Efficient computation of low-rank Gaussian process models for surface and image registration

J. Dölz and T. Gerig, M. Lüthi, H. Harbrecht and T. Vetter

Departement Mathematik und Informatik
Fachbereich Mathematik
Universität Basel
CH-4051 Basel

Preprint No. 2017-01
February 2017

www.math.unibas.ch

Efficient Computation of Low-Rank Gaussian Process Models for Surface and Image Registration

Jürgen Dölz and Thomas Gerig[†‡], Marcel Lüthi[‡], Helmut Harbrecht and Thomas Vetter[†‡]

***Abstract*—Gaussian Process Morphable Models (GPMMs) are a unifying approach to non-rigid surface and image registration, where a deformation prior is defined using a Gaussian process. By a simple exchange of the covariance function we can formulate a wide variety of different deformation priors, such as spline-based models, free-form deformations or statistical shape and deformation models. How well the method works in practical applications depends crucially on how well a low-rank approximation of the Gaussian process can be computed. In this article we propose the use of the pivoted Cholesky decomposition for this task. This method makes it possible to efficiently compute a low-rank approximation for very large point sets, such as given by 3D meshes or 3D image grids, with a rigorously controlled approximation error. Compared to the current state of the art, which is based on the Nyström method, the approximation error is controllable and can be specified by a user-defined threshold. Further we propose a computationally more efficient and greedy alternative to currently used Karhunen-Loève expansion. This makes it possible to compute more accurate model approximations at the same computational costs. Detailed experiments from the registration of high quality human face scans and medical CT images containing the forearm with Ulna and Radius demonstrate the efficiency of the method and the computational advantages over the Nyström method.**

I. Introduction

A common approach in medical image analysis and computer vision is analysis by synthesis: An image is analyzed by synthesizing it using a generative model [1], [2]. The resulting model-parameters are then used to understand the content of the target image. Popular examples of analysis by synthesis in medical image analysis are atlas (or template) matching approaches [3], [4], [5], or statistical shape and appearance models [6], [7], [8]. The main idea behind all these methods is the following: It is assumed that any object Γ_T to be analyzed can be written as a deformed version of a reference object $\Gamma_R \subset \Omega$ for some deformation $\mathbf{u}^\star \colon \Omega \to \mathbb{R}^3$:

$$\Gamma_T = \{\mathbf{x} + \mathbf{u}^\star(\mathbf{x}) \colon \mathbf{x} \in \Gamma_R\}.$$

Usually, a non-rigid registration approach is used to find the right deformation $\mathbf{u}^\star$, for a given surface or image representations of Γ_R and Γ_T. The crucial question for practical applications is how to model the family of possible deformations $\mathbf{u}$. Popular choices include spline-based deformations [9], basis-functions derived from differential operators [10], [11], radial basis functions [12], [13], or basis functions learned from example datasets using PCA [14], [5], [7]. Recently, Lüthi et al. proposed to model the deformations as a Gaussian process $\mathcal{GP}(\boldsymbol{\mu}, \mathbf{K})$ with mean function $\boldsymbol{\mu} \colon \Omega \to \mathbb{R}^3$ and covariance (or kernel) function $\mathbf{K} \colon \Omega \times \Omega \to \mathbb{R}^{3\times 3}$ [15], [16], [17]. In this view, all the above mentioned models correspond to special choices of the covariance function and it becomes easy to combine characteristics of the individual models [16], or to incorporate additional prior knowledge by, for example, enforcing mirror symmetries [18], landmark constraints [19], or by making models spatially varying [17]. While in general these Gaussian process priors are infinite dimensional, and thus difficult to use in practical algorithms, Lüthi et al. [15] exploit that all the models make strong smoothness assumptions, and hence the corresponding Gaussian process $\mathbf{u} \sim \mathcal{GP}(\boldsymbol{\mu}, \mathbf{K})$ can be well approximated using a truncated Karhunen-Loève-expansion [20]:

$$\mathcal{GP}(\boldsymbol{\mu}, \mathbf{K}) \sim \mathbf{u} \approx \boldsymbol{\mu} + \sum_{i=1}^{M} \alpha_i \sqrt{\lambda_i} \boldsymbol{\phi}_i,\ \alpha_i \sim \mathcal{N}(0,1),$$

where $(\lambda_i, \boldsymbol{\phi}_i)$ are eigenvalue/eigenfunction pairs of the integral operator associated to the covariance function $\mathbf{K}$. Using this representation, any deformation $\mathbf{u}$ becomes a linear combination of the eigenfunctions $\boldsymbol{\phi}_i$ and can be written as

$$\mathbf{u}(\boldsymbol{\alpha}, \cdot) = \boldsymbol{\mu}(\cdot) + \sum_{i=1}^{M} \alpha_i \sqrt{\lambda_i} \boldsymbol{\phi}_i(\cdot).$$

As in this approach all the modelling assumptions are incorporated into the eigenfunctions $\boldsymbol{\phi}_i$, the accurate computation of these eigenfunctions becomes most crucial. In this work, we propose a mathematically rigorous and computationally efficient method for computing this eigen-decomposition based on the pivoted Cholesky decomposition [21], [22]. The pivoted Cholesky decomposition is ideally suited for this task as in surface and image registration, the natural discretization given by the surface triangulation or the image grid leads to a huge eigenvalue problem to be solved. The pivoted Cholesky decomposition avoids to ever span the full covariance matrix, and thus can work with millions of points for computing the approximation.

Compared to the Nyström method, which was originally proposed for this purpose by Lüthi et al. [15], our method has several advantages. 1) The Nyström method requires the choice of a small subset of points on which the eigen-decomposition is computed [23]. The choice of this point-set greatly influences the approximation accuracy, but for general domains it is not clear how to optimally choose these points. In contrast, our algorithm always selects the optimal points automatically. 2) As there is no rigorous way to control the

[†] Author names in alphabetical order.

[‡] J. Dölz and T. Gerig, M. Lüthi, H. Harbrecht and T. Vetter are with the Department of Mathematics and Computer Science, University of Basel, Switzerland.

approximation error, the Nyström method always computes as many eigenfunctions as the user indicates. In the pivoted Cholesky decomposition, the basis functions are computed iteratively, where in every iteration the next basis function is chosen with respect to a computationally efficient pivoting strategy, until a given accuracy threshold is reached. This property makes the approximation error of the analytically defined model fully controllable. 3) Thanks to the lower computational complexity, our approach enables to compute a much larger number of leading eigenfunctions, and thus to approximate models that are not feasible using the Nyström method. This is in particular important in image registration, where the domain of the deformation is large and consequently a higher number of basis functions need to be computed to obtain a good approximation.

The paper is structured as follows: In Section III we describe how to build prior models over deformation fields using Gaussian Process Morphable Models, and how these models can be used for surface and image registration. In Section IV we describe the pivoted Cholesky decomposition and how it can be used to solve large-scale eigenvalue problems. A computationally more efficient greedy basis for the deformation fields is discussed in Section VII. How this theory can be applied to compute prior models for surface and image registration is discussed in Section V and Section VI, where we also compare the approximation quality of models built with the new method and the Nyström method. In the last part of this paper, we show registration experiments on forearm CT images and compare the results of the different models to a state-of-the art registration method.

The core algorithms of the method proposed are implemented in the open-source project **Scalismo** [24].

II. Related Work

The Gaussian Process Morphable Model (GPMM) framework, on which our work is based on, can be seen as the unification of different concepts. On the one hand, statistical shape models (SSM) can be extended with additional flexibility using kernel functions. On the other hand the models are used as statistical priors for surface and image registration. In the context of surface registration most similar is the work by [25], which specifically concentrates on 3D surface registration with an iterative closest point optimization method. In the context of the extension of SSM with additional flexibility, [26] proposed to extend the sample SSM with a smoothness prior. However, the method needs to span the full covariance matrix, which is computationally infeasible on a large set of points. Also the work by Grenander et al. [1] contains similarities to our approach, as they propose to use a basis function representation to span the model space. However, in all these works, the basis functions have to be known analytically [10], or the initial model needs to be of finite rank [27]. In [28] and also in [25] the covariance function is not approximated, which is only feasible for compact kernels with small correlation lengths.

In the context of Gaussian processes and the computation of low-rank approximations to covariance matrices, the pivoted Cholesky decomposition is an established algorithm, cf. e.g. [23], [21], [22], [29]. Having the low-rank approximation at hand, it has been shown in [21] that the eigenpairs of the covariance matrix can be obtained approximately by solving an eigenvalue problem which has the dimension of the rank of the low-rank approximation.

