - \frac{n}{n+1} \frac{1}{1 + d_n^X}\right) \geq 1\right).$$

Now, given  $X, Y, Z$  three random variables, we have the event inclusion

$$\{X + Y + Z \geq 1\} \subset \{X \geq 1/3\} \cup \{Y \geq 1/3\} \cup \{Z \geq 1/3\}.$$

(If both  $X < 1/3$  and  $Y < 1/3$ , then  $Z \geq 1 - X - Y \geq 1/3$ .) In particular, using a union bound,

$$\mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell) \leq \mathbb{P}\left(\frac{q''_{n,\alpha}}{n} \geq \frac{1}{3}\right) + \mathbb{P}\left(\sum_{j=1}^n (b_X^n)_j^2 \geq \frac{1}{3}\right) + \mathbb{P}\left(1 - \frac{n}{n+1} \frac{1}{1 + d_n^X} \geq \frac{1}{3}\right). \quad (158)$$

We now show that the three random variables in the equation above converge to 0 in probability, which will imply that  $\mathbb{P}(\mathcal{F}_\alpha^n = \mathbb{R}^\ell) \rightarrow 0$  when  $n \rightarrow \infty$ . First, from Lemma B.6, we have that  $q''_{n,\alpha}$  converges a.s. to a constant, so a.s.,  $q''_{n,\alpha}/n \rightarrow 0$  when  $n \rightarrow \infty$ , in particular

$$\mathbb{P}\left(\frac{q''_{n,\alpha}}{n} \geq \frac{1}{3}\right) \xrightarrow{n \rightarrow \infty} 0. \quad (159)$$

Next, denoting  $\lambda_{\min,n} = \min \text{Spec}(\Sigma_{n,\lambda=0}^{11})$ , we have

$$\begin{aligned} 0 \leq d_n^X &= \frac{1}{n+1} \mathbb{E}[X_{n+1}^c (\Sigma_{n,\lambda}^{11})^{-1} X_{n+1}^c] \leq \frac{1}{n+1} \frac{1}{\lambda + \lambda_{\min,n}} \|X_{n+1} - \bar{X}_n\|^2 \\ &\leq \frac{2}{(n+1)(\lambda + \lambda_{\min,n})} (\|X_{n+1}\|^2 + \|\bar{X}_n\|^2) = \frac{2\|X_{n+1}\|^2}{(n+1)(\lambda + \lambda_{\min,n})} + \frac{2\|\bar{X}_n\|^2}{(n+1)(\lambda + \lambda_{\min,n})}. \end{aligned} \quad (160)$$

From the SLLN and the continuity of the eigenvalues of self adjoint matrices ([20], Problem 1 p. 198), we have a.s.,  $\lambda_{\min,n} \rightarrow \lambda_{\min}$  where  $\lambda_{\min} = \min \text{Spec}(\Sigma^{11})$ . Thus the second term in equation converges a.s. to 0. Moreover,  $X_{n+1} \stackrel{d}{=} X_1$ . Thus, from Slutsky's lemma and the continuous mapping theorem,

$$\frac{2\|X_{n+1}\|^2}{(n+1)(\lambda + \lambda_{\min,n})} \xrightarrow[n \rightarrow \infty]{d} 0 \times \|X_1\|^2 = 0.$$

Because the limit above is constant, the convergence above also holds in probability. Thus,  $d_n^X \rightarrow 0$  in probability, when  $n \rightarrow \infty$ . Hence, assuming  $n \geq 3$ ,

$$\mathbb{P}\left(1 - \frac{n}{n+1} \frac{1}{1 + d_n^X} \geq \frac{1}{3}\right) = \mathbb{P}\left(d_n^X \geq \frac{3}{2} \frac{n}{n+1} - 1\right) \leq \mathbb{P}\left(d_n^X \geq \frac{9}{8} - 1\right) = \mathbb{P}\left(d_n^X \geq \frac{1}{8}\right) \xrightarrow{n \rightarrow \infty} 0. \quad (161)$$

Finally, using that  $(b_X^n)_{n+1}^2 = w_{n+1}^2 = n/(n+1)$  (see Lemma B.5),

$$\|b_X^n\|^2 = (b_X^n)^\top b_X^n = (b_X^n)_{n+1}^2 + \sum_{j=1}^n (b_X^n)_j^2 = \frac{n}{n+1} + \sum_{j=1}^n (b_X^n)_j^2. \quad (162)$$

Now, denote  $\mathbf{1}_n$  the vector of ones of  $\mathbb{R}^n$ , and observe that  $\mathbf{B}_n^X = \pi_{\mathbf{1}_n}^\perp \mathbf{M}$  where  $\mathbf{M} \in \mathcal{M}_{n,k}$  is given by  $\mathbf{M}_{ij} = (X_i)_j$ . Observe further that  $w = c \times (-1 \dots -1 \ n)^\top = c \times (-\mathbf{1}_n^\top \ n)^\top$ , where  $c = (n+1)^{-1} \|v\|^{-1}$ . Using that  $\pi_{\mathbf{1}_n}^\perp \mathbf{1}_n = 0$  and that the matrix  $\pi_{\mathbf{1}_n}^\perp$  is symmetric,

$$w^\top \begin{pmatrix} \mathbf{B}_n^X (\boldsymbol{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c \\ 0 \end{pmatrix} = -c \left( \mathbf{1}_n^\top \pi_{\mathbf{1}_n}^\perp \right) \mathbf{M} (\boldsymbol{\Sigma}_{n,\lambda}^{11})^{-1} (X_{n+1}^c) = 0.$$

Hence,

$$\begin{aligned} \|b_X^n\|^2 &= \left( \frac{\|v\|}{n} \begin{pmatrix} \mathbf{B}_n^X (\boldsymbol{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c \\ 0 \end{pmatrix} - w \right)^\top \left( \frac{\|v\|}{n} \begin{pmatrix} \mathbf{B}_n^X (\boldsymbol{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c \\ 0 \end{pmatrix} - w \right) \\ &= \frac{\|v\|^2}{n} (X_{n+1}^c)^\top (\boldsymbol{\Sigma}_{n,\lambda}^{11})^{-1} \frac{(\mathbf{B}_n^X)^\top \mathbf{B}_n^X}{n} (\boldsymbol{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c + \|w\|^2 \\ &\leq \frac{\|v\|^2}{n} (X_{n+1}^c)^\top (\boldsymbol{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c + 1 = d_n^X + 1. \end{aligned} \quad (163)$$

We used that  $n^{-1} (\mathbf{B}_n^X)^\top \mathbf{B}_n^X (\boldsymbol{\Sigma}_{n,\lambda}^{11})^{-1} \preceq \mathbf{I}_k$  in equation (163). Combining equations (162) and (163),

$$0 \leq \sum_{j=1}^n (b_X^n)_j^2 \leq d_n^X + \frac{1}{n+1}, \quad (164)$$

and thus  $\sum_{j=1}^n (b_X^n)_j^2 \rightarrow 0$  in probability, when  $n \rightarrow \infty$ . As a consequence,

$$\mathbb{P} \left( \sum_{j=1}^n (b_X^n)_j^2 \geq \frac{1}{3} \right) \xrightarrow{n \rightarrow \infty} 0. \quad (165)$$

Combining equations (158), (159), (161) and (165), we finally obtain that

$$\mathbb{P}(\mathcal{F}_n^\alpha = \mathbb{R}^\ell) \xrightarrow{n \rightarrow \infty} 0, \quad (166)$$

which finishes the proof.  $\square$

*Proof of Proposition 4.1.* We clearly have, from the SLLN,

$$\widehat{\boldsymbol{\Sigma}}_{n,\lambda} = \frac{1}{n} \mathbf{B}_n^\top \mathbf{B}_n + \lambda \mathbf{I}_p \xrightarrow[n \rightarrow \infty]{a.s.} \boldsymbol{\Sigma}_\lambda. \quad (167)$$

In particular, since we chose  $\lambda > 0$ , the continuous mapping theorem implies that

$$\mathcal{A}_n \xrightarrow[n \rightarrow \infty]{a.s.} \mathcal{A}_\infty. \quad (168)$$

Next, the fact that  $q_{n,\alpha} \rightarrow q_{1-\alpha}^\mathcal{E}$  almost surely is deferred in Lemma B.6. This fact also implies that  $q_{n,\alpha}/n \xrightarrow[n \rightarrow \infty]{a.s.} 0$ . Thus, using Slutsky's lemma,

$$\rho_{n,\alpha} \xrightarrow[n \rightarrow \infty]{d} q_{1-\alpha}^\mathcal{E} - X_c^\top (\boldsymbol{\Sigma}_\lambda^{11})^{-1} X_c = \rho_{\infty,\alpha}. \quad (169)$$

Likewise, the convergence in distribution of  $Z_0^n$  is also obtained via Slutsky's lemma. Next, from the continuous mapping theorem, the volume converges in distribution to the random limit volume given by

$$\text{Vol}(\mathcal{E}_\alpha^\infty) = v_\ell \det((\rho_{\infty,\alpha})_+ \mathcal{A}_\infty)^{1/2} = v_\ell \sqrt{\det(\boldsymbol{\Sigma}_\lambda / \boldsymbol{\Sigma}_\lambda^{11})} \left( q_{1-\alpha}^\mathcal{E} - X_c^\top (\boldsymbol{\Sigma}_\lambda^{11})^{-1} X_c \right)_+^{\ell/2}. \quad (170)$$

The limit probability in equation (55) corresponds to  $\rho_{\infty,\alpha} < 0$ . We now show that it is  $\leq \alpha$ . For this, we introduce the simplified notations

$$\boldsymbol{\Sigma}_\lambda = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{C} \end{pmatrix}, \quad \mathbf{A} = \boldsymbol{\Sigma}_\lambda^{11}. \quad (171)$$

For conciseness, denote  $A = V_c^\top \Sigma_\lambda^{-1} V_c$ ,  $B = X_c^\top (\Sigma_\lambda^{11})^{-1} X_c$  and  $F_A, F_B$  their CDF: we first show that  $A \geq B$ . From equation (90),

$$\begin{aligned} A &= X_c^\top \mathbf{A}^{-1} X_c + (R_c - \mathbf{B}^\top \mathbf{A}^{-1} X_c)^\top (\Sigma_\lambda / \mathbf{A})^{-1} (R_c - \mathbf{B}^\top \mathbf{A}^{-1} X_c) \\ &= B + (R_c - \mathbf{B}^\top \mathbf{A}^{-1} X_c)^\top (\Sigma_\lambda / \mathbf{A})^{-1} (R_c - \mathbf{B}^\top \mathbf{A}^{-1} X_c). \end{aligned} \quad (172)$$

