Received: 29 July 2022 | Revised: 17 September 2022 | Accepted: 11 October 2022

DOI: 10.1002/sta4.516

ORIGINAL ARTICLE

WILEY

Novel whitening approaches in functional settings

Marc Vidal [1,2,3] | Ana M. Aguilera [2]

[1]IPEM, Ghent University, Ghent, Belgium

[2]Department of Statistics and Operations Research, Institute of Mathematics, University of Granada, Granada, Spain

[3]Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Saxony, Germany

Correspondence
Ana M. Aguilera, University of Granada, Department of Statistics and Operations Research, Faculty of Sciences, Campus de Fuentenueva s/n, 18071, Granada, Spain.
Email: aaguiler@ugr.es

Funding information
Ministerio de Ciencia e Innovación, Grant/Award Number: PID2020-113961GB-I00; Methusalem, Vlaamse regering

Whitening is a critical normalization method to enhance statistical reduction via reparametrization to unit covariance. This article introduces the notion of whitening for random functions assumed to reside in a real separable Hilbert space. We compare the properties of different whitening transformations stemming from the factorization of a bounded precision operator under a particular geometrical structure. The practical performance of the estimators is shown in a simulation study, providing helpful insights into their optimization. Computational algorithms for the estimation of the proposed whitening transformations in terms of basis expansions of a functional data set are also provided.

KEYWORDS
correlation operator, cross-covariance operator, functional independent component analysis, Mahalanobis distance, sphering, whitening operator

1 | INTRODUCTION

There are many situations in modern applied sciences where data exhibit intricate structures with patterns difficult to capture using common reduction methods. Centring and whitening (or sphering) are natural preprocessing steps to facilitate the inspection of latent sources going beyond second-order correlations. Whitening is a symmetric spheric transformation based on the factorization of inverse covariance matrices that maps a random variable to orthogonality. Due to the inherent rotational freedom of this transformation, there exist infinitely many possible ways of sphering. A considerable body of literature has emerged developing optimal whitening techniques and its theoretical properties; early works include Johnson (1966), Price and Nicewander (1977) and Li and Zhang (1998). Most notably, Eldar and Oppenheim (2003) introduced a procedure using infinite-dimensional features induced by a positive definite kernel to enhance the goodness of fit of the transformation via mean-squared error evaluation between the sphered and original data. Seghouane and Saad (2014) derived an efficient algorithm for whitening high-dimensional data with a reduced-rank approximation that uses Lanczos vectors. Furthermore, the constriction of rotational freedom to a measure of cross-covariance and cross-correlation between the original and sphered data has been developed to asses the degree of resemblance and compression achieved in the transformation (Garthwaite et al., 2012; Johnson, 1966; Kessy et al., 2018; Price & Nicewander, 1977).

Functional data analysis is a consolidated branch of modern statistics with active research in methodological developments for sampling units modelled as functions, surfaces, images or other similar objects varying over a continuum; see, for example, Horváth and Kokoszka (2012), Hsing and Eubank (2015), Aneiros et al. (2019), Cuevas (2014), Goia and Vieu (2016) and Li et al. (2022) for a survey on recent developments in the field. Here, we develop a natural extension of typical whitening procedures for data in the form of random functions or curves as well as discussing their optimization. As functional data are inherently infinite-dimensional, generalizations of multivariate objects such as inverses become an issue that complicates the extension of certain statistical techniques. In that sense, a whitening transformation corresponds to an inverse problem given that covariance operators are not invertible, thus requiring further regularization or assumptions of finite-dimensional space dependency. As an alternative, the presented approach recasts the range of the covariance operator into a weaker norm to accommodate the definition of a whitening operator and exploit its properties in infinite-dimensional spaces.

The motivation behind this article is to discuss the optimization and the use of whitening. The importance of this transformation lies at the core of invariant coordinate selection (Ilmonen et al., 2012; Tyler et al., 2009) and independent component analysis (Nordhausen & Oja, 2018) as well as their functional counterparts (Archimbaud et al., 2022; Vidal et al., 2021; Virta et al., 2020). Whitening enforces statistical independence and allows certain rotational freedom to enhance the estimation of components with non-Gaussian kurtosis or skewness, which are commonly encountered in economic, neuroscientific and biomechanical data. In the latter case, it has been shown to improve the outcomes of variable selection in function-on-scalar regression (Chen et al., 2016). Recent developments in machine learning also demonstrate the utility of whitening in predictive models for image recognition (Chen et al., 2020).

Extending whitening methods to functional settings encompasses exploiting the richness of the topological features of the data (smoothness, continuity and contiguity) over time or some other domain to obtain more robust whitening representations. Our approach accommodates functional whitening on a general high-dimensional framework in which we have resorted to basis function systems as a staple for its computational implementation. Other alternatives such as kernel regressors can certainly be considered. The proposed whitening procedures might have further potential applications for spatio-temporal functional data when stationary and isotropic assumptions are not satisfied; see, for example, Mateu and Giraldo (2021) and Blake et al. (2022) for a review of current spherical approaches. A spatio-temporal stochastic process is a particular case of a functional variable with values in a Hilbert space of three-argument functions defined on the three-dimensional spatio-temporal domain. Therefore, the theoretical formulation of functional whitening operators is valid in this context, and the estimation from discrete-time-space observations could be done by projection on to the tensor function space generated by three bases (one basis for each argument of the spatio-temporal domain) as proposed in Aguilera-Morillo et al. (2016) and Durbán et al. (2022).

The remainder of the paper is organized as follows: in Sections 2 and 3, we introduce the definition of a whitening transformation in function spaces and provide particular examples of these class of operators. Section 4 discusses the optimality of these transformations, which is further proven in a simulation study in Section 6. Computational details are provided in Section 5. To conclude, in Section 7, we briefly discuss some of the implications of our work and related issues a pre-whitening transformation entails. All technical proofs and additional numerical results are included in Appendix A and Online Materials.

2 | THE WHITENING OPERATOR

Let $\mathbb{H}$ be a separable Hilbert space of functions on a compact interval $T \subset \mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle : \mathbb{H} \times \mathbb{H} \to \mathbb{R}$ and norm $\|\cdot\|$. A functional random variable X taking values in $\mathbb{H}$ with $E\|X\|^2 <$ has a mean function $\mu = E(X)$ and a covariance operator $\Gamma = E[(X-\mu)\otimes(X-\mu)]$, where $f\otimes g$ is the induced tensor product in $\mathbb{H}$ defined by $(g\otimes f)h = \langle h,g\rangle f$. Under these assumptions, it is well known that Γ is positive definite, self-adjoint, trace-class (their eigenvalues are summable) and therefore compact (see, for example, Hsing & Eubank, 2015, § 7.2). Then, Γ admits the spectral representation

$$\Gamma = \sum_{j=1} \lambda_j (\gamma_j \otimes \gamma_j) = \sum_{j=1} \lambda_j P_{\gamma_j}, \tag{1}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ is its set of non-negative eigenvalues converging to zero and $\{\gamma_j\}$ an orthonormal basis of corresponding eigenfunctions. We employ the notation $P_{\gamma_j} = \gamma_j \otimes \gamma_j$ for the projection of $\mathbb{H}$ on to the one-dimensional eigenspace spanned by γ_j. Thorough the text, we assume that Γ has strictly positive eigenvalues; hence, Γ is injective. Furthermore, as Γ is a self-adjoint positive operator, consider there exists the operator $\Gamma^{1/2}$ such that $(\Gamma^{1/2})^2 = \Gamma$.

