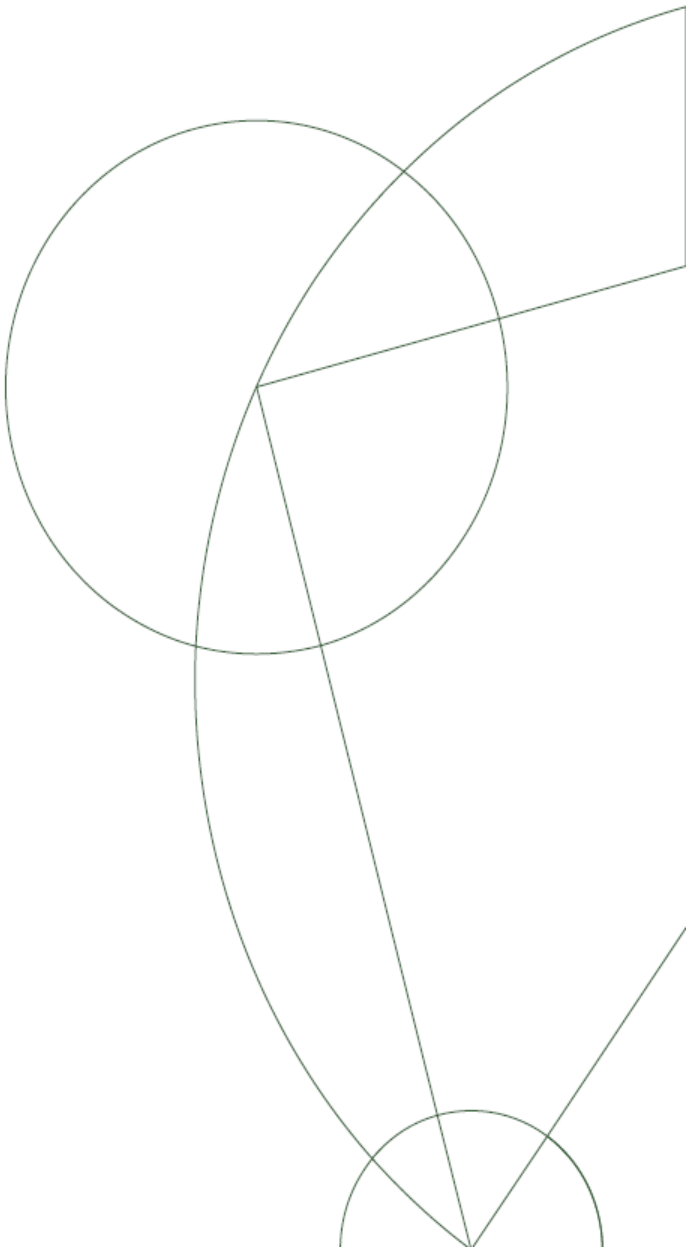
FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

PhD Thesis

Alessandro Crimi

Statistical Analysis of Vertebral Shapes



Academic Advisor : prof. Mads Nielsen

Co-advisor : Dr. Martin Lillholm

To the people who helped and inspired me in this work,
and for those to whom this work will be useful.
In memory of Giuseppe Romeo.

Contents

Acronyms

AUC Area Under the roc Curve

ASM Active Shape Model

CI Confidence Interval

CT Computed Tomography

BMD Bone Mineral Density

BNLO Balanced Nested Leave Out

CTX C-Telopeptide collagen type 1 cross-linked

DSA Discriminative shape alignment

DXA Dual Energy X-ray Absorptiometry

FPR False Positive Rate

GPA Generalized Procrustus Alignment

HRT Hormone Replacement Therapy

LDA Linear Discriminant Analysis

MAP-G Gaussian Maximum A Posteriori estimation

MAP-IWIS Inverted Wishart Maximum A Posteriori estimation

MAP-UIWIS Uncommitted Inverted Wishart Maximum A Posteriori estimation

ML Maximum Likelihood

ML-T Maximum Likelihood Tikhonov regularized

MRI Magnetic Resonance Imaging

NTX N-Telopeptide collagen type 1 cross-linked

OR Odds Ratio

PCA Principal Components Analysis

PDM Point Distribution Model

QM Quantitative Morphometric method

ROC Receiver Operating Characteristic

SD Standard Deviation

SQ Semi-Quantitative method

SVD Singular Value Decomposition

WHO World Health Organisation

Abstract

This thesis proposes and evaluates some algorithms for shape based analysis, and their application in clinical studies related to osteoporosis and vertebral osteoporotic fracture. The main objective of these algorithms is to improve the investigation of such and other diseases. The reported studies were performed using radiographs of the lumbar region. In particular, manual annotations on these radiographs performed by experts are used. There are two main contributions. Firstly, a study for assessing vertebral fracture risk using shape analysis is described. Secondly a methodology for estimating covariance matrices when few samples are available is proposed. This last issue is a common challenge in statistical tasks, and it is also present in the aforementioned shape based framework.

Fracture risk assessment is currently based on a series of factors such as BMD, age, use of corticosteroids, and previous fractures. A clinical study employed a shape based method for predicting the risk of vertebral fractures. The described approach can be used combined with the traditional risk factors allowing a more reliable prediction of vertebral fractures. The reported longitudinal case-control study involved postmenopausal women, the patients of the case group sustained at least one incident lumbar fracture, while the patients of the control group maintained skeletal integrity over a 7.3-year period. The patients of both groups were fracture-free at baseline. The investigation focused on the ability of the shape based analysis in discriminating these two groups at baseline where no fractures were present. Lateral lumbar radiographs of the patients were taken twice, and an expert radiologist graded the fracture status of their spines. On the acquired radiographs, a radiologist and two additional x-ray technicians performed independently the annotations of the spines. A statistical shape model from the spine shape annotations was built and performed a classification on the baseline data, identifying possible small changes in the spine shape that can result in future fractures. This technique led to an AUC of 0.70 ± 0.013 for the radiologist annotations. The reported study presented a typical drawbacks of clinical studies: the small sample size. The rest of the thesis is focused on this aspect, proposing possible solutions.

In several statistical tasks, such as shape based fracture prediction, the estimation of covariance matrices is a crucial step. When using few samples of a high dimensional representation, e.g. of shapes, the standard Maximum Likelihood estimation (ML) of the covariance matrix is often rank deficient, and may lead to imprecise results. Regularization by prior knowledge using Maximum A Posteriori (MAP) esti-

mates is proposed, and a comparison between ML, Tikhonov regularization and MAP estimates using a number of priors is reported. The covariance estimates are evaluated on both synthetic and real data, and the estimates' influence is considered on a missing-data reconstruction task. Here, high resolution vertebra and cartilage models are reconstructed from incomplete and lower dimensional representations. In the reported experiments, the Bayesian methods outperformed the traditional ML method and Tikhonov regularization, reducing to 1 mm the mean reconstruction error for the vertebral shapes, and to 20% mean error for the cartilage sheets, using only 10 samples for both experiments. The use of these MAP methods can also improve the results of other statistical tasks such as the shape based fracture prediction study.

Chapter 1

Introduction

This thesis focus on statistical methodologies based on shape model, and their application to osteoporotic vertebral fracture. This chapter comprises the motivation of this work, an outline of the thesis and an overview of the involved contributions.

1.1 Motivation

Osteoporosis is a bones disease leading to an increased risk of fracture. It represents a major public health threat which afflicts individuals, especially postmenopausal women. It has been estimated that approximately one in three women and one in twelve men over the age of 50 have osteoporosis [72]. Millions of fractures are diagnosed worldwide annually, mostly involving the lumbar vertebrae, hip, wrist, and ribs. Consequently, the burden sustained by the medical system related to osteoporotic fractures is heavy and increasing rapidly. In fact, the total cost for the year 2000 in Europe has been estimated at 31 billion Euros [68].

Vertebral fractures are the most common fractures, and as the proportion of the elderly population grows, they are becoming increasingly common. Osteoporotic fractures can occur in absence of trauma or after a minimal trauma, and in this case they are also called *fragility fractures*. Early diagnosis of osteoporosis allows preventive therapeutic treatments [89]. The most common measure of bone quality is the Bone Mineral Density (BMD) which measures the amount of matter per cubic centimeter of bones. This measure is relevant since there is a statistical association between poor bone density and higher probability of fracture [90]. Other methodologies used in vertebral clinical analysis are morphometric indices obtained by analyzing radiographs [30, 45, 83].

Although half of all osteoporotic fractures are vertebral, they are just the tip of the iceberg, since their presence usually increases the risk of any further vertebral fractures [86] and the risk of any subsequent hip fracture [7]. Hip fractures are serious harms in old age and they are correlated to patient death within a year [87] and disabling conditions [32]. Therefore, early assessment of vertebral fracture risk is crucial for

reducing the general number of fractures through appropriate treatments for patients, as it will improve the quality of life for patients; and subsequently decrease the related expenses to health system.

Shape based prediction of osteoporotic fractures, like other statistical studies, requires a large amount of data. More specifically, a sufficient number of high quality samples compared to the dimensionality of the shapes is needed. This can be a challenging situation since often not so many data are available.

Some algorithms have been proposed [31, 40, 53, 57, 64, 103, 108, 113] with the aim of tackling this issue, mainly through the regularization of covariance matrices, since precise covariance matrices influence the quality of the prediction and other tasks. However, some aspects of regularization can still be investigated. In the next chapters, a study regarding prediction of vertebral fractures among populations is reported, and a methodology for estimating covariance matrices when few samples are available is also discussed.

1.2 Outline

The main contents of this work are presented in five chapters: Chapter 2 gives a general clinical background of osteoporosis and osteoporotic fractures, Chapter 3 describes some of the most common shape models used in medical imaging, and Chapter 4 is related to prediction of osteoporosis using some of these models. Chapter 5 presents a possible solution for the common problem of few available samples in statistical tasks, while Chapter 6 contains the conclusions. The following gives a brief description of each chapter.

- Chapter 2 discusses the clinical background. This includes a general description of osteoporosis and its epidemiology. Diagnosis, prevention and established treatments are also discussed. Moreover, well-known osteoporotic fracture measurements are described, such as quantitative morphometric indices [30,83], and semi-quantitative morphometric indices [45].

- Chapter 3 contains a survey of the most common deformable shape models used in computer vision and then focuses on methods related to fracture assessment. With the improvement of image processing technologies, in parallel to the established morphometric indices [30, 44, 83], the interest of diagnostic computer based methods is growing. The initial part of the chapter gives an historical introduction of active contour shapes, explaining how statistical shape models became successful methodologies in medical imaging. The second half describes some of the computer based methods used for vertebral fracture assessment, which can be considered as the ancestors of the prediction algorithm described in Chapter 4. The chapter also introduces the common problematic use of these methods when few samples are available.

- Chapter 4 presents a study of vertebral fracture prediction using a statistical shape model. It contains a description of the population used in our experi-

ments, the statistical analysis, and the used methodology. Two different approaches for aligning spine shapes are discussed. Here, a shape based analysis is used for fracture prediction purposes rather than diagnostic as described in the previous chapter.

- Chapter 5 deals with the crucial problem of having a small sample size in statistical analysis. Epidemiological studies, such as the case described in Chapter 4, often encounter the limitation of few available samples. A novel method for Bayesian regularization of covariance matrices of shape based data is described.
- Chapter 6 contains the discussions about vertebral fracture risk assessment and the Bayesian covariance estimation.

1.3 Contributions

The thesis has two major contributions:

- **Prediction of vertebral fractures:** A study which uses a shape based method for predicting vertebral fractures is reported. A case-control longitudinal study was performed, analyzing a shape based morphometric index, which combined with the traditional risk factors allows a more reliable prediction of vertebrae fracture. The investigated methodology can facilitate a population selection in clinical trials, helping identifying patients with high fracture risk.
- **Novel method for covariance estimation:** Shape based analysis and other statistical tasks often require the estimation of covariance matrices as a pivotal step. When the dimensionality of the shape space is high, and the available number of samples is small, then the estimation of a covariance matrix with standard methods can be poor. A novel method for overcoming this common scenario in clinical trials is proposed. The method is based on the involvement of prior knowledge during the covariance estimation.

Chapter 2

Osteoporotic Vertebral Fractures

This chapter gives an overview of the clinical aspects of osteoporosis and osteoporotic vertebral fractures. The first part of the chapter comprises an introduction to the disease describing its risk factors, treatments and prevention. The remainder of the chapter briefly describes the most common diagnostic methods for osteoporotic vertebral fractures. In particular, quantitative and semi-quantitative morphometric indices are reviewed.

2.1 Osteoporosis

Osteoporosis is a disease characterized by a decrease in density and quality of bone which can lead to fractures. The main biological reason lies in the deficiency of estrogens which occurs with age. Bone usually has a hard external cortical coating enclosing a complex net of trabecular structures. Those structures are not static. They are the result of the interaction between two kinds of cells, called osteoclasts and osteoblasts. The former resorb bones from the skeleton, while the latter are responsible for bone formation. With age, especially in post-menopausal women, the level of estrogens drops, compromising the equilibrium between the actions of osteoclasts and osteoblasts, and weakening the bone structures. Figure 2.1 shows the difference in micro-structure between normal and osteoporotic bone. Osteoporosis can weak bone structures causing fractures, called osteoporotic fractures, which occur mainly in hip, wrist and spine, with the incidence relative to ages depicted in Figure 2.2. The most harmful fractures are the hip fractures, since correlation between hip fractures and patient death within one year [87], and long-term disability [32] have been reported. Vertebral fractures are the most common and represent a risk factor for future hip fractures [7]. However, the presence of vertebral fractures can be overlooked, since vertebral fractures can be asymptomatic [26]. Therefore, estimates of vertebral fractures must be obtained in other ways. It can be argued that osteoporotic fractures are the result of the process previously described rather than an abrupt change, though the initial stage of deterioration is hard to detect.

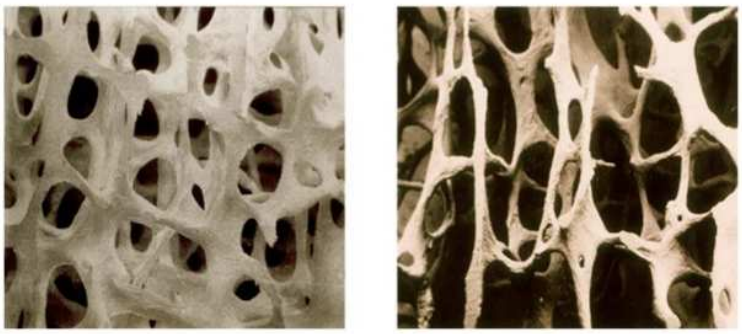


Figure 2.1: Comparison of healthy trabecular structures on the left and osteoporotic trabecular structures on the right. Image from Consensus development conference (1993), Am J Med 94:64d, courtesy of IOF.