Since  $(\Sigma_\lambda / \mathbf{A}) \succ 0$ , the equation above shows that  $A \geq B$  and thus  $F_A \leq F_B$ . Hence, denoting  $Q_A$  the quantile function of  $A$  and using that  $Q_A(p) \leq x \iff p \leq F_A(x)$ ,

$$\begin{aligned} \mathbb{P}(\mathcal{E}_\alpha^\infty = \emptyset) &= \mathbb{P}(X_c^\top (\Sigma_\lambda^{11})^{-1} X_c > q_{1-\alpha}^\mathcal{E}) = 1 - \mathbb{P}(X_c^\top (\Sigma_\lambda^{11})^{-1} X_c \leq q_{1-\alpha}^\mathcal{E}) \\ &= 1 - F_B(Q_A(1-\alpha)) \leq 1 - F_A(Q_A(1-\alpha)) \leq \alpha. \end{aligned} \quad (173)$$

To finish, if  $\lambda = 0$ , we assume that  $\min \text{Spec}(\Sigma) > 0$ . From the continuity of the smallest eigenvalue over the set of Hermitian matrices [20], Problem 1 p. 198, and equation (167), we deduce that almost surely,  $\min \text{Spec}(\hat{\Sigma}_n) > 0$  for  $n$  large enough ( $n$  depends on the sample). Thus, equation (168) also holds, from the continuity of the map  $\mathbf{A} \mapsto \mathbf{A}^{-1}$  over the set of invertible matrices of size  $p$ . Since  $\min \text{Spec}(\Sigma^{11}) \geq \min \text{Spec}(\Sigma) > 0$  ([20], Theorem 4.3.15), the same argument shows that  $(\hat{\Sigma}_n^{11})^{-1} \rightarrow (\Sigma^{11})^{-1}$  almost surely. The rest of the proof is identical to the case  $\lambda > 0$ .  $\square$

*Proof of Proposition 4.2.* Under the assumptions, we have  $S := X_c^\top (\Sigma^{11})^{-1} X_c \sim \chi^2(k)$ ,  $V_c^\top \Sigma^{-1} V_c \sim \chi^2(k+\ell)$  and  $q_{1-\alpha}^\mathcal{E} = F_{\chi^2(k+\ell)}^{-1}(1-\alpha)$ . We begin with computing  $\mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^\infty)^q]$ . Denoting  $t = q_{1-\alpha}^\mathcal{E}$ , we have that  $\mathbb{E}[\text{Vol}(\mathcal{E}_\alpha^\infty)^q] = v_\ell^q \det(\Sigma / \Sigma^{11})^{q/2} \mathbb{E}[(t - S)_+^{q\ell/2}]$ . Setting  $C_k = 2^{k/2} \Gamma(k/2)$ , the expectation is further simplified as

$$\begin{aligned} \mathbb{E}[(t - S)_+^{q\ell/2}] &= \int_{\mathbb{R}^+} (t - s)_+^{q\ell/2} \frac{s^{k/2-1} e^{-s/2}}{C_k} ds = C_k^{-1} t^{q\ell/2} \int_0^t (1 - s/t)_+^{q\ell/2} s^{k/2-1} e^{-s/2} ds \\ &= C_k^{-1} t^{q\ell/2} \int_0^1 (1 - v)^{q\ell/2} (tv)^{k/2-1} e^{-tv/2} t dv \\ &= C_k^{-1} t^{(k+q\ell)/2} \int_0^1 (1 - v)^{q\ell/2} v^{k/2-1} e^{-tv/2} dv \\ &= C_k^{-1} B\left(\frac{k}{2}, \frac{q\ell}{2} + 1\right) t^{(k+q\ell)/2} \Phi_{\text{Beta}(\frac{k}{2}, \frac{q\ell}{2} + 1)}(it/2) \\ &= \frac{\Gamma(q\ell/2 + 1)}{2^{k/2} \Gamma((k+q\ell)/2 + 1)} t^{(k+q\ell)/2} {}_1F_1\left(\frac{k}{2}, \frac{k+q\ell}{2} + 1, -\frac{t}{2}\right). \end{aligned} \quad (174)$$

Above,  $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$  is the Euler Beta function, and  $\Phi_\mu(t) = \int e^{itx} \mu(dx)$  is the characteristic function of a given measure  $\mu$ . Here, we recognize the characteristic function of the Beta distribution  $\text{Beta}(k/2, q\ell/2 + 1)$  [22], p 218. For the proof of the convergence statement, we begin with observing that  $\lfloor \alpha(n+1) \rfloor > \alpha(n+1) - 1$ . From equation (149), when  $n \geq 2(1-\alpha)/\alpha$ ,

$$q_{n,\alpha} \leq \frac{np}{\alpha(n+1) - 1} = \frac{np}{\alpha n - (1-\alpha)} = \frac{p}{\alpha - (1-\alpha)/n} \leq \frac{p}{\alpha - \frac{(1-\alpha)\alpha}{2(1-\alpha)}} = \frac{2p}{\alpha}. \quad (175)$$

From this we deduce that, for  $n$  large enough so that  $1 - (1 + 2p/\alpha)/n > 1/2$ ,

$$\begin{aligned} \rho_{n,\alpha} &= \left( \frac{n(q_{n,\alpha} + 1)}{n - (q_{n,\alpha} + 1)} - 1 - (X_c^{n+1})^\top (\hat{\Sigma}_{n,\lambda}^{11})^{-1} X_{n+1}^c \right)_+ \leq \left( \frac{n(q_{n,\alpha} + 1)}{n - (q_{n,\alpha} + 1)} \right)_+ \\ &\leq \frac{q_{n,\alpha} + 1}{1 - (q_{n,\alpha} + 1)/n} \leq \frac{1 + 2p/\alpha}{1 - (1 + 2p/\alpha)/n} \leq 2(1 + 2p/\alpha). \end{aligned} \quad (176)$$

Hence, using that  $\det(\hat{\Sigma}_n / \hat{\Sigma}_n^{11}) \leq \det(\hat{\Sigma}_n^{22})$ , we obtain that

$$\text{Vol}(\mathcal{E}_\alpha^n) = v_\ell \det(\hat{\Sigma}_n / \hat{\Sigma}_n^{11})^{1/2} (\rho_{n,\alpha})_+^{\ell/2} \lesssim \det(\hat{\Sigma}_n^{22})^{1/2}. \quad (177)$$

Now, recall from [34], Theorem 3.4.8, that  $\det(\hat{\Sigma}_n^{22})$  is equal, in distribution, to  $\det(\Sigma^{22}) U_0 \times \dots \times U_{\ell-1}$ , where  $nU_i \sim \chi^2(n-i)$  and the  $U_i$  are independent. In particular, for all  $q > 0$  [21], p 420,

$$\mathbb{E}[\det(\hat{\Sigma}_n^{22})^q] = \det(\Sigma^{22})^q \prod_{i=0}^{\ell-1} n^{-q} \mathbb{E}[(nU_i)^q] = \det(\Sigma^{22})^q \prod_{i=0}^{\ell-1} \frac{2^q \Gamma(q + (n-i)/2)}{n^q \Gamma((n-i)/2)}. \quad (178)$$

But, when  $s \rightarrow +\infty$  and  $q$  is fixed,  $\Gamma(s+q)/\Gamma(s) \sim s^q$  ([1], Section 6.1.39). Applying this result to each term in the product above, in the regime where  $n \rightarrow +\infty$ , we obtain that for all  $q > 0$ ,  $\sup_n \mathbb{E}[\det(\hat{\Sigma}_n^{22})^q] < +\infty$ . From [47], Example 2.21, and equation (54), all the moments of  $\text{Vol}(\mathcal{E}_\alpha^n)$  converge toward those of  $\text{Vol}(\mathcal{E}_\alpha^\infty)$ . In particular, this proves equation (61).  $\square$

*Proof of Proposition 4.3.* The convergence of  $q'_{n,\alpha}$  is given by Lemma B.6, equation (105). It is also proved in Proposition 3.7 that, if  $\mathbb{E}[\|X_1\|^2] < +\infty$ , then  $d_n^X \rightarrow 0$  in probability (see e.g. equation (161)), so that from the definition of  $t_n$  (equation (45)),  $t_n \rightarrow 1$  in probability. Thus, from the definition of  $\rho'_{n,\alpha}$  (equation (47)), we obtain that  $\rho'_{n,\alpha} \rightarrow q'_{1-\alpha}$  in probability.  $\square$

*Proof of Proposition 4.4.* Decompose  $\Sigma$  as in equation (49). As in Propositions 4.1 and 4.3, set  $V_c := V_1 - \mathbb{E}[V_1]$ ,  $V_c = (X_c^\top R_c^\top)^\top$  and  $T_c = R_c - \Sigma^{21}(\Sigma^{11})^{-1}X_c$ . Then

$$Y_1 - \hat{Y}_1 \in \mathcal{F}_\alpha^\infty \iff R_1 \in \mathcal{F}_\alpha^\infty \iff T_c^\top (\Sigma/\Sigma^{11})^{-1}T_c \leq q_{1-\alpha}^\mathcal{F}. \quad (179)$$

Under the assumption that  $V_c \sim \mathcal{N}(0, \Sigma)$ , using the bilinearity of  $\text{Cov}(\cdot, \cdot)$ ,

$$\text{Cov}(T_c, X_1) = \text{Cov}(T_c, X_c) = \text{Cov}(R_c, X_c) - \Sigma^{21}(\Sigma^{11})^{-1}\text{Cov}(X_c, X_c) = \Sigma^{21} - \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{11} = 0.$$

Because  $V_c$  is a Gaussian random vector, this implies that  $T_c$  and  $X_1$  are independent. In particular, from equation (179) and the definition of  $q_{1-\alpha}^\mathcal{F}$ ,

$$\mathbb{P}(Y_1 - \hat{Y}_1 \in \mathcal{F}_\alpha^\infty | X_1 = x) = \mathbb{P}(T_c^\top \mathcal{A}_\infty^{-1}T_c \leq q_{1-\alpha}^\mathcal{F} | X_1 = x) = \mathbb{P}(T_c^\top \mathcal{A}_\infty^{-1}T_c \leq q_{1-\alpha}^\mathcal{F}) = 1 - \alpha,$$

which finishes the proof.  $\square$

*Proof of Proposition 4.5.* Denote  $\bar{R}_n := n^{-1} \sum_{i=1}^n R_i$  and  $\bar{X}_n := n^{-1} \sum_{i=1}^n X_i$ , and write

$$\begin{aligned} \mathbb{E}[\tilde{Y}_{n+1}] - \mathbb{E}[Y_1] &= \mathbb{E}[\hat{Y}_{n+1} - Y_{n+1}] + \mathbb{E}[\bar{R}_n] + \mathbb{E}[\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}(X_{n+1} - \bar{X}_n)] \\ &= -\mathbb{E}[R_{n+1}] + \mathbb{E}[\bar{R}_n] + \mathbb{E}[\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}X_{n+1}] - \mathbb{E}[\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}\bar{X}_n] \\ &= \mathbb{E}[\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}]\mathbb{E}[X_{n+1}] - \mathbb{E}[\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}\bar{X}_n] \\ &= \mathbb{E}[\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}(\mathbb{E}[X_1] - \bar{X}_n)]. \end{aligned} \quad (180)$$