Whereas these works are restricted to the low-rank approximation of matrices, is has been analyzed in [30] how the continous eigenvalue problem can be efficiently discretized and solved by the pivoted Cholesky decomposition by the use of finite elements. In [31] the authors employ the pivoted Cholesky decomposition to compute a low-rank factorization of kernel functions in terms of function skeletons. Since one can add another basis function to the low-rank factorization without recomputing the others, they call the obtained basis the "Newton" basis, in analogy to Newton interpolation. This kernel based approach has been extended in [32] to compute a Karhunen-Loève expansion if radial basis functions are used for the spatial discretization.

III. Gaussian Process Morphable Models

A. Modeling Deformation Priors

Gaussian Process Morphable Models (GPMM), which have been introduced in Lüthi et al. [15], provide the strength to define prior models for registration analytically in advance. Using a matrix-valued Gaussian process, vector fields are defined continuously on a domain $\Omega \subset \mathbb{R}^3$ independent of the discretization. The vector fields act as the non-rigid transformation of the reference object $\Gamma_R \subset \Omega$, which could be any geometric object or grid defined in Ω.

We define a Gaussian process $\mathcal{GP}(\boldsymbol{\mu}, \mathbf{K})$ with mean function $\boldsymbol{\mu}\colon \Omega \to \mathbb{R}^3$ and covariance function $\mathbf{K}\colon \Omega \times \Omega \to \mathbb{R}^{3\times 3}$. Then, any deformation $\mathbf{u}$ sampled from $\mathcal{GP}(\boldsymbol{\mu}, \mathbf{K})$, gives rise to a new surface by warping the reference surface Γ_R:

$$\Gamma = \{\mathbf{x} + \mathbf{u}(\mathbf{x})\colon \mathbf{x} \in \Gamma_R\}.$$

Similar to the PCA representation of a statistical shape model, a Gaussian process $\mathcal{GP}(\boldsymbol{\mu}, \mathbf{K})$ can be represented in terms of an orthonormal set of basis functions $\{\boldsymbol{\phi}_i\}_{i=1}^{\infty}$

$$\mathbf{u}(\mathbf{x}, \boldsymbol{\alpha}) \sim \boldsymbol{\mu}(\mathbf{x}) + \sum_{i=1}^{\infty} \alpha_i \sqrt{\lambda_i} \boldsymbol{\phi}_i(\mathbf{x}),\ \alpha_i \in \mathcal{N}(0,1), \quad (1)$$

where $(\lambda_i, \boldsymbol{\phi}_i)$ are the eigenpairs of the integral operator

$$\mathcal{T}_{\mathbf{K}}\mathbf{f}(\cdot) := \int_{\Omega} \mathbf{K}(\cdot, \mathbf{x})\mathbf{f}(\mathbf{x})\, \mathrm{d}\rho(\mathbf{x}) \quad (2)$$

with $\rho(\mathbf{x})$ denoting a measure. The representation (1) is known as the Karhunen-Loève expansion of the Gaussian process [20]. Since the random coefficients α_i are uncorrelated, the variance of $\mathbf{u}$ is given by the sum of the variances of the individual components. Consequently, the eigenvalue λ_i corresponds to the variance explained by the i-th component. This suggests that, if the λ_i decay sufficiently quickly, we can, instead of (1), use the low-rank approximation

$$\mathbf{u}_M(\mathbf{x}, \boldsymbol{\alpha}) \sim \boldsymbol{\mu}(\mathbf{x}) + \sum_{i=1}^{M} \alpha_i \sqrt{\lambda_i} \boldsymbol{\phi}_i(\mathbf{x}). \quad (3)$$

The resulting model is a finite dimensional, parametric model with M components, similar to a standard statistical model. The expected error of this approximation is given by the tail sum

$$\sum_{i=M+1}^{\infty} \lambda_i. \tag{4}$$

Estimates for the decay of the λ_i exist and show that the tail sum is reasonably small even for small M, provided that the covariance function $\mathbf{K}$ is sufficiently smooth, cf. [33]. Note, however, that there is no restriction that $\mathbf{K}$ needs to be the sample covariance matrix. Any valid positive semi-definite covariance function can be used.

B. Image and Surface Registration

The modeled concepts and the parametric model are then turned into a registration algorithm. We have to define a reference image or 3D surface Γ_R and a target data-set Γ_T. Also we have to define a distance measure $\mathcal{D}$ between the objects. In the surface registration setting, the distance function is often defined as the distance between a point on the reference and its corresponding closest point on the target surface [15]. Together with the distance measure, we can formulate the registration problem as

$$\underset{\mathbf{u}\in\mathcal{F}_\mathbf{K}}{\arg\min}\, \mathcal{D}[\Gamma_R, \Gamma_T, \mathbf{u}] + \eta\|\mathbf{u}\|^2_\mathbf{K}, \tag{5}$$

where $\|\cdot\|_\mathbf{K}$ denotes the norm of the kernel functions's reproducing Hilbert space $\mathcal{F}_\mathbf{K}$ and η is a regularization parameter. Replacing $\mathbf{u}$ by its low-rank approximation $\mathbf{u}_M$ from (3), we can restate the problem in the parametric form

$$\underset{\alpha_1,\dots,\alpha_M}{\arg\min}\, \mathcal{D}\left[\Gamma_R, \Gamma_T, \boldsymbol{\mu} + \sum_{i=1}^{M} \alpha_i\sqrt{\lambda_i}\boldsymbol{\phi}_i\right] + \eta\sum_{i=1}^{M}\alpha_i^2, \tag{6}$$

which can be optimized with common methods, such as gradient descent.

IV. Solving Large Scale Eigenvalue Problems

A. The Nyström Method

Any given covariance function yields an integral operator (2), whose corresponding continuous eigenvalue problem is given by

$$\big(\mathcal{T}_\mathbf{K}\boldsymbol{\phi}_m\big)(\mathbf{x}) = \lambda_m\boldsymbol{\phi}_m(\mathbf{x}). \tag{7}$$

In order to solve the eigenvalue problem numerically, it has to be discretized, i.e., it has to be transformed into a finite dimensional problem

$$\mathbf{C}\boldsymbol{\phi}_{m,N} = \lambda_{m,N}\boldsymbol{\phi}_{m,N} \tag{8}$$

with $\boldsymbol{\phi}_{m,N} \in \mathbb{R}^N$ and $\mathbf{C} \in \mathbb{R}^{N\times N}$. The Nyström method performs this step by discretizing the integral operator $\mathcal{T}_\mathbf{K}$ from (2) by a sampling approach. It relies on some random samples $\mathbf{x}_1, \dots, \mathbf{x}_N$ drawn according to ρ, to approximate

$$\int_\Omega \mathbf{K}(\cdot,\mathbf{x})\mathbf{f}(\mathbf{x})\,\mathrm{d}\rho(\mathbf{x}) \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{K}(\cdot,\mathbf{x}_i)\mathbf{f}(\mathbf{x}_i), \tag{9}$$

cf. [23] and the references therein. Although the estimator can be inaccurate, it comes with minimal assumptions on the measure ρ, which makes it highly attractive for problems with little information. If the reference domain is for example given by a set of vertices, appropriate samples can be drawn from this set.

Evaluating (9) at the sample points and multiplying with N yields, similar to (8), the finite dimensional eigenvalue problem

$$\mathbf{C}_{\text{Nystr}}\boldsymbol{\phi}_{m,N} = N\lambda_m\boldsymbol{\phi}_{m,N} \tag{10}$$

with the matrix

$$\mathbf{C}_{\text{Nystr}} = \big[\mathbf{K}(\mathbf{x}_i,\mathbf{x}_j)\big]_{i,j=1}^{N}$$

and the point values

$$\boldsymbol{\phi}_{m,N} \approx \big[\boldsymbol{\phi}_m(\mathbf{x}_i)\big]_i, \quad i = 1,\dots,N.$$

Combining (2), (7) and (9), the eigenfunctions can then be evaluated at any given point by

$$\boldsymbol{\phi}_m(\mathbf{x}) \approx \lambda_{m,N}^{-1}\sum_{i=1}^{N}\mathbf{K}(\mathbf{x},\mathbf{x}_i)\big(\boldsymbol{\phi}_{m,N}\big)_i. \tag{11}$$

Probabilistic error bounds for the eigenpairs exist and show that the accuracy increases with the number of sample points, cf. [34].