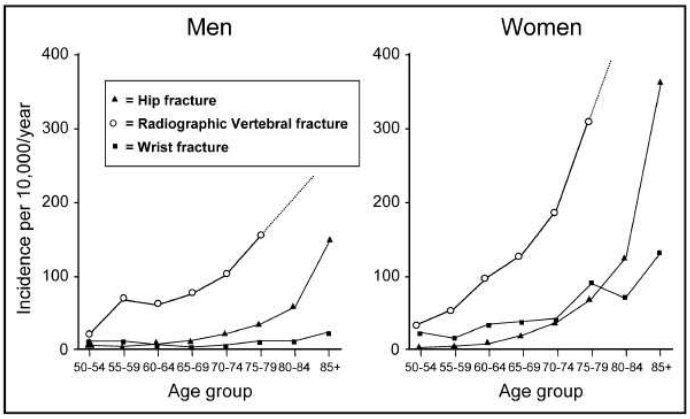


Figure 2.2: Incidence of osteoporotic fractures (Image courtesy of Denis-Cooper 2005).

With the advances in the health-care system and the current demographic trends, it is expected an increase of people suffering from this disease, with consequent impact on the people quality of life and the health-care system cost [110]. In this context, diagnosis and prevention are important, since early detection of osteoporosis can allow more effective treatments to patients.

2.1.1 Bone Mineral Density

Aerial bone mineral density (BMD) is the measure of the amount of matter per square centimeter of bones, mainly obtained using Dual-energy X-ray absorptiometry (DXA) [78]. Since 1994, the World Health Organisation (WHO) considers BMD as the standard diagnostic criteria of osteoporosis [69]. BMD is also relevant for fracture prediction, because bones with low BMD have high probability of fracture [59, 90]. According to the WHO criteria [69], the BMD measurement can be used to classify bones into three major categories depending on the t-score, e.i. depending on how many standard deviations a measurement is below an estimated mean from a population of healthy pre-menopausal Caucasian women.

- Normal: t-score defined as -1 or higher.
- Osteopenia: t-score between -1 and -2.5.
- Osteoporosis: t-score defined as -2.5 or lower.

Moreover, it is important to bear in mind that there are different common sites to measure BMD, such as lumbar spine, total hip, and femoral neck; and all can give different readings.

2.1.2 Risk factors

One of the most common causes of deterioration in bone structure is the deficiency in estrogen production, which occurs mainly after the menopause in women.

Along with the natural process of aging, there are other causes which can increase the risk of osteoporosis, and are included in a WHO tool for fracture prediction [76]. Physical inactivity and some drug therapies based on corticosteroids have the possible side effects of accelerating bone loss. Moreover, there are other lifestyle factors increasing the risk of osteoporosis such as smoking, excessive intake of alcohol, malnutrition, and lack of calcium or vitamin D [68]. Maintenance of bone mass requires attention to these factors and regular physical exercise, which promotes bone development and remodeling. As mentioned in the previous section, the presence of an osteoporotic fracture implies an increased risk of further osteoporotic fractures. This is one of the main motivation of the works described in this thesis, since a single vertebral fracture usually anticipates further vertebral fractures or hip fractures which can be fatal [7]. Therefore, vertebral fractures can be considered as a risk-factor for more disabling osteoporotic condition.

2.1.3 Treatments

Like other diseases, there are two ways for coping with osteoporotic fractures and osteoporosis: prevention and treatment. Obviously these two issues are related to each other. A primary non-invasive way of preventing osteoporosis, and therefore osteoporotic fractures, is to reduce the risk factors mentioned in the previous section.

Regarding treatments, several drugs have been produced and investigated. Increasing the quality of fracture risk assessment in early diagnosis holds a pivotal role, since it can allow a prompt use of these treatments.

Hormone Replacement Therapy (HRT) was the standard treatment for osteoporosis in the past. In several experiments it proved to have a positive impact reducing the number of osteoporotic fractures [116]. However, several side effects such as breast cancer, thromboembolism, and cardiovascular diseases were experienced [100].

Bisphosphonates are oestoclasts inhibitors which limit the bone resorption. Despite their efficacy, they also have several side effects such as gastrointestinal disturbances [8] and osteonecrosis of the jaw and hip [88]. Osteonecrosis is the bone death caused by poor blood supply and can lead to very painful and potentially disabling conditions.

More recent drugs, such as Teriparatide rhPTH(1-34) and SERM, were proposed. In clinical studies with patients already undergoing an osteoporotic fracture, rhPTH(1-34) reduced the incidence of further fractures by 30% to 65% [117], and the use of SERM showed a decrease of bone turnover below concerning levels, an increase of BMD in the lumbar spine by 2%, and a reduction of vertebral fracture risk by 40% [104]. So far, no severe side effects were found for these approaches.

Another possible treatment is Calcitonin, a hormone which inhibits bone resorption by osteoclasts and promotes bone formation by osteoblasts. Some studies suggested that, compared to calcium alone, Calcitonin given intranasally reduced the rates of fracture by two thirds in elderly women with moderate osteoporosis [92]. Also for Calcitonin no severe side effects were encountered so far.

2.2 Vertebral fractures

The human vertebral column, also called the spine, comprises 24 articulating vertebrae and 9 fused vertebrae constituting the sacrum and coccyx, as shown in Figure 2.3. The articulating vertebrae have bodies, which are cylinder-like structures positioned one over the other, forming a chain which supports the body allowing some flexibility. The human spine has several main curvatures corresponding to sections termed cervical, thoracic, lumbar, sacrum and coccyx. The uppermost section is the cervical region composed of 7 vertebrae (termed C1-C7 from the uppermost to the lowest), 12 vertebrae are present in the thoracic region (T1-T12), and 5 vertebrae in the lumbar region (L1-L5). The sacrum contains 5 fused vertebrae while the coccyx contains 4 fused vertebrae. One of the goals of this thesis is the prediction of vertebral fractures, analyzing slight differences in the shape of the spine or in the vertebrae themselves. In particular, lumbar vertebrae and the lowest thoracic vertebra T12 are analyzed. A choice primarily made for facilitating the comparison between the study described in Chapter 4 and previous studies using the same cohort [79].

Many possible features are obtainable from radiographs of vertebral bodies such as bone density, textural information [62], and anthropometric parameters [60]. In particular, useful information can be extrapolated from the geometric analysis of the three vertical heights and the two horizontal endplates identified by landmarks [60] as

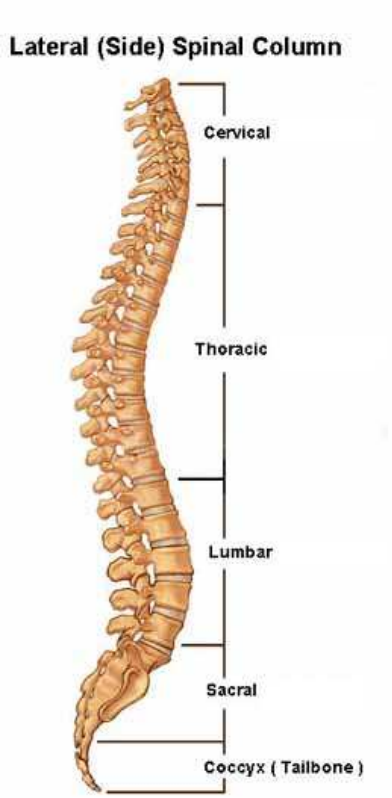


Figure 2.3: An illustration of a spine with its regions (Image courtesy of SpineUniverse Marketing Vertical Health LLC.)

depicted in Figure 2.4. As it will be described in the next sections, this information is used in several morphometric analyses for quantifying the presence or absence of osteoporotic fractures.

When trabecular structures are compromised and vertebral fractures occur, either the central endplates deform making the vertebra biconcave, or one of the corners collapses making the vertebral structure wedge-like. Furthermore, if all the heights collapse, crush fractures occur. Figure 2.5 (a) and (b) show a radiograph of a spine with healthy vertebrae and a radiograph of a spine with a fractured L4. The depicted spine presents a biconcave fracture, an overview of this and other vertebral fractures is reported in Figure 2.6.

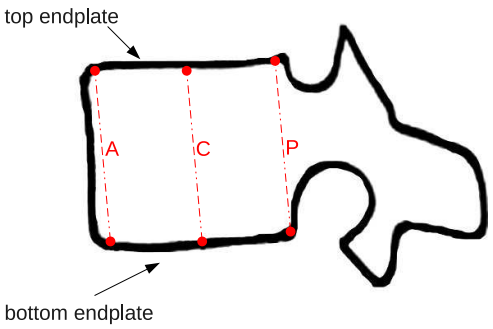


Figure 2.4: A drawn vertebra with anterior (A), central (C) and posterior (P) heights obtained linking vertically the six representative points.

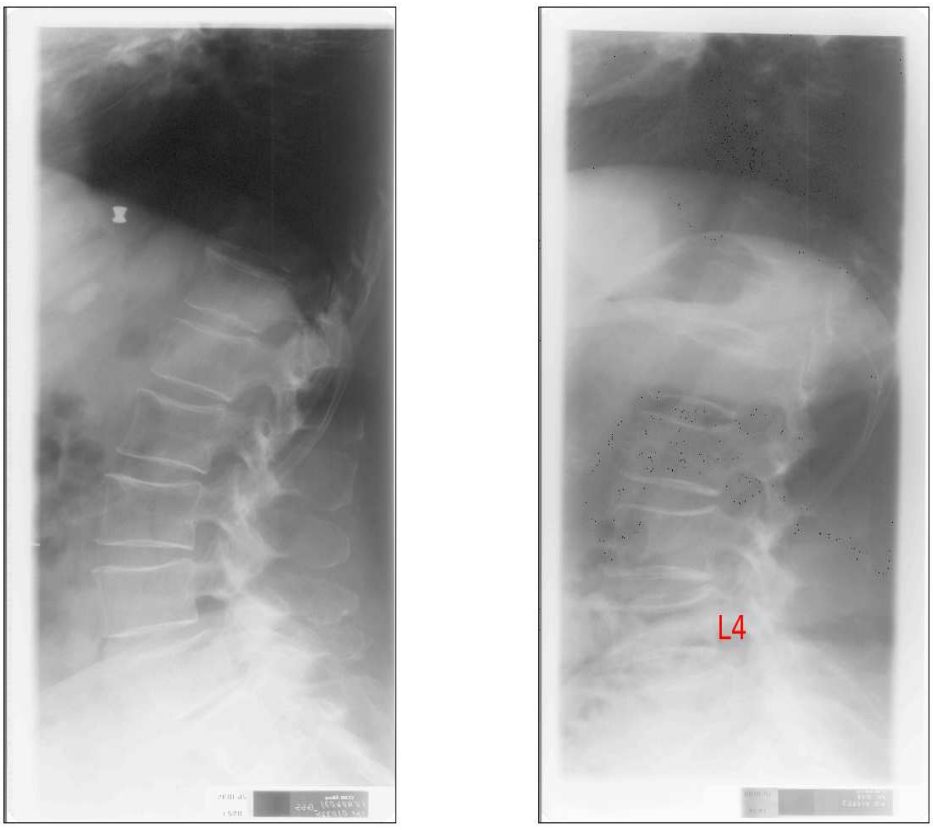
2.3 Current and future vertebral fracture assessment

Diagnosis and prediction of osteoporosis is not a trivial process. Regarding fracture assessment several indices, such as the morphometric measurements from radiographs, are used. Computed radiography is now the "gold" standard, but it can have problems associated with projective geometry. Some studies [47, 115] have shown that MRI-based morphometry can be a competitive modality, especially for identifying spinal pathologies, such as infections or disc degeneration. However, the cost of a MRI-based investigation is still high compared to a simpler radiograph investigation, which remains the most common modality in this context.

For prediction, the WHO recommends a fracture risk assessment based on a series of factors [76]. The following two subsections explain the most common methods for present fracture assessment (diagnosis), while the last subsection describes some clinical procedures for risk assessment (prediction). Shape based analysis and its usefulness in vertebral fracture risk assessment are described in the next chapter, and the proposed methods for increasing the quality of risk assessment are described in Chapter 5 to 7.

2.3.1 Quantitative Morphometry

Morphometric methods are classified in Quantitative Morphometric methods (QM) and Semi-Quantitative methods (SQ). The first are objective measurements of vertebrae in radiograph, while the latter are based on the subjective judgment of a radiologist. One of the most known QM methods is the Eastell-Melton criteria [30, 85], which quantifies vertebral fracture based on the ratios of anterior, middle or neighbouring posterior heights and posterior heights of vertebral bodies, compared with the expected standard value for the particular vertebrae in exam. This assessment is strictly related to a specific threshold and takes into account also the top and



(a) (b)

Figure 2.5: (a) An healthy spine. (b) A spine with a biconcave fracture at L4. Figure 2.6 illustrates other kind of fractures which can occur.

lower neighboring vertebrae of each vertebra in a spine. In the original version of the criteria [85], only percentage of reductions were considered, while in the adjusted version [30] the reductions are evaluated while taking into account the standard deviation of the heights. The method is prone to false positive fractures, especially when the neighboring vertebrae to the vertebra in exam are also fractured. With the aim of reducing this drawback, McCloskey [83] proposed some improvements to the Eastell-Melton criteria which reduced the number of false positives. The method requires the fulfillment of two criteria to detect a vertebral fracture: a reduction of the ratios as in the Eastell's approach and a reduction in the ratios between the heights and their predicted posteriors. For each vertebra of each patient, a mean predicted posterior height is obtained from the predicted posterior heights of four adjacent vertebrae, avoiding possible issues related to fractured neighboring vertebrae.

2.3.2 Semi-Quantitative Morphometry

A different approach is given by the SQ methods, which are based on the subjective interpretation of expert radiologists. For example, the Genant semi-quantitative method [44] quantifies the presence or absence of vertebral fractures according to the apparent reduction of height in the vertebral bodies, and to specific deformation of the vertebral endplates.

In particular, fractures are classified according to their grade of severity as mild, moderate, and severe; and to the kind of deformation form which is generated: wedge, endplate, and crush. A summary of this classification is depicted in Figure 2.6. This figure shows, respectively for different grades, on the left column wedge fractures, on the central column endplate fractures and on the right column crush fractures. Although the height reduction is the main criterion, the Genant method also suggests taking in consideration changes at the endplate and cortical margin, and misalignment along the spine.

This subjective analysis can lead to a smaller false positive rate compared to QM approaches since it takes into account geometric artifacts in radiographs and does not consider non-osteoporotic fractures. Non-osteoporotic fractures, such as trauma fractures, are clearly recognized by the visual inspection of a radiologist, while a QM analysis can consider them indiscriminately since only the height reductions are taken into account. However, the subjectivety of SQ methods can introduce discrepancy among radiologists in case of mild fractures. With the aim of reducing the discrepancy among radiologist assessments, a series of steps to follow while using this method was proposed [65].

2.3.3 Fracture risk assessment

This section reports the standard clinical evaluation for fracture risk and other osteoporosis risk measurements which can implicitly be related to fracture risk.

Frax [76] is the most well-known approach for fracture prediction, it is a prognostic tool based on information such as BMD, age, use of corticosteroids, previous fractures and dietary habits. Its webinterface is depicted in Figure 2.7.