In equation (180), we used that  $\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}$  and  $X_{n+1}$  are independent. Hence (explanation below),

$$\|\mathbb{E}[\tilde{Y}_{n+1}] - \mathbb{E}[Y_1]\|_2^2 \leq \mathbb{E}[\|\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}(\mathbb{E}[X_1] - \bar{X}_n)\|_2]^2 \quad (181)$$

$$\leq \mathbb{E}[\|\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_{\text{Op}}\|(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}(\mathbb{E}[X_1] - \bar{X}_n)\|_2]^2 \quad (182)$$

$$\lesssim \mathbb{E}[\|\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_F\|(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}(\mathbb{E}[X_1] - \bar{X}_n)\|_2]^2 \quad (183)$$

$$\lesssim \mathbb{E}[\|\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_F^2] \times \mathbb{E}[\|(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}(\mathbb{E}[X_1] - \bar{X}_n)\|_2^2]. \quad (184)$$

We used the triangle inequality in equation (181), the operator norm such that  $\|\mathbf{M}x\|_2 \leq \|\mathbf{M}\|_{\text{Op}}\|x\|_2$  in equation (182), the equivalence of  $\|\cdot\|_{\text{Op}}$  with the Frobenius norm  $\|\mathbf{M}\|_F = \text{Tr}(\mathbf{M}\mathbf{M}^\top)^{1/2}$  in equation (183), and the Cauchy-Schwarz inequality in equation (184). But, since  $\hat{\Sigma}_{n,\lambda}^{22} - \hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}\hat{\Sigma}_n^{12} = \hat{\Sigma}_{n,\lambda}/\hat{\Sigma}_{n,\lambda}^{11} \succcurlyeq 0$ ,

$$\begin{aligned} \mathbb{E}[\|\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_F^2] &= \mathbb{E}[\text{Tr}(\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}\hat{\Sigma}_n^{12})] \leq \mathbb{E}[\text{Tr}(\hat{\Sigma}_{n,\lambda}^{22})] \\ &\leq \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\text{Tr}((R_i - \bar{R}_n)(R_i - \bar{R}_n)^\top)] + \lambda\ell \\ &\leq \mathbb{E}[\|R_1 - \bar{R}_n\|_2^2] + \lambda\ell = \frac{n-1}{n}\text{Tr}(\Sigma^{22}) + \lambda\ell. \end{aligned} \quad (185)$$

Moreover, using that  $\|(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_F^2 = \text{Tr}((\hat{\Sigma}_{n,\lambda}^{11})^{-1}) \leq \text{Tr}(\lambda^{-1}\mathbf{I}_k) = k/\lambda \lesssim 1$  conjointly with the operator and Frobenius norms,

$$\begin{aligned} \mathbb{E}[\|(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}(\mathbb{E}[X_1] - \bar{X}_n)\|_2^2] &\lesssim \mathbb{E}[\|(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_F^2 \|\mathbb{E}[X_1] - \bar{X}_n\|_2^2] \\ &\lesssim \mathbb{E}[\|\mathbb{E}[X_1] - \bar{X}_n\|_2^2] = \text{Tr}(\Sigma)/n. \end{aligned} \quad (186)$$

Equations (184), (185) and (186) then yield equation (67). We now show equation (68). Assuming that there exists  $q > 1$  such that  $\mathbb{E}[\|V_1\|^{4q}] < +\infty$ , we first show that the sequence  $\|\tilde{Y}_{n+1} - Y_{n+1}\|_2^2$  is uniformly integrable. Write

$$\begin{aligned}\tilde{Y}_{n+1} - Y_{n+1} &= \hat{Y}_{n+1} - Y_{n+1} + \bar{R}_n + \hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}(X_{n+1} - \bar{X}_n) \\ &= \bar{R}_n - R_{n+1} + \hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}(X_{n+1} - \bar{X}_n).\end{aligned}\tag{187}$$

Hence, starting from equation (187) and following the steps leading to equation (184),

$$\begin{aligned}\|\tilde{Y}_{n+1} - Y_{n+1}\|_2^{2q} &\lesssim \left(\|\bar{R}_n - R_{n+1}\|_2 + \|\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_F \|(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}(X_{n+1} - \bar{X}_n)\|_2\right)^{2q} \\ &\lesssim \|\bar{R}_n - R_{n+1}\|_2^{2q} + \|\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_F^{2q} \|(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}(X_{n+1} - \bar{X}_n)\|_2^{2q}.\end{aligned}\tag{188}$$

Taking the expectation and using the Cauchy-Schwarz inequality,

$$\mathbb{E}[\|\tilde{Y}_{n+1} - Y_{n+1}\|_2^{2q}] \lesssim \mathbb{E}[\|\bar{R}_n - R_{n+1}\|_2^{2q}] + \mathbb{E}[\|\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_F^{4q}]^{1/2} \mathbb{E}[\|X_{n+1} - \bar{X}_n\|_2^{4q}]^{1/2}.$$

Using the convexity of  $(\cdot)^{2q}$  and  $(\cdot)^{4q}$ , we have that  $\mathbb{E}[\|\bar{R}_n - R_{n+1}\|_2^{2q}] \lesssim \mathbb{E}[\|R_1\|_2^{2q}]$  and  $\mathbb{E}[\|X_{n+1} - \bar{X}_n\|_2^{4q}] \lesssim \mathbb{E}[\|X_1\|_2^{4q}]$ . Likewise,

$$\begin{aligned}\mathbb{E}[\|\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}\|_F^{4q}] &= \mathbb{E}[\text{Tr}(\hat{\Sigma}_n^{21}(\hat{\Sigma}_{n,\lambda}^{11})^{-1}\hat{\Sigma}_n^{12})^{2q}] \leq \mathbb{E}[\text{Tr}(\hat{\Sigma}_{n,\lambda}^{22})^{2q}] \\ &\leq \mathbb{E}\left[\left(\ell\lambda + \frac{1}{n} \sum_{i=1}^n \text{Tr}((R_i - \bar{R}_n)(R_i - \bar{R}_n)^\top)\right)^{2q}\right] \\ &\leq \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^n (\|R_i - \bar{R}_n\|^2 + \ell\lambda)\right)^{2q}\right] \\ &\leq \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (\|R_i - \bar{R}_n\|^2 + \ell\lambda)^{2q}\right] \\ &\leq \mathbb{E}[(\|R_1 - \bar{R}_n\|^2 + \ell\lambda)^{2q}] \lesssim \mathbb{E}[\|R_1\|_2^{4q}] + \lambda^{2q}.\end{aligned}\tag{189}$$

Hence,  $\sup_n \mathbb{E}[\|\tilde{Y}_{n+1} - Y_{n+1}\|_2^{2q}] < +\infty$ . Setting  $S := \Sigma^{21}(\Sigma_\lambda^{11})^{-1}X_1 - R_1$  and observing from Proposition 4.1 that

$$\tilde{Y}_{n+1} - Y_{n+1} \xrightarrow[n \rightarrow +\infty]{d} S - \mathbb{E}[S],\tag{190}$$

we may use [47], Example 2.21, so that  $\|\tilde{Y}_{n+1} - Y_{n+1}\|_2^2$  is uniformly integrable, with

$$\mathbb{E}[\|\tilde{Y}_{n+1} - Y_{n+1}\|_2^2] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[\|S - \mathbb{E}[S]\|^2] = \text{Tr}(\text{Cov}(S)).\tag{191}$$

Finally,

$$\begin{aligned}\text{Cov}(S) &= \Sigma^{22} + \Sigma^{21}(\Sigma_\lambda^{11})^{-1}\Sigma^{11}(\Sigma_\lambda^{11})^{-1}\Sigma^{12} - 2\Sigma^{21}(\Sigma_\lambda^{11})^{-1}\Sigma^{12} \\ &= \Sigma^{22} + \lambda\mathbf{I}_\ell - \lambda\mathbf{I}_\ell + \Sigma^{21}(\Sigma_\lambda^{11})^{-1}(\Sigma^{11} + \lambda\mathbf{I}_k - \lambda\mathbf{I}_k)(\Sigma_\lambda^{11})^{-1}\Sigma^{12} \\ &\quad - 2\Sigma^{21}(\Sigma_\lambda^{11})^{-1}\Sigma^{12} \\ &= \Sigma_\lambda^{22} - \Sigma^{21}(\Sigma_\lambda^{11})^{-1}\Sigma^{12} - \lambda(\mathbf{I}_\ell + \Sigma^{21}(\Sigma_\lambda^{11})^{-1}\Sigma^{12}) \\ &= \Sigma_\lambda/\Sigma_\lambda^{11} - \lambda(\mathbf{I}_\ell + \Sigma^{21}(\Sigma_\lambda^{11})^{-1}\Sigma^{12}) = \mathbf{M}_\lambda.\end{aligned}\tag{192}$$

This finishes to prove that  $\mathbb{E}[\|\tilde{Y}_{n+1} - Y_{n+1}\|_2^2] \rightarrow \text{Tr}(\mathbf{M}_\lambda)$ . Using the same previous arguments for extra-diagonal entries of  $\text{Cov}(\tilde{Y}_{n+1} - Y_{n+1})$  and equation (67), we obtain equation (68).  $\square$