B. Computational Considerations

To make the approximation step from (7) to (8) as accurate as possible, the Nyström approximation needs a large number of samples, i.e. $N \gg 1$, and leads thus a $N \times N$, dense eigenvalue problem. Also the finite element scheme introduced in Section VI-A and most discretization methods for integral operators lead to large, dense eigenvalue problems (8) in order to achieve a high approximation accuracy. Solving a dense, $N\times N$ eigenvalue problem amounts to a complexity of $\mathcal{O}(N^3)$, which is computationally infeasible for $N \gg 1$.

To overcome this problem, the authors of [15] have remarked that only the largest M eigenvalues of (8) and their eigenvectors have to be computed to obtain an approximation to (3). They have thus assembled the eigenvalue problem (8) for a feasible size of N and have computed a predetermined number M of eigenpairs by using a randomized SVD, cf. e.g. [35] and the references therein, in complexity $\mathcal{O}(MN^2)$. While this reduces the complexity of the eigenvalue problem, it is a priori unclear how to choose M in order to have the tail sum (4) under control. Morover, the overall complexity of the problem in memory and computation time is still at least $\mathcal{O}(N^2)$.

For positive semi-definite matrices, as covariance matrices, the pivoted Cholesky decomposition is an efficient tool to reduce the complexity in memory and computation time to $\mathcal{O}(NM^2)$, where M is adaptively chosen during the decomposition to maintain a given approximation accuracy.

C. Dimension Reduction of the Eigenvalue Problem

Although the system matrix of the discrete eigenvalue problem (8) is dense, it still has a low dimensional structure in the sense of a low-rank approximation. In fact, the decay of the eigenvalues of the integral operator (2) has been well investigated in [33], where it has been proven that the eigenvalues satisfy the decay estimate

$$\lambda_m \leq Cm^{-2p/d}.$$

Here, p is some parameter which increases with the smoothness of the kernel. It is therefore evident that there exists a reasonably sized M such that the tail sum (4) is sufficiently small. As the eigenvalues of the discrete eigenvalue problem will have a similar decay, the system matrix $\mathbf{C}$ has a corresponding low-rank approximation of rank $M \ll N$.

Having a low-rank factorization $\mathbf{C} \approx \mathbf{L}_M \mathbf{L}_M^\intercal$ of rank M at hand, one can drastically reduce the dimension of the eigenvalue problem (8) by substituting the low-rank approximation into (8), such that the eigenvalue problem becomes

$$\mathbf{L}_M \mathbf{L}_M^\intercal \mathbf{v}_{m,N} = \lambda_{m,N} \mathbf{v}_{m,N}.$$

Exploiting the fact that $\mathbf{L}_M \mathbf{L}_M^\intercal$ has the same eigenvalues as $\mathbf{L}_M^\intercal \mathbf{L}_M$, we obtain an equivalent eigenvalue problem

$$\mathbf{L}_M^\intercal \mathbf{L}_M \check{\mathbf{v}}_{m,N} = \lambda_{m,N} \check{\mathbf{v}}_{m,N}, \tag{12}$$

which has the reduced dimension $M \ll N$ and can thus be solved by standard eigensolvers for dense matrices. The eigenvectors $\boldsymbol{\phi}_{m,N}$ are then given by

$$\boldsymbol{\phi}_{m,N} \approx \mathbf{v}_{m,N} = \mathbf{L}_M \check{\mathbf{v}}_{m,N}. \tag{13}$$

Thus, given a low-rank approximation $\mathbf{C} \approx \mathbf{L}_M \mathbf{L}_M^\intercal$, the solution of the dense small eigenvalue problem (12) and the computation of the eigenvectors by (13) can be accomplished in complexity $\mathcal{O}(M^3)$ and in $\mathcal{O}(NM^2)$, respectively.

However, in order to maintain the complexity $\mathcal{O}(NM^2)$ for the solution of the eigenvalue problem (8), we have to find a low-rank factorization of $\mathbf{C}$ without fully even assembling it, since this would amount to at least N^2 operations.

D. Low-rank Approximation with the Pivoted Cholesky Decomposition

For general matrices, a low-rank approximation can for example be obtained by fully assembling the matrix and then applying a (randomized) SVD. However, even the assembly has already complexity $\mathcal{O}(N^2)$, which makes the computation and the storage of a low-rank factorization prohibitively expensive for large N. Which is why we aim at a low-rank factorization in the first place. If we invest our a priori knowledge about $\mathbf{C}$, i.e. that it is a positive semi-definite matrix, we can obtain a low-rank approximation for any given rank M by the *pivoted Cholesky decomposition* without fully assembling it. Even more, the algorithm does not only give us a rank M approximation for some given M, we can also prescribe some tolerance in the *trace-norm* for positive semi-definite matrices to which the algorithm automatically finds a low-rank factorization, such that the approximation error is below that tolerance. Especially, it automatically detects a rank M which is required to fullfil that tolerance. The pivoted Cholesky decomposition, cf. [21], [22], is given in Algorithm 1.

Algorithm 1 The pivoted Cholesky decomposition

Input: Matrix $[\mathbf{C}]_{i,j=1}^N$, error tolerance $\varepsilon > 0$
Output: Low-rank approximation $\mathbf{C}_M = \sum_{i=1}^M \boldsymbol{\ell}_i \boldsymbol{\ell}_i^\intercal$ with $\operatorname{trace}(\mathbf{C} - \mathbf{C}_M) \leq \varepsilon \cdot \operatorname{trace}(\mathbf{C})$
Set $M = 1$
Set $\mathbf{d} = \operatorname{diag}(\mathbf{C})$ and $error = \|d\|_{\ell^1}$
Initialize $\boldsymbol{\pi} = [1, 2, \ldots, N]$
while $error > \varepsilon$ **do**
  Set $i = \arg\max\{d_{\pi_j} : j = M, M+1, \ldots, N\}$
  Swap π_M and π_i
  Set $\ell_{M,\pi_M} = \sqrt{d_{\pi_M}}$
  for $M + 1 \leq i \leq N$ **do**
    Compute
    $$\ell_{M,\pi_i} = \left(a_{\pi_M,\pi_i} - \sum_{j=1}^{M-1} \ell_{j,\pi_M} \ell_{j,\pi_i} \right) / \ell_{M,\pi_M}$$
    Update $d_{\pi_i} = d_{\pi_i} - \ell_{M,\pi_M} \ell_{M,\pi_M}$
  end for
  Compute $error = \sum_{i=1}^N d_{\pi_i}$
  Set $M = M + 1$
end while

As can be seen from the algorithm, it is purely algebraic, i.e. it only relies on the access to the specific matrix entries, which, in order to keep computational efficiency, can be computed only when required by the algorithm.

Mathematically, the idea of the algorithm is that for symmetric semi-positive definite matrices, the maximal element always lies on the diagonal and the trace is a norm on semi-positive definite matrices, giving the algorithm a rigorous error estimate. For this purpose, only the diagonal of the covariance matrix has to be precomputed, whereas all other entries only have to be computed if they appear in the low rank factorization. Especially, the algorithm automatically finds a rank M to reach a prescribed accuracy ε in complexity $\mathcal{O}(NM^2)$.

It remains to explain how to obtain a discrete correlation matrix from a Gaussian process.

V. Building Prior Models for Surface Registration

A. Spatial Discretization

For surface registration, the surface is usually represented as a set of vertices. There is thus almost no structure available, which makes the Nyström method decribed in Section IV-A the method of choice. While the usual approach is to sample the points for the discretization from the vertices defining the surface, cf. e.g. [15], [23], the pivoted Cholesky decomposition, under mild smoothness assumptions on the kernel function, enables us to solve the eigenvalue problem (10) easily for millions of points. We can thus compute the eigenfunctions directly on all mesh points and can thus completely avoid the sampling on the surface and its corresponding uncertain error. Consequently, the pivoted Cholesky decomposition gives

a completely deterministic solution, even for more advanced Nyström schemes as in Appendix A.

The next section demonstrates how the pivoted Cholesky decomposition improves the accuracy of the Nyström scheme.