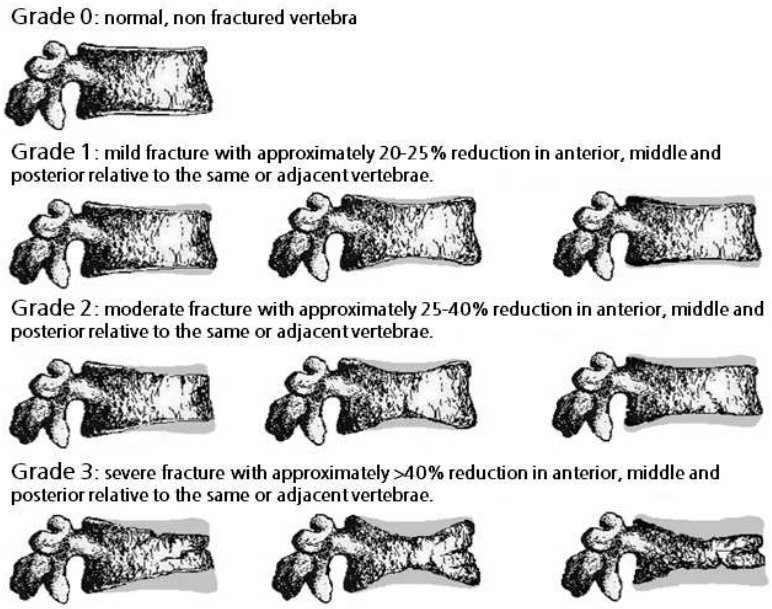


Figure 2.6: The grades of the standard Genant semi-quantitative method. Am J Med 94:64d, courtesy of IOF.

Although BMD is considered the most important indicator also for future fracture risk, since there is a relationship between loss of BMD and fracture risk [59, 90], a single score is only a reliable indication of the current bone density status, and several acquisitions would be required to identify bone loss rate. Volumetric BMD was recently proposed in several studies related to osteoporotic fractures [52]. The methodology is capable of quantifying the mineral density using quantitative Computed Tomography (qCT) in a much more precise way than standard BMD. This measurement quantifies the fracture risk according to skeletal site, type of fracture predicted, and population assessed [15, 52].

To predict osteoporosis (and therefore implicitly fractures) in a single reading, other solutions were proposed. Experiments using urinary or serum collagen type 1 cross-linked C-telopeptide (CTX) [16, 99], and urinary collagen type 1 cross-linked N-telopeptide (NTX) [63] gave interesting results. These collagens are markers of bone resorption, since high levels of CTX and NTX indicate excessive bone turnover and can be related to fracture risk [111]. It has been noted that women with high serum CTX have a 25% probability of hip fractures within 4-5 years, and women with low BMD and high CTX have a 54% probability of hip fractures within 4-5 years [42]. In parallel to these bio-markers, Ruegsegger introduced [56] a measure also based on qCT, where 3D structural parameters are derived from measurements of spatial structures such as the "caves" given by the trabecular bone structures [93].

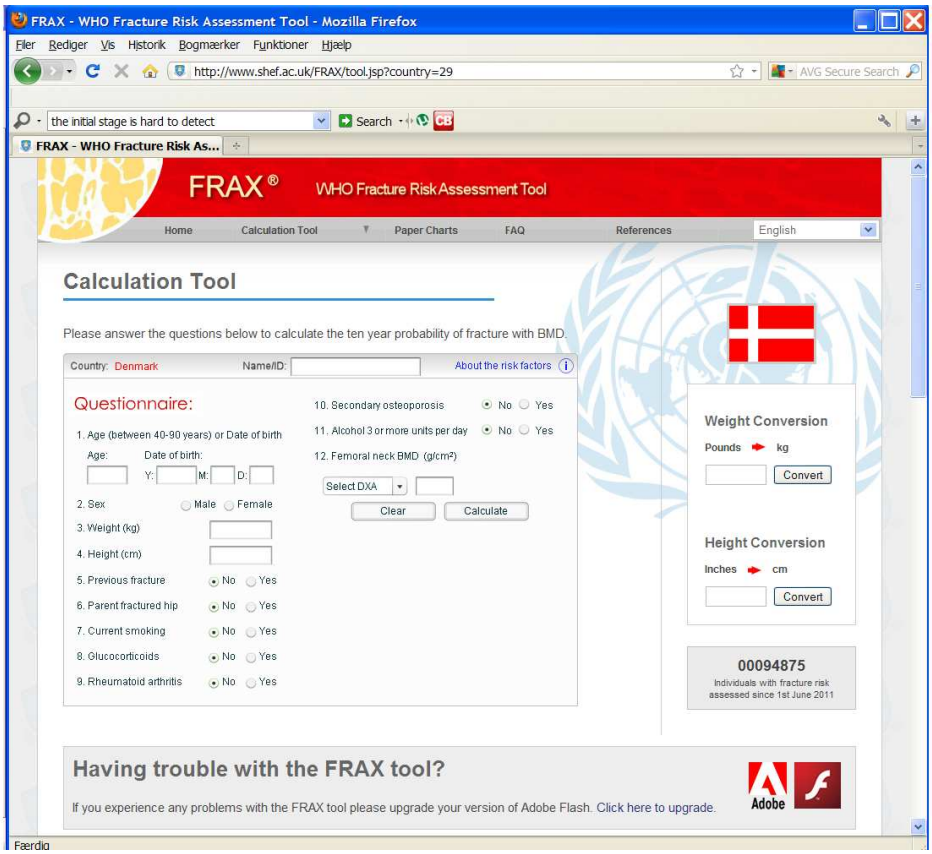


Figure 2.7: The webinterface of the Frax tool, http://www.shef.ac.uk/FRAX/.

The method estimates a volume-based local thickness by fitting a maximal sphere to every point in the structure. The spheres give a measure of the width of the trabecular structures which is then quantified as a mean plate thickness. The presence of small radius spheres is a good indicator of healthy trabecular structure and therefore of healthy bones.

Despite the promising results, clinical prediction of osteoporosis is not based on the evaluation of these bio-markers or on CT measurements, and the Frax tool is still the standard criterion for the evaluation of fracture risk.

2.4 Longitudinal case-control studies

In trials investigating osteoporosis and other clinical studies, a set of procedures is conducted to collect reliable information. This section introduces some terminology

related to these clinical procedures which will be used in the next chapters.

A case-control study is an epidemiological study where the patients are observed in order to determine their status to the exposure or non exposure to a specific event [101]. The case group is a non-randomized group of subjects either presenting a disease or being exposed to an event, while the control group is the matched reference group of individuals without symptoms or exposition to an event. The possible relationship of a suspected characteristic or risk factor, related to the disease or event in exam, is investigated through the comparison of these two groups [101]. An example of case group is a group of women smoking tobacco, and an example of control group is another matched group of non smoking woman, and the consequent correlation to either cardiovascular diseases or lung cancer is the target of the study.

A longitudinal study is the investigation of several observations of the same variables at different time points [101]. For instance, a breast cancer screening or a BMD measurement performed every 2 years in postmenopausal women. Generally, the initial observation is called "baseline" and the subsequent observations are "follow-up" observations.

In the study reported in Chapter 4, those two concepts are used jointly. In fact, the reported study analyzes two groups of women where the patients in the case group experienced at least a vertebral fracture, and the patients in the control group maintained their spine integrity. The longitudinal aspect of the study is referred to the fact that both patient groups had the spine status and lumbar vertebrae acquired twice. In particular, the patients of both groups were fracture free at baseline.

Chapter 3

Statistics on shapes in Computer Vision and Medical Imaging

If you need statistics to understand the results of your experiment, then you should have designed a better experiment.
Ernst Rutherford

The previous chapter contained the clinical background for osteoporosis and osteoporotic fractures. Although BMD and the other risk factors included in Frax represent the standard procedure for prediction of osteoporotic fractures, there is still room for further investigation on morphometric index for this purpose, and this work is concerned about the use of shape based methodologies for this purpose. This chapter describes briefly computer based shape representation in general, shape model for fracture risk assessment, and related statistical challenges such as few available samples. These subjects are included to provide some understanding of the field, for giving some terminology and since the methodology presented in the following chapters can be seen as an evolution of these methods.

3.1 Shapes in Computer Vision and Medical imaging

Paraphrasing D.G.Kendall, *a shape of an object is all the geometrical information that remains when location, scale, and rotational effects are filtered out from an object* [73]. However, people generally identify a shape of an object in an image simply with its contour oriented in a common center of frame. Several computational representations exist, some of them are based on Fourier descriptors [48], parametric descriptors [74], or landmark representations [19]. A survey describing these and other shape representations is given in [121]. Data regarding vertebral shapes is mainly available as manual annotations on x-ray images [60], and similar point based representations obtained by segmentation [22]. Therefore, the focus of this thesis is mainly on landmark representations.

In landmark based models, a geometrical representation can be constructed by the concatenation of a defined set of points/landmarks $p_1, p_2, ..., p_n$ lying on the boundary, assuming that these points correspond to relevant features. It can be argued that an object boundary is given ideally by infinitely many points (or better by a very high dimensional representation considering the limited resolution of an image), and reducing to few landmarks can cause some loss of information. However, small number of landmarks can be a sufficient representation for comparisons of shapes [105]. For example in medical imaging, this information can be used for registering MRI and other acquisition modalities [36], and for giving a qualitative description of human organs [12, 28] as well. Some work on shape modeling have been focused on learning shape models from training samples which can adapt to similar unseen shapes allowing some flexibility. These methods are called deformable template methods or deformable model methods, and their flexibility allows the necessary freedom to match the biological variation of structures and organisms in images. A survey of deformable methods and their extensions is reported in [84]. Since some of these methods are generally based on statistical analysis, the need of a sufficient amount of data for the statistics is necessary. If the available samples are few relatively to the dimensionality of the shape, then this will effect the performance. The second part of this chapter introduces this issue and some possible solutions, which will be discussed more in detail in Chapter 5.

3.1.1 Deformable template models and active contours

There are several approaches of deformable templates [84]. The oldest work is probably the spring-loaded template described in the 1973 pictorial structure of Fischler and Elshlager [35], also called "constellation model". In a pictorial structure, an object is represented by several parts arranged in a deformable configuration. The single parts are modeled independently, and the deformable configuration is given by the connections between pairs of parts. This qualitative description is useful for generic recognition problems [33], but it does not analyzes the object boundaries.

More recent deformable template algorithms, more focused on object boundary, are the active contour methods [71, 91]. Generally, an active contour model uses splines guided by an external and an internal energy minimization. The external energy modifies the active contour by taking into account features such as edges in an image, while the internal energy defines the manner in which the contour is allowed to move in the process of transformation (e.g. the level of smoothness). The prior knowledge which can be associated to this model is limited to the constraints related to the internal energy. The most well-known example of these guided splines approach is the "snakes" algorithm [71]. Snakes are planar deformable contours capable of slithering around an image, and they have been successfully applied in several tasks, and also to medical images [96, 118].

3.1.2 Active Shape Model and Point Distribution Model

A different approach from the snakes algorithm is given by the so called Active Shape Model (ASM) [19], which can be seen as a deformable template model where the prior knowledge is based on statistical information from a training data-set without information of contour smoothness. In practice, constraints, such as relevant anatomical landmarks, are associated to a shape (and texture) model, where two aspects regarding the landmarks are noteworthy. Firstly, the landmarks must be consistent in all the images, i.e., the landmarks need to correspond in different images. Secondly, those landmark based representations need to be centered to a common frame of reference, reducing the extrinsic variation of the shape in images due to positioning [73]. This alignment process is typically performed using the Generalized Procrustus Analysis (GPA) [11], which removes translation, scaling and rotation from a set of shapes. Landmarks can be defined manually as manual annotations or obtained automatically using segmentation algorithms.

ASMs are based on another statistical model called Point Distribution Model (PDM) [19], where every shape can be approximated to a mean plus some residual variation. This residual variation is given in the covariance matrix of shapes coordinates, since it represents the possible correlation between landmark coordinates in shapes.

3.1.3 Deformable shape loci and m-rep

Fritsch et al. [41] have devised a methodology called deformable shape loci, where medial loci, or cores, define the underlying structure of shape edges, allowing a robust to noise redefinition of the edges. The model also employs a probabilistic prior model relating relevant shape features and their spatial relationships. This idea allowed the formulation of a geometric framework for object modeling in medical image analysis based on a sampled medial representation called an m-rep [37]. An approach which provides a coordinate system for a more accurate representation of 3D image data.

This model is used in several medical imaging tasks but is not currently used on osteoporotic vertebral shape. It is, however, described since the experiments reported in Chapter 5 are also performed on knee cartilage sheets using this representation.

3.2 Computer based fracture assessment

In parallel with the established clinical methods for fracture assessment described in the previous chapter, some semi-automatic and automatic computer-based approaches using shape (and appearance) models [3,22–24,97,98,106] were proposed. These shape based methods rely either on manual annotations of each vertebra on a radiograph or on the result of segmentation algorithms:

The approach presented by Roberts et al. [3, 97, 98, 106] constructs a statistical shape model from vertebral shapes from DXA images and employs a linear discriminant classifier for identifying the fracture status. The verebral representation is given

by dense landmarks on the vertebral boundary obtained either by manual annotations [97] or semi-automatic segmentation [98] and an appearance model given by the textures comprised by the landmarks. A more detailed explanation is given in [20].

Another method also based on statistical shape models was proposed by De Brujine et al. [22–24]. Here, the landmarks are obtained as a discretization of a dense set of points non equally distributed. The method uses conditional shape models trained on a set of healthy spines, for designing an hypothetical vertebral shape for each vertebra in the image. The difference between the true shape and the artificially constructed shape is an indicator of shape abnormality. The conditional shape model for designing the vertebra uses pairwise statistics with the neighboring vertebrae, considering the probabilities of each vertebra jointly. Moreover, the method uses a discriminant classifier for grading the fracture, giving a continuous measure of deformity, and introduces the problem of needing a large sample size for classification tasks.

These two approaches have been investigated for diagnostic purposes, while little work was done so far for fracture risk assessment. Previous computer-based studies involved normalized vertebral heights [27, 79] and curvature analysis [95] for analyzing slight deformities which can cause future fractures. The normalized heights approaches [27, 79] try to identify possible pre-fracture deformities searching irregularity among the vertebral heights. In [79] vertebral heights were assumed to be normally distributed, following a bi-variate normal distribution and classified according to it. In [27], the correlation between aging and height reductions is argued. More specifically, the sum of anterior heights is proposed as a biomarker for ostoporotic vertebral fractures. Instead, the curvature approach [95] argues that local irregularities in the alignment of the lumbar vertebrae contribute to increase the fracture risk. The motivation lies in the hypothesis that there is a correlation between curvature of the lumbar spine and irregularity of the vertebrae alignment, which will lead to challenges to maintain structural integrity. The study presented in Chapter 4 is a methodology which comprises some of these aspects for future fracture prediction.

3.3 Regularization for statistical shape models

Active shape model and other deformable templates methods require the estimation of covariance matrices as one of the initial steps. When the dimensionality of the shape space is high, and the available number of samples is small, then the estimation of a covariance matrix with standard methods can be poor. In osteoporosis studies and in other clinical contexts, this circumstance can occur, since the number of patients available can be very small compared to the required dimensionality of the vertebral and spine shape.