*Proof of Proposition 4.6.* For equation (69), we follow the proof of Théorème 8.9, p. 90 of the lecture notes [4]. We denote  $F$  the CDF of  $\|R_1\|^2$  and  $F_n$  the empirical CDF obtained from the iid sample  $(\|R_1\|, \dots, \|R_n\|)$ . We can write

$$\begin{aligned}|F(\beta_{n,\alpha}) - F(q_{1-\alpha}(\|R_1\|^2))| &\leq |F(\beta_{n,\alpha}) - F_n(\beta_{n,\alpha})| + |F_n(\beta_{n,\alpha}) - F(q_{1-\alpha}(\|R_1\|^2))| \\ &\leq \|F - F_n\|_\infty + \left| \frac{\lceil (1-\alpha)(n+1) \rceil}{n} - (1-\alpha) \right| \rightarrow 0 \text{ a.s.,}\end{aligned}\tag{193}$$

where we used the Glivenko-Cantelli theorem in equation (193). Hence, a.s.,  $F(q'_{n,\alpha})$  converges to  $F(q_{1-\alpha}(\|R_1\|^2)) = 1 - \alpha$ . We obtain equation (69) from the continuity of the quantile function of  $\|R_1\|^2$  on a neighbourhood of  $1 - \alpha$ . We then deduce equation (70) from the continuous mapping theorem, as  $\text{Vol}(\mathcal{B}_\alpha^n) = v_\ell(\beta_{n,\alpha})^{\ell/2}$ .  $\square$



*Proof of Proposition 4.7.* Let  $B$  be the centered ball of  $\mathbb{R}^\ell$ , of radius  $t$ . Let  $A \subset \mathbb{R}^\ell$  be a measurable set such that  $\lambda(A) = \lambda(B)$ , where  $\lambda$  stands for the Lebesgue measure of  $\mathbb{R}^\ell$ . Since  $A = (A \cap B) \cup (A \cap B^c)$  where the union is disjoint (and symmetrically for  $B$ ), we also have that

$$\lambda(A \cap B^c) = \lambda(B \cap A^c). \quad (194)$$

Now, let  $(\delta_1, \dots, \delta_\ell) \in \mathbb{R}_+^\ell$  be such that  $\delta_1 \dots \delta_\ell = 1$ , and  $s \in \mathbb{R}^\ell$ . Observe that  $\sum_{i=1}^\ell \delta_i (z_i + s_i)^2 = (z+s)^\top \mathbf{D}_\delta (z+s)$  where  $\mathbf{D}_\delta$  is the diagonal matrix such that  $(\mathbf{D}_\delta)_i = \delta_i$ . The set  $E_\delta := \{z \in \mathbb{R}^\ell : (z+s)^\top \mathbf{D}_\delta (z+s) \leq t\}$  is an ellipsoid with volume  $\lambda(E_\delta) = v_\ell \det(\mathbf{D}_\delta)^{-1/2} t^\ell = v_\ell t^\ell = \lambda(B)$ . Hence,

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^\ell T_i^2 \leq t\right) - \mathbb{P}\left(\sum_{i=1}^\ell \delta_i (T_i + s_i)^2 \leq t\right) &= \int_{\mathbb{R}^\ell} \mathbb{1}_{z^\top z \leq t} g(\|z\|^2) dz \\ &\quad - \int_{\mathbb{R}^\ell} \mathbb{1}_{(z+s)^\top \mathbf{D}_\delta (z+s) \leq t} g(\|z\|^2) dz \\ &= \int_B g(\|z\|^2) dz - \int_{E_\delta} g(\|z\|^2) dz \\ &= \int_{B \cap E_\delta^c} g(\|z\|^2) dz - \int_{E_\delta \cap B^c} g(\|z\|^2) dz. \end{aligned} \quad (195)$$

But since  $g$  is decreasing,  $B$  is a superlevel set of  $g$ :  $g(\|z\|^2) \geq g(t^2)$  for  $z \in B$  and  $g(\|z\|^2) \leq g(t^2)$  for  $z \in B^c$ . Thus, from equations (194) and (195),

$$\mathbb{P}\left(\sum_{i=1}^\ell T_i^2 \leq t\right) - \mathbb{P}\left(\sum_{i=1}^\ell \delta_i (T_i + s_i)^2 \leq t\right) \geq g(t) (\lambda(B \cap E_\delta^c) - \lambda(E_\delta \cap B^c)) = 0. \quad (196)$$

Hence, for all  $t > 0$ ,  $\mathbb{P}(\sum_{i=1}^\ell \delta_i (T_i + s_i)^2 \leq t) \leq \mathbb{P}(\sum_{i=1}^\ell T_i^2 \leq t)$ . Since  $f \leq g$  implies  $g^{-1} \leq f^{-1}$ , we obtain that for all  $\alpha \in (0, 1)$ ,

$$q_{1-\alpha}\left(\sum_{i=1}^\ell T_i^2\right) \leq q_{1-\alpha}\left(\sum_{i=1}^\ell \delta_i (T_i + s_i)^2\right). \quad (197)$$

This finishes the proof.  $\square$

*Proof of Lemma 4.8.* Given a random vector  $X$ , we denote  $\Phi_X(t) := \mathbb{E}[\exp(it^\top X)]$ . Given that the distribution of  $T$  is spherical, introduce  $\psi$  the function such that  $\Phi_T(t) = \psi(\|t\|^2)$ ,  $t \in \mathbb{R}^p$ . Since  $V_1 = \mu + \Sigma^{1/2}T$ , we have  $\Phi_{V_1}(t) = \exp(it^\top \mu) \psi(t^\top \Sigma t)$  [39], p 34. We also denote  $T' := (T_1, \dots, T_\ell)^\top \in \mathbb{R}^\ell$ . Decomposing  $\mu = (\mu_X^\top \mu_R^\top)^\top$  where  $\mu_R \in \mathbb{R}^\ell$ , we have for  $t' \in \mathbb{R}^\ell$  [39], p 34-35,

$$\Phi_{R_1}(t') = \Phi_{V_1}((0, t')) = e^{i(t')^\top \mu_R} \psi((t')^\top \Sigma^{22} t'), \quad (198)$$

$$\Phi_{T'}(t') = \Phi_T((0, t')) = \psi(\|t'\|^2). \quad (199)$$

Hence,  $R_1 = \mu_R + (\Sigma^{22})^{1/2} \tilde{T}$  for some  $\tilde{T}$  with a spherical distribution. It is also clear from the equations above that  $\Phi_{\tilde{T}}(t') = \psi(\|t'\|^2) = \Phi_{T'}(t')$ , so that  $\tilde{T}$  and  $T'$  are equal in distribution. Write  $\Sigma^{22} = \mathbf{P} \mathbf{D} \mathbf{P}^\top$  an eigendecomposition of  $\Sigma^{22}$ , where  $\mathbf{P}$  is orthogonal and  $\mathbf{D}$  is diagonal,  $\mathbf{D}_{ii} = \lambda_i$ . Then, setting  $s := \mathbf{D}^{-1/2} \mathbf{P}^\top \mu$  and using that  $\mathbf{P}^\top T' \stackrel{d}{=} T'$  from the sphericity of  $T'$ ,

$$\begin{aligned} \|R_1\|^2 &= \|\mu_R + \mathbf{P} \mathbf{D}^{1/2} \mathbf{P}^\top \tilde{T}\|^2 \stackrel{d}{=} \|\mu_R + \mathbf{P} \mathbf{D}^{1/2} \mathbf{P}^\top T'\|^2 \stackrel{d}{=} \|\mu_R + \mathbf{P} \mathbf{D}^{1/2} T'\|^2 \\ &= \|\mathbf{D}^{1/2} (s + T')\|^2 = \sum_{i=1}^\ell \lambda_i (T_i + s_i)^2. \end{aligned} \quad (200)$$

This finishes the proof.  $\square$

*Proof of Proposition 4.9.* Using Lemma 4.8,  $q_{1-\alpha}(\|R_1\|^2) = q_{1-\alpha}(\sum_{i=1}^\ell \lambda_i(T_i + s_i)^2)$  where the  $\lambda_i$  are the eigenvalues of  $\Sigma^{22}$  and from Proposition 4.6,

$$\left(\frac{\text{Vol}(\mathcal{E}_\alpha^\infty)}{\text{Vol}(\mathcal{B}_\alpha^\infty)}\right)^{2/\ell} \leq \det(\Sigma/\Sigma^{11})^{1/\ell} \frac{q_{1-\alpha}(\sum_{i=1}^{k+\ell} T_i^2)}{q_{1-\alpha}(\sum_{i=1}^\ell \lambda_i(T_i + s_i)^2)} \quad (201)$$

$$\begin{aligned} &\leq c_\alpha(k, \ell) \left(\frac{\det(\Sigma/\Sigma^{11})}{\det(\Sigma^{22})}\right)^{1/\ell} \frac{q_{1-\alpha}(\sum_{i=1}^\ell T_i^2)}{\det(\Sigma^{22})^{-1/\ell} q_{1-\alpha}(\sum_{i=1}^\ell \lambda_i(T_i + s_i)^2)} \\ &\leq c_\alpha(k, \ell) \left(\frac{\det(\Sigma/\Sigma^{11})}{\det(\Sigma^{22})}\right)^{1/\ell} \frac{q_{1-\alpha}(\sum_{i=1}^\ell T_i^2)}{q_{1-\alpha}(\sum_{i=1}^\ell \delta_i(T_i + s_i)^2)} \leq c_\alpha(k, \ell) \left(\frac{\det(\Sigma/\Sigma^{11})}{\det(\Sigma^{22})}\right)^{1/\ell}, \end{aligned} \quad (202)$$

where  $\delta_i = \lambda_i/(\lambda_1 \dots \lambda_\ell)^{1/\ell}$  and where we applied Proposition 4.7 in the last inequality. We have thus proved Proposition 4.9 when  $\lambda = 0$ . We now assume that  $\lambda > 0$ . For this, we follow the steps that lead to equation (202), so that

$$\begin{aligned} \text{Vol}(\mathcal{E}_{\alpha, \lambda}^\infty)^{2/\ell} &= v_\ell^{2/\ell} \det(\Sigma_\lambda/\Sigma_\lambda^{11})^{1/\ell} (q_{1-\alpha}(V_c^\top \Sigma_\lambda^{-1} V_c) - X_c^\top (\Sigma_\lambda^{11})^{-1} X_c)_+ \\ &\leq v_\ell^{2/\ell} \det(\Sigma_\lambda/\Sigma_\lambda^{11})^{1/\ell} q_{1-\alpha}(V_c^\top \Sigma_\lambda^{-1} V_c) \\ &\leq v_\ell^{2/\ell} \det(\Sigma_\lambda/\Sigma_\lambda^{11})^{1/\ell} q_{1-\alpha}(V_c^\top \Sigma^{-1} V_c) \\ &\leq v_\ell^{2/\ell} \det(\Sigma_\lambda/\Sigma_\lambda^{11})^{1/\ell} q_{1-\alpha}\left(\sum_{i=1}^{k+\ell} T_i^2\right) \leq \det(\Sigma_\lambda/\Sigma_\lambda^{11})^{1/\ell} c_\alpha(k, \ell) \text{Vol}(\mathcal{B}_\alpha^\infty)^{2/\ell}. \end{aligned} \quad (203)$$