B. Comparison of Generalisations for Surface GPMM

We use a data-set consisting of 39 registered face scans to represent the ground-truth and as target surfaces for the registration. The data-set is registered using the method proposed by [36] which includes additional constraints to cope with artifacts and noise of the raw 3D scans. To establish a fair comparison between the fitting accuracy of the methods, we create an experiment where only the generalisation ability of the models is evaluated. The model for this experiment is defined with a scalar multi-scale B-spline kernel, which is introduced by Opfer [37]. Given a univariate third order B-spline b_3 and the function $\psi(\mathbf{x}) = b_3(x_1)b_3(x_2)b_3(x_3)$, the kernel reads

$$k_{\text{BSp}}(\mathbf{x},\mathbf{y}) = \sum_{j=\underline{l}}^{\overline{j}} \sum_{k\in\mathbb{Z}^d} 2^{2-j}\psi\big(2^j\mathbf{x}-k\big)\psi\big(2^j\mathbf{y}-k\big).$$

According to [37], this results in a valid, positive definite kernel function on multiple scales. In our experiment, we define the levels from $\underline{l} = -5$ to $\overline{l} = -2$ and refer to [15] for details on matrix valued B-spline kernels.

Taking the Nyström method proposed by [15] as a reference, we use the same kernel function to approximate three different parametric models. For the reference we sample an uniform subset of 1000 points and approximate 1000 eigenfunctions, which, due to the expensive interpolation (11) to extend the eigenfunctions to all mesh points, amounts to the borderline of feasibility. With the pivoted Cholesky method we create a model with a similar amount of basis-functions with a tolerance of $\varepsilon = 0.05$ (1200 basis functions) and a more accurate model with a tolerance of $\varepsilon = 0.01$ (2200 basis functions). Especially, we are able to sample on all grid points and can avoid the expensive interpolation of the eigenfunctions. Since the face data-set is registered and thus in correspondence with the reference, we can define a direct projection in the model space as

$$\underset{\alpha_1,\dots,\alpha_M}{\arg\min}\, \mathcal{D}_c\left[\Gamma_R, \Gamma_T, \boldsymbol{\mu} + \sum_{i=1}^{M} \alpha_i\sqrt{\lambda_i}\boldsymbol{\phi}_i\right],$$

where the distance function $\mathcal{D}_c$ consists of the squared euclidean distance of every point $\mathbf{x}$ with its corresponding point on the registered target. This is a least squares problem and the optimal solution for this problem can be computed in closed form solution, which is shown in [38].

The 39 registered human faces are projected into all three models and the average point-to-point distance is measured and illustrated in Figure 1. The Cholesky approximated model with $\varepsilon = 0.05$ performs similar as the Nyström model. However, when the accuracy of the Cholesky model is increased to $\varepsilon = 0.01$, we are, due to the lack of the expensive interpolation, able to approximate more basis functions, which leads to a much smaller error.

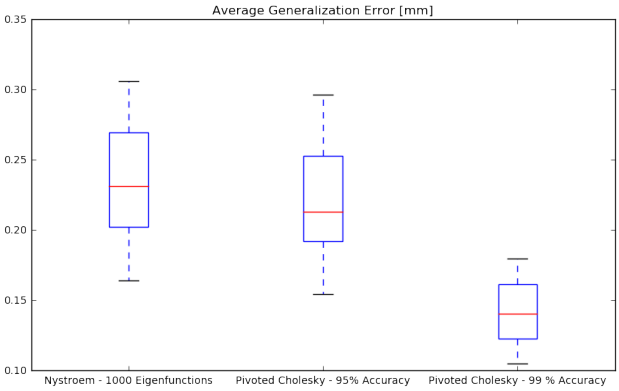

Fig. 1. Comparison of the generalization ability of three differently build models. The error is measured with the distance closest to the target surface in $[mm]$. The GPMM approximated with the Nyström method on 1000 eigenfunctions is comparable with a model built with the pivoted Cholesky method with $\varepsilon = 0.05$ (1200 basis functions). However, the approximate model with $\varepsilon = 0.01$ accuracy (2200 basis functions) is superior and is able to generalize better to the ground truth registration data-set.

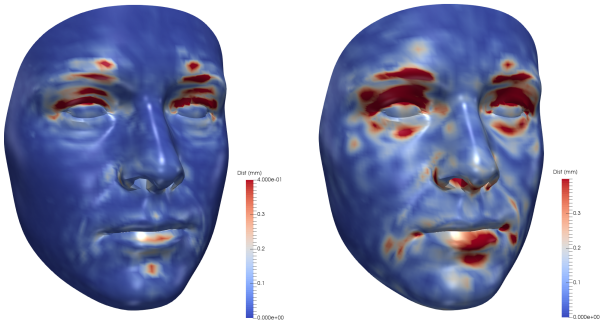

Fig. 2. The average closest point error to the registered ground-truth data-set. On the **left** the error of pivoted Cholesky model with $\varepsilon = 0.01$ is visualized. On the **right** side we see the average distance error with the model approximated with Nyström. The color bar goes from blue (small error) to red (larger error). Especially in the eye and lip region, the large improvement of a more accurate model is visible.

In Figure 2, the average error is visualized on the reference face surface. We can see that the defined Cholesky model with $\varepsilon = 0.01$ error has less regions with large errors than the Nyström model on the right side.

VI. Prior Models for Image Registration

A. Spatial Discretization

In contrast to surface registration, where the computational domain is only given by vertices, the computational domains in image registration provide more structure to exploit for the discretization. We can thus employ more advanced methods to discretize the continuous eigenvalue problem (7) and shall follow the approach of [30] and show how to use the pivoted Cholesky decomposition for the computation of the Karhunen-Loève expansion with finite elements. In principle, a large class of finite element spaces can be treated, cf. [30], but for ease of representation we restrict ourselves to piecewise linear

finite elements on a uniform rectangular grid and the measure $\rho(\mathbf{x}) = 1$ and refer to Appendix B for the more general case.

Finite element schemes for functions with values in three dimensions rely on a finite dimensional subspace $\mathbf{V}_N \subset \left[L^2(\Omega)\right]^3$ with basis $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_N\}$ to represent the eigenfunctions of the Karhunen-Loève expansion. To construct such a finite dimensional space, we consider a uniform rectangular grid $\mathcal{Q}_h$ on Ω where each cell has a size of $h_1 \times h_2 \times h_3$. To each vertex $\mathbf{x}_1, \ldots, \mathbf{x}_N$ we assign a function φ_i with the property

$$\varphi_i(\mathbf{x}_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad i, j = 1, \ldots, N, \tag{14}$$

where on each cell $Q_h \in \mathcal{Q}_h$, the basis function φ_i is a trilinear polynomial, i.e.

$$\begin{aligned} \varphi_i(\mathbf{x})\big|_{Q_h} = a_1 \quad &+ a_2 x_1 \quad + a_3 x_2 \quad + a_4 x_1 x_2 \quad + \\ &a_5 x_3 + a_6 x_1 x_3 + a_7 x_2 x_3 + a_8 x_1 x_2 x_3. \end{aligned}$$

Here, the coefficients are uniquely determined such that (14) holds. This means especially that the φ_i are only nonzero in the eight cells with vertex $\mathbf{x}_i$. Note especially that all φ_i are linearly independent, so we can define $V_h \subset L^2(\Omega)$ as the vector space spanned by the basis $\varphi_1, \ldots, \varphi_N$. A finite dimensional subspace of $\left[L^2(\Omega)\right]^3$ is then given by $\mathbf{V}_h = V_h \times V_h \times V_h$.

Having a finite dimensional subspace at hand and employing a well known mass lumping scheme yields, cf. e.g. [39] and Appendix B, the eigenvalue problem

$$\mathbf{C}_{\text{FEM}} \boldsymbol{\phi}_{m,N} = h^2 \lambda_{m,N} \boldsymbol{\phi}_{m,N} \tag{15}$$

with $h = \sqrt{h_1^2 + h_2^2 + h_3^2}$, system matrix

$$\mathbf{C}_{\text{FEM}} = \left[\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)\right]_{i,j=1}^N$$

and approximate eigenfunctions

$$\boldsymbol{\phi}_m(\mathbf{x}) \approx \boldsymbol{\phi}_{m,N}(\mathbf{x}) = \sum_{i=1}^N \left(\boldsymbol{\phi}_{m,N}\right)_i \boldsymbol{\varphi}_i(\mathbf{x}).$$

As the error estimates for the Nyström approach are not purely deterministic, they do not allow a direct connection between the threshold of the pivoted Cholesky decomposition and the approximation scheme. This is different for finite element schemes, where [30] gives precise error estimates on how to change the threshold if the grid size is refined. In our case, if the threshold is chosen as $\varepsilon \sim h^{\min(p,2)}$, the solution of the eigenvalue problem computed with the pivoted Cholesky decomposition will essentially behave the same as solving the dense eigenvalue problem (15).