In particular, if we consider the standard MLE method, the obtained covariance matrix may be rank deficient, implying that some eigenvalues have magnitude close to zero, and that the corresponding eigenvectors are arbitrary. This limitation encouraged the introduction of more robust covariance matrix estimators [31, 40, 53, 57, 64, 103, 108, 113]. Regularization, and specifically shrinkage, have

been proposed to improve estimates of covariance matrices. Shrinkage methods aim at improving estimates by shrinking the estimation towards zero or in the general setting towards some specific value. The oldest shrinkage methods were proposed in the early 60's [108, 113], and afterwards a series of similar methods appeared [31, 40, 53, 57, 64, 103]. These methods generally depend on the proper choice of a mixing parameter, which is often selected to maximize the expected accuracy of the shrunken estimator by cross-validation [40, 57] or by using an analytical estimate of the shrinkage intensity [31, 53, 64, 103]. Regularized estimators have been shown to outperform the standard ML estimator for small sample sizes [103]. Most commonly a simple form of Tikhonov regularization is used [57], where non-zero values are added to the diagonal elements of the covariance matrix as $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{reg}} = \hat{\boldsymbol{\Sigma}} + \alpha^2 \boldsymbol{I}$. Here, α^2 is the mentioned mixing parameter. This regularization parameter is not likely to be known in advance, and finding its optimal value can be cumbersome and computationally demanding [40]. The analytical methods for covariance estimation [31, 53, 64, 103] propose several ways to estimate analytically the mixing parameter. These methods estimate the optimal value analyzing the sample to sample variance of entries in $\hat{\boldsymbol{\Sigma}}$ as

$$\hat{\alpha}^2 = \frac{n}{(n-1)^2} \frac{\sum_{i,j=1}^{d} var_k(z_{ij}(k))}{\sum_{i \neq j} s_{ij}^2 + \sum (s_{ii} - \nu)}, \tag{3.1}$$

where $z_{ij}(k)$ is the difference between the centered $i-th$ data element and the centered $j-th$ data element, n the number of samples, s_{ij} is the $ij-th$ element of the sample covariance matrix, and ν is the mean eigenvalues. However, either using a cross-validation for finding the optimal value or the automatic estimation, the method does not correct the case of non-optimally estimated eigenvectors.

Probabilistic PCA considers PCA as the Maximum Likelihood solution of a probabilistic latent variable model [114]. This method finds the non-corrupted eigenvalues/eigenvectors, but it is not always able to regularize the covariance matrix optimally for the point of view of eigenvectors as well. A similar method was proposed by Allassonniere et al. [1], where kernels in Hilbert space are used for estimating the covariance matrix of a generative model.

In Chapter 5, some alternatives to the aforementioned regularization methods are proposed.

Chapter 4

Shape-based Assessment of Vertebral Fracture Risk in Postmenopausal Women using Discriminative Shape Alignment

This chapter is based on

- "Shape Based Assessment of Vertebral Fracture Risk using Discriminative Shape Alignment in Postmenopausal Women" Alessandro Crimi, Marco Loog, Marleen de Brujine, Mads Nielsen and Martin Lillholm. Academic Radiology, in press.
- "Morphometric Assessment of Vertebral Fracture Risk in Postmenopausal Women using Vertebral Shape Models and Discriminative Shape Alignment." Alessandro Crimi, Marco Loog, Marleen de Bruijne, Mads Nielsen, Martin Lillholm. Abstract Annual meeting American Society for Bone and Mineral Research 2011 San Diego, California, USA 2011.

Rationale and Objectives. Risk assessment of future osteoporotic vertebral fractures is currently mainly based on risk factors, such as Bone Mineral Density, age, prior fragility fractures, and smoking. It can be argued that an osteoporotic vertebral fracture is not exclusively an abrupt event but the result of a decaying process. For evaluating the fracture risk, a shape based classifier, identifying possible small pre-fracture deformities, may be constructed.
Materials and Methods. During a longitudinal case-control study, a large population of postmenopausal women, fracture-free at baseline, were followed. The 22 women who sustained at least one lumbar fracture by follow-up represented the case group. The control group comprised 91 women who maintained skeletal integrity, and matched the case group according to the standard osteoporosis risk factors. On radiographs, a

radiologist and two technicians independently performed manual annotations of the vertebrae, and a fracture prediction using shape features extracted from the baseline annotations was performed. This was implemented using posterior probabilities from a standard linear classifier.
Results. The classifier tested on our population quantified the vertebral fracture risk, giving statistical significant results: AUC = 0.71 ± 0.013 and Odds Ratio = 4.9 with [2.94 ; 8.05] 95% C.I. for the radiologist annotations.
Conclusion. The shape based classifier provided meaningful information for the prediction of vertebral fractures. The approach was tested on an osteoporosis risk factors matched case-control group. Therefore, the method can be considered as an additional bio-marker, which combined with the traditional risk factors can improve population selection, e.g., in clinical trials, identifying patients with high fracture risk.

4.1 Introduction

Osteoporosis is a disease, characterized by a decrease in density and quality of bone, potentially leading to fractures. Vertebral fractures are the most common type of osteoporotic fractures [110], and they may occur in absence of trauma or after minimal trauma. In this case they are also called fragility fractures. Moreover, the presence of vertebral fractures increases the risk of hip fractures [7] and additional vertebral fractures [86]. A growing percentage of elderly people develop osteoporosis and experience osteoporotic fractures, and consequently the related burden sustained by the medical system is heavy and increasing rapidly [110]. The total cost for the year 2000 in Europe has been estimated to 31.7 billion Euros and it is foreseen to grow up to 76 billion Euros by the year 2050 [68], while the estimated expenses for USA the year 2005 were 19 billion Dollars [14]. Therefore, early assessment of fracture risk is crucial for reducing the number of fractures through appropriate treatments, for improving the life quality of patients, and reducing costs to the health system.

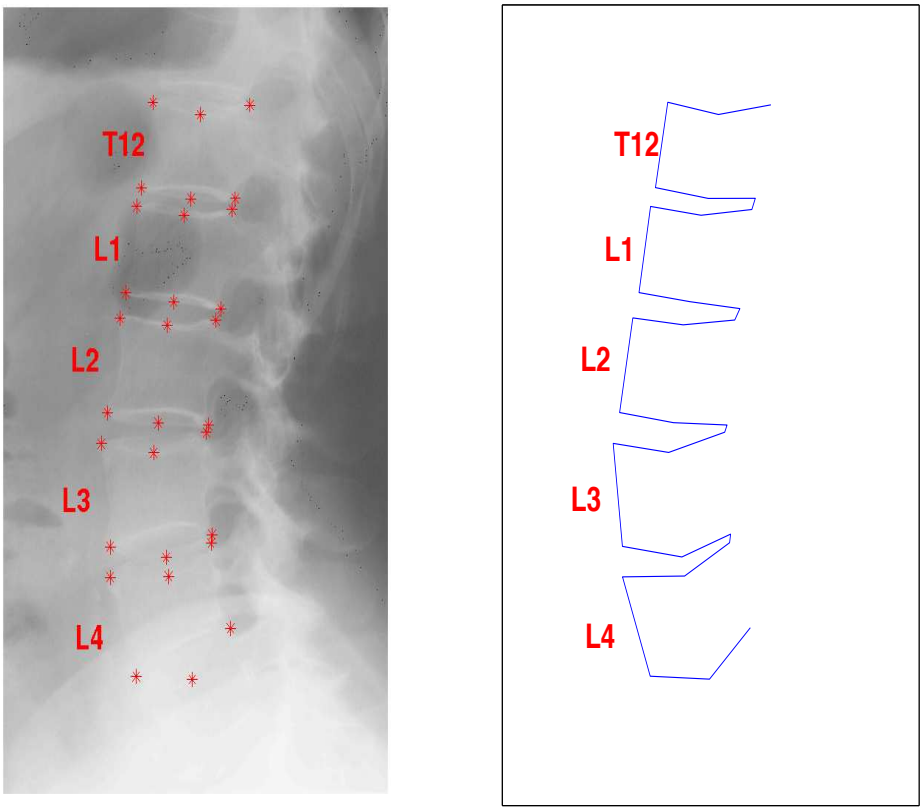
Several clinical factors can provide information on fracture risk. These can be Bone Mineral Density (BMD), age, prior fragility fractures, smoking, use of corticosteroids, excessive consumption of alcohol and rheumatoid arthritis [68]. The World Health Organization (WHO) has formalized the analysis of such risk factors in the FRAX tool [76], which gives a probability of hip fractures over a 10-year period and a probability of a major osteoporotic fracture (located either in spine, forearm, hip or shoulder) over a 10-year period as well. The FRAX tool is available as a web-service and as a paper version with guided charts [39].

Diagnostic morphometric approaches include Semi-Quantitative methods (SQ) performed using subjective judgment of an expert radiologist [46], and fully Quantitative Morphometry methods (QM) based on the measure of relative height reductions in vertebrae [30, 83]. In particular, each vertebra on a radiograph had the four corners and the mid-points of the endplates annotated according to Hurxthal criteria [60]. Figure 4.1(a) shows the overlay of a typical lumbar spine radiograph, and its rela-

tive annotation. Both methods have advantages and disadvantages. Using only the height measurements of a QM method, subtle shape information could be lost or non-osteoporotic fractures, such as trauma fractures, could be indiscriminately considered. Some of this information is used in SQ approaches. Hence, the Genant SQ method has become one of the common standards for fracture assessment even if it can introduce discrepancy among radiologists, particularly for mild fractures [45]. With the aim of reducing the discrepancy among radiologist assessments, a series of steps to follow while using this method was proposed [65]. In parallel with these established clinical methods, some semi-automatic and automatic computer-based approaches were proposed for quantifying vertebral fractures using a shape model [3, 23, 24, 97, 98]. The approach presented by Roberts et al. [97, 98] constructed a statistical shape model from a vertebral representation (and an appearance model), and employed a linear discriminant classifier for identifying the fracture status. Another method also based on statistical shape models was proposed by De Brujine et al. [23, 24]. The method used conditional shape models trained on a set of healthy spines, for designing an hypothetical vertebral shape for each vertebra in the image. The difference between the true shape and the artificially constructed shape was an indicator of shape abnormality, and therefore of possible fracture.

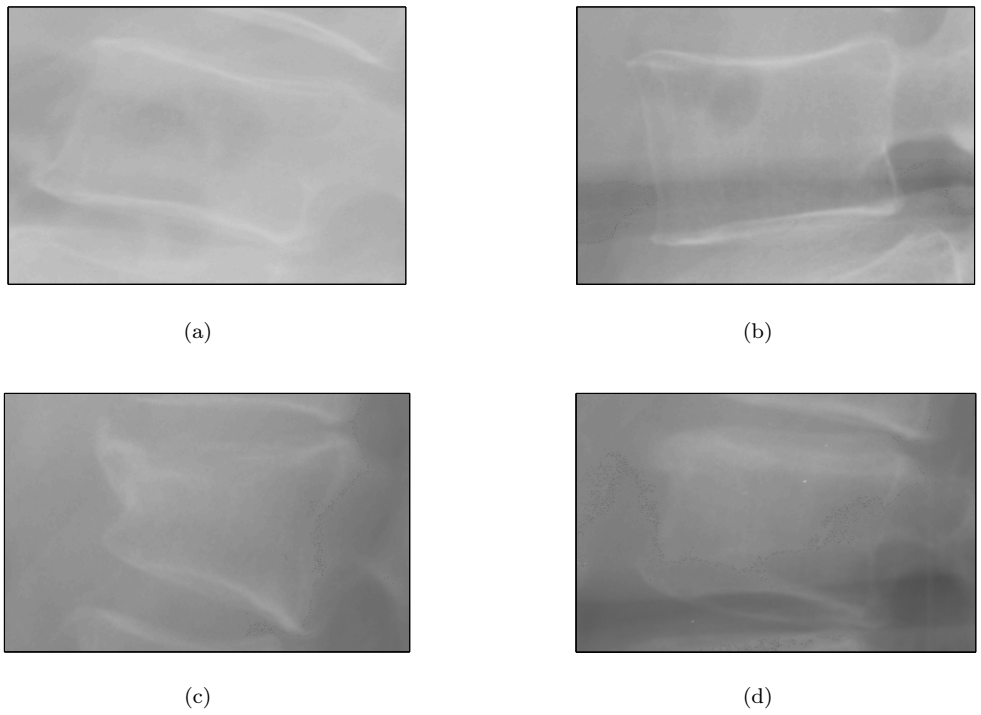
Regarding prediction of osteoporotic vertebral fractures based on morphometric indices, less investigation has been carried out so far. Previous studies involved normalized height measurements [27, 79], curvature analysis [95] and single vertebral shape analysis [80] for analyzing slight deformities which can cause future fractures. Some studies about the relation between vertebral dimensions and fracture risk were also conducted [102].

The aim of the method described in this manuscript was to summarize both the aspect of spine curvature and vertebral heights in an overall shape representation, giving a more complete model for predicting vertebral fractures. A similar approach has been carried out for the prediction of incident hip fractures [5, 51]. In practice, the method employed shape models, similarly as previously described for [97], but performing prediction of osteoporotic fracture. The prediction is carried out by examining baseline data where no fractures were present, identifying possible small pre-fracture deformities of the spine shape that can result in future fracture. This set-up has been already proposed [79, 80, 95] and it is here re-evaluated deeply considering spine shapes of the lumbar region, evaluating inter-annotator reproducibility, and using a sophisticated cross-validation methodology. It can be argued that an osteoporotic fracture should not be considered exclusively as an abrupt change in a vertebra, but as a part of a slow deforming process involving a decrease in density and quality in the bones, and the consequent bio-mechanics of the whole spine [13]. Since the related changes in spine shapes and vertebral heights could be slow and not very pronounced, they could not be always visible to the naked eye. Figure 4.2(a) and Figure 4.2(b) depict respectively a baseline example of a vertebra that fractured by follow-up and a vertebra that did not. Figure 4.2(c) and Figure 4.2(d) show the follow-up radiographs, acquired about 7 years later, of the same vertebrae respectively in Figure 4.2(a) and 4.2(b).



(a) (b)

Figure 4.1: On the left (a), overlay of the lumbar region in an x-ray image (case baseline) and its 6 point annotation given by the stars. On the right (b), the spine shape that can be obtained from the concatenation of the 2D-landmarks. The annotations were performed considering coherently the lowest endplate of the vertebrae. The contrast in the reproduction above is most likely too low to clearly visualize all the lower boundaries. For the experiments reported in this study, a spine shape was used as the collection of these landmarks without any interpolation.



(a) (b)

(c) (d)

Figure 4.2: (a) Example of baseline vertebra which fractured by follow-up, (b) example of baseline vertebra which did not fracture by follow-up, (c) and (d) are follow-up radiographs of the same vertebrae depicted respectively in (a) and (b).