Definition 1. The whitening operator Ψ transforms a functional variable X into a new element $\mathbb{X} = \Psi(X-\mu)$ with zero mean and covariance operator being exactly the identity inside $\mathbb{H}$.

We note that there is no convention in how mean-centring should be performed, if before or after the transformation. To ease the notation, in the sequel, it will be assumed that $\mu = 0$.

A natural way to produce a whitening operator is via factorization of precision operators (the inverse of Γ), which suggests the expression $\Psi^*\Psi = \Gamma^{-1}$, where Ψ^* is the adjoint of Ψ. A priori, major drawbacks might arise in this context, as the precision operator turns out to be unbounded and, in general, X does not belong to its domain (see, for example, Mas, 2006, § 2.2). One should therefore proceed with care, as even considering certain kinds of regularization, $\Psi(X)$ may not exist in $\mathbb{H}$, that is, $||\Psi(X)||^2 =$. We show in the next section how Ψ can also be correctly validated to lead to a number of whitening operators in the functional setting.

3 | FUNCTIONAL WHITENING PROCEDURES

A way to make the inverse problem more manageable is to consider the subspace of $\mathbb{H}$ induced by the eigenelements of Γ, defined as

$$\mathbb{M} = \left\{ h \in \mathbb{H} : \sum_{j=1} \frac{\langle h,\ \gamma_j \rangle^2}{\lambda_j} < \right\}. \tag{2}$$

The completion of $\mathbb{M}$ is given under the inner product

$$\langle f, g \rangle_{\mathbb{M}} = \sum_{j=1} \lambda_j^{-1} \langle f, \gamma_j \rangle \langle g, \gamma_j \rangle = \langle \Gamma^{1/2\dagger} f, \Gamma^{1/2\dagger} g \rangle \, f,\ \ g \in \mathbb{M}, \tag{3}$$

where $\Gamma^{1/2\dagger}$ is the Moore–Penrose inverse of $\Gamma^{1/2}$ (Hsing & Eubank, 2015, § 3.5.7). The space $\mathbb{M}$ can be seen as a reproducing kernel space, commonly encountered in functional data analysis (Kupresanin et al., 2010). Note that, the range of Γ becomes closed under a weaker norm derived from (3), hence allowing the precision to exist in $\mathbb{M}$. This is admissible while assuming that the decay of the absolute value of the coefficients $\langle h, \gamma_j \rangle$ converges faster to zero relative to the eigenvalues λ_j, which is known as Picard's criterion (Engl et al., 2000, § 2.2). Then, we can use the inner product (3) to construct a space of isotropic functions (i.e., their covariance operator satisfies the identity), so that the space ends up having certain Gaussian appearance in the same sense as a Cameron-Martin space (Bogachev, 1998, p. 44, § 2.4). Recalling Definition 1, "inside" explicitly refers to the closure of the identity under this framework, as the sample paths of $\mathbb{X}$ are then conceived as living at the boundary.

Bearing the above in mind, a whitening operator can be generally defined as a two-step transformation, with representation $\Gamma^{1/2\dagger} : \mathbb{M} \to \mathbb{M}$, restricted to map elements in the range of Γ. Therefore, the other part of the mapping comprises a projection on to the space generated by $\{\gamma_j\}_{j=1}$, so that X becomes entirely determined by the covariance operator Γ before whitening. Unless otherwise stated, in the following sections we reset X to the range space of Γ, so that the proposed whitening transformations map elements of the kind $\sum_{j=1} \langle \cdot, \gamma_j \rangle \gamma_j$. With this, we hereafter define the whitening operator $\Psi_{\gamma \otimes \gamma} = \Gamma^{1/2\dagger}$ whose spectral decomposition can be straightforwardly written as

$$\Psi_{\gamma \otimes \gamma} = \Gamma^{1/2\dagger} = \sum_{j=1} \lambda_j^{-1/2} (\gamma_j \otimes \gamma_j). \tag{4}$$

The above transformation is a direct extension of the popular zero phase component analysis whitening procedure proposed by Bell and Sejnowski (1997). According to the following Proposition, rotational freedom becomes apparent, leading to a family of whitening operators we describe bellow.

Proposition 1. The covariance operator of $\mathbb{X} = \mathcal{U} \Psi_{\gamma \otimes \gamma}(X)$ satisfies the identity in $\mathbb{M}$ for any unitary transformation $\mathcal{U}$ in $\overline{\text{ran}}(\Gamma^{1/2})$, the closure of the range space of $\Gamma^{1/2}$.

A slight modification in (4), alternatively produces the non-symmetric whitening operator

$$\Psi_{\gamma \otimes e} = \sum_{j=1} \lambda_j^{-1/2} (\gamma_j \otimes e_j), \tag{5}$$

where $\{e_j\}$ is a fixed orthonormal basis of $\mathbb{H}$. In fact, here we see the role of operator $\mathcal{U}$ as the agent of sending γ_j to e_j. This functional whitening procedure follows the principles of Friedman (1987), which only considers a single rotation of the covariance matrix eigenvectors.

To further extend the class of whitening operators, one can consider a succinct form of decorrelation by defining the diagonal operator $\mathcal{V} = \text{diag}(\Gamma) = \sum_{k=1} P_{e_k} \Gamma P_{e_k}$, where $P_{e_k} = (e_k \otimes e_k)$. The operator $\mathcal{V}$ is not unique as it depends on an arbitrary orthonormal basis of $\mathbb{H}$. In other words, there is no privileged orthonormal basis on $\mathbb{H}$ to define $\mathcal{V}$, and for each one of them, different operators can be obtained. This way, the notion of standardization in multivariate analysis can be extended to the functional case by the operator $\mathcal{V}^{1/2\dagger}$, where $\mathcal{V}^{1/2\dagger}$ is the Moore–Penrose inverse of $\mathcal{V}^{1/2}$. Further usefulness of this operator will become clear in the following sections.