B. Spatially Varying Kernel Models

In is experiment we compare the low-rank approximation methods on covariance functions, where the correlation length varies depending on predefined regions in the domain Ω. These type of covariance functions allow for the specification of different kind of smoothness depending on the region. In practice this is especially useful for modeling different tissue properties in medical images. Since the Nyström approximation only computes the eigenfunctions on a uniformly sampled subset of points, it might well approximate coarse correlations, but it is likely to miss small deformation regions if the subset is not densely sampled. As an experimental setup we define a coarse kernel function $k_c(\mathbf{x}, \mathbf{y})$ and also a fine kernel $k_f(\mathbf{x}, \mathbf{y})$, which are both defined as Gaussian kernels with $\sigma = 100$ for k_c and $\sigma = 15$ for k_f. Together with a function $t \colon \Omega \to (0, 1)$, which activates the fine kernel on a predefined region in Ω, we formulate a spatially varying kernel as

$$k_s(\mathbf{x}, \mathbf{y}) = k_c(\mathbf{x}, \mathbf{y}) + t(\mathbf{x}) k_f(\mathbf{x}, \mathbf{y}) t(\mathbf{y}). \tag{16}$$

For this experiment we choose average sized volume as the image domain Ω and place a 3D reference surface inside this domain. We approximated a the previously mentioned kernels on the surface in the volume with $\varepsilon = 0.01$. Random samples of this model act as the ground-truth for our experiment since the image deformation models should provide exactly the same flexibility in the region of the surface. Therefore we define t as the region around the segmented ground-truth surface of the reference image. We built the model with the Nyström method by uniformly sampling 1000 points on the domain Ω. The second model is built with the pivoted Cholesky method and a tolerance of $\varepsilon = 0.01$. To evaluate the flexibility of the two volume models, we compare how much detail is approximated in the finest region where k_f is active. We compare the two models by projecting the random samples of the surface model into the two volume models and calculate the remaining distance error. Since the Nyström model sampled a subset of 1000 points its parametric model rank is fixed 1000. In contrast to the pivoted Cholesky method, the Nyström method lacks the ability to automatically select the right number of basis functions to approximate the given kernel well. The pivoted Cholesky method automatically selected 3002 basis functions to approximate the kernel. In Figure 3 the remaining error of both models is compared. The Nyström method seems to approximate only the coarse correlations. On the other hand, the pivoted Cholesky method approximates these regions with the tolerance of $\varepsilon = 0.01$ error well.

VII. A Simple Greedy Basis

The conversion of the infinite dimensional optimization problem in the reproducing kernel Hilbert space from (5) to a finite dimensional optimization problem in (6) is historically motivated by a PCA of the prescribed kernel function, which leads to the Karhunen-Loève expansion in the continuous case. This automatically captures the most significant features of the infinite dimensional problem into the finite dimensional problem.

However, the Karhunen-Loève expansion has more structure than actually needed for the optimization in (6). In the following, we consider the Karhunen-Loève expansion computed by the pivoted Cholesky decomposition as described in Chapter IV-C and assume that the Karhunen-Loève expansion is not further truncated for the optimization.

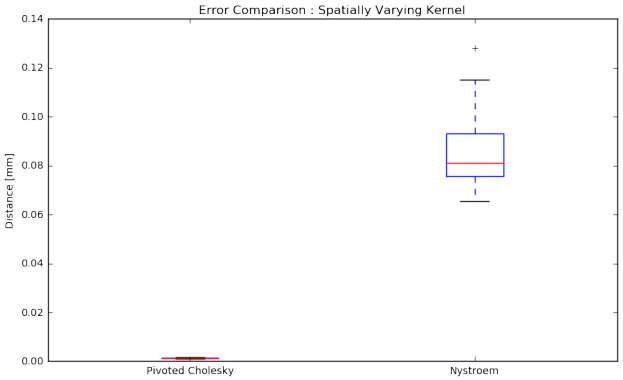

Fig. 3. In this Figure a comparison between two model approximations is shown with a spatially-varying kernel. Both models were approximated on the the same volume and the same kernel function. 50 deformation samples from a well approximated ground-truth model are projected in Nyström model and also in the pivoted Cholesky model. We can see that the pivoted Cholesky generalizes well to these samples in contrast to the Nyström model.

Denoting the i-th eigenfunction discretized by N degrees of freedom obtained by the pivoted Cholesky decomposition by $\boldsymbol{\phi}_{i,N}$, we have

$$\mathbf{K}(\mathbf{x},\mathbf{y}) \approx \sum_{i=1}^{M} \lambda_{i,N}\boldsymbol{\phi}_{i,N}(\mathbf{x})\boldsymbol{\phi}_{i,N}(\mathbf{y}).$$

Abbreviating $\boldsymbol{\Sigma}_{N,M} = \operatorname{diag}(\lambda_{1,N},\dots,\lambda_{M,N})$ and

$$\boldsymbol{\Phi}_{N,M}(\mathbf{x}) = \left[\boldsymbol{\phi}_{1,N}(\mathbf{x})\middle|\dots\middle|\boldsymbol{\phi}_{M,N}(\mathbf{x})\right],$$

the Karhunen-Loève expansion can be written in matrix notation as

$$\mathbf{K}(\mathbf{x},\mathbf{y}) \approx \boldsymbol{\Phi}_{N,M}(\mathbf{x})\boldsymbol{\Sigma}_{N,M}\boldsymbol{\Phi}_{N,M}(\mathbf{y})^{\mathsf{T}}.$$

By associating (13) with its corresponding functions, we deduce that

$$\boldsymbol{\Phi}_{N,M}(\mathbf{x}) = \mathbf{L}_M(\mathbf{x}) \underbrace{\left[\check{\mathbf{v}}_{1,N}\middle|\dots\middle|\check{\mathbf{v}}_{M,N}\right]}_{=:\check{\boldsymbol{\Phi}}_{N,M}}.$$

It thus holds

$$\mathbf{K}(\mathbf{x},\mathbf{y}) \approx \mathbf{L}_M(\mathbf{x})\check{\boldsymbol{\Phi}}_{N,M}\boldsymbol{\Sigma}_{N,M}\check{\boldsymbol{\Phi}}^{\mathsf{T}}_{N,M}\mathbf{L}^{\mathsf{T}}_M(\mathbf{y}),$$

which effectively means that the approximate eigenfunctions of the Karhunen-Loève expansion are orthogonal transformations of the Newton basis $\mathbf{L}_M$ obtained by the pivoted Cholesky decomposition. The spanned space of the singular vectors of the M largest eigenvalues and the M first columns of the pivoted Cholesky decomposition are thus the same. Omitting the rescaling of the Newton basis by the singular values and denoting the function associated with the vector $\boldsymbol{\ell}_i$ by $\boldsymbol{\ell}_{i,N}$, one can rewrite the expansion (3) as

$$\mathbf{u}_M(\mathbf{x},\boldsymbol{\alpha}) \sim \boldsymbol{\mu}(\mathbf{x}) + \sum_{i=1}^{M} \alpha_i \boldsymbol{\ell}_{i,N}(\mathbf{x}),$$

see also [30], and the optimization (6) as

$$\underset{\alpha_1,\dots,\alpha_M}{\arg\min}\, \mathcal{D}\left[\Gamma_R, \Gamma_T, \boldsymbol{\mu} + \sum_{i=1}^{M} \alpha_i \boldsymbol{\ell}_{i,N}\right] + \eta \sum_{i=1}^{M} \alpha_i^2.$$

We can thus directly work with the *Newton basis* given by the column vectors of the low-rank approximation of the pivoted Cholesky decomposition, cf. [31], and totally omit the solution of any eigenvalue problems.

The Newton basis offers some new interesting possibilities, since one can easily expand the basis if a higher accuracy is needed. If the basis vectors need to be orthonormal, one can apply an orthonormalization method like the Gram-Schmidt algorithm, cf. e.g. [40].