4.2 Materials and Methods

4.2.1 Data

In this study, we used a case-control cohort of post-menopausal women obtained as follows. A large population of 4062 Caucasian postmenopausal women were followed for a mean period of 7.3 years during a previous study [4], and a reanalysis of a relatively small subset of this cohort was used for this study. For each patient, BMD at the spine were assessed, and lateral radiographs of the spine were acquired. The lumbar radiographs and BMD measurements were taken twice, once in 1992-95 (baseline) and again in 2000-01 (follow-up), and the vertebrae were graded for their fracture status by an experienced radiologist, according to the semi-quantitative method of Genant [46]. The Genant semi-quantitative method quantifies the presence or absence of vertebral fractures according to the apparent reduction of height in the vertebral bodies, and to specific deformations of the vertebral endplates. Any patients whose images contained non-osteoporotic deformities and/or fractures were excluded; whether these were real

or caused by projection errors. Patients, whose radiographs had the area surrounding the T12 vertebra too dark, were also removed. The reason for these dark/low contrast regions is related to the acquisition protocol. The distance between the focal plane and the film was kept constant at 1.2m and the central beam was directed to L2. Due to biological variation among the patients some of the radiographs were not clear near the edges where T12 is typically depicted. The fracture status of the spines was re-evaluated by a computer algorithm using a modification of Genant's methodology with a strict measured threshold, where a vertebra was considered fractured if either of the ratios between any of the anterior, middle, and posterior heights was 0.8 or smaller. In particular, the ratios were obtained dividing the biggest height by the smallest height of the vertebra in exam and the threshold was previously defined in [79]. This additional step was deliberately used for filtering out subjects where there was borderline disagreement between the SQ and QM based fracture assessments, preventing the algorithm from being influenced by such a disagreement. All subjects were fracture-free at the time of the first acquisition according to both the SQ and QM method. The patients removed due to the disagreement between the QM and SQ method were 11 cases and 7 controls. The 22 patients of this population, who sustained at least one incident osteoporotic lumbar fracture by follow-up, were selected as the case group. While the control group contained 91 patients who maintained skeletal integrity from baseline to follow-up. The patients in the control group were matched to the patients in the case group according to traditional osteoporosis risk factors such as age, height, weight, and spine BMD. The match was performed in the study of Pettersen et al. [95] over the sub-set of patients that were fracture-free at follow-up, selecting approximately 5 random control patients for each case patient to ensure an acceptable total number of subjects.
Written consent was obtained from each participant according to the Helsinki Declaration II. The study was approved by the local ethics committee. The methodology described in this manuscript was considered in a case-control population matched for the standard osteoporotic risk factor, to evaluate the predictive value of the methodology independently from the other risk factor [4].

The used images were digitized lateral radiographs, obtained using a Vidar DosimetryPro Advantage scanner at 570 DPI resulting in images of 9651×4008 pixels in 12-bit gray scale with a pixel size of $44.6\mu m \times 44.6\mu m$. Each image was used to define annotations with 6 points [60] for each vertebra as in Figure 4.1(a). A first set of annotations was performed by radiologist PP. To facilitate inter-annotator validation the same radiographs were also annotated independently by two trained technicians (JP and AO). The annotated vertebrae were four of the five lumbar vertebrae L1-L4 and the lowest thoracic T12. The L5 vertebra was not taken in consideration due to the fact that it is often covered by the hip and therefore difficult to see and annotate.

4.2.2 Methods

Our aim was to evaluate whether an unseen spine has a high fracture risk or not. For this purpose, we designed a machine learning algorithm, which, given a set of spines

as described in the previous section, was capable of assessing the fracture risk of an unseen spine representation. Only baseline spine data were used, since the prediction was supposed to be performed when all the spines were still considered fracture free. The algorithm comprised a training phase reported in the following first four steps, and a classification phase regarding the unseen spine described in the fifth step:

1. Compute the spine mean landmark shapes $\boldsymbol{\mu}$ or a landmark reference shapes $\boldsymbol{\mu_{ref}}$ from the spine shapes of the used cohort (see Subsection 4.2.2 and 4.2.2).
2. Align all the spine shapes to the common reference of frame given by $\boldsymbol{\mu}$ or $\boldsymbol{\mu_{ref}}$ (see Subsection 4.2.2 and 4.2.2).
3. Build a statistical shape model from the landmark coordinates of the spine shapes, and represent the shapes using this model (see Subsection 4.2.2).
4. Train a classifier using the shape information obtained from step 3 (see Subsection 4.2.2).
5. Quantify the fracture risk of the unseen shape in an unbiased way, using the alignment, the statistical model and classifier trained in step 2, 3 and 4.

In particular, the shape alignment was performed using two different approaches, the well-known Generalized Procrustes Analysis (GPA) [11, 49] described in Subsection 4.2.2, and the supervised approach of Discriminative Shape Alignment (DSA) [80] described in Subsection 4.2.2, where the alignment took into consideration to which group each training shape belongs.

Shape Alignment

Vertebral bodies were not usually depicted in the same orientation or scale in radiographs mainly due to biological variation and positioning. Therefore, the shapes were aligned to a coherent representation before further analysis.

In our experiments, we used the GPA technique, which removed translational, rotational, and scaling differences from the shapes preserving the intrinsic geometry. GPA estimates a mean shape $\boldsymbol{\mu}$ from a data-set and aligns all the shapes to this mean. These two steps are repeated several times until convergence criteria are satisfied. The process is based on the minimization of the Procrustes distance between two shapes as

$$d_{GPA}(\boldsymbol{x_i}, \boldsymbol{x_j}) = ||\boldsymbol{W_i x_i} - \boldsymbol{x_j}||^2, \tag{4.1}$$

where $\boldsymbol{W_i}$ is a similarity transformation matrix, which can be found either by using a closed form solution [28] or in an iterative fashion [10], and x_i and x_j are specific shapes.

Point Distribution Model

After the shapes were aligned to a common frame of reference, the remaining variance with respect to the estimated mean shape can be modeled using a statistical methodology called Principal Components Analysis (PCA) [66], allowing the construction of a Point Distribution Model (PDM) [19]. PCA is an orthogonal linear transformation which projects the values of given variables into a new coordinate system. The basis components of this new space are called *principal components*, and they define the allowable shapes. Such components are ordered in a descending order according to the variance of the observed variables. Selecting a subset s of the total q basis with $s \leq q$, the projection can reduce the dimensionality of the data, and therefore the noise, allowing possible better performance in some tasks [6].

In detail, the projected shape features were obtained from the 2D-landmark annotations in the following way: each shape was described as a vector $\boldsymbol{x}$. Those n shapes were afterwards aligned, mean-subtracted and collected in a sample matrix $\boldsymbol{X} \in \mathbb{R}^{q \times n}$. In our experiments, q was equal to 60 since we used 30 2D-landmarks. The covariance matrix $\boldsymbol{\Sigma} = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T$ can be decomposed as

$$\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T, \tag{4.2}$$

where $\boldsymbol{V} = [\boldsymbol{v}_1|\ldots|\boldsymbol{v}_q]$ is the column matrix of eigenvectors, and $\boldsymbol{\Lambda} = \text{diag}\left([\lambda_1, \ldots, \lambda_q]\right)$ is the diagonal matrix of corresponding eigenvalues [66]. These eigenvectors and eigenvalues could be used for representing the shapes into the new coordinate system as

$$\boldsymbol{Y} = \sqrt{\boldsymbol{\Lambda}^{-1}}\boldsymbol{V}^T\boldsymbol{X}. \tag{4.3}$$

This equation represents the linear transformation PCA is performing on the data $\boldsymbol{X}$ to yield the transformed $\boldsymbol{Y}$.

Through equation (4.3), PDM [19] represents each spine shape as the sum of the mean shape and an approximation of the s largest principal components as

$$\tilde{\boldsymbol{x}} = \boldsymbol{V}\text{diag}\left(\left[\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_s}, 0, \ldots, 0\right]\right)\boldsymbol{y} + \boldsymbol{\mu}, \tag{4.4}$$

where $\tilde{\boldsymbol{x}}$ is the approximation of $\boldsymbol{x}$ using this model, $\boldsymbol{y}$ is the column vector in $\boldsymbol{Y}$ corresponding to the column vector in $\boldsymbol{X}$, and $\text{diag}\left(\left[\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_s}, 0, \ldots, 0\right]\right)$ is the diagonal matrix containing a subset of eigenvalues. The optimal number of used components s represents a pivotal parameter which needs to be estimated. We describe in the next Section 4.2.3 how this selection was performed in our experiments.

Linear Discriminant Classifier

The main goal of our framework was to evaluate whether there was a correlation between a spine shape and a high risk of future vertebral fractures. For this purpose, representing the spine shapes as described in the previous section allowed their representations in a space of s dimensions, where a Linear Discriminant Analysis

(LDA) [29] can be used. The LDA classifier was trained according to the corresponding categorical class labels of each shape. In practice, label values of -1 to control (or non future fractured vertebrae), and $+1$ to case (or future fractured vertebrae) were assigned.

LDA is a statistical methodology searching for a linear combination of features or measurements which characterize or separate two or more classes of objects. LDA separated the two classes by assuming that the conditional probability density functions are given by the normal distributions $N(\boldsymbol{\mu_{case}}, \boldsymbol{\Sigma})$ and $N(\boldsymbol{\mu_{control}}, \boldsymbol{\Sigma})$, where those distributions were defined by the empirical mean of the case and control groups respectively, and by the pooled covariance estimated from the samples of both groups. Our aim was to evaluate how high the probability was of a new spine shape $\boldsymbol{x}$ to belong to the case group, and therefore to sustain a fracture. Hence, after the approximation of $\boldsymbol{x}$ to $\tilde{\boldsymbol{x}}$, the risk score was assessed as

$$P_{case}(\tilde{\boldsymbol{x}}|\boldsymbol{\mu_{case}}, \boldsymbol{\Sigma}). \tag{4.5}$$

By normalizing the score as $\overline{P_{case}} = P_{case}/(P_{case} + P_{control})$, the probability of fracture risk in the interval between 0 and 1, where values close to 1 represented high risk of vertebra fracture and 0 low risk. Figure 4.3 depicts, from left to right, some spine shapes ordered from low to high fracture risk as defined by the described probability. The picture illustrates how difficult is to see the slight differences in the vertebrae shapes and spine curvature without accurate measurements.

Figure 4.3: Examples of baseline spine shape, ordered from left to right accordingly to their probability of belonging to the case group based on our fracture risk score. For the experiments reported in this study, the spine shapes were used as the collection of the landmarks without any interpolation.

Discriminative Shape Alignment

The algorithm described in Subsection 4.2.2 aligned all the shapes in a fully unsupervised way, without taking into account whether the shapes belonged to the future

fractured group or not. Using the class label information during the alignment can benefit the performance of a subsequent classification [80], which was our final goal. The idea was to use a linear regression for the iterative alignment, involving the class labels $\{k_i\}$, obtaining the so called DSA [80]. Given n shapes $\{\boldsymbol{x_i}\}$ and their labels $\{k_i\}$ indicating a future fractured spine or not, it is possible to find a reference shape $\boldsymbol{mu_{ref}}$ to which all the $\{\boldsymbol{x_i}\}$ shapes should be aligned, such that the two shape classes are as far as possible from each other. To perform this supervised alignment two processes need to be merged: a discriminative linear regression [55] which discriminates the shapes and an analytic solution of the Procrustes alignment.
The discriminative regression solution can be obtained by finding the vector $\boldsymbol{a^T} \in \mathbb{R}^n$ that minimizes

$$\Sigma_{i=1}^{N}(\boldsymbol{a^T x_i} - k_i)^2, \tag{4.6}$$

where $\boldsymbol{a^T}$ is the discriminative linear regression vector.
Representing a shape x_i as a complex normalized vector $z_i \in \mathbb{C}$, it is possible to align it to the mean shape $\boldsymbol{m} \in \mathbb{C}$ [28] by the following operation:

$$\boldsymbol{z_i^{'}} = \boldsymbol{z_i z_i^H m}. \tag{4.7}$$

Here, z_i^H represents the Hermitian of the shape vector. To use this alignment in conjunction with the minimization of equation (4.6), the complex vectors must be converted to real values yielding the expression:

$$\Sigma_{i=1}^{N}(\boldsymbol{a^T Z_i \mu_{ref}} - k_i)^2, \tag{4.8}$$

where $\boldsymbol{Z_i} = \begin{pmatrix} Re(z_i z_i^H) & -Im(z_i z_i^H) \\ Im(z_i z_i^H) & Re(z_i z_i^H) \end{pmatrix}$. The algorithm is initiated with the mean shape as the $\boldsymbol{\mu_{ref}}$, and iteratively updates the reference shape in conjunction with the projection vector $\boldsymbol{a^T}$. Expression (4.8) cannot be solved in closed-form since it represents an intertwined linear regression problem [80].

Therefore, an iterative schema should minimize equation (4.8) alternating two regressions by every time keeping fixed either $\boldsymbol{a^T Z_i}$ or $\boldsymbol{\mu_{ref}}$.

Since our spine shape representation was up to $q = 60$ dimensions and only 113 spine shapes were available, the two regressions can define an ill-posed problem [113]. Hence, in each alignment iteration regularization [54, 57] had to be introduced [80]. Therefore, the iterative schema was defined by the following steps:

1. Compute $\boldsymbol{a^T} = \boldsymbol{K}(\boldsymbol{F^H F} + r\boldsymbol{I})^{-1}\boldsymbol{F^H}$.
2. Re-project the shapes using the updated $\boldsymbol{a^T}$ and update the matrix $\boldsymbol{F}$.
3. Update the mean as $\boldsymbol{\mu_{ref}} = (\boldsymbol{F^H F} + r\boldsymbol{I})^{-1}\boldsymbol{F^H K}$.
4. Re-compute the $\boldsymbol{F}$ matrix.

Here, $\boldsymbol{F}$ is the multidimensional matrix containing all the $\boldsymbol{Z_i}$ matrices of all the shapes multiplied by $\boldsymbol{\mu_{ref}}$, $\boldsymbol{K}$ the vector containing all the labels k_i, and r is a

regularization-parameter which controlled the degree of regularization. The estimation of the regularization parameter was similar to the search of the optimal number of eigenmodes and it is described in the following section.

Previous investigations [80] demonstrated that this discriminative alignment can reduce the number of false positive compared to the traditional GPA.

4.2.3 Experiments

The experiments were performed in a cross-validation manner over the entire data-set, considering independently the three sets of annotations. Each spine annotation was randomly selected in turn to be the test shape, the remaining shapes from the data-set were used once again in a cross validation manner to obtain the optimal parameters subsequently tested on the test shape, where optimal is referred to the parameters yielding the highest Area Under the ROC-curve (AUC), as explained in Subsection 4.2.4. In this way, the alignment, the projection and the classifier were trained excluding the current test shape. For each classified spine test shape, we then recorded the classifier measures of fracture risk. The final performance was also measured using the AUC. This partitioning was called *Balanced-Nested- Leave-Out* (BNLO), since it maintains a balance between the classes removing at each iteration one shape from the case group and the proportional from the control group (5 for the current cohort). The algorithm can be performed for several random samples where one single shape could be scored several times while different shapes from the other class were removed randomly. For the GPA version of our algorithm the only needed parameter was the number of used eigenmodes s as in equation (5.5), while for the DSA version, the parameter space is also given by a regularization parameter during the optimization. The steps involved in the BNLO algorithm for the DSA case are summarized in the following Pseudocode 1. The BNLO algorithm comprised a two nested layers of cross-validation. For n random shapes, the outer layer selected a test shape and scored it using a tuned shape model and classiffier after a tuning phase which occurred in the inner nested layer. The inner layer computed the AUC of further n random shapes for all the parameter values in an unbiased way, removing once again iteratively a test shape from a class and the proportional of the other before the training, and testing the inner-layer test shapes on the trained shape model and classiffier. The parameters yielding the highest AUC obtained here were selected for tuning the shape model and classifier in the outer layer. At this point, the outer-layer test shape removed in the begin can be tested. The probability scores obtained in the outer layer were used to define the AUC and Odds Ratio (OR).