In view of equation (203), we show that the map  $\lambda \mapsto \det(\Sigma_\lambda/\Sigma_\lambda^{11})$  is strictly increasing. For this, denote  $0 \leq \lambda_1 \leq \dots \leq \lambda_p$  the eigenvalues of  $\Sigma$ , and  $0 \leq \lambda_1^1 \leq \dots \leq \lambda_k^1$  the eigenvalues of  $\Sigma^{11}$ . Then, setting  $f_i(\lambda) := (\lambda_i + \lambda)/(\lambda_i^1 + \lambda)$  for  $i = 1, \dots, k$ ,

$$\det(\Sigma_\lambda/\Sigma_\lambda^{11}) = \frac{\det(\Sigma_\lambda)}{\det(\Sigma_\lambda^{11})} = \prod_{i=1}^k \left(\frac{\lambda_i + \lambda}{\lambda_i^1 + \lambda}\right) \prod_{i=k+1}^{k+\ell} (\lambda + \lambda_i) = \prod_{i=1}^k f_i(\lambda) \prod_{i=k+1}^{k+\ell} (\lambda + \lambda_i). \quad (204)$$

But from the interlacing theorem ([20], Theorem 4.3.15), for all  $i \in \{1, \dots, k\}$ ,  $\lambda_i \leq \lambda_i^1$ . Hence,  $f_i'(\lambda) = (\lambda_i^1 - \lambda_i)/(\lambda_i^1 + \lambda)^2 \geq 0$ , and from equation (204), the map  $\lambda \mapsto \det(\Sigma_\lambda/\Sigma_\lambda^{11})$  is strictly increasing (and goes to  $+\infty$  when  $\lambda \rightarrow +\infty$ ). As a consequence, if the inequality in equation (77) is strict, then there exists a unique  $\lambda_0 > 0$  verifying

$$\det(\Sigma_{\lambda_0})/\det(\Sigma_{\lambda_0}^{11}) = \det(\Sigma_{\lambda_0}/\Sigma_{\lambda_0}^{11}) = c_\alpha(k, \ell)^{-\ell} \det(\Sigma^{22}), \quad (205)$$

which corresponds to the announced criterion.  $\square$

*Proof of Proposition 4.10.* We use notations exclusive to this proof. Let  $T = (X^\top \ Y^\top)^\top \in \mathbb{R}^p$  have a spherical distribution, with density

$$f_T(x, y) = g(\|x\|^2 + \|y\|^2), \quad x \in \mathbb{R}^k, \ y \in \mathbb{R}^\ell.$$

Then  $Y$  has a spherical distribution with density

$$f_Y(y) = \int_{\mathbb{R}^k} g(\|x\|^2 + \|y\|^2) dx =: g_Y(\|y\|^2). \quad (206)$$

Define now  $U \in \mathbb{R}^k$ ,  $V \in \mathbb{R}^\ell$  and  $Z \in \mathbb{R}^\ell$  by

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix} + \Sigma^{1/2} T, \quad Z := (\Sigma/\Sigma^{11})^{-1/2} (V - \mu_v - \Sigma^{21} (\Sigma^{11})^{-1} (U - \mu_u)).$$

We now show that  $Z \stackrel{d}{=} Y$ . For this, observe first that the density of  $(U^\top \ V^\top)^\top$  is

$$f_{(U, V)}(u, v) = \det(\Sigma)^{-1/2} g\left(\begin{pmatrix} u - \mu_u \\ v - \mu_v \end{pmatrix}^\top \Sigma^{-1} \begin{pmatrix} u - \mu_u \\ v - \mu_v \end{pmatrix}\right).$$

Setting  $w = v - \mu_v - \Sigma^{21}(\Sigma^{11})^{-1}(u - \mu_u)$ , we have from equation (90) that

$$\begin{pmatrix} u - \mu_u \\ v - \mu_v \end{pmatrix}^\top \Sigma^{-1} \begin{pmatrix} u - \mu_u \\ v - \mu_v \end{pmatrix} = (u - \mu_u)^\top (\Sigma^{11})^{-1} (u - \mu_u) + w^\top (\Sigma / \Sigma^{11})^{-1} w. \quad (207)$$

Taking the definition of  $w$  above as a function of  $u$  and  $v$ , we have, for any bounded measurable function  $h$  (explanation below),

$$\mathbb{E}[h(Z)] = \int_{\mathbb{R}^k} \int_{\mathbb{R}^\ell} h((\Sigma / \Sigma^{11})^{-1/2} w) f_{(U,V)}(u, v) dv du \quad (208)$$

$$= \det(\Sigma)^{-1/2} \int_{\mathbb{R}^k} \int_{\mathbb{R}^\ell} h((\Sigma / \Sigma^{11})^{-1/2} w) g((u - \mu_u)^\top (\Sigma^{11})^{-1} (u - \mu_u) + w^\top (\Sigma / \Sigma^{11})^{-1} w) dv du \quad (209)$$

$$= \det(\Sigma)^{-1/2} \det(\Sigma / \Sigma^{11})^{1/2} \int_{\mathbb{R}^k} \int_{\mathbb{R}^\ell} h(z) g((u - \mu_u)^\top (\Sigma^{11})^{-1} (u - \mu_u) + \|z\|^2) dz du \quad (210)$$

$$= \int_{\mathbb{R}^\ell} h(z) \left( \int_{\mathbb{R}^k} g(\|x\|^2 + \|z\|^2) dx \right) dz. \quad (211)$$

We used the transfer lemma in equation (208), equation (207) in equation (209), the change of variable  $z = (\Sigma / \Sigma^{11})^{-1/2} w$  (where  $u$  is fixed) in the inner integral in equation (210), and the change of variable  $x = (\Sigma^{11})^{-1/2} (u - \mu_u)$  and Fubini's theorem in equation (211). Because equation (211) holds for all bounded measurable function  $h$ , the density of  $Z$  is that of  $Y$ , hence  $Z \stackrel{d}{=} Y$  and

$$\|Z\|^2 \stackrel{d}{=} \|Y\|^2. \quad (212)$$

Observe finally that the equation above still holds if  $Y$  is replaced with any subvector of  $T$  of dimension  $\ell$ , since the distribution of  $T$  is spherical.

We now go back to the vector  $V_1 = (X_1^\top R_1^\top)^\top$  of Proposition 4.10, which we assume to be of the form  $V_1 = \mu + \Sigma^{1/2} T$ , where  $T = (T_1, \dots, T_p)^\top$  has a spherical distribution with a non-increasing generator  $g$ . Using the notation of Proposition 4.3 for  $T_c$  and  $\mathcal{A}_\infty$ , equation (212) and Lemma 4.8 respectively show that

$$T_c^\top \mathcal{A}_\infty^{-1} T_c \stackrel{d}{=} \sum_{i=1}^{\ell} T_i^2, \quad \text{and} \quad \|R_1\|^2 \stackrel{d}{=} \sum_{i=1}^{\ell} \lambda_i (T_i + s_i)^2,$$

for some vector  $s \in \mathbb{R}^\ell$  and where the  $\lambda_i$  are the eigenvalues of  $\Sigma^{22}$ . Then, denoting  $\delta_i := \det(\Sigma^{22})^{-1/\ell} \lambda_i$ , and using that  $\det(\Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12}) \leq \det(\Sigma^{22})$ ,

$$\begin{aligned} \left( \frac{\text{Vol}(\mathcal{F}_\alpha^\infty)}{\text{Vol}(\mathcal{B}_\alpha^\infty)} \right)^{2/\ell} &= \det(\Sigma / \Sigma^{11})^{1/\ell} \frac{q_{1-\alpha}(T_c^\top \mathcal{A}_\infty^{-1} T_c)}{q_{1-\alpha}(\|R_1\|^2)} \\ &= \left( \frac{\det(\Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12})}{\det(\Sigma^{22})} \right)^{1/\ell} \frac{q_{1-\alpha}(\sum_{i=1}^{\ell} T_i^2)}{q_{1-\alpha}(\sum_{i=1}^{\ell} \delta_i (T_i + s_i)^2)} \end{aligned} \quad (213)$$

$$\leq \frac{q_{1-\alpha}(\sum_{i=1}^{\ell} T_i^2)}{q_{1-\alpha}(\sum_{i=1}^{\ell} \delta_i (T_i + s_i)^2)}. \quad (214)$$

We now wish to apply Proposition 4.7 on the spherical vector  $Y := (T_1, \dots, T_\ell)$  to conclude that the ratio above is smaller than 1. For this we need to check that the generator of  $Y$  is non-increasing. Since  $g$  is assumed non-increasing we can use that, for all  $0 \leq r \leq r'$  and  $x \in \mathbb{R}^k$ ,  $g(\|x\|^2 + r) - g(\|x\|^2 + r') \geq 0$  so that, using equation (206),

$$g_Y(r) - g_Y(r') = \int_{\mathbb{R}^k} g(\|x\|^2 + r) - g(\|x\|^2 + r') dx \geq 0.$$

Thus Proposition 4.7 applies to equation (214), which finishes the proof.  $\square$

*Proof of Proposition 4.11.* For all  $i$ , define  $\delta_i := \lambda_i / (\lambda_1 \cdots \lambda_\ell)^{1/\ell}$  and  $(G_1, \dots, G_\ell)$  to be a random vector with independent components, such that  $G_i$  follows the gamma distribution  $\gamma(\delta_i, \delta_i^{-1})$ . Next, introduce  $T := (T_1, \dots, T_\ell)^\top$  such that

$$T_i = \varepsilon_i \sqrt{G_i} - (1 - \varepsilon_i) \sqrt{G_i}, \quad (215)$$

where  $\varepsilon_i \sim \text{Bernoulli}(1/2)$ ,  $(\varepsilon_1, \dots, \varepsilon_\ell)$  are mutually independent and are independent of  $(G_1, \dots, G_\ell)$ . Then  $(T_1, \dots, T_\ell)$  are independent,  $\mathbb{E}[T_i] = 0$  and  $T_i^2 = G_i \sim \gamma(\delta_i, \delta_i^{-1})$ , so that  $\text{Var}(T_i) = \delta_i \delta_i^{-1} = 1$ . Next, define  $\mathbf{D}$  to be the diagonal matrix such that  $\mathbf{D}_{ii} = \lambda_i^{1/2}$  and define  $R_1 := \mathbf{D}T$ . Then  $\mathbb{E}[R_1] = 0$ , and because  $\text{Var}(T_i) = 1$ , we have  $\mathbf{\Sigma} := \text{Cov}(R_1) = \mathbf{D}^2$ , and

$$\|\mathbf{\Sigma}^{-1/2} R_1\|^2 = R_1^\top \mathbf{\Sigma}^{-1} R_1 = \|T\|^2 = \sum_{i=1}^{\ell} T_i^2, \quad (216)$$

$$\det(\mathbf{\Sigma})^{-1/\ell} \|R_1\|^2 = \det(\mathbf{\Sigma})^{-1/\ell} T^\top \mathbf{D}^2 T = \sum_{i=1}^{\ell} \delta_i T_i^2. \quad (217)$$