VIII. Image Registration Experiments

A. Experimental Setup

In this registration experiment, we mainly evaluate the following points: 1) We compare the prior models approximated with the proposed method to the currently used Nyström method. 2) The registration results of the proposed models are compared to a state-of-the-art multi-scale B-spline registration algorithm, which is implemented in Elastix [41]. 3) We compare the eigenbasis model to the proposed Newton basis model and show that the resulting registration error is comparable. All the registrations are done with the same data-set consisting of 27 CT images of the human forearm. The surface Ulna and the Radius have been manually segmented by experts to provide a ground-truth measure. The data has been rigidly aligned to an arbitrary data-set using four landmarks and are provided in a resolution of $800 \times 800 \times 500$. To establish a fair comparison, we chose the same kernel function to define the deformation prior as in [15] that performed best in their experiments. To evaluate the accuracy of our experiments, we computed the average squared distance error of the registered result to the ground-truth segmentations of the provided 27 CT images.

B. Model Approximation Comparison

In this experiment we evaluate how the proposed Cholesky method performs in comparison to the originally proposed Nyström method. In Figure 4, the registration accuracy of the differently approximated models is visualized. If we approximate the pivoted Cholesky model with an error of $\varepsilon = 0.01$ and only keep the first 300 eigenfunctions, the registration results are similar to the Nyström approximated model. However, when the Cholesky model is approximated with more eigenfunctions (2000), an increase in accuracy is still visible in the results.

C. Newton Basis Experiment

In this experiment we show a registration result comparison between models spanned by the approximated eigenfunctions and the Newton basis directly computed from the Cholesky decomposition. Since the eigenbasis of the first model is composed by an orthogonal transformation of the Newton basis, both models span the same subspace. In this experiment we evaluate if both type of basis functions are suitable for the optimization algorithm for registration. In Figure 5, the registration accuracies of both models are presented and lead, as expected, to comparable results.

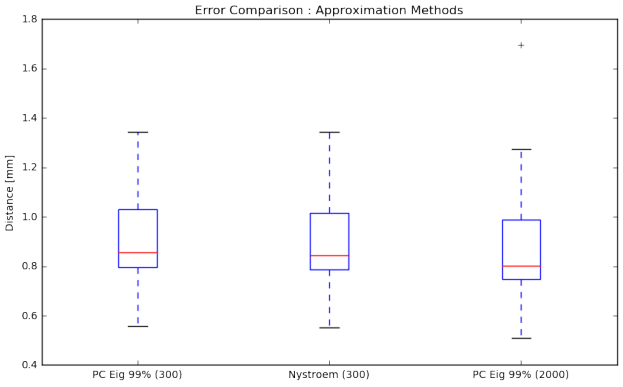

Fig. 4. A comparison between different model approximations using the registration accuracy of the Ulna registration. Our proposed Cholesky model with $\varepsilon = 0.01$ error is comparable with the Nyström method with 300 eigenfunctions. For the third model we kept all the approximated eigenbasis, which leads to a slight improvement.

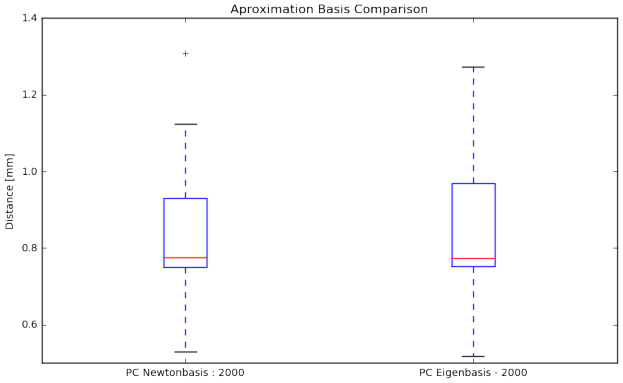

Fig. 5. Registration error comparison between a model built with an eigenbasis and a Newton basis model. The resulting registration accuracy of both models is similar, which leads to the conclusion that both type of basis seem well suited the optimization.

D. Registration Method Evaluation

In this last experiment we compare the proposed image deformation models to a state-of-the-art B-spline registration algorithm, which is implemented in Elastix [41]. In Figure 6, our registration approach is compared to the B-spline registration included in Elastix [41]. When the B-spline method is optimized on a single B-spline scale, the proposed method is comparable, but slightly more robust. However, the multi-scale B-spline optimization also improves the results. This is because the optimization procedure is specially adapted to exploit the multi-scale structure of the B-splines to make it less prone to get stuck in local optima. In GPMMs the optimization algorithm is generic and is not dedicated to a specific deformation prior. To become more robust towards local optima, one would instead strengthen the prior of the deformations. As Lüthi et al. [15] have shown, this can be done in multiple ways, as for example, making it spatially varying, symmetric or to include landmarks. In Figure 6 we show exemplary how adding landmark constraints to restrict the prior to only deformations that match the landmarks, lead

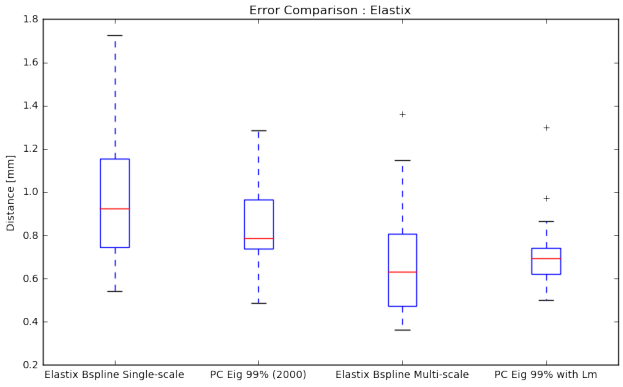

Fig. 6. A comparison of the built registration models with a state of the art B-spline method, which is implemented in Elastix. Our model is comparable with the single scale approach. However, also with the inclusion of landmark to guide the optimization procedure the method is also comparable to the multi-scale version.

to similar registration accuracy, and at the same time much more robust results.

IX. Conclusion

We have presented a novel low-rank approximation method for the Gaussian Process Morphable Model framework (GPMM). The framework enables the possibility to build complex prior deformation models for image and surface registration. An important part of the framework is to parametrize the model to make the optimization numerically feasible. We showed that the proposed approximation method has significant advantages over the originally proposed method: 1) The method enables full control over the approximation error. 2) The greedy algorithm is able to stop at a predefined accuracy. 3) The memory efficient algorithm enables the approximation of much larger point-sets, which leads to high-resolution models. We theoretically and experimentally evaluated the proposed algorithm in the context of surface registration and image registration. The proposed algorithm is mathematically sound and error bounds are provided for image and surface registration. Also we propose the possibility to omit the SVD calculation and use the directly calculated Newton basis from the pivoted Cholesky decomposition. This leads to the removal of a large calculation bottleneck without the loss of accuracy. Since these Newton basis are computed greedily it allows for iterative refinement. Future work would be to use the proposed basis functions to build a multi-scale registration. Especially in medical image registration, this method would improve the robustness of the registration method to avoid local minima. We showed the accuracy of the models in the context of human face surface registration evaluated on ground-truth registrations. Also, we demonstrated the applicability of the method in the context of medical image registration, where the human forearm was registered. We showed that the method is competitive to state of the art registration methods. The core algorithms proposed in this work are published open source in the *Scalismo* framework [24].

APPENDIX

A. Advanced Nyström Schemes

Nyström schemes are suitable if the eigenfunctions of the Karhunen-Loève expansion are only required in certain predetermined points $\mathbf{x}_1, \dots, \mathbf{x}_N$. For this purpose, the integral operator (2) is approximated by a quadrature formula

$$\int_\Omega \mathbf{K}(\cdot, \mathbf{x})\mathbf{f}(\mathbf{x}) \,\mathrm{d}\rho(\mathbf{x}) \approx \sum_{i=1}^N \omega_i \mathbf{K}(\cdot, \boldsymbol{\xi}_i)\mathbf{f}(\boldsymbol{\xi}_i)$$

with quadrature points $\boldsymbol{\xi}_i$ and weights ω_i. The discrete eigenvalue problem then reads

$$\mathbf{C}_{\text{Nystr}}\hat{\boldsymbol{\phi}}_{m,N} = \lambda_{m,N}\hat{\boldsymbol{\phi}}_{m,N}$$

with the system matrix

$$\mathbf{C}_{\text{Nystr}} = \left[\omega_j \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)\right]_{i,j=1}^N$$

and the point values

$$\hat{\boldsymbol{\phi}}_{m,N} \approx \left[\boldsymbol{\phi}_m(\mathbf{x}_i)\right]_i, \quad i = 1, \dots, N.$$

Note that the system matrix $\mathbf{C}_{\text{Nystr}}$ is not symmetric in general. Assuming positive quadrature weights, i.e. $\omega_i > 0$, defining

$$\mathbf{M}_{\text{Nystr}} = \operatorname{diag}(\sqrt{\omega_1}, \dots, \sqrt{\omega_N})$$

and setting $\boldsymbol{\phi}_{m,N} = \mathbf{M}_{\text{Nystr}}\hat{\boldsymbol{\phi}}_{m,N}$ yields a symmetric, generalized eigenvalueproblem

$$\mathbf{M}_{\text{Nystr}}\mathbf{C}\mathbf{M}_{\text{Nystr}}^\intercal\boldsymbol{\phi}_{m,N} = \lambda_{m,N}\mathbf{M}_{\text{Nystr}}\boldsymbol{\phi}_{m,N}$$

with the matrix

$$\mathbf{C} = \left[\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)\right]_{i,j=1}^N, \tag{17}$$

see also [39]. As it turns out, the finite element scheme yields an eigenvalue problem with a similar structure.