The algorithm was meant to be run for a specific number of randomly selected shapes, where the selection is performed with replacement. However, to speed up the process, the algorithm can be applied in parallel and the number of random samples distributed over the parallel repetitions. In our experiments, for example, instead of using 3200 random samples in both cross-validation layers in a single run, 10 parallel repetitions of 1000 samples each were used, obtaining the proportional of $(3.200 \times 3.200) \approx 10 \times (1.000 \times 1.000)$ iterations. The algorithm comprises three major

parts: alignment, training of the classifier, and risk scoring. Only the alignment is different in the two versions of the algorithm. For a single alignment, the GPA algorithm employs 10 ms while the DSA algorithm requires 170 ms. The training and testing employ respectively 30 ms and 2 ms. These timings are obtained times on a 2.2 GHz dual core Intel processor. These operations are repeated thousands of times during the cross-validation described above.

4.2.4 Statistical analysis

The patients of the control group were selected matching them to the patients in the case group according to the standard risk factors for osteoporosis. To validate this match, the p-values representing the two populations comparison were computed using the Wilcoxon rank sum test as reported in the next section, along with the mean values and standard deviations for the major biological information. Non-significant p-values guaranteed the similarity of the two classes. The performance of the prediction algorithm was validated computing the AUC for each set of annotations, the p-values of such curves compared to chance using the Delong test [25], and the Odds Ratio (OR) of the classification with 95% confidence interval. P-values smaller than 0.05 were considered significant. The use of the cross-validation approach described in the previous subsection assured unbiased results. However, the BNLO algorithm employed more samples than "really" present in the data-set, since we used the same samples several times, causing very small misleading p-values due to the normalization during the Delong test. To avoid this unwanted side effect, we averaged the scores with respect to the same test shape in exam. Furthermore, the BNLO algorithm was run for 10 iterations (see Subsection 4.2.3) and the scores averaged with respect to the same test shape also over the different repetitions. The inter-annotator reproducibility was assessed using the coefficient of variation (CV) calculated as follows: $CV\% = (\sigma_{height}/\mu_{height}) * 100$. This measure was used to assess the three heights variability of the five vertebrae. The root mean square (RMS)-CV was used to aggregate over vertebrae and subjects yielding one RMS-CV for the anterior, middle, and posterior heights respectively.

4.3 Results

The patients of the two classes were successfully matched for BMI, BMD, age, height, weight, physical activities and smoking habits as reported in Table 4.1. The table also comprises the median values for the two groups and the mean and standard deviation between parentheses. Figure 4.4 depicts the ROC-curves for the three distinct sets of annotations using the two different alignment methods. Figure 4 (a) and Table 2, reporting the results for the GPA version of the algorithm, show significant values using the PP set of annotation and not statistically significant values using the other sets of annotations. The curves depicted in Figure 4(b) and the values of Table 3 show better or relatively similar results using the discriminative alignment approach, compared to the results obtained using GPA; with this setting the AUCs were significant

Pseudocode 1 BNLO algorithm using the DSA version:

```
n_class = 2 {case,control}.
n_random_shapes = 1000 (selected with replacement)
n_eigenvalues = maximum dimension(in our case 60)
regularization_parameters = linspace(0,1,100).
n_parallel_repetition = 10

Repeat for n_parallel_repetition
  For all the n_random_shapes samples do
    1. Select a random test shape x, and remove
       the proportional random shapes U from the other class.

    For all the n_random_shapes samples do
      2. Select another test shape w and  proportional J shapes
         of the other class from the data-set {case, control}\{x,U}.

         For all the regularization_parameters
          For all n_eigenvalues
            3. Train a shapemodel  using {case, control}\{x,w,U,J}.
            4. Train a classifier using the features obtained in step 3.
            5. Score the test shape w.
          end
         end

    end of the parameter evaluation phase

    6. Select the optimal number of eigenvalues
       and regularization_parameters giving the max AUC.
    7. Train a shapemodel using {case,control}\{x,U} and the optimal
       value given in the previous step 6.
    8  Train the classifier using the resulting features of step 7.
    9. Score the test shape of the shape x using the shapemodel and
       classifier obtained trained in steps 7 and 8.

  end of test phase
end of parallel repetition

10. Average the scores of the previous loop  with respect to the test shapes.
11. Use all the scores for computing a ROC-curve, AUC and p-values.
```

for all sets of annotations.

The inter-annotator reproducibility of the annotations has been quantified using the CV of the three heights: 1.17% for the anterior, 2% for the posterior, and 0.64% for the central height. These measurements express the different heights of the 5 vertebrae using the RMS of the CV.

Feature	**cases(n=22)**	**controls(n=91)**	**p-value**
Age ($year$)	64.7 (66.9 ± 5.4)	65.4 (66.6 ± 5.9)	p= 0.98
Height (cm)	162 (163 ± 4.6)	161 (161 ± 6.0)	p= 0.21
Weight (kg)	64.5 (68.3 ± 11.7)	65.4 (65.4 ± 8.4)	p=0.53
BMI (kg/m^2)	24.8 (25.5 ± 4.0)	24.6 (25.0 ± 3.2)	p=0.72
SpineBMD (g/cm^2)	0.79 (0.81 ± 0.14)	0.84 (0.86 ±0.14)	p=0.20

Table 4.1: Biological data used for matching the case and control group. The values of the first two columns are median values for the two groups with the mean and standard deviation between parentheses. The last column reports the Wilcoxon rank sum test p-values for the comparison between the two groups

Dataset	**AUC**	**p-value**	**OR - C.I. 95%**
PP	0.72 ± 0.009	p=0.017	5.85 [3.52 ; 9.71]
JP	0.54 ± 0.089	p=0.352	1.04 [0.62 ; 1.66]
AO	0.53 ± 0.061	p=0.384	1.03 [0.60 ; 1.45]

Table 4.2: AUCs, p-values and Odds-ratios obtained using the GPA alignment methodology and different sets of annotations. The reported results are mean values with respect to the same test-shape for 10 iterations using 1000 samples of the BNLO cross-validation protocol. The reported p-values were obtained comparing the scores to chance using the Delong test.

4.4 Discussion

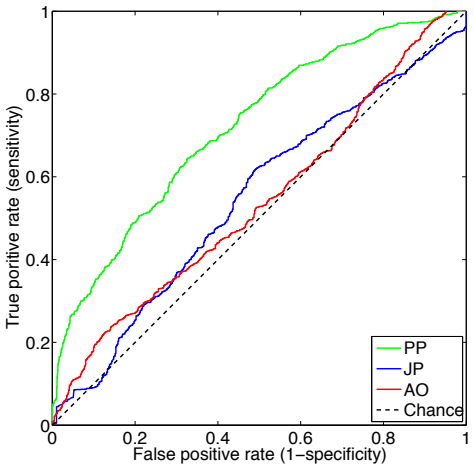
This manuscript described a morphometric index based on a statistical shape model, for predicting vertebral fractures in the lumbar region. It can be hypothesized [27,98] that with age the spine curvature and vertebrae shapes undergo minimal changes. These changes in the spine structure can produce musculoskeletal forces influencing the spine itself, and leading in the long run to one or more fractures in the spine [120]. The described shape based framework was able to capture slight shape and curvature variations related to possible future fractures. The reported longitudinal case-control study involved 113 post-menopausal women from a larger population. The control and case group data-set were matched with respect to traditional osteoporosis risk factors and obtained from a larger population. The subjects of both groups were fracture-free at baseline. During the study, radiographs of the lumbar region were acquired

Dataset	AUC	p-value	OR - C.I. 95%
PP	0.71 ± 0.013	p=0.019	4.87 [2.94 ; 8.05]
JP	0.62 ± 0.025	p=0.043	2.04 [1.21 ; 3.36]
AO	0.60 ± 0.046	p=0.047	1.92 [1.06 ; 2.87]

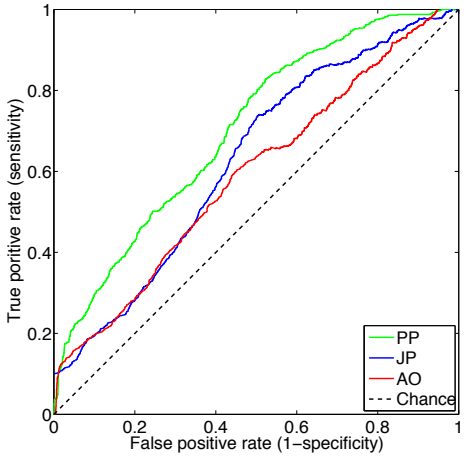
Table 4.3: AUCs, p-values and Odds-ratios obtained using the DSA alignment methodology and different sets of annotations. The reported results are mean values with respect to the same test-shape for 10 iterations using 1000 samples of the BNLO cross-validation protocol. The reported p-values were obtained comparing the scores to chance using the Delong test.

and spine fracture status assessed using one semi-quantitative and one quantitative morphometric method. The surprising small number of patients is the consequence of strict selection criteria such as the aforementioned absence of fracture at baseline, the consensus assessment between the QM and SQ methods, the absence of non-osteoporotic fractures and corticosteroids use. The spine shapes obtained from those annotations were used, in an unbiased way, as features for an LDA classifier identifying the spines with high risk of fracture.

The DSA version of the described algorithm was capable of quantifying the fracture risk of spines for all three used sets of annotations. However, the algorithm performed better on the radiologist annotations than on the annotations performed by the trained technicians. In particular, even the GPA based algorithm performed well on the PP data-set, but the experiments on the JP and AO set of annotations needed the DSA version of the algorithm to reach statistically significant results. The discrepancy can reflect the fact that radiologist annotations are qualitatively more precise than technician annotations. This sensitivity to annotation quality has also been seen in another study using a different morphometric index tested on almost the same data-set [79]. In this previous study, a methodology based on normalized vertebral heights slightly outperformed our method for the radiologist annotations but gave similar results regarding the technician annotations. These observations lead to the suggestion to perform further shape based predictions using radiologist annotations. Despite the difference in results compared to the method described in [79], possibly based on the fact that such a method is specifically designed to most deformed vertebral heights, the approach reported in this manuscript can be used in other prediction and classification contexts, where label information of the training data-set is available, since it is based on the general concept of shape. Since the DSA algorithm included the class label information, it gave more robust results compared to the GPA in our experiments. However, the DSA algorithm is computationally more demanding. The results reported in this paper, although promising, should be validated in future studies to establish applicability on other populations. Moreover,



(a)



(b)

Figure 4.4: ROC curves obtained using the BNLO algorithm and (a) the GPA alignment, (b) the DSA alignment. The curves are constructed using all the scores without averaging the scores with respect to the same test shape. The same random samples were used for both versions of the algorithm.

the described algorithm was tested on the lumbar vertebrae L1-L4 and the thoracic vertebra T12 which represents the junction of two different regions, and therefore interesting from the point of view of fragility of the spine structure [82]. This choice

was primarily made for facilitating the comparison with studies on al most the same cohort [79]. The L5 vertebra, which was not included in the presented model due to the fact that it is often covered by the hip, could also improve the shape analysis. This can represent a potential limitation of the study. The extension from the lumbar region to the rest of the spine was attractive, but as reported in [79], the biological variation between lumbar and thoracic vertebral shapes is quite large, and possibly the transition from the use on lumbar vertebrae to the use on thoracic vertebrae would not be trivial. The presence and absence of vertebral fractures was established by both a radiologist using Genant's SQ method [45] and subsequently by a computer algorithm using measures of the vertebral heights with a strict fracture threshold. As a future work, it would be relevant to validate the described fracture-risk algorithm with respect to changing the "ground truth" fracture assessment methodology to other established methods, such as the Melton or McCloskey methods [30, 65, 83].

Our method was tested on a population matched according to the common osteoporosis risk factors, giving relative discrimination between the case and control spine shapes. Therefore, the algorithm can be considered as a further morphometric bio-marker, which combined with the traditional risk factors allows a more reliable prediction of vertebral fractures. The matched population was used to evaluate the predictive value of the methodology independently from the other risk factor [4]. A comparison between the described methodology and the standard osteoporotic risk factors in populations with less restriction is relevant and considered as a future work. However, this approach could facilitate a population selection in clinical trials, reducing the number of required patients, and helping identifying patients with high fracture risk, who could benefit from treatments.

Chapter 5

Maximum A Posteriori Estimation of Linear Shape Variation with Application to Vertebra and Cartilage Modeling

This chapter is based on the manuscripts

- "Maximum A Posteriori Bayes Estimation of Linear Shape Variation with Application to Vertebra and Cartilage Modeling." Alessandro Crimi, Martin Lillholm, Anarta Ghosh, Marleen de Bruijne, Erik Dam, Mads Nielsen, Jon Sporring. IEEE Transaction in Medical Imaging, vol.30(8) pages 1514-1526, August 2011.
- "Bayes estimation of shape model with application to vertebrae boundaries." Alessandro Crimi, Anarta Ghosh, Jon Sporring, and Mads Nielsen. In Proceedings of the SPIE: Medical Imaging 2009, Orlando, Florida, USA, 2009.
- "Prior knowledge regularization in statistical medical image tasks." Alessandro Crimi, Jon Sporring, Marleen de Bruijne, Martin Lillholm, Mads Nielsen. PMMIA workshop in conjunction with MICCAI 2009,Imperial college London, UK, 2009.

The estimation of covariance matrices is a crucial step in several statistical tasks. Especially when using few samples of a high dimensional representation of shapes, the standard Maximum Likelihood estimation (ML) of the covariance matrix can be far from the truth, is often rank deficient, and may lead to unreliable results. In this paper, we discuss regularization by prior knowledge using Maximum A Posteriori (MAP) estimates. We compare ML to MAP using a number of priors and to Tikhonov regularization. We evaluate the covariance estimates on both synthetic and

real data, and we analyze the estimates' influence on a missing-data reconstruction task, where high resolution vertebra and cartilage models are reconstructed from incomplete and lower dimensional representations. Our results demonstrate that our methods outperform the traditional ML method and Tikhonov regularization.

5.1 Introduction

Estimation of the covariance matrix is an initial and pivotal step of Principal Component Analysis (PCA) [94], Factor Analysis (FA) [112], some versions of regression, and many other statistical tasks. If the sample size is small, and the number of considered variables is large, then the estimation of the covariance matrix can be poor. Regularization and specifically shrinkage have been proposed to improve estimates of covariance matrices.