We now apply Theorem 1.2 from [37], in order to describe the tail behaviour of both random variables above. For this, observe that  $\delta_i T_i^2 \sim \gamma(\delta_i, 1)$ . The Laplace transforms of  $\sum_{i=1}^{\ell} T_i^2$  and  $\sum_{i=1}^{\ell} \delta_i T_i^2$  are then given by

$$\mathcal{L}_1(\omega) := \mathcal{L}\left(\sum_{i=1}^{\ell} T_i^2\right)(\omega) = \prod_{i=1}^{\ell} \frac{1}{(1 + \delta_i^{-1} \omega)^{\delta_i}}, \quad (218)$$

$$\mathcal{L}_2(\omega) := \mathcal{L}\left(\sum_{i=1}^{\ell} \delta_i T_i^2\right)(\omega) = \prod_{i=1}^{\ell} \frac{1}{(1 + \omega)^{\delta_i}}. \quad (219)$$

To apply Theorem 1.2 from [37], we need to identify the abscissa of convergence of  $\mathcal{L}_1$  and  $\mathcal{L}_2$ . The abscissa of convergence of a (probability) measure  $\nu$  over  $[0, \infty)$  is the unique scalar  $\sigma_0 \in \mathbb{R} \cup \{-\infty\}$  such that the integral  $f(z) = \int_{[0, \infty)} e^{-zt} \nu(dt)$  converges for  $\text{Re}(z) > \sigma_0$ , diverges for  $\text{Re}(z) < \sigma_0$  and has a singularity at  $\sigma_0$  [37], p 267. For  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , it is clear from equations (218) and (219) that their abscissa of convergence are given by  $\sigma_1 := -\min_i \delta_i < 0$  and  $\sigma_2 := -1$  respectively. We now apply [37], Theorem 1.2, on  $\mathcal{L}_1$  and  $\mathcal{L}_2$  (we check the conditions for applying this result at the end of the proof). This yields

$$\lim_{t \rightarrow +\infty} \frac{\log \mathbb{P}(\|\mathbf{\Sigma}^{-1/2} R_1\|^2 > t)}{\log \mathbb{P}(\det(\mathbf{\Sigma})^{-1/\ell} \|R_1\|^2 > t)} = \frac{\sigma_1}{\sigma_2} = \frac{-\min_i \delta_i}{-1} = \min_i \delta_i < 1. \quad (220)$$

Indeed, if  $\min_i \delta_i \geq 1$ , then we would have  $\lambda_1 = \dots = \lambda_\ell$  which contradicts our assumption. Because both logarithms above are negative, this yields

$$\log \mathbb{P}(\|\mathbf{\Sigma}^{-1/2} R_1\|^2 > t) > \log \mathbb{P}(\det(\mathbf{\Sigma})^{-1/\ell} \|R_1\|^2 > t) \quad \text{for large } t. \quad (221)$$

We then take the exponential of the equation above, we compare the resulting CDFs (this reverses the ordering w.r.t. “>”) and we compare the inverses of those CDFs (this reverses back the ordering). We finally obtain that for  $\alpha$  small enough,

$$q_{1-\alpha}(\|\mathbf{\Sigma}^{-1/2} R_1\|^2) > q_{1-\alpha}(\det(\mathbf{\Sigma})^{-1/\ell} \|R_1\|^2). \quad (222)$$

In particular, for  $\alpha$  small enough,  $\text{Vol}(\mathcal{E}_\alpha^\infty) > \text{Vol}(\mathcal{B}_\alpha^\infty)$ , which is the announced result. We now check the two technical conditions of [37], Theorem 1.2, on  $\mathcal{L}_2$  (the proof is similar for  $\mathcal{L}_1$ ). First, for  $\lambda > 0$ ,

$$\lambda \log \mathcal{L}_2(\sigma_2 + \lambda) = -\left(\sum_{i=1}^{\ell} \delta_i\right) \lambda \log \lambda \xrightarrow{\lambda \rightarrow 0^+} 0. \quad (223)$$

Second, for all  $0 < \lambda_1 \leq \lambda_2$  sufficiently small,

$$\frac{\mathcal{L}_2(\sigma_2 + \lambda_2)}{\mathcal{L}_2(\sigma_2 + \lambda_1)} = \prod_{i=1}^{\ell} \left(\frac{\lambda_1}{\lambda_2}\right)^{-\delta_i} = \left(\frac{\lambda_2}{\lambda_1}\right)^{-\gamma}, \quad (224)$$

with  $\gamma := \sum_{i=1}^{\ell} \delta_i > 0$ . From [37], equation (1.3) and Lemma 3.1, this finishes the proof.  $\square$

*Proof of Lemma B.6.* In this proof, we denote  $m := \mathbb{E}[V_1]$ ,  $\hat{m}_n := n^{-1} \sum_{i=1}^n V_i$  for notational conciseness. We also denote

$$T_i = \mathbf{\Sigma}_\lambda^{-1/2}(V_i - m), \quad \hat{T}_i = \hat{\mathbf{\Sigma}}_{n,\lambda}^{-1/2}(V_i - \hat{m}_n), \quad \nu_n = \frac{1}{n} \sum_{j=1}^n \delta_{T_j}, \quad \hat{\nu}_n = \frac{1}{n} \sum_{j=1}^n \delta_{\hat{T}_j}. \quad (225)$$

Note that we are interested in the quantiles of  $(\|\widehat{T}_1\|^2, \dots, \|\widehat{T}_n\|^2)$ , since  $q_{n,\alpha} = np_{(n_\alpha)}$  and  $np_{i,n} = \|\widehat{T}_i\|^2$ . Below we borrow the empirical process notation,  $\nu_n(f) = n^{-1} \sum_{j=1}^n f(T_j)$ . The proof is carried out by studying the empirical characteristic function. Let  $t \in \mathbb{R}^p$ . From the SLLN,

$$\nu_n(e^{i\langle t, \cdot \rangle}) = \frac{1}{n} \sum_{j=1}^n e^{i\langle t, T_j \rangle} \xrightarrow{a.s.} \mathbb{E}[e^{i\langle t, T_1 \rangle}]. \quad (226)$$

Thus, for all  $t \in \mathbb{R}^p$ ,  $\nu_n(e^{i\langle t, \cdot \rangle}) \rightarrow \mathbb{E}[e^{i\langle t, T_1 \rangle}]$  a.s.. We then prove below that thanks to the continuity of the characteristic functions and the separability of  $\mathbb{R}^p$ , we can interchange the “a.s.” and “ $\forall t \in \mathbb{R}^p$ ” to obtain that almost surely,

$$\forall t \in \mathbb{R}^p, \quad \nu_n(e^{i\langle t, \cdot \rangle}) \xrightarrow{n \rightarrow \infty} \mathbb{E}[e^{i\langle t, T_1 \rangle}]. \quad (227)$$

Let us prove this fact. Set  $\mathcal{D} = \mathbb{Q}^p$ , which is a countable dense subset of  $\mathbb{R}^p$ . From equation (226), for all  $t \in \mathcal{D}$ , there exists  $\Omega_t$  with  $\mathbb{P}(\Omega_t) = 1$  such that for all  $\omega \in \Omega_t$ ,  $\nu_n(e^{i\langle t, \cdot \rangle}) \xrightarrow{a.s.} \mathbb{E}[e^{i\langle t, T_1 \rangle}]$ . We set  $\Omega' := \cap_{t \in \mathcal{D}} \Omega_t$ , which has probability one and aim at proving that for all  $\omega \in \Omega'$ , equation (227) holds. Let  $\omega \in \Omega'$ ,  $t \in \mathbb{R}^p$ ,  $\varepsilon > 0$  and  $t' \in \mathcal{D}$  to be chosen later. We can write

$$\left| \nu_n(e^{i\langle t, \cdot \rangle}) - \mathbb{E}[e^{i\langle t, T_1 \rangle}] \right| \leq \left| \nu_n(e^{i\langle t, \cdot \rangle}) - \nu_n(e^{i\langle t', \cdot \rangle}) \right| + \left| \nu_n(e^{i\langle t', \cdot \rangle}) - \mathbb{E}[e^{i\langle t', T_1 \rangle}] \right| + \left| \mathbb{E}[e^{i\langle t', T_1 \rangle}] - \mathbb{E}[e^{i\langle t, T_1 \rangle}] \right|.$$

In the last term above,  $t'$  can be chosen such that this term is smaller than  $\varepsilon/3$  from the continuity of  $t \mapsto \mathbb{E}[e^{i\langle t, T_1 \rangle}]$  and the density of  $\mathcal{D}$ , and the second term is smaller than  $\varepsilon/3$  for all  $n$  large enough, since equation (226) applies for  $t' \in \mathcal{D}$ . Finally, for the first term,

$$\left| \nu_n(e^{i\langle t, \cdot \rangle}) - \nu_n(e^{i\langle t', \cdot \rangle}) \right| \leq \frac{1}{n} \sum_{j=1}^n \left| e^{i\langle t, T_j \rangle} - e^{i\langle t', T_j \rangle} \right| = \frac{2}{n} \sum_{j=1}^n \left| \sin \left( \frac{\langle t - t', T_j \rangle}{2} \right) \right| \leq \left( \frac{1}{n} \sum_{j=1}^n \|T_j\| \right) \|t - t'\|. \quad (228)$$

From the SLLN, the sum above is bounded by some  $M_\omega > 0$  for  $n$  large enough, and thus  $t'$  can also be chosen (as a function of  $\omega$ ) such that this term is smaller than  $\varepsilon/3$ . We have thus proved that almost surely, equation (227) holds.