Two whitening procedures with appealing properties were defined in Kessy et al. (2018) by constraining the arbitrariness of the transformation to inherent autocorrelations. Next, we will suppose that $\mathcal{R} : \mathbb{M} \to \mathbb{M}$ defined as $\mathcal{R} = \mathcal{V}^{1/2\dagger} \Gamma \mathcal{V}^{1/2\dagger}$ is a compact operator, boundedly invertible, with associated singular system $\{\rho_j,\ \varphi_j\}_{j=1}$. The operator $\mathcal{R}$ closely resembles the usual correlation matrix and its spectral decomposition leads to an eigenspace that will be of use in combination to $\mathcal{V}^{1/2\dagger}$. Thus, there is no loss of generality in assuming that

$$\Psi_{\varphi \otimes \varphi} = \mathcal{R}^{1/2\dagger} \mathcal{V}^{1/2\dagger} = \left\{ \sum_{j=1} \rho_j^{-1/2} (\varphi_j \otimes \varphi_j) \right\} \mathcal{V}^{1/2\dagger},$$

or analogously to (5),

$$\Psi_{\varphi\otimes e}=\left\{\sum_{j=1}\rho_j^{-1/2}(\varphi_j\otimes e_j)\right\}\mathcal{V}^{1/2\dagger}$$

satisfies the usual properties of a whitening transformation in the sense of Definition 1. Both operators decline the span of Γ by its diagonal, merging it with the spectral decomposition of $\mathscr{R}$, which leads to a non-symmetric operator. In turn, the operator $\Psi_{\varphi\otimes\varphi}$ is up to permutation or sign changes but also invariant under unitary transformations within the subspace spanned by the eigenvectors of $\mathscr{R}$. Again, we can write $\mathcal{U}\Psi_{\varphi\otimes\varphi}$, where $\mathcal{U}$ denotes a unitary operator in the range of $\mathscr{R}^{1/2\dagger}\mathcal{V}^{1/2\dagger}$.

Triangular factorization of self-adjoint and positive operators might provide us another whitening procedure closely related to the Cholesky decomposition of the precision matrix. Due to a result of Krein (see Theorem 3.4.5 in Balakrishnan, 1976), the usual precision operator is factored as

$$(\mathbb{I}+\Gamma)^{-1}=(\mathbb{I}-\Delta^*)(\mathbb{I}-\Delta),$$

where Δ is a triangular Volterra operator on $\mathbb{H}=L^2(T)$, Δ^* is its adjoint, and $\mathbb{I}$ is the identity operator. The factorization in (3) leads to the whitening operator $\Psi_\Delta=(\mathbb{I}-\Delta)$. Nevertheless, we will restrict ourselves to a common Cholesky decomposition of the precision operator in the finite dimensional setting, as shown in Section 5. Optimal whitening with triangular operators presents further technical difficulties requiring a separate study, in part, because of the great variety of them.

Remark 1. To alleviate numerical instabilities when Picard's condition may not hold, a Tikhonov solution of the kind $(\Gamma+\alpha\mathbb{I})^{1/2\dagger}$ can be used; see, for example, Berrendero et al. (2020). This regularization acts as a smoothed spectral truncation on the lower order eigenelements (small eigenvalues), bounding from below the precision by means of the identity operator and a real parameter α. In this sense, rates of convergence for these transformations could be thought comparable with the ones suggested in Caponera and Panaretos (2022) for the autocorrelation operator. However, a Tikhonov regularization approach does not assure that Definition 1 holds, making it necessary to seek alternative strategies to solve the problem (or to discard the suitability of whitening). Interestingly, the projection of $X\in\mathbb{H}$ on to span$(\{\gamma_j\})$, $j\in\{1,\ldots,q\}$, that is, $X^{(q)}=P_\gamma^{(q)}X$, where q is fixed to minimize $E||X-X^{(q)}||^2$, provides a natural mechanism of regularization by the second-order structure of the variable. This regularization procedure combined with an additional roughness penalty based on the d-order derivative on γ_j yields to a bi-smoothed approximation to whitening (Vidal et al., 2021), which in turn can enhance posterior analyses of low-dimensional structures in high-dimensional settings. Other concomitant methods to solve ill-posed problems for functional data are reported in Kraus and Stefanucci (2019). See Engl et al. (2000) for a general reference on the subject.

4 | OPTIMAL FUNCTIONAL WHITENING

Similarly to the multivariate case, optimality in a functional whitening transformation can be identified in two different ways. The first one corresponds to a problem aiming to find a component wise transformation that is closer to the original curves using some measure of adjustment or resemblance. The second one is related to the ability of the whitening operator to compress the original functional variable and retain the maximum information content.

The adjustment between the original projected and the whitened functional variable is usually measured by the minimization of

$$E\|X-\mathbb{X}\|^2=\mathrm{tr}(\Gamma)-2\mathrm{tr}(\Gamma_{X\mathbb{X}})+\mathrm{tr}(\Gamma_{\mathbb{X}}),$$

where $\Gamma_{X\mathbb{X}}=E(X\otimes\mathbb{X})$ is the cross-covariance operator between X and $\mathbb{X}$. As $2\mathrm{tr}(\Gamma_{X\mathbb{X}})$ is the only dependence between the original and the whitened variable, the minimization problem can be reduced to the maximization of $\mathrm{tr}(\Gamma_{X\mathbb{X}})$. We implicitly assume that $\mathbb{X}$ falls in a space with the usual inner product.

Proposition 2. The whitening operator $\Psi_{\gamma\otimes\gamma}$ is the unique transformation that minimizes the quadratic distance $E\|X-\mathbb{X}\|^2$.

The least-squares problem in Proposition 2, however, is restrictive in the sense that it only allows to quantify the goodness of fit of the whitening transformation without not being further explanatory of correlations or level of compression. For a correlation-based similarity objective, a scale-invariant measure is usually required. Sangalli et al. (2009) introduced an analogue of Pearson's uncentred correlation coefficient for first

derivatives, which might be useful in this context. The proposed measure is invariant under strictly increasing affine transformations, and thus, it might serve to evaluate the resemblance in terms of correlations between both variables. Here, we consider a functional extension of the criteria used in Kessy et al. (2018, § 6· 2), consisting of the minimization of the mean squared error between the standardized functional variable and the whitened one. The operator $\mathcal{V}^{1/2^\dagger}$ scales the original variable without removing correlations, allowing to construct a scale invariant measure without being necessary to compute derivatives. The optimality objective is then expressed as

$$E\|\mathcal{V}^{1/2^\dagger}(X)-\mathbb{X}\|^2=\mathrm{tr}\{\Gamma_{\mathcal{V}^{1/2^\dagger}(X)}\}-2\mathrm{tr}\{\Gamma_{\mathcal{V}^{1/2^\dagger}(X)\mathbb{X}}\}+\mathrm{tr}(\Gamma_{\mathbb{X}}),$$

which corresponds to the maximization of $\mathrm{tr}\{\Gamma_{\mathcal{V}^{1/2^\dagger}(X)\mathbb{X}}\}$.