Shrinkage methods aim at improving estimates by shrinking the estimation towards zero or in the general setting towards some specific value. The oldest shrinkage methods were proposed in the early 60's [108,113], and afterwards a series of important methods appeared [31,40,53,57,64,103]. These methods depend on the proper choice of a mixing parameter, and the mixing parameter is often selected to maximize the expected accuracy of the shrunken estimator by cross-validation [40,57] or by using an analytical estimate of the shrinkage intensity [31, 53, 64, 103]. Regularized estimators have been shown to outperform the standard Maximum Likelihood (ML) estimator for small sample sizes [103]. It is most common to use a simple form of ridge regression [57], where nonzero values are added to the diagonal elements of the covariance matrix. Ridge regression was originally introduced by Tikhonov for solving an underdetermined system of linear equations [113]. Friedman presented a restricted version of this regularization, linking the mixing parameter to the minimization of misclassification risk during discriminant analysis. Generally in regularization, non-zero values are added to the diagonal elements of a covariance matrix as $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{reg}} = \hat{\boldsymbol{\Sigma}} + \alpha^2 \boldsymbol{I}$, where α^2 is the discussed mixing parameter. It was shown that this mixing parameter is not likely to be known in advance, and finding its optimal value can be computationally demanding [40].

Probabilistic PCA considers the Maximum Likelihood solution of a probabilistic latent variable model. This method finds the non-corrupted eigenvalues/eigenvectors, but it is not always able to regularize the covariance matrix optimally [114]. A similar method was proposed by Allassonniere et al. [1], where kernels in Hilbert space are used for estimating the covariance matrix of a generative model.

Stein estimators [31, 53, 64, 103] estimate the mixing parameter using just the sample covariance matrix. However, these methods rely on the estimation of a specific cost function.

An overview of methods optimized for data in non-linear subspaces is given in [77], particularly we wish to highlight Minimum Noise Fraction [50], Minimum Autocorrelation Factor [109], and Principal Geodesic Analysis (PGA) [38]. In these cases, PCA can be used as local, linear approximations.

In this paper, we investigate regularization by Maximum A Posteriori (MAP) estimates of covariance matrices in a PCA setting. A number of priors are discussed for this MAP method, and the usefulness of the priors is evaluated against known ground truth covariance matrices. Our underlying motivation is to improve the accuracy of medical diagnosis and prognosis of diseases, such as Osteoporosis (OP) and Osteoarthritis (OA). As an example application, we demonstrate how low resolution vertebral shape outlines can be interpolated into high-resolution full contours using a shape model based on a MAP estimate of the point distribution model's intrinsic covariance matrix.

5.1.1 Principle Component Analysis and Maximum Likelihood Covariance estimation (ML)

In statistical shape analysis, a shape model is typically built using a set of training examples often represented as a sample matrix [17, 18]: A shape $\boldsymbol{x} \in \mathbb{R}^q$ is a vector of ζ labeled landmarks, $\boldsymbol{p} \in \mathbb{R}^l$,

$$\boldsymbol{x} = [\boldsymbol{p}_1^T | \boldsymbol{p}_2^T | \dots | \boldsymbol{p}_\zeta^T]^T, \tag{5.1}$$

and the collection of n shapes is collected in a sample matrix, $\boldsymbol{X} \in \mathbb{R}^{q \times n}$,

$$\boldsymbol{X} = [\boldsymbol{x}_1 | \boldsymbol{x}_2 | \dots | \boldsymbol{x}_n]. \tag{5.2}$$

As an example, consider 30 shapes each consisting of 6 landmarks in 2 dimensional X-ray images, then $\boldsymbol{p}_i \in \mathbb{R}^2$, $\boldsymbol{x} \in \mathbb{R}^{12}$, and $\boldsymbol{X} \in \mathbb{R}^{12 \times 30}$.

PCA [94] also known as *Karhunen-Loeve transformation* [70] and *Hotelling transformation* [58] is a widely used technique for applications such as dimensionality reduction, data compression, and data visualization [66]. PCA is an orthogonal projection of the data onto a lower dimensional linear subspace, such that the variance of the projected data is maximized sequentially along each new coordinate axis. The linear subspace is not always a precise model of the variability in the data: In the case of shapes, the shapes themselves may intrinsically lie on a non-linear submanifold. Furthermore, shape normalization using, e.g., Procrustes' method [11, 49] for neutral position, size, and orientation will also typically project the shapes onto a non-linear submanifold. As an example, size normalization, such as $\overline{\boldsymbol{x}} = \boldsymbol{x} / \left\|\boldsymbol{x}\right\|_2$ will project the samples onto a $q-1$ dimensional hypersphere, as illustrated in Figure 5.1 for 2-dimensional data. In this case, PCA may be considered an approximation of the projected variability in the tangent plane under suitable choice of mean. In part, this limitation can be addressed by exploiting the metric of the embedding manifold using techniques, such as, PGA [38]. Nevertheless, we only consider regularized linear models, since our focus is on cases, where the number of samples is small compared to the dimensionality of the embedding space, $n \ll q$, and since we do not assume prior knowledge about any underlying shape manifolds.

PCA models the cloud of (normalized) shapes $\{\boldsymbol{x}_i\}_{i=1\dots n}$ as a Gaussian distribution with mean $\boldsymbol{\mu} = E[\boldsymbol{x}]$ and covariance $\boldsymbol{\Sigma} = E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T]$. In this paper, we

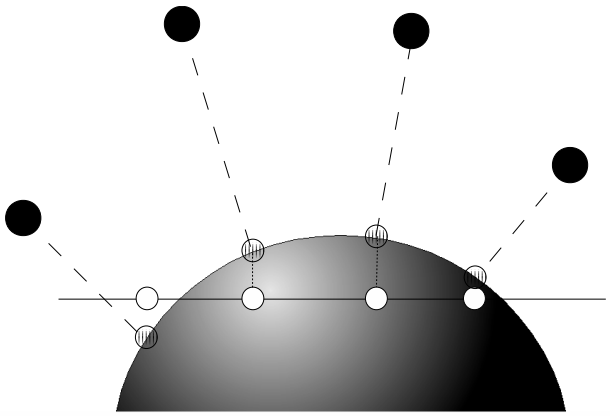


Figure 5.1: Size normalization of samples (filled circles) projects onto a non-Euclidean hypersphere (striped circles). A linear model hereof further projects onto a line (empty circles).

assume that $\mathbf{\Sigma}$ has full rank except for shape normalization, e.g., when normalizing for translation, scaling and rotation, then $\mathbf{\Sigma}$ has rank $q-l-1$. The covariance matrix $\mathbf{\Sigma}$ is symmetric and positive semi definite, which is why we may write its eigenvalue decomposition as,

$$\mathbf{\Sigma} = \boldsymbol{V}\mathbf{\Lambda}\boldsymbol{V}^T, \tag{5.3}$$

where $\boldsymbol{V} = [\boldsymbol{v}_1|\ldots|\boldsymbol{v}_q]$ is the column matrix of eigenvectors, and $\mathbf{\Lambda} = \text{diag}\left([\lambda_1, \ldots, \lambda_q]\right)$ is the diagonal matrix of corresponding eigenvalues. Eigenvectors and values are assumed sorted according to $|\lambda_i| \geq |\lambda_j|$ for $i < j$. Transforming the shapes onto a new coordinate system $\mathbf{\Xi} = [\boldsymbol{\xi}_1|\boldsymbol{\xi}_2|\ldots|\boldsymbol{\xi}_n]$ as,

$$\mathbf{\Xi} = \sqrt{\mathbf{\Lambda}^{-1}}\boldsymbol{V}^T(\boldsymbol{X} - \boldsymbol{\mu}\mathbf{1}_n^T), \tag{5.4}$$

where $\mathbf{1}_n = [1, \ldots, 1]$ is a column vector of n ones, such that the outer product $\boldsymbol{\mu}\mathbf{1}_n = [\boldsymbol{\mu}|\ldots|\boldsymbol{\mu}]$ is a matrix of n $\boldsymbol{\mu}$s, and where the inverse of zero eigenvalues are set to zero. We may consider the shape variations as ordered by variance in an uncorrelated coordinate system, and we may write an approximation of a shape $\boldsymbol{x}$ by the t absolutely largest eigenvalues as,

$$\tilde{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{V}\,\text{diag}\left(\left[\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_t}, 0, \ldots, 0\right]\right)\boldsymbol{\xi}, \tag{5.5}$$

where $\boldsymbol{\xi}$ is column vector in $\mathbf{\Xi}$ corresponding to the column vector in $\boldsymbol{X}$. The quadratic error of the approximation will be $\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2^2 = \sum_{i=t+1}^{q} \lambda_i$.

An often used estimate of the covariance matrix is the ML estimator [2]: The shapes $\boldsymbol{x}$ are assumed to be independent and to have identical Gaussian density

function, yielding the following joint distribution

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n | \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \frac{1}{(2\pi)^{\frac{(nq)}{2}} |\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \right), \quad (5.6)$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. This distribution is often called the likelihood of $\{\boldsymbol{x}_i\}_{i=1\ldots n}$, and the point of maximum likelihood for varying $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is derived for completeness in Appendix 5.5 and is found to be,

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i, \quad (5.7a)$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^T. \quad (5.7b)$$

The covariance estimate is slightly biased, but for large n the bias is negligible. The speed of the convergence of the estimator is inversely proportional to the number of shapes in a data set. However, faster convergence may be achieved, when prior knowledge is available.

5.1.2 Tikhonov Regularization of Covariance Estimates (ML-T)

Regularization is one method for introducing prior knowledge and is often used as a tool for improving the estimate, when the number of samples is small compared to the dimensionality of the embedding space, $n \ll q$. It is most common to use a simple form of ridge regression or Tikhonov regularization [57,113], where nonzero values are added to the diagonal elements of the covariance matrix,

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{reg}} = \hat{\boldsymbol{\Sigma}} + \alpha^2 \boldsymbol{I}, \quad (5.8)$$

where $\alpha \in \mathbb{R}$. We refer to this method as ML-T. By diagonalizing $\hat{\boldsymbol{\Sigma}}$, it is seen that ML-T only affects the eigenvalues of $\hat{\boldsymbol{\Sigma}}_{reg}$ adding α^2 to the eigenvalues of $\hat{\boldsymbol{\Sigma}}$ regardless of their value. Adding a constant to zero eigenvalues is useful for regularizing an inversion but may not be for analyzing the subspace spanned by the data. It is our experience, that for $n \ll q$, the low absolute eigenvalues require relatively more regularization than the high absolute eigenvalues. Varying α as function of absolute eigenvalue is not easily done using ML-T. In this paper, we discuss Bayes estimation of the covariance matrix using the Maximum A Posteriori (MAP) method. We discuss two priors closely tied to the Wishart distribution, which allows for varying degree of regularization depending on absolute eigenvalue. Finally, we discuss a more general Gaussian prior, which allow for prior knowledge on off diagonal interactions.

5.2 Maximum A Posteriori Covariance Estimation

If we consider $\mathbf{\Sigma}$ a random variable in the space of symmetric, positive semi definite matrices, then Bayes theorem states that

$$p(\mathbf{\Sigma}|\boldsymbol{S}) = \frac{p(\boldsymbol{S}|\mathbf{\Sigma})p(\mathbf{\Sigma})}{p(\boldsymbol{S})}, \tag{5.9}$$

where $p(\mathbf{\Sigma}|\boldsymbol{S})$ is the posterior, $p(\boldsymbol{S}|\mathbf{\Sigma})$ the likelihood, $p(\mathbf{\Sigma})$ the prior, and $p(\boldsymbol{S})$ is the evidence. We may postulate prior knowledge as prior densities and calculate the Maximum A Posteriori (MAP) estimate. As a limiting case, MAP converges to ML as the prior converges to the improper uniform distribution.

For all Bayes estimators of covariance matrices presented in this paper, we consider the same likelihood distribution $p(\boldsymbol{S}|\mathbf{\Sigma}, n)$ of the sample covariance matrix, $\boldsymbol{S} \in \mathbb{R}^{q\times q}$,

$$\boldsymbol{S} = (\boldsymbol{X} - \boldsymbol{\mu}\mathbf{1}_n^T)(\boldsymbol{X} - \boldsymbol{\mu}\mathbf{1}_n^T)^T. \tag{5.10}$$

A Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$ on $\boldsymbol{x}$ induces a Wishart distribution on the likelihood of $\boldsymbol{S}$ [2, Chapter 7],

$$p(\boldsymbol{S}|\mathbf{\Sigma}, n) = \frac{|\boldsymbol{S}|^{(n-q-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\boldsymbol{S}\right)\right)}{2^{nq/2}\,|\mathbf{\Sigma}|^{n/2}\,\Gamma_q\left(\frac{n}{2}\right)}, \tag{5.11}$$

where Γ_q is the multivariate Gamma function,

$$\Gamma_q\left(\frac{n}{2}\right) = \pi^{q(q-1)/4}\prod_{i=1}^{q}\Gamma\left(\frac{n}{2} - \frac{i-1}{2}\right). \tag{5.12}$$

The zero mean distribution that best fits the samples $\boldsymbol{S}$ is found as the point of maximum log likelihood for varying $\mathbf{\Sigma}$ as,

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\boldsymbol{S}, \tag{5.13}$$

identical to (5.7b). A common prior for covariance matrices is the Inverted Wishart to be discussed in the following.

5.2.1 Inverted Wishart Prior (MAP-IW)

A prior giving a simple MAP estimate is the inverted Wishart distribution [2, Chapter 7],

$$p(\mathbf{\Sigma}|\mathbf{\Psi}, m) = \frac{|\mathbf{\Psi}|^{m/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Psi}\mathbf{\Sigma}^{-1}\right)\right)}{2^{mq/2}\,|\mathbf{\Sigma}|^{(m+q+1)/2}\,\Gamma_q\left(\frac{m}{2}\right)}, \tag{5.14}$$

where $\mathbf{\Psi}$ and $m \geq q$ are parameters of the distribution. The inverted Wishart density originates as the distribution of $\boldsymbol{S}^{-1}$, when $\boldsymbol{x}$ is Gaussian. Since the evidence is

independent on $\boldsymbol{\Sigma}$, we find that the point of MAP is

$$\hat{\boldsymbol{\Sigma}} = \arg\max_{\boldsymbol{\Sigma}} p(\boldsymbol{S}|\boldsymbol{\Sigma}, n) p(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, m) \tag{5.15a}$$

$$= \arg\max_{\boldsymbol{\Sigma}} \frac{c\,|\boldsymbol{\Psi}|^{m/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left((\boldsymbol{S}+\boldsymbol{\Psi})\boldsymbol{\Sigma}^{-1}\right)\right)}{|\boldsymbol{\Sigma}|^{(n+m+q+1)/2}} \tag{5.15b}$$

using

$$c = \frac{|\boldsymbol{S}|^{(n-q-1)/2}}{2^{(n+m)q/2}\Gamma_q\left(\frac{n}{2}\right)\Gamma_q\left(\frac{m}{2}\right)}. \tag{5.16}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n+m+q+1}\left(\boldsymbol{S}+\boldsymbol{\Psi}\right). \tag{5.17}$$

Regarding the structure of (5.17), we find that it can be similar to Tikhonov regularization (5.8), when $\boldsymbol{\Psi} = \alpha^2\boldsymbol{I}$, $\alpha^2 \simeq 1/n$, and $n \gg m+q+1$. In any case, the parameter $\boldsymbol{\Psi}$ should be set by prior knowledge, e.g., as a measure of the uncertainty in landmark detection by a human operator, or using a geometric distance measure such as (5.32) also investigated in [107].