B. Advanced Finite Element Schemes

Having a finite dimensional subspace at hand yields, cf. e.g. [39], the generalized eigenvalue problem

$$\mathbf{C}_{\text{FEM}}\boldsymbol{\phi}_{m,N} = \lambda_{m,N}\mathbf{M}_{\text{FEM}}\boldsymbol{\phi}_{m,N} \tag{18}$$

with system matrices

$$\mathbf{C}_{\text{FEM}} = \left[\left(\mathcal{T}_{\mathbf{K}}\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i\right)_{[L^2_\rho(D)]^3}\right]_{i,j=1}^N,$$
$$\mathbf{M}_{\text{FEM}} = \left[\left(\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i\right)_{[L^2_\rho(D)]^3}\right]_{i,j=1}^N,$$

$\mathcal{T}_{\mathbf{K}}$ denoting the integral operator from (2), and the approximate eigenfunctions

$$\boldsymbol{\phi}_m(\mathbf{x}) \approx \boldsymbol{\phi}_{m,N}(\mathbf{x}) = \sum_{i=1}^N \left(\boldsymbol{\phi}_{m,N}\right)_i \boldsymbol{\varphi}_i(\mathbf{x}).$$

It thus remains to explain how to assemble these matrices.

Since the basis functions $\boldsymbol{\varphi}_i$ are non-zero only on a few elements, the mass matrix $\mathbf{M}_{\text{FEM}}$ is sparse. Inserting the definition of $\mathcal{T}_{\mathbf{K}}$ into the definition of $\mathbf{C}_{\text{FEM}}$, we obtain

$$\mathbf{C}_{\text{FEM}} = \left[\int_D \int_D \mathbf{K}(\mathbf{x}, \mathbf{y})\boldsymbol{\varphi}_j(\mathbf{y})\boldsymbol{\varphi}_i^\intercal(\mathbf{x}) \,\mathrm{d}\rho(\mathbf{y}) \,\mathrm{d}\rho(\mathbf{x})\right]_{i,j=1}^N.$$

In order to compute this integral, it is very common in finite element methods to replace $\mathbf{K}$ by its interpolation $\mathbf{K}_h$ in the finite element space, i.e. we approximate

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) \approx \sum_{i,j=1}^N \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)\boldsymbol{\varphi}_i(\mathbf{x})\boldsymbol{\varphi}_j^\intercal(\mathbf{y}).$$

Inserting this approximation into the definition of $\mathbf{C}_{\text{FEM}}$ yields

$$\mathbf{C}_{\text{FEM}} = \mathbf{M}_{\text{FEM}}\mathbf{C}\mathbf{M}_{\text{FEM}}^\intercal$$

with the matrix $\mathbf{C}$ defined as for the Nyström scheme in (17). The eigenvalue problem (18) thus turns into

$$\mathbf{M}_{\text{FEM}}\mathbf{C}\mathbf{M}_{\text{FEM}}^\intercal\boldsymbol{\phi}_{m,N} = \lambda_{m,N}\mathbf{M}_{\text{FEM}}\boldsymbol{\phi}_{m,N}. \tag{19}$$

C. Connection between the two schemes

The two schemes can lead to the very same eigenvalue problem. In implementations of finite element schemes, there are almost always quadrature formulas involved. Using piecewise linear ansatz functions and replacing the integrals by a trapezoidal rule yields a diagonal matrix $\mathbf{M}_{\text{FEM}}$ (this is also referred to as "mass lumping"). The definition of $\mathbf{M}_{\text{Nystr}}$ then amounts to quadrature weights to a quadrature formula with the vertices of the finite element mesh as evaluation points. The two schemes are thus equivalent in this specific case.

D. Dimension Reduction of the Eigenvalue Problem

Again, having a low-rank factorization $\mathbf{C} \approx \mathbf{L}_M\mathbf{L}_M^\intercal$ of rank M at hand, one can reduce the dimension of the eigenvalue problems (19). For ease of notation, we do not distinguish between $\mathbf{M}_{\text{FEM}}$ and $\mathbf{M}_{\text{Nystr}}$ and consider the eigenvalue problem

$$\mathbf{M}\mathbf{C}\mathbf{M}^\intercal\boldsymbol{\phi}_{m,N} = \lambda_{m,N}\mathbf{M}\boldsymbol{\phi}_{m,N}. \tag{20}$$

By substituting the low-rank approximation $\mathbf{C} \approx \mathbf{L}_M\mathbf{L}_M^\intercal$ and $\mathbf{v}_{m,N} = \mathbf{M}^{1/2}\boldsymbol{\phi}_{m,N}$ into (20), the eigenvalue problem becomes

$$\mathbf{M}^{1/2}\mathbf{L}_M\mathbf{L}_M^\intercal(\mathbf{M}^{1/2})^\intercal\mathbf{v}_{m,N} = \lambda_{m,N}\mathbf{v}_{m,N}.$$

Exploiting the fact that $\mathbf{M}^{1/2}\mathbf{L}_M\mathbf{L}_M^\intercal(\mathbf{M}^{1/2})^\intercal$ has the same eigenvalues as $\mathbf{L}_M^\intercal(\mathbf{M}^{1/2})^\intercal\mathbf{M}^{1/2}\mathbf{L}_M = \mathbf{L}_M^\intercal\mathbf{M}\mathbf{L}_M$, we obtain an equivalent eigenvalue problem

$$\mathbf{L}_M^\intercal\mathbf{M}\mathbf{L}_M\tilde{\mathbf{v}}_{m,N} = \lambda_{m,N}\tilde{\mathbf{v}}_{m,N}.$$

This modified eigenvalue problem has again dimension $M \ll N$ and can thus be solved by standard eigensolvers for dense matrices.

ACKNOWLEDGMENT

This work has been funded as part of two Swiss National Science foundation projects in the context of the projects *SNF153297* and *SNF156101*. We thank Andreas Morel-Forster and Volker Roth for interesting and enlightning discussions. A special thanks goes to Ghazi Bouabene and Christoph Langguth for their work on the *Scalismo* software, in which all the methods are implemented.

REFERENCES

[1] U. Grenander and M. I. Miller, "Computational anatomy: An emerging discipline," *Quarterly of applied mathematics*, vol. 56, no. 4, pp. 617–694, 1998.

[2] A. Yuille and D. Kersten, "Vision as bayesian inference: analysis by synthesis?" *Trends in cognitive sciences*, vol. 10, no. 7, pp. 301–308, 2006.

[3] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: a survey," *Medical image analysis*, vol. 24, no. 1, pp. 205–219, 2015.

[4] M. B. Cuadra, V. Duay, and J.-P. Thiran, "Atlas-based segmentation," in *Handbook of Biomedical Imaging*. Springer, 2015, pp. 221–244.

[5] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3d statistical deformation models using non-rigid registration," in *MICCAI '01: Medical Image Computing and Computer-Assisted Intervention*, 2001, pp. 77–84.

[6] T. Heimann, B. Van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes *et al.*, "Comparison and evaluation of methods for liver segmentation from ct datasets," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 8, pp. 1251–1265, 2009.

[7] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press, 1999, pp. 187–194.

[8] T. F. Cootes, C. Beeston, G. J. Edwards, and C. J. Taylor, "A unified framework for atlas matching using active appearance models," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 1999, pp. 322–333.

[9] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *Medical Imaging, IEEE Transactions on*, vol. 18, no. 8, pp. 712–721, 1999.

[10] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 376–387, 1991.

[11] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 6, pp. 526–534, 2001.