Proposition 3. The whitening operator $\Psi_{\varphi\otimes\varphi}$ is the unique transformation that minimizes the quadratic distance $E\|\mathcal{V}^{1/2^\dagger}(X)-\mathbb{X}\|^2$.

Robustness in small local changes is not necessarily guaranteed when a whitening transformation is based on minimal least squared adjustment. To measure the degree of compression of a whitening transformation, Kessy et al. (2018) used the row sum of squared cross-covariance and cross-correlations between the components of the whitened and the original vector. Then, a monotonically decreasing condition on the resultant variance is established to be maximized. As one might suspect, a similar approximation can be developed in the functional data context.

First, note that the operators $\Gamma_{X\mathbb{X}}$,$\Gamma_{\mathbb{X}X}$ are not self-adjoint, whereas $\Gamma_{X\mathbb{X}}$ is the adjoint of $\Gamma_{\mathbb{X}X}$. Define then the compound operator $\Gamma_{X\mathbb{X}}\circ\Gamma_{X\mathbb{X}}$, which is self-adjoint and compact. Formerly, a straightforward way to measure how the whitening operator effectively compresses the original functional variable in terms of a cross-covariance relation might be as

$$\sigma_{\mathrm{cov}}=\sup_{\|e_k\|=1}\langle e_k,\Gamma_{\mathbb{X}X}\circ\Gamma_{X\mathbb{X}}(e_k)\rangle.$$

We can similarly proceed for the cross-correlation operator, now defined in the same sense of Kupresanin et al. (2010) as $\mathscr{R}_{X\mathbb{X}}=\Gamma^{1/2^\dagger}\Gamma_{X\mathbb{X}}\tilde{\Gamma}^{1/2^\dagger}$ with $\tilde{\Gamma}=E(\mathbb{X}\otimes\mathbb{X})$. Thus, if the aim is to maximize the compression under a cross-correlation measure, we look for the maximization of the rate

$$\sigma_{\mathrm{corr}}=\sup_{\|e_k\|=1}\langle e_k,\mathscr{R}_{\mathbb{X}X}\circ\mathscr{R}_{X\mathbb{X}}(e_k)\rangle.$$

In our simulation study, we show that the whitening operators $\Psi_{\gamma\otimes e}$ and $\Psi_{\varphi\otimes e}$ maximize the proposed rates of compression.

5 | FINITE DIMENSIONAL APPROXIMATION

To generalize the practical application of our theoretical precepts, in this section, we show how to estimate whitening transformations from a sample of functions approximated with any basis system. Let $X_i\,(i=1,\ldots,n)$ be n independent copies of X not directly observable. The curves X_i are reconstructed from a vector of measurements collected in a finite set of time points $t_{i0}, t_{i1}, \ldots, t_{im_i}$, contaminated with additive independent errors, that is, $X_{ik}=X_i(t_{ik})+\epsilon_{ik}$, $k\in\{0,\ldots,m_i\}$. The observations $X_1,\ldots,X_n$ are assumed in a q-dimensional space $\mathbb{H}^{(q)}$ of $L^2(T)$ spanned by a collection of basis functions $\phi=(\phi_1,\ldots,\phi_q)^\top$ not necessarily orthonormal in the usual sense. For two functions $f=\phi^\top\mathfrak{f}$ and $g=\phi^\top\mathfrak{g}$, the inner product is defined by $\langle f,g\rangle=\mathfrak{f}^\top\mathcal{G}\mathfrak{g}$ where $\mathcal{G}\in\mathbb{R}^{q\times q}$ is the Gram matrix of inner products between pairs of basis functions. Then, X_i can be expressed as the vector valued function $X(t)=A\phi(t)$ where $A\in\mathbb{R}^{n\times q}$ is a matrix of coefficients and $\phi(t)=(\phi_1(t),\ldots\phi_q(t))^\top$.

The q-dimensional sample covariance operator $\Gamma^{(q)}$ is defined for any $f\in\mathbb{H}^{(q)}$ as $\Gamma^{(q)}(f)=\langle C^{(q)}(s,\cdot),f\rangle$ where $C^{(q)}$ is the covariance kernel function of X admitting the following matrix representation

$$\begin{aligned}C^{(q)}(s,t)&=n^{-1}X(s)^\top X(t)\\&=\phi(s)^\top\mathcal{G}^{-1/2}(n^{-1}\mathcal{G}^{1/2}A^\top A\mathcal{G}^{1/2})\mathcal{G}^{-1/2}\phi(t)\\&=\hat{\phi}(s)^\top\Sigma_{A\mathcal{G}^{1/2}}\hat{\phi}(t).\end{aligned}$$

Then, the coordinates of $C^{(q)}$ can be expressed in terms of an orthonormalized basis $\hat{\phi}(t)=\mathcal{G}^{-1/2}\phi(t)$ as $n^{-1}\mathcal{G}^{1/2}A^\top A\mathcal{G}^{1/2}=\Sigma_{A\mathcal{G}^{1/2}}$. The matrix $\Sigma_{A\mathcal{G}^{1/2}}\in\mathbb{R}^{q\times q}$ has the eigendecomposition $\Sigma_{A\mathcal{G}^{1/2}}=U\Lambda U^\top$ where U are eigenvectors and Λ is a diagonal matrix with entries the eigenvalues of $\Sigma_{A\mathcal{G}^{1/2}}$. The eigenfunctions $\{\gamma_j\}$ are then defined as $\gamma(t)=B\phi(t)$ with $B=\mathcal{G}^{-1/2}U$. Further, consider the decomposition $\Sigma_{A\mathcal{G}^{1/2}}=D^{1/2}R_{A\mathcal{G}^{1/2}}D^{1/2}$, where $D=\mathrm{diag}(\Sigma_{A\mathcal{G}^{1/2}})$ and $R_{A\mathcal{G}^{1/2}}$ is a matrix capturing the correlations of $\Sigma_{A\mathcal{G}^{1/2}}$ with eigendecomposition $R_{A\mathcal{G}^{1/2}}=V\Theta V^\top$. With this, we next derive the whitening procedures described in Section 3 for functional data. Normalization is omitted for the sake of clarity.