We wish to use the Inverted Wishart to model independent landmark noise, therefore the structure of the true covariance matrix is assumed to be block diagonal (one block per landmark), and hence $\boldsymbol{\Psi}$ must also be block diagonal. Further, we observe that for typical ML estimates of $\hat{\boldsymbol{\Sigma}}$ for $n \ll q$, the close to zero eigenvalues are particularly poorly estimated. We speculate that this could be caused by over-fitting of the first eigenvectors to the data leaving the directions of lesser variance to be underestimated: The first eigenvalue and corresponding eigenvector is typically estimated through the direction of largest variance in a least squares sense. With only few available samples, this direction of largest variance and the corresponding eigenvalue will, on average, be overestimated. This is straightforward to validate empirically through repeated ML estimation on small samples from a known covariance structure; if the samples afford a direction with variance larger than the known covariance this will always be chosen by the ML estimate. This will, on average, bias the average size of the estimate of the initial eigenvalue. Thus and in contrast to ML-T, we propose to boost small eigenvalues more than large eigenvalues of $\boldsymbol{\Sigma}$, by setting $\boldsymbol{\Psi}$ to be blockwise proportional to the per landmark inverse maximum likelihood estimate of its covariance matrix. As an example, for 2-dimensional landmarks distributed as $\boldsymbol{\Sigma}_i$, then $\boldsymbol{\Psi} = \mathrm{diag}\left(\boldsymbol{\Sigma}_1^{-1}, \ldots, \boldsymbol{\Sigma}_\zeta^{-1}\right)$,

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n+m+q+1}\left(\boldsymbol{S} + \alpha^2 \mathrm{diag}\left(\boldsymbol{S}_1^{-1}, \ldots, \boldsymbol{S}_\zeta^{-1}\right)\right), \tag{5.18}$$

where $\alpha \in \mathbb{R}$ is a free mixing parameter. We call this the MAP-IW. This is not a proper prior, since it depends on the measurements. Furthermore, it is noted that when $\alpha \neq 0$, both ML-T and MAP-IW enforces full rank. We defer discussion of its usefulness to the experimental section.

5.2.2 Uncommitted Inverted Wishart Prior (MAP-UIW)

The MAP-IW (5.18) uses an improper prior that depends on an estimate of the covariance of each landmark, which in total may be considered an approximation of $\boldsymbol{\Sigma}$. We propose to include $\boldsymbol{\Psi} = \alpha^2 \boldsymbol{\Sigma}^{-1}$ directly in the estimation, in which case the maximization results in a system of quadratic equations,

$$\hat{\boldsymbol{\Sigma}}^2 = \frac{1}{n + 2m + q + 1} \left(\boldsymbol{S}\hat{\boldsymbol{\Sigma}} + 2\alpha^2 \boldsymbol{I} \right), \tag{5.19}$$

as derived in Appendix 5.5. This is denoted the MAP-UIW method. As explained above, the MAP-UIW enforces full rank of the estimate. We solve this iteratively using the fix point iteration,

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \hat{\boldsymbol{\Sigma}}_t - \delta \left(\hat{\boldsymbol{\Sigma}}_t^2 - \frac{1}{n + 2m + q + 1} \left(\boldsymbol{S}\hat{\boldsymbol{\Sigma}}_t + 2\alpha^2 \boldsymbol{I} \right) \right), \tag{5.20}$$

where δ is a sufficiently small constant to avoid divergence. Symmetry of the numerical solution may be guaranteed by taking the average with the transpose, or by only performing calculations for the lower left triangular matrix including the diagonal. We have experimented with both solutions and found no significant differences. For simplicity, we prefer the average. Further, we have experimented with a number of different initial points and the method appears to be robust, however, in this work we use ML-T as the initial point $\hat{\boldsymbol{\Sigma}}_0$.

An alternative solution to (5.19) is obtained by a different factorization of the equation (5.48):

$$\hat{\boldsymbol{\Sigma}}^2 = \frac{1}{n + 2m + q + 1} \left(\sqrt{\hat{\boldsymbol{\Sigma}}} \boldsymbol{S} \sqrt{\hat{\boldsymbol{\Sigma}}} + 2\alpha^2 \boldsymbol{I} \right), \tag{5.21}$$

this solution theoretically solves the symmetrization issue and employed as a fix point iteration converges with fewer steps than (5.20). However, there are some numerical drawbacks: First, the need of a SVD decomposition during the square root computation of the matrices slows down the algorithm. Second, since we perform iterations using a fixed step size, and the space of positive definite matrices (pd) is not a linear vector space, even this elegant solution can in practice step outsize pd-space. Therefore, in our experiments we prefer to use (5.20).

5.2.3 Gaussian Prior (MAP-G)

An alternative prior is the Gaussian distribution in covariance space,

$$p(\hat{\boldsymbol{\Sigma}} | \boldsymbol{B}, s) = \frac{1}{(2\pi s^2)^{-\frac{q(q+1)}{2}}} \exp \left(- \frac{\left\| \hat{\boldsymbol{\Sigma}} - \boldsymbol{B} \right\|_2^2}{2s^2} \right), \tag{5.22}$$

where $\boldsymbol{B} \in \mathbb{R}^{q \times q}$ and $s \in \mathbb{R}$ represent the mean and variance of the covariance matrix distribution. The variance s controls the trade-off between the prior and the ML

estimate. A large value of s would make the prior uniform and would yield the ML estimate. On the other hand, a very small s would yield a delta spike around the mean value and thus bias the estimate heavily towards the mean of the Gaussian. Ideally, the Gaussian prior should be imposed on the space of symmetric, positive definite matrices, e.g., as approximated by the exponential of the norm of the differences of logarithms of matrices [34], but this makes the equation prohibitively complicated, hence we consider (5.22) a Euclidean approximation. The MAP of (5.22) is derived in Appendix 5.5 and is found to be a system of third degree polynomials,

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\left(\boldsymbol{S} - \frac{2}{s^2}\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{B})^T\hat{\boldsymbol{\Sigma}}\right). \tag{5.23}$$

We seek its solution by the following iterative scheme,

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \hat{\boldsymbol{\Sigma}}_t + \delta\left(\hat{\boldsymbol{\Sigma}}_t - \frac{1}{n}\left(\boldsymbol{S} - \frac{2}{s^2}\hat{\boldsymbol{\Sigma}}_t(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}} - \boldsymbol{B})^T\hat{\boldsymbol{\Sigma}}_t\right)\right), \tag{5.24}$$

which we have found to converge for $t \to \infty$, when using iterative line search in δ, enforcing symmetry by averaging, and using the ML-T as initial point $\hat{\boldsymbol{\Sigma}}_0$. Matrix $\boldsymbol{B}$ is a free positive semi-definite matrix, which is the embodiment of prior knowledge. Unfortunately, covariance matrices depend on the coordinate system, however, the matrix $\tilde{\boldsymbol{B}}$ may be specified in a coordinate system of choice and rotated into the specific coordinate system by a suitable rotation matrix $\boldsymbol{R}$ as $\boldsymbol{B} = \boldsymbol{R}\tilde{\boldsymbol{B}}\boldsymbol{R}^T$. In this section we have considered 3 different priors to be imposed on our covariance matrix estimates. It is important to bear in mind that the ML-T assumes independence between all elements in the shape, and MAP-IW assume independence between landmarks. Hence both are priors of spatial noise. The MAP-UIW does not have assumption of independence, and MAP-G may be used to directly steer the estimate towards preferred shape variations. Hence both are general priors of shape variations. The classical view of the Inverse Wishart prior has a similar number of parameters as our Gaussian prior. There is, however, a clear difference in how these parameters are used: For the Inverse Wishart prior given $\boldsymbol{\Psi}$, the covariance matrix of maximum probability is the solution to $\mathrm{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-\mathbf{1}}\right) = 0$, i.e., the j'th column in $\boldsymbol{\Sigma}^{-\mathbf{1}}$ must be perpendicular to the j'th row in $\boldsymbol{\Psi}$. Hence, if we wish to align the Inverse Wishart prior to favor a particular covariance matrix $\overline{\boldsymbol{\Sigma}}$, then we should select a $\boldsymbol{\Psi}$ from the $(q-1)(q-2)\ldots 1 = (q-1)!$ dimensional manifold fulfilling the above equation, none of which intuitively relate to $\overline{\boldsymbol{\Sigma}}$. Conversely for the Gaussian prior, maximum probability is where $||\boldsymbol{\Sigma} - \boldsymbol{B}||_2^2 = 0$, and in order to favor a particular covariance matrix $\overline{\boldsymbol{\Sigma}}$, we simply and intuitively should choose $\boldsymbol{B} = \overline{\boldsymbol{\Sigma}}$. Furthermore, the additive structure of the Inverse Wishart prior estimators is directly comparable to Tikhonov regularization. In light of the difficulties of large dimensional manifolds described above, we argue that the constrained as well as the uncommitted Wishart priors are the practical extensions of Tikhonov regularization.

5.3 Experimental setup

We evaluate the performance of the previously described methods for estimating covariance matrices: Maximum Likelihood (ML), ML regularized by Tikhonov regularization (ML-T), MAP with an inverted Wishart prior (MAP-IW), MAP with an uncommitted inverted Wishart prior (MAP-UIW), and finally MAP with a Gaussian prior (MAP-G). The evaluation is focused on two main tasks: Firstly, we compare the distance from a known ground-truth covariance matrix to the estimated covariance matrices for synthetic data. Secondly, we evaluate a reconstruction task from low to high-resolution shapes. The details of our validation protocol are described in the following subsections. Section 5.3.1 describes the metric used for comparing matrices, Section 5.3.2 discusses the reconstruction of high dimensional shape representations, while Section 5.3.3 presents the parameter optimization scheme used in the two validation tasks. Section 5.3.4 describes the mean value $\boldsymbol{B}$ of the Gaussian prior, which we use for our experiments. The results of this validation protocol and the experimental setups are described in sections 5.4.1 and 5.4.2 respectively for synthetic data and for vertebra and cartilage shape reconstruction.

5.3.1 Matrix comparison

The first part of the validation protocol is a comparison between the ground-truth covariance matrix and different covariance matrix estimates. A matrix norm is required for this purpose. A suitable norm is the Frobenius or 2-norm,

$$d(S_1, S_2) = \|\boldsymbol{S_1} - \boldsymbol{S_2}\|_2 \,, \tag{5.25}$$

where $\boldsymbol{S_1}$ can be the ground truth matrix and $\boldsymbol{S_2}$ a matrix estimated with one of the methods described in this paper. Another useful norm is the log-Euclidean distance between two matrices [34], which approximates the distance between the matrices along the geodesics in the space of symmetric, positive definite matrices, $d(S_1, S_2) = \|\log(\boldsymbol{S}_1) - \log(\boldsymbol{S}_2)\|_2$, where $\log(\boldsymbol{S}) = \boldsymbol{V}\log(\boldsymbol{\Lambda})\boldsymbol{V^T}$, when $\boldsymbol{S} = \boldsymbol{V\Lambda V^T}$. With our data, we found no qualitative differences between these norms, and in Section 5.4.1 we present our results using the Frobenius norm.

5.3.2 Increasing shape resolution

The direct matrix comparison experiments are not well suited for real data, where ground truth covariance matrices are not available. Therefore, we focus our attention on the example application of increasing shape resolution.

Accurate manual annotation of, e.g., full vertebrae boundaries is tedious and time consuming. Likewise for studies of cartilage shape, more detailed models may lead to more accurate results. These methodologies are described more in detail in Section 5.4.2. Increasing the shape resolution from a low dimensional manual annotation can be a viable alternative.

Letting $\boldsymbol{y}$ be an incomplete or lower dimensional shape vector of dimensionality $k < q$, we would like to find the corresponding higher resolution shape $\boldsymbol{x} \in \mathbb{R}^q$ such that:

$$\boldsymbol{y} = \boldsymbol{L}\boldsymbol{x}, \tag{5.26}$$

where $\boldsymbol{L} : \mathbb{R}^q \to \mathbb{R}^k$ is a linear mapping. For our experiments the matrix $\boldsymbol{L}$ is a projection matrix. The lower dimensional shape $\boldsymbol{y}$ is aligned to $\boldsymbol{L}\boldsymbol{\mu}$ by the standard Procrustes method. Since the system is under-determined, the solution is not uniquely defined. Following Blanz and Vetter [9], we estimate $\boldsymbol{x}$ by minimizing the functional

$$E(\boldsymbol{x}) = \|\boldsymbol{L}\boldsymbol{x} - \boldsymbol{y}\|_2^2 . \tag{5.27}$$

Since $\boldsymbol{x}$ belongs to the shape model (5.5), the aforementioned functional transforms to:

$$E(\boldsymbol{\xi}) = \|\boldsymbol{L}\boldsymbol{\mu} + \boldsymbol{Q}\boldsymbol{\xi} - \boldsymbol{y}\|_2^2 , \tag{5.28}$$

where $\boldsymbol{\xi}$ is the parameter vector of the shape model (5.5), and $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{V}_t\boldsymbol{\Lambda}_t$, where $\boldsymbol{V}_t$ is the eigenvectors and $\boldsymbol{\Lambda}_t = \operatorname{diag}(\lambda_i)$ the eigenvalues corresponding to the t principal eigenmodes of the covariance matrix. This covariance matrix is estimated using one of the methods described in the previous sections. $\boldsymbol{Q}$ is a $k \times q$ matrix of rank at most k. The corresponding reconstruction is [9]:

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{V}_t\boldsymbol{\Lambda}_t\boldsymbol{H}\boldsymbol{W}^{+}\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{y}, \tag{5.29}$$

where $\boldsymbol{U}$, $\boldsymbol{H}$ and $\boldsymbol{W}$ are defined by the singular value decomposition of $\boldsymbol{Q}$, $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{W}\boldsymbol{H}^{\boldsymbol{T}}$, and the diagonal matrix is inverted as,

$$\boldsymbol{W}^{+}_{ii} = \begin{cases} \frac{1}{w_i} & \text{if } w_i > 0 \\ 0 & \text{else.} \end{cases} \tag{5.30}$$

and w_i being the i^{th} singular value of $\boldsymbol{W}$. By construction, (5.29) is invariant to scaling of the underlying regularized covariance matrix with non-zero real constants.

5.3.3 Parameter optimization

Some of the covariance matrix estimators described in Section 5.2 require optimization of parameters such as α, m, and s, and we use cross-validation [43, 75] as follows: For given data, we 20 times randomly select a non-overlapping train and test set of a specified size. For each pair, we examine a range of parameter values, and for each parameter value, we estimate the generalization error using leave-one-out on the training set. The optimal parameter value is chosen to minimize the generalization error on the training set. The error on the corresponding test set is then calculated using the optimal parameter value. Finally, we report the average performance over all test sets. Figure 5.2 shows the 20 different training error curves as a function of the mixing parameter α for MAP-UIW as well as their detected minima. The same approach is used for optimizing the parameters of the other methods. For the reconstruction evaluation, we note that the parameters of some priors are multiplicative constants and do not require optimization as described previously.