We now prove that equation (227) also holds for  $\widehat{\nu}_n$ , and work similarly as above. For this, observe that for all  $t \in \mathbb{R}^p$ ,

$$\left| \widehat{\nu}_n(e^{i\langle t, \cdot \rangle}) - \nu_n(e^{i\langle t, \cdot \rangle}) \right| \leq \frac{1}{n} \sum_{j=1}^n \left| e^{i\langle t, \widehat{T}_j \rangle} - e^{i\langle t, T_j \rangle} \right| = \frac{2}{n} \sum_{j=1}^n \left| \sin \left( \frac{\langle t, \widehat{T}_j - T_j \rangle}{2} \right) \right| \leq \frac{\|t\|}{n} \sum_{j=1}^n \|\widehat{T}_j - T_j\|. \quad (229)$$

Furthermore, using the matrix operator norm such that  $\|Mx\| \leq \|M\|_{\text{Op}} \|x\|$ ,

$$\begin{aligned} \|\widehat{T}_j - T_j\| &= \left\| \left( \Sigma_\lambda^{-1/2} - \widehat{\Sigma}_{n,\lambda}^{-1/2} \right) V_j - \Sigma_\lambda^{-1/2} m + \widehat{\Sigma}_{n,\lambda}^{-1/2} \widehat{m}_n \right\| \\ &\leq \left\| \Sigma_\lambda^{-1/2} - \widehat{\Sigma}_{n,\lambda}^{-1/2} \right\|_{\text{Op}} \|V_j\| + \left\| \Sigma_\lambda^{-1/2} m - \widehat{\Sigma}_{n,\lambda}^{-1/2} \widehat{m}_n \right\|. \end{aligned}$$

From equation (167), the continuous mapping theorem and the SLLN, for all  $t \neq 0$  (the case  $t = 0$  is trivial),

$$\frac{1}{\|t\|} \left| \widehat{\nu}_n(e^{i\langle t, \cdot \rangle}) - \nu_n(e^{i\langle t, \cdot \rangle}) \right| \leq \left\| \Sigma_\lambda^{-1/2} - \widehat{\Sigma}_{n,\lambda}^{-1/2} \right\|_{\text{Op}} \left( \frac{1}{n} \sum_{j=1}^n \|V_j\| \right) + \left\| \Sigma_\lambda^{-1/2} m - \widehat{\Sigma}_{n,\lambda}^{-1/2} \widehat{m}_n \right\| \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (230)$$

Note that for  $\widehat{\Sigma}_{n,\lambda}^{-1/2} \rightarrow \Sigma_\lambda^{-1/2}$  to hold a.s., we either require that  $\lambda > 0$  or  $\min \text{Spec} \Sigma > 0$ , so that we can apply the same arguments as those at the end of the proof of Proposition 4.1. Also, the a.s. convergence of equation (230) does not depend on  $t$ , as the bound on the right-hand side of equation (230) does not depend on  $t$ . Hence, almost surely,

$$\forall t \in \mathbb{R}^p, \quad \left| \widehat{\nu}_n(e^{i\langle t, \cdot \rangle}) - \nu_n(e^{i\langle t, \cdot \rangle}) \right| \xrightarrow{n \rightarrow \infty} 0. \quad (231)$$

Combining equation (231) with equation (227), we finally obtain that almost surely,

$$\forall t \in \mathbb{R}^p, \quad \widehat{\nu}_n(e^{i\langle t, \cdot \rangle}) \xrightarrow{n \rightarrow \infty} \mathbb{E}[e^{i\langle t, T_1 \rangle}]. \quad (232)$$

Applying Lévy's theorem sample by sample for  $\omega \in \Omega'$ , we deduce that almost surely, in the Prokhorov metric for the weak convergence of measures over  $\mathbb{R}^p$ ,

$$\hat{\nu}_n \xrightarrow[n \rightarrow \infty]{} \nu_{T_1}. \quad (233)$$

Above,  $\nu_{T_1}$  is the probability measure of  $T_1$ , defined over  $\mathbb{R}^p$ . Applying pushforward integration with the continuous map  $N(x) = \|x\|^2$ , almost surely (in the Prokhorov metric for probability measures over  $\mathbb{R}$ ),

$$N_{\#}\hat{\nu}_n = \frac{1}{n} \sum_{j=1}^n \delta_{\|\hat{T}_j\|^2} \xrightarrow[n \rightarrow \infty]{} N_{\#}\nu_{T_1} =: \nu_{\|T_1\|^2}. \quad (234)$$

Hence, almost surely, for all  $t \in \mathbb{R}$  which is a continuity point of  $F_{\|T_1\|^2}$ , where  $F_{\|T_1\|^2}$  denotes the CDF of  $\|T_1\|^2$ ,

$$\hat{F}_n(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}(\|\hat{T}_j\|^2 \leq t) \xrightarrow[n \rightarrow \infty]{} F_{\|T_1\|^2}(t). \quad (235)$$

We now denote  $Q_{\|T_1\|^2}$  the quantile function of  $\|T_1\|^2$  and  $\hat{Q}_n$  the empirical quantile function of  $(\|\hat{T}_1\|^2, \dots, \|\hat{T}_n\|^2)$ . From equation (235) and [47], Lemma 21.2, we have that, almost surely, for all  $\beta \in (0, 1)$  which is a continuity point of  $Q_{\|T_1\|^2}$ ,

$$\hat{Q}_n(\beta) = \hat{F}_n^{-1}(\beta) \xrightarrow[n \rightarrow \infty]{} Q_{\|T_1\|^2}(\beta). \quad (236)$$

In particular, this holds for  $\beta = 1 - \alpha$ , since we have assumed that  $1 - \alpha$  was a continuity point of  $Q_{\|T_1\|^2}$ . To finish, we need to show that equation (236) also holds for the order statistics  $q_{n,\alpha} = \hat{Q}_n(\frac{n+1}{n}(1 - \alpha))$ . For this, from the assumptions, let  $\mathcal{V}$  be a neighbourhood of  $1 - \alpha$  such that  $Q_{\|T_1\|^2}$  is continuous on  $\mathcal{V}$ , and let  $M > 0$  be such that  $1 - \alpha + 1/M \in \mathcal{V}$ . There exists  $n$  large enough such that  $\frac{n+1}{n}(1 - \alpha) \leq 1 - \alpha + 1/M$ . Hence, from the monotony of the empirical quantile, we have that for all such  $n$ ,

$$\hat{Q}_n(1 - \alpha) \leq \hat{Q}_n\left(\frac{n+1}{n}(1 - \alpha)\right) \leq \hat{Q}_n(1 - \alpha + 1/M). \quad (237)$$

From equation (236) and the continuity of  $Q_{\|T_1\|^2}$  at both  $1 - \alpha$  and  $1 - \alpha + 1/M$ , almost surely,

$$Q_{\|T_1\|^2}(1 - \alpha) \leq \liminf_n \hat{Q}_n\left(\frac{n+1}{n}(1 - \alpha)\right) \leq \limsup_n \hat{Q}_n\left(\frac{n+1}{n}(1 - \alpha)\right) \leq Q_{\|T_1\|^2}(1 - \alpha + 1/M).$$

Since this equation holds for all  $M > 0$  large enough, we have that  $\liminf_n \hat{Q}_n(\frac{n+1}{n}(1 - \alpha)) = \limsup_n \hat{Q}_n(\frac{n+1}{n}(1 - \alpha))$ , hence the sequence  $(\hat{Q}_n(\frac{n+1}{n}(1 - \alpha)))_n$  is convergent and almost surely,

$$q_{n,\alpha} = \hat{Q}_n\left(\frac{n+1}{n}(1 - \alpha)\right) \xrightarrow[n \rightarrow \infty]{} Q_{\|T_1\|^2}(1 - \alpha) = q_{1-\alpha}^{\mathcal{E}}. \quad (238)$$

This finishes the proof of equation (103). We now prove equation (104). For this, introduce  $m^X = \mathbb{E}[X_1]$ ,  $\hat{m}_n^X := n^{-1} \sum_{j=1}^n X_j$ , and

$$U_i := (\Sigma_{\lambda}^{11})^{-1/2}(X_i - m^X), \quad \hat{U}_i := (\hat{\Sigma}_{n,\lambda}^{11})^{-1/2}(X_i - \hat{m}_n^X), \quad (239)$$

$$K_i := \begin{pmatrix} T_i \\ U_i \end{pmatrix}, \quad \hat{K}_i := \begin{pmatrix} \hat{T}_i \\ \hat{U}_i \end{pmatrix}, \quad \mu_n = \frac{1}{n} \sum_{j=1}^n \delta_{K_j}, \quad \hat{\mu}_n = \frac{1}{n} \sum_{j=1}^n \delta_{\hat{K}_j}. \quad (240)$$

We quickly show why we still have that almost surely,

$$\forall t \in \mathbb{R}^{p+k}, \quad \hat{\mu}_n(e^{i\langle t, \cdot \rangle}) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[e^{i\langle t, K_1 \rangle}]. \quad (241)$$

First, equation (227) still holds when replacing  $T_i$  with  $K_i$  because the  $K_i$  remain iid, and the separability of  $\mathbb{R}^{p+k}$  and the continuity of  $t \mapsto \mathbb{E}[e^{i\langle t, K_1 \rangle}]$  still hold : that is,

$$\forall t \in \mathbb{R}^{p+k}, \quad \mu_n(e^{i\langle t, \cdot \rangle}) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[e^{i\langle t, K_1 \rangle}]. \quad (242)$$

We now show that equation (231) still holds when replacing  $T_i$  with  $K_i$  and  $\widehat{T}_i$  with  $\widehat{K}_i$ . We start from the bound

$$\|\widehat{K}_j - K_j\| = \sqrt{\|\widehat{T}_i - T_i\|^2 + \|\widehat{U}_i - U_i\|^2} \leq \|\widehat{T}_i - T_i\| + \|\widehat{U}_i - U_i\|. \quad (243)$$

This bound implies that (see equation (229)) for all  $t \in \mathbb{R}^{p+k}, t \neq 0$ ,

$$\frac{1}{\|t\|} \left| \widehat{\mu}_n(e^{i\langle t, \cdot \rangle}) - \mu_n(e^{i\langle t, \cdot \rangle}) \right| \leq \frac{1}{n} \sum_{j=1}^n \|\widehat{K}_j - K_j\| \leq \frac{1}{n} \sum_{j=1}^n \|\widehat{T}_j - T_j\| + \frac{1}{n} \sum_{j=1}^n \|\widehat{U}_j - U_j\|. \quad (244)$$

We can then use the same arguments on the two sums above as the ones that lead to equation (230), to obtain that a.s.,

$$\forall t \in \mathbb{R}^{p+k}, \quad \left| \widehat{\mu}_n(e^{i\langle t, \cdot \rangle}) - \mu_n(e^{i\langle t, \cdot \rangle}) \right| \xrightarrow[n \rightarrow \infty]{} 0. \quad (245)$$