Proposition 4. Let us consider the orthonormalized basis $e(t) = \hat{\phi}(t)$ of $\mathbb{H}^{(q)}$. Using the functional representation $X(t) = A\mathcal{G}^{1/2}\hat{\phi}_j(t)$, the coefficients of each functional whitening operator proposed in Section 3 are obtained by their respective multivariate whitening procedures of the orthonormalized coefficient matrix $A\mathcal{G}^{1/2}$ as follows:

$$\begin{aligned}
\Psi_{\gamma\otimes\gamma}\{X(t)\} &= (A\mathcal{G}^{1/2})\Sigma^{-1/2}_{A\mathcal{G}^{1/2}}\hat{\phi}(t),\\
\Psi_{\gamma\otimes e}\{X(t)\} &= (A\mathcal{G}^{1/2})U\Lambda^{-1/2}\hat{\phi}(t),\\
\Psi_{\varphi\otimes\varphi}\{X(t)\} &= (A\mathcal{G}^{1/2})D^{-1/2}R^{-1/2}_{A\mathcal{G}^{1/2}}\hat{\phi}(t),\\
\Psi_{\varphi\otimes e}\{X(t)\} &= (A\mathcal{G}^{1/2})D^{-1/2}V\Theta^{-1/2}\hat{\phi}(t),\\
\Psi_{\Delta}\{X(t)\} &= (A\mathcal{G}^{1/2})L\hat{\phi}(t),
\end{aligned}$$

where L is the solution to the Cholesky factorization $LL^{\top} = \Sigma^{-1}_{A\mathcal{G}^{1/2}}$.

6 | NUMERICAL ANALYSIS

Our simulation study compares, on a simple asset, the performance of all functional whitening procedures discussed in Section 3. Specifically, we examine the behaviour and optimization of these estimators using a B-spline basis system under different conditions.

The synthetic data were generated from a Gaussian stochastic process with a quadratic covariance function $\text{cov}(X_i, X_j) = \exp\{-(X_i - X_j)^2/2\ell^2\}$, where the hyperparameter is fixed to $\ell = 15$. Namely, we reproduce a set of random curves Y_{ik}, $i \in \{1, \ldots, n\}; k \in \{1,\ldots,m\}$ on a fine grid of $m = 50$ equally spaced time points in the unit interval and with $n = 180,500,1000$ observations of the form

$$Y_{ik} = \mu(t_{ik}) + X_i(t_{ik}) + \epsilon_{ik}, \epsilon_{ik} \overset{\text{iid}}{\sim} \text{N}(0,0.4),$$

where the mean is defined as $\mu(t) = \sin(4\pi t/m)$. To evaluate accuracy of the resulting estimates, we introduce a measure of interference signal ratio ISR adapted to the case. Denote the latent process as $Y^{\bullet}_{ik} = \mu(t_{ik}) + X_i(t_{ik})$, then, assuming that the whitened curves are mean centred, the proposed ISR objective function minimizes

$$\text{ISR} = n^{-1}\sum_{i=1}^{n}\left[Y^{\bullet}_i - \{\chi_i(t_{ik}) + \mu(t_{ik})\}\right]^2,$$

where $\chi_i = \{\mathcal{V}^{1/2^{\dagger}}\}^{-1}(\mathbb{X}_i)$ is a set of rescaled curves. Notice that we use an inverse scaling regardless of invariant rotations, as the aim is to preserve such inferred information on the de-standardized curves. If the inverse mapping $\Psi^{-1}(\mathbb{X}_i) = X_i$ provides a return to the original curves, the proposed method only reverts the whitening transformation to the original data scale. For real data sets, the ISR measure is not applicable as the profile $X_i(t)$ is unknown.

These simulations have been repeated 1000 times so that for each iteration, a sample of size n is simulated and the five whitening transformations are applied to each of the 1000 samples. Then, the mean on the 1000 iterations is computed for each precision measure. The results for $n = 180$ are shown in Table 1 and Figure 1. For $n = 500$ and 1000, results are comparable and can be found in the Online Materials. The suitability of the five proposed functional whitening procedures is evaluated by all optimization methods discussed in Section 4. The operators $\Psi_{\gamma\otimes\gamma}$, $\Psi_{\varphi\otimes\varphi}$

TABLE 1 Simulation results for the different whitening transformations using a B-spline basis system of dimension $q = \{13,25\}$ and $n = 180$ (mean values for 1000 iterations)

	ISR		$\text{tr}(\Gamma_{\mathbb{X}})$		$\text{tr}\{\Gamma_{\mathcal{V}^{1/2^{\dagger}}(\mathbb{X})}\}$		σ_{cov}		σ_{corr}	
Operator	$\phi^{(13)}$	$\phi^{(25)}$	$\phi^{(13)}$	$\phi^{(25)}$	$\phi^{(13)}$	$\phi^{(25)}$	$\phi^{(13)}$	$\phi^{(25)}$	$\phi^{(13)}$	$\phi^{(25)}$
$\Psi_{\gamma\otimes\gamma}$	0.888	1.307	7.108	7.239	3.580	4.977	5.456	2.685	1.159	1.120
$\Psi_{\gamma\otimes e}$	1.093	1.611	5.322	5.355	2.736	3.746	8.759	8.225	2.214	3.920
$\Psi_{\varphi\otimes\varphi}$	0.875	1.279	7.072	7.214	3.606	5.000	5.162	2.540	1.000	1.000
$\Psi_{\varphi\otimes e}$	1.091	1.615	5.376	5.352	2.760	3.741	8.718	8.168	2.283	3.936
Ψ_{Δ}	0.968	1.392	5.335	5.427	2.779	3.807	3.564	3.116	1.768	2.387