[12] L. Zagorchev and A. Goshtasby, "A comparative study of transformation functions for nonrigid image registration," *IEEE transactions on image processing*, vol. 15, no. 3, pp. 529–538, 2006.

[13] A. M. Siddiqui, A. Masood, and M. Saleem, "A locally constrained radial basis function for registration and warping of images," *Pattern Recognition Letters*, vol. 30, no. 4, pp. 377–390, 2009.

[14] T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, and others, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, 1995.

[15] M. Lüthi, C. Jud, T. Gerig, and T. Vetter, "Gaussian process morphable models," *arXiv preprint arXiv:1603.07254*, 2016.

[16] M. Lüthi, C. Jud, and T. Vetter, "A unified approach to shape model fitting and non-rigid registration," in *Machine Learning in Medical Imaging*. Springer, 2013, pp. 66–73.

[17] T. Gerig, K. Shahim, M. Reyes, T. Vetter, and M. Lüthi, "Spatially varying registration using gaussian processes," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*. Springer, 2014, pp. 413–420.

[18] M. Lüthi, A. Forster, T. Gerig, and T. Vetter, "Gaussian process morphable models," in *Statistical Shape and Deformation Analysis - Methods, Implementation and Applications*, G. Zheng, S. Li, and G. Szekely, Eds., 2017, to appear.

[19] M. Lüthi, C. Jud, and T. Vetter, "Using landmarks as a deformation prior for hybrid image registration," *Pattern Recognition*, pp. 196–205, 2011.

[20] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*. Springer, 2004, vol. 3.

[21] H. Harbrecht, M. Peters, and R. Schneider, "On the low-rank approximation by the pivoted cholesky decomposition," *Applied Numerical Mathematics*, vol. 62, no. 4, pp. 428 – 440, 2012.

[22] L. Foster, A. Waagen, N. Aijaz, M. Hurley, A. Luis, J. Rinsky, C. Satyavolu, M. J. Way, P. Gazis, and A. Srivastava, "Stable and efficient gaussian process calculations," *Journal of Machine Learning Research*, vol. 10, no. Apr, pp. 857–882, 2009.

[23] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. Springer, 2006.

[24] "Scalismo - scalable image analysis and shape modelling," http://github.com/unibas-gravis/scalismo.

[25] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. L. Yuille, "Robust estimation of nonrigid transformation for point set registration," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[26] Y. Wang and L. H. Staib, "Boundary finding with prior shape and smoothness models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, 2000.

[27] S. C. Joshi, A. Banerjee, G. E. Christensen, J. G. Csernansky, J. W. Haller, M. I. Miller, and L. Wang, "Gaussian random fields on submanifolds for characterizing brain surfaces," in *Information Processing in Medical Imaging*. Springer, 1997, pp. 381–386.

[28] C. Jud, N. Mori, and P. C. Cattin, "Sparse kernel machines for discontinuous registration and nonstationary regularization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 9–16.

[29] N. H. Beebe and J. Linderberg, "Simplifications in the generation and transformation of two-electron integrals in molecular calculations," *International Journal of Quantum Chemistry*, vol. 12, no. 4, pp. 683–705, 1977.

[30] H. Harbrecht, M. Peters, and M. Siebenmorgen, "Efficient approximation of random fields for numerical applications," *Numerical Linear Algebra with Applications*, vol. 22, no. 4, pp. 596–617, 2015.

[31] M. Pazouki and R. Schaback, "Bases for kernel-based spaces," *Journal of Computational and Applied Mathematics*, vol. 236, no. 4, pp. 575 – 588, 2011.

[32] G. Santin and R. Schaback, "Approximation of eigenfunctions in kernel-based spaces," *Advances in Computational Mathematics*, vol. 42, no. 4, pp. 973–993, 2016.

[33] M. Griebel and H. Harbrecht, "Approximation of bi-variate functions: singular value decomposition versus sparse grids," *IMA Journal of Numerical Analysis*, vol. 34, no. 1, pp. 28–54, 2014.

[34] L. Rosasco, M. Belkin, and E. D. Vito, "On learning with integral operators," *The Journal of Machine Learning Research*, vol. 11, pp. 905–934, 2010.

[35] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.

[36] B. Amberg, S. Romdhani, and T. Vetter, "Optimal step nonrigid icp algorithms for surface registration," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[37] R. Opfer, "Multiscale kernels," *Advances in Computational Mathematics*, vol. 25, no. 4, pp. 357–380, 2006.

[38] T. Albrecht, M. Lüthi, T. Gerig, and T. Vetter, "Posterior shape models," *Medical image analysis*, vol. 17, no. 8, pp. 959–973, 2013.

[39] W. Hackbusch, *Integral Equations: Theory and Numerical Treatment*. Basel: Birkhäuser, 1995, vol. 4.

[40] G. Golub and C. Van Loan, *Matrix Computations*, 4th ed. Johns Hopkins University Press, 2012.

[41] S. Klein, M. Staring, K. Murphy, M. Viergever, J. Pluim *et al.*, "Elastix: a toolbox for intensity-based medical image registration," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 196–205, 2010.

LATEST PREPRINTS

No.	Author: *Title*
2016-11	**I. Hedén, S. Zimmermann** *The decomposition group of a line in the plane*
2016-12	**J. Ballani, D. Kressner, M. Peters** *Multilevel tensor approximation of PDEs with random data*
2016-13	**M. J. Grote, M. Kray, U. Nahum** *Adaptive eigenspace method for inverse scattering problems in the frequency domain*
2016-14	**H. Harbrecht, M. Peters, M. Schmidlin** *Uncertainty quantification for PDEs with anisotropic random diffusion*
2016-15	**F. Da Lio, L. Martinazzi** *The nonlocal Liouville-type equation in R and conformal immersions of the disk with boundary singularities*
2016-16	**A. Hyder** *Conformally Euclidean metrics on R^n with arbitrary total Q-curvature*
2016-17	**G. Mancini, L. Martinazzi** *The Moser-Trudinger inequality and its extremals on a disk via energy estimates*
2016-18	**R. N. Gantner, M. D. Peters** *Higher order quasi-Monte Carlo for Bayesian shape inversion*
2016-19	**C. Urech** *Remarks on the degree growth of birational transformations*
2016-20	**S. Dahlke, H. Harbrecht, M. Utzinger, M. Weimar** *Adaptive wavelet BEM for boundary integral equations: Theory and numerical experiments*
2016-21	**A. Hyder, S. Iula, L. Martinazzi** *Large blow-up sets for the prescribed Q-curvature equation in the Euclidean space*
2016-22	**P. Habegger** *The norm of Gaussian periods*
2016-23	**P. Habegger** *Diophantine approximations on definable sets*

Preprints are available under *https://math.unibas.ch/research/publications*

LATEST PREPRINTS

No.	Author: *Title*
2016-24	**F. Amoroso, D. Masser** *Lower bounds for the height in Galois extensions*
2016-25	**W. D. Brownawell, D. W. Masser** *Zero estimates with moving targets*
2016-26	**H. Derksen, D. Masser** *Linear equations over multiplicative groups, recurrences, and mixing III*
2016-27	**D. Bertrand, D. Masser, A. Pillay, U. Zannier** *Relative Manin-Mumford for semi-abelian surfaces*
2016-28	**L. Capuano, D. Masser, J. Pila, U. Zannier** *Rational points on Grassmannians and unlikely intersections in tori*
2016-29	**C. Nobili, F. Otto** *Limitations of the background field method applied to Rayleigh-Bénard convection*
2016-30	**W. D. Brownawell, D. W. Masser** *Unlikely intersections for curves in additive groups over positive characteristic*
2016-31	**M. Dambrine, H. Harbrecht, M. D. Peters, B. Puig** *On Bernoulli's free boundary problem with a random boundary*
2016-32	**H. Harbrecht, J. Tausch** *A fast sparse grid based space-time boundary element method for the nonstationary heat equation*
2016-33	**S. Iula** *A note on the Moser-Trudinger inequality in Sobolev-Slobodeckij spaces in dimension one*
2016-34	**C. Bürli, H. Harbrecht, P. Odermatt, S. Sayasone, N. Chitnis** *Mathematical analysis of the transmission dynamics of the liver fluke, Opisthorchis viverrini*
2017-01	**J. Dölz and T. Gerig, M. Lüthi, H. Harbrecht and T. Vetter** *Efficient computation of low-rank Gaussian process models for surface and image registration*

Preprints are available under *https://math.unibas.ch/research/publications*