Combining equations (242) and (245), we obtain equation (241). Equation (241) implies that in the Prokhorov metric, almost surely,

$$\widehat{\mu}_n \xrightarrow[n \rightarrow \infty]{} \mu_K$$

Using pushforward integration with the continuous map  $(x, y) \mapsto \|x\|^2 - \|y\|^2$ , we obtain again that almost surely, for all  $t \in \mathbb{R}$  which is a continuity point of  $F_{\|T_1\|^2 - \|U_1\|^2}$ , where  $F_{\|T_1\|^2 - \|U_1\|^2}$  denotes the CDF of  $\|T_1\|^2 - \|U_1\|^2$ ,

$$\widehat{F}_n(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}(\|\widehat{T}_j\|^2 - \|\widehat{U}_j\|^2 \leq t) \xrightarrow[n \rightarrow \infty]{} F_{\|T_1\|^2 - \|U_1\|^2}(t). \quad (246)$$

In particular, using the same arguments as for  $q_{n,\alpha}$ , we obtain that a.s.,

$$q''_{n,\alpha} \xrightarrow[n \rightarrow \infty]{} Q_{\|T_1\|^2 - \|U_1\|^2}(1 - \alpha) = q_{1-\alpha}^{\mathcal{F}}. \quad (247)$$

We finish with proving equation (105). For this, observe that for all  $i \in \{1, \dots, n\}$ ,

$$p''_{i,n} \leq p'_{i,n} \leq p''_{i,n} + \max_{1 \leq j \leq n} n(b_X^n)_j^2, \quad (248)$$

hence

$$q''_{n,\alpha} \leq q'_{n,\alpha} \leq q''_{n,\alpha} + \max_{1 \leq j \leq n} n(b_X^n)_j^2. \quad (249)$$

Using the same arguments as in the proof of Proposition 3.4, the assumption that  $\mathbb{E}[\|X_1\|^{4q}] < +\infty$  for some  $q > 1$  implies that

$$\max_{1 \leq j \leq n} n(b_X^n)_j^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (250)$$

Combining equations (247) and (249), and using that a.s. convergence implies convergence in probability, we have that  $q'_{n,\alpha} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} q_{1-\alpha}^{\mathcal{F}}$ , which finishes the proof.  $\square$

**Acknowledgments** The authors would like to sincerely thank Mathieu Riou, Elie Goudout and Franck Barthe for fruitful discussions, and F. Barthe's proof of Proposition 4.7 in particular.

**Funding** This work was funded by the project ROMEO (ANR-21-ASIA-0001) from the ASTRID joint program of the French National Research Agency (ANR) and the French Innovation and Defence Agency (AID).



## References

- [1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. National Bureau of Standards Applied Mathematics Series, No. 55. U. S. Government Printing Office, Washington, DC, 1964.
- [2] A. N. Angelopoulos and S. Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023.
- [3] V. N. Balasubramanian, S.-S. Ho, and V. V. (Eds.). *Conformal Prediction for Reliable Machine Learning: theory, adaptations and applications*. Morgan Kaufmann, 2014.
- [4] J.-B. Bardet and F. Malrieu. Probabilités et statistique pour l’agrégation externe, 2018.
- [5] S. Bates, E. Candès, L. Lei, Y. Romano, and M. Sesia. Testing for outliers with conformal p-values. *Ann. Statist.*, 51(1):149 – 178, 2023.
- [6] S. Braun, L. Aolaritei, M. I. Jordan, and F. Bach. Minimum volume conformal sets for multivariate regression. arXiv preprint arXiv:2503.19068, 2025.
- [7] E. Burnaev and V. Vovk. Efficiency of conformalized ridge regression. In *Conference on Learning Theory*, pages 605–622, 2014.
- [8] S. Chatterjee and A. S. Hadi. Influential observations, high leverage points, and outliers in linear regression. *Statist. Sci.*, 1(3):379–416, 1986. With discussion.
- [9] Y. Chikuse. *Statistics on special manifolds*, volume 174 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2003.
- [10] A. R. Cohen Brandwein and W. E. Strawderman. Minimax estimation of location parameters for spherically symmetric unimodal distributions under quadratic loss. *Ann. Statist.*, 6(2):377–416, 1978.
- [11] P. G. Constantine, E. Dow, and Q. Wang. Active subspace methods in theory and practice: applications to kriging surfaces. *SIAM J. Sci. Comput.*, 36(4):A1500–A1524, 2014.
- [12] D. Devetyarov and I. Nouretdinov. Prediction with confidence based on a random forest classifier. In *Artificial Intelligence Applications and Innovations: 6th IFIP WG 12.5 International Conference, AIAI 2010, Larnaca, Cyprus, October 6-7, 2010. Proceedings 6*, pages 37–44. Springer, 2010.
- [13] J. Diquigiovanni, M. Fontana, and S. Vantini. Conformal prediction bands for multivariate functional data. *J. Multivariate Anal.*, 189:Paper No. 104879, 14, 2022.
- [14] T. D. Downs. Orientation statistics. *Biometrika*, 59:665–676, 1972.
- [15] I. L. Dryden and K. V. Mardia. *Statistical shape analysis with applications in R*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, second edition, 2016.
- [16] R. M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.*, 1:290–330, 1967.
- [17] M. Fontana, G. Zeni, and S. Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.
- [18] F. Gamboa, E. Goudout, I. Henderson, A. Mazoyer, and M. Riou. Conformal prediction for general time series. *Working paper*, 2025.
- [19] F. Gamboa, I. Henderson, A. Mazoyer, and O. Roustant. Conformal prediction on the sobolev space  $h^a(\mu)$ . *Working paper*, 2025.
- [20] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

- [21] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions. Vol. 1.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1994. A Wiley-Interscience Publication.
- [22] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions. Vol. 2.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1995. A Wiley-Interscience Publication.
- [23] C. Johnstone and B. Cox. Conformal uncertainty sets for robust optimization. In L. Carlsson, Z. Luo, G. Cherubin, and K. An Nguyen, editors, *Proceedings of the Tenth Symposium on Conformal and Probabilistic Prediction and Applications*, volume 152 of *Proceedings of Machine Learning Research*, pages 72–90. PMLR, 08–10 Sep 2021.
- [24] Y. Kato, D. M. Tax, and M. Loog. A review of nonconformity measures for conformal prediction in regression. In H. Papadopoulos, K. A. Nguyen, H. Boström, and L. Carlsson, editors, *Proceedings of the Twelfth Symposium on Conformal and Probabilistic Prediction with Applications*, volume 204 of *Proceedings of Machine Learning Research*, pages 369–383, 2023.
- [25] M. G. Kim. Sources of high leverage in linear regression model. *arXiv preprint arXiv:2006.04024*, 2020.
- [26] I. Klebanov, B. Sprungk, and T. J. Sullivan. The linear conditional expectation in Hilbert space. *Bernoulli*, 27(4):2267–2299, 2021.
- [27] R. Kulik and P. Soulier. *Heavy-tailed time series*. Springer, 2020.
- [28] J. Lei, M. G’Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.*, 113(523):1094–1111, 2018.
- [29] J. Lei and L. Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 76:71–96, 2014.
- [30] Z. Liang, M. Sesia, and W. Sun. Integrative conformal p-values for out-of-distribution testing with labelled outliers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2024.
- [31] Z. Lin, S. Trivedi, and J. Sun. Conformal prediction intervals with temporal dependence. *Transactions on Machine Learning Research*, 2022.
- [32] G. Lugosi and M. Matabuena. Uncertainty quantification in metric spaces. *arXiv preprint arXiv:2405.05110*, 2024.
- [33] P. C. Mahalanobis. On the generalized distance in statistics. *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, 80:pp. S1–S7, 2018.
- [34] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate analysis*. Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York-Toronto, 1979.
- [35] S. Messoudi, S. Destercke, and S. Rousseau. Copula-based conformal prediction for multi-target regression. *Pattern Recognition*, 120:108101, 2021.
- [36] S. Messoudi, S. Destercke, and S. Rousseau. Ellipsoidal conformal inference for multi-target regression. In U. Johansson, H. Boström, K. An Nguyen, Z. Luo, and L. Carlsson, editors, *Proceedings of the Eleventh Symposium on Conformal and Probabilistic Prediction with Applications*, volume 179 of *Proceedings of Machine Learning Research*, pages 294–306. PMLR, 24–26 Aug 2022.
- [37] A. Mimica. Exponential decay of measures and Tauberian theorems. *J. Math. Anal. Appl.*, 440(1):266–285, 2016.
- [38] I. Molchanov. *Theory of random sets*, volume 87 of *Probability Theory and Stochastic Modelling*. Springer-Verlag, London, 2017. Second edition of [MR2132405].
- [39] R. J. Muirhead. *Aspects of multivariate statistical theory*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1982.

- [40] K. W. Ng, G.-L. Tian, and M.-L. Tang. *Dirichlet and related distributions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2011. Theory, methods and applications.
- [41] H. Papadopoulos. Inductive conformal prediction: Theory and application to neural networks. In *Tools in artificial intelligence*. Citeseer, 2008.
- [42] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2006.
- [43] Y. Romano, E. Patterson, and E. Candès. Conformalized quantile regression. *Advances in neural information processing systems*, 32, 2019.
- [44] K. Stankevičiūtė, A. M. Alaa, and M. van der Schaar. Conformal time-series forecasting. In *Advances in Neural Information Processing Systems*, volume 34, pages 6216–6228, 2021.
- [45] G. Thurin, K. Nadjahi, and C. Boyer. Optimal transport-based conformal prediction. arXiv preprint arXiv:2501.18991v2, 2025.
- [46] D. E. Tyler. A distribution-free  $M$ -estimator of multivariate scatter. *Ann. Statist.*, 15(1):234–251, 1987.
- [47] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.
- [48] V. Vovk. Conditional validity of inductive conformal predictors. In *Proceedings of the Asian Conference on Machine Learning*, volume 25, pages 475–490, 2012.
- [49] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.
- [50] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a random world*. Springer, 2005.
- [51] S. S. Wilks. Determination of sample sizes for setting tolerance limits. *Ann. Math. Statistics*, 12:91–96, 1941.
- [52] S. S. Wilks. Statistical prediction with special reference to the problem of tolerance limits. *Ann. Math. Statistics*, 13:400–409, 1942.
- [53] C. Xu, H. Jiang, and Y. Xie. Conformal prediction for multi-dimensional time series by ellipsoidal sets. *arXiv preprint arXiv:2403.03850*, 2024.
- [54] H. Zhou and H.-G. Müller. Conformal inference for random objects. arXiv preprint arXiv:2405.00294, 2024.