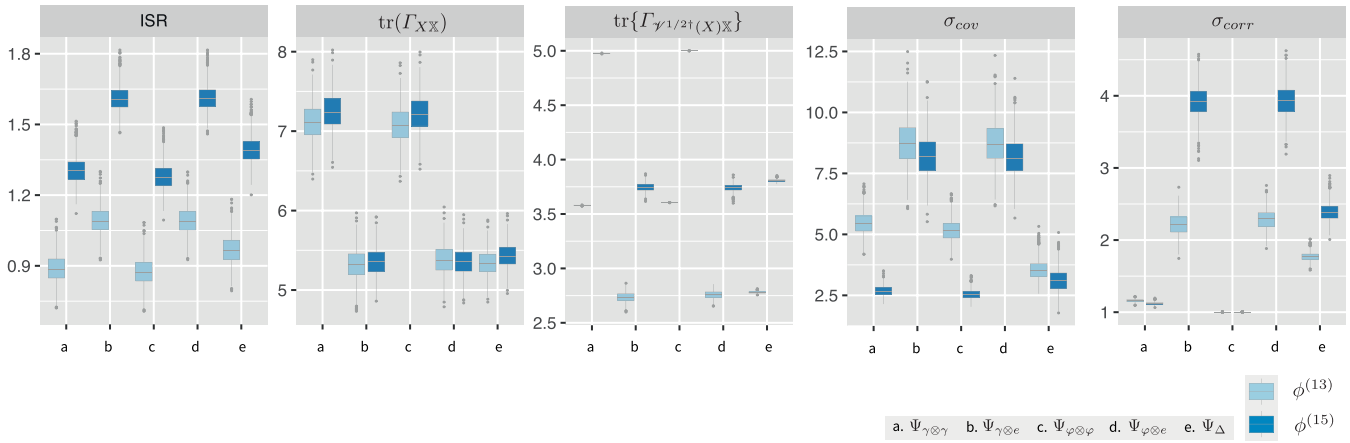

FIGURE 1 Boxplots of the simulations results for $n = 180$ and 1000 iterations

achieve, respectively, maximum cross-covariance and cross-correlation degree of closeness in the mean squared error sense. It is worth mentioning that $\Psi_{\varphi\otimes\varphi}$ attains the ISR minimal fit, supporting the fact that a correlated-based approach can lead to an improved data reduction in terms of robustness. On the other end, the rate of compression is maximized for the operators $\Psi_{\gamma\otimes e}$, $\Psi_{\varphi\otimes e}$ according to their respective cross-covariance and cross-correlation optimization. However, these operators produce low rates for every similarity index and behave as confounding factors that might harm the robustness of the final results. This has to be carefully considered as the performance of a further analysis may be ill-conditioned on the initial procedure. Finally, the operator Ψ_{Δ} takes a reasonably intermediate position for the proposed optimization objectives. These results are parallel to those in Kessy et al. (2018).

7 | DISCUSSION

This article introduces a theoretical framework and a practical estimation algorithm for whitening procedures in function spaces. Despite the extended use of whitening due to, for instance, fast convergence in learning algorithms, some caveats may be worth to comment. In Rousseeuw and Leroy (1987, pp. 271–273), the authors exemplify the problems a pre-whitening step entails, which often causes spurious associations of outliers as inliers and vice versa. This raises the question under what circumstances whitening may and may not be indicated. We do know that whitening enforces statistical independence, an essential property to uncover hidden patterns in the data that are not directly observable with traditional methods. However, losing the correlatedness property of functional data over time carries some consequences: if one could assume Gaussianity in the original space, then no interesting projections in a non-Gaussian sense would be found and rather one would obtain white noise with no practical use. The prior application of robust techniques to detect and remove unusual functional observations with influential behavior (see, for example, Alemán-Gómez et al., 2022; Arribas-Gil & Romo, 2014; Navarro-Esteban & Cuesta-Albertos, 2021) or a functional principal component data reduction, can eventually enhance the suitability of some statistical techniques based on a whitening transformation.

Indeed, the superiority of functional whitening lies in the continuity and the inherent smoothness of the data. Smoothing techniques offer precise control over the noise, which is favorable to presume robust sphericity. Overall, the proposed estimators can improve the applicability of subsequent reduction techniques under certain conditions. A whitening transformation based on a correlation measure performs notably well when the aim is to maximize the similarity with the original curves. However, the statistical interpretability of transformations that maximally compress the empirical functions is ambiguous, and routinely it may not be a reliable solution unless some high-frequency components of X are aimed to be preserved. Some preliminary tests show that, in such cases, regularization via penalized estimators might provide great performance in functional classification.

ACKNOWLEDGEMENTS

We wish to thank Prof. Jasson Vindas for providing valuable comments and suggestions. This research was partially supported by the Mathusalem funding from the Flemish Government, the project FQM-307 of the Government of Andalusia (Spain) and by the project PID2020-113961GB-I00 of the Spanish Ministry of Science and Innovation (also supported by the FEDER programme). We also acknowledge the financial support of the Consejería de Conocimiento, Investigación y Universidad, Junta de Andalucía (Spain) and the FEDER programme for the project A-FQM-

66-UGR20 and the IMAG-María de Maeztu grant CEX2020-001105-M/AEI/10.13039/501100011033. Funding for open access charge: Universidad de Granada / CBUA.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

All R (R Development Core Team, 2020) source code and simulation scripts with the results for the full simulation study in Section 6 are available through the GitHub repository: https://github.com/m-vidal/functional-whitening.

ORCID

Marc Vidal https://orcid.org/0000-0002-1084-3242
Ana M. Aguilera https://orcid.org/0000-0003-2425-6716

REFERENCES

Aguilera-Morillo, M. C., Durbán, M., & Aguilera, A. M. (2016). Prediction of functional data with spatial dependence: A penalized approach. *Stochastic Environmental Research and Risk Assessment*, *31*, 7–22.
Alemán-Gómez, Y., Arribas-Gil, A., Desco, M., Elías, A., & Romo, J. (2022). Depthgram: Visualizing outliers in high-dimensional functional data with application to fMRI data exploration. *Statistics in Medicine*, *41*(11), 2005–2024.
Aneiros, G., Cao, R., Fraiman, R., Genest, C., & Vieu, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*, *170*, 3–9.
Archimbaud, A., Boulfani, F., Gendre, X., Nordhausen, K., Ruiz-Gazen, A., & Virta, J. (2022). ICS for multivariate functional anomaly detection with applications to predictive maintenance and quality control. *Econometrics and Statistics*.
Arribas-Gil, A., & Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, *15*(4), 603–619.
Balakrishnan, A. V. (1976). *Applied functional analysis*. New York: Springer-Verlag.
Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, *37*(23), 3327–3338.
Berrendero, J. R., Bueno-Larraz, B., & Cuevas, A. (2020). On Mahalanobis distance in functional settings. *Journal of Machine Learning Research*, *21*, 1–33.
Blake, L. R., Porcu, E., & Hammerling, D. M. (2022). Parametric nonstationary covariance functions on spheres. *Stat*, *11*(1), e468.
Bogachev, V. I. (1998). *Gaussian measures*. New York: American Mathematical Society.
Caponera, A., & Panaretos, V. M. (2022). On the rate of convergence for the autocorrelation operator in functional autoregression. *Statistics & Probability Letters*, *189*, 109575.
Chen, Y., Goldsmith, J., & Ogden, R. T. (2016). Variable selection in function-on-scalar regression. *Stat*, *5*, 101–88.
Chen, Z., Bei, Y., & Rudin, C. (2020). Concept whitening for interpretable image recognition. *Nature Machine Intelligence*, *2*, 772–782.
Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, *147*, 1–23.
Durbán, M., Lee, D.-J., Aguilera-Morillo, M. C., & Aguilera, A. M. (2022). From spatiotemporal smoothing to functional spatial regression: a penalized approach, *Geostatistical functional data analysis*: John Wiley & Sons, Ltd, pp. 403–423.
Eldar, Y. C., & Oppenheim, A. V. (2003). MMSE whitening and subspace whitening. *IEEE Transactions on Information Theory*, *72*, 1846–1851.
Engl, H., Hanke, M., & Neubauer, A. (2000). *Regularization of inverse problems*. Boston: Kluwer Academic Publishers.
Friedman, J. H. (1987). Exploratory projection pursuit. *Journal of the American Statistical Association*, *82*, 249–266.
Garthwaite, P. H., Critchley, F., Anaya-Izquierdo, K., & Mubwandarikwa, E. (2012). Orthogonalization of vectors with minimal adjustment. *Biometrika*, *99*, 787–798.
Goia, A., & Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *Journal of Multivariate Analysis*, *146*, 1–6.
Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. New York: Springer.
Hsing, T., & Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. Chichester: Wiley.
Ilmonen, P., Oja, H., & Serfling, R. (2012). On invariant coordinate system (ICS) functionals. *International Statistical Review*, *80*, 93–110.
Johnson, R. M. (1966). The minimal transformation to orthonormality. *Psychometrika*, *31*, 61–66.
Kessy, A., Lewin, A., & Strimmer, K. (2018). Optimal whitening and decorrelation. *The American Statistician*, *72*, 309–314.
Kraus, D., & Stefanucci, M. (2019). Classification of functional fragments by regularized linear classifiers with domain selection. *Biometrika*, *106*, 161–180.
Kupresanin, A., Shin, H., King, D., & Eubank, R. L. (2010). An RKHS framework for functional data analysis. *Journal of Statistical Planning and Inference*, *140*(12), 3627–3637.
Lee, K. Y., Li, B., & Zhao, H. (2016). On an additive partial correlation operator and nonparametric estimation of graphical models. *Biometrika*, *103*, 513–530.
Li, G., & Zhang, J. (1998). Sphering and its properties. *Sankhya A*, *60*, 119–133.
Li, Y., Qiu, Y., & Xu, Y. (2022). From multivariate to functional data analysis: Fundamentals, recent developments, and emerging areas. *Journal of Multivariate Analysis*, *188*, 104806.
Mas, A. (2006). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis*, *98*, 1231–1261.
Mateu, J., & Giraldo, R. (2021). *Geostatistical functional data analysis*, Wiley Series in Probability and Statistics: Wiley.
Navarro-Esteban, P., & Cuesta-Albertos, J. A. (2021). High-dimensional outlier detection using random projections. *TEST*, *30*, 908–934.
Nordhausen, K., & Oja, H. (2018). Independent component analysis: A statistical perspective. *WIREs Computational Statistics*, *10*, 1–23.
Price, J. M., & Nicewander, W. A. (1977). Maximally correlated orthogonal composites and oblique factor analytic solutions. *Psychometrika*, *42*, 439–442.

R Development Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org

Rousseeuw, P. J., & Leroy, A. (1987). *Robust regression and outlier detection*. New York: Wiley.

Sangalli, L. M., Secchi, P., Vantini, S., & Veneziani, A. (2009). A case study in exploratory functional data analysis: Geometrical features of the internal carotid artery. *Journal of the American Statistical Association*, *104*, 37–48.

Seghouane, A.-K., & Saad, Y. (2014). Prewhitening high-dimensional fMRI data sets without eigendecomposition. *Neural Computation*, *26*, 907–919.

Tyler, D. E., Critchley, F., Dümbgen, L., & Oja, H. (2009). Invariant co-ordinate selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*, 549–592.

Vidal, M., Rosso, M., & Aguilera, A. M. (2021). Bi-smoothed functional independent component analysis of EEG data. *Mathematics*, *9*, 1243.

Virta, J., Li, B., Nordhausen, K., & Oja, H. (2020). Independent component analysis for multivariate functional data. *Journal of Multivariate Analysis*, *176*, 1–19.

How to cite this article: Vidal, M., & Aguilera, A. M. (2023). Novel whitening approaches in functional settings. *Stat*, *12*(1), e516. https://doi.org/10.1002/sta4.516

APPENDIX A: TECHNICAL PROOFS

We operate with random variables in the range of the operator Γ restricted to its null space (the eigenvalues of Γ are then assumed strictly positive). Under this assumption, we can consider elements of the kind $\sum_{i=1}\langle\cdot,\gamma_i\rangle\gamma_{i,}$ that can be whitened using the bounded operator $\Gamma^{1/2\dagger}$. The necessary and sufficient condition for a whitened functional variable to exist is that

$$\|\mathbb{X}\|^2=\sum_{j=1}\lambda_j^{-1}|\langle\sum_{i=1}\langle X,\ \ \gamma_i\rangle\gamma_{i,}\ \gamma_j\rangle|^2=\sum_{j=1}\frac{|\langle X,\ \gamma_j\rangle|^2}{\lambda_j}<, \tag{A1}$$

where $\mathbb{X}=\Gamma^{1/2\dagger}(X)$ and X is a random element in the range space of Γ we denote by $\mathbb{H}(\Gamma)$. Note that Condition (A1) cannot be reached when $\langle X,\gamma_j\rangle^2=\lambda_j$, or for $c_j\to c>0$, $\langle X,\gamma_j\rangle^2=\lambda_j c_j$ (Mas, 2006), nor in more general spaces since in all these cases integrability breaks down. In addition, the use of certain types of regularization (see Section 7) do not help the proposed definition of functional whitening to be fulfilled. As a result, condition (A1) sets the groundwork that support our next proof.

Proof of Proposition 1. The covariance operator of the whitened variable $\mathbb{X}=\Psi_{\gamma\otimes\gamma}(X)$ can be written as

$$E(\mathbb{X}\otimes\mathbb{X})=E[\Gamma^{1/2\dagger}(X)\otimes\Gamma^{1/2\dagger}(X)].$$

By the tensor product properties $\Gamma^{1/2\dagger}E(X\otimes X)\Gamma^{1/2\dagger *}$, and since $\Gamma^{1/2\dagger}=\Gamma^{1/2\dagger *}$, then

$$\Gamma^{1/2\dagger}\Gamma\Gamma^{1/2\dagger}=P_{\overline{\text{ran}}(\Gamma^{1/2})},$$

where $P_{\overline{\text{ran}}(\Gamma^{1/2})}$ is the projection operator on to the closure of the range space of Γ. Indeed, the operator $P_{\overline{\text{ran}}(\Gamma^{1/2})}$ is compact and it equals to the identity inside $\mathbb{H}$. As long as $\Psi_{\gamma\otimes\gamma}(X)$ has covariance operator the identity it is also invariant under unitary affine transformations. Therefore, for a random element $X\in\mathbb{H}(\Gamma)$ and a location function $b\in\mathbb{H}(\Gamma)$, if we write

$$\begin{aligned}\Gamma^{1/2\dagger}(\mathcal{U}X+b)\ &=E[(\mathcal{U}X+b)\otimes(\mathcal{U}X+b)]^{1/2\dagger}\\&=E[(\mathcal{U}X\otimes\mathcal{U}X)]^{1/2\dagger},\end{aligned}$$

following the same tensor product rule as above, we have that

$$E[\mathcal{U}(X\otimes X)\mathcal{U}^*]^{1/2\dagger}=\mathcal{U}^{1/2\dagger}\Gamma^{1/2\dagger}\mathcal{U}^{1/2\dagger *},$$

This shows that unitary equivariance holds for the operator $\Psi_{\gamma\otimes\gamma} = \Gamma^{1/2\dagger}$ in the closure of the range space of $\Gamma^{1/2}$. □

Proof of Proposition 2. By the tensor product properties, the trace of $\Gamma_{\mathbb{X}\mathbb{X}}$ reduces to

$$\begin{aligned}\mathrm{tr}(\Gamma_{\mathbb{X}\mathbb{X}}) &= \mathrm{tr}\{E(X\otimes\mathbb{X})\}\\ &= \mathrm{tr}\{E(X\otimes X)(\mathcal{U}\Psi_{\gamma\otimes\gamma})^*\}\\ &= \mathrm{tr}\{\mathcal{U}\Gamma^{1/2\dagger}\Gamma\}\\ &= \mathrm{tr}(\mathcal{U}\Gamma^{1/2}),\end{aligned}$$

since $(\mathcal{U}\Psi_{\gamma\otimes\gamma})^* = \mathcal{U}\Psi_{\gamma\otimes\gamma}$. Then, we need to find a unitary transformation that maximizes $\mathrm{tr}(\mathcal{U}\Gamma^{1/2})$. From the Spectral Theorem we have that $\mathcal{U}\Gamma^{1/2} = \mathcal{U}\left(\sum_{j=1}\lambda_j^{1/2}P_{\gamma_j}\right)$. If we write $P_\gamma = \sum_{j=1}P_{\gamma_j}$, then $\mathcal{U}P_\gamma$ continues to converge to the identity for $\mathcal{U}$ being an isometry on $\mathrm{ran}(\Gamma^{1/2})$. This means that $\mathrm{tr}(\mathcal{U}\Gamma^{1/2})$ is maximized at $\mathcal{U} = P_\gamma$. Then, the corresponding optimal whitening operator that minimizes the mean squared error between the original and the whitened functional variable is $P_\gamma\Gamma^{1/2\dagger} = \Psi_{\gamma\otimes\gamma}$. □

Proof of Proposition 3. We can proceed similarly as in Proposition 2 noting that

$$\mathrm{tr}\{E(\mathcal{V}^{1/2\dagger}(X)\otimes\mathbb{X})\} = \mathrm{tr}(\mathcal{U}\mathcal{V}^{1/2\dagger}\Gamma^{1/2}), \tag{A2}$$

Then, from Lee et al. (2016) § 3.1, one can deduce that $\Gamma^{1/2} = (\mathcal{V}^{1/2}\mathscr{R}\mathcal{V}^{1/2})^{1/2}$ and therefore,

$$\begin{aligned}\mathcal{U}\mathcal{V}^{1/2\dagger}\Gamma^{1/2} &= \mathcal{U}\mathcal{V}^{1/2\dagger}(\mathcal{V}^{1/2}\mathscr{R}\mathcal{V}^{1/2})^{1/2}\\ &= \mathcal{U}(\mathcal{V}^{1/2\dagger}\mathcal{V}^{1/2}\mathscr{R}\mathcal{V}^{1/2}\mathcal{V}^{1/2\dagger})^{1/2}\\ &= \mathcal{U}\mathscr{R}^{1/2}.\end{aligned}$$

Similarly as in Proposition 2, it is easy to see that $\mathrm{tr}(\mathcal{U}\mathscr{R}^{1/2})$ is maximized at $\mathcal{U} = P_\varphi$, where P_φ denotes the projection operator on to the space spanned by the eigenfunctions of $\mathscr{R}$. Then, the corresponding optimal whitening operator that minimizes the mean squared error between the original and the standardized functional variable is $P_\varphi\mathscr{R}^{1/2\dagger}\mathcal{V}^{1/2\dagger} = \Psi_{\varphi\otimes\varphi}$. □

Proof of Proposition 4. Given the basis expansion of the functional sample, $X(t) = A\phi(t)$, a general whitened sample admits the basis expansion $\mathbb{X}(t) = \tilde{A}\hat{\phi}(t)$, with coefficients matrix $\tilde{A} = A\mathcal{G}^{1/2}W$ with $WW^{\mathrm{top}} = W^{\mathrm{top}}W = \Sigma^{-1}_{A\mathcal{G}^{1/2}}$. Then, the covariance kernel of the whitened sample is

$$\tilde{C}(s,t) = \hat{\phi}(s)^{\mathrm{top}}W^{\mathrm{top}}\Sigma_{A\mathcal{G}^{1/2}}W\hat{\phi}(t) = \phi(s)^{\mathrm{top}}\mathcal{G}^{-1}\phi(t).$$

Note that when the space is approximated with an orthonormal basis, then $\mathcal{G} = I_q$. As the covariance kernel is Hermitian and positive-definite, the uniformly converging spectral expansions are obtained for both the kernel and its associated operator (Mercer's Theorem). Therefore, from our result, we deduce that the covariance operator of $\mathbb{X}$ is exactly the identity in the topology of the space. That is, $\tilde{C}(f) = f, \forall f \in \mathbb{H}^{(q)}$.

Consider the first whitening procedure $\mathbb{X}(t) = \Psi_{\gamma\otimes\gamma}\{X(t)\} = \Gamma^{1/2\dagger}\{X(t)\}$. Taking into account that the covariance operator can be expressed as $\Gamma\{X(t)\} = A\mathcal{G}^{1/2}\Sigma_{A\mathcal{G}^{1/2}}\hat{\phi}(t)$, we obtain that $\mathbb{X}(t) = A\mathcal{G}^{1/2}\Sigma^{-1/2}_{A\mathcal{G}^{1/2}}\hat{\phi}(t)$.

The proof for the second whitening approach $\mathbb{X}(t) = \Psi_{\gamma\otimes e}\{X(t)\}$ is straightforward by considering the orthonormal basis $e_j(t) = \hat{\phi}_j(t), j \in \{1,\ldots,q\}$.

The third and fourth whitened samples $\mathbb{X}(t) = \Psi_{\varphi\otimes\varphi}\{X(t)\}$ and $\mathbb{X}(t) = \Psi_{\varphi\otimes e}\{X(t)\}$ are respectively obtained as the first and second ones applied to the standardized functional sample $Y(t) = \mathcal{V}^{1/2\dagger}\{X(t)\} = A\mathcal{G}^{1/2}D^{-1/2}\hat{\phi}(t)$.

Finally, the fifth whitening approach is obtained in terms of the Cholesky factorization of the inverse covariance matrix of the coefficients matrix $\Sigma_{A\mathcal{G}^{1/2}}$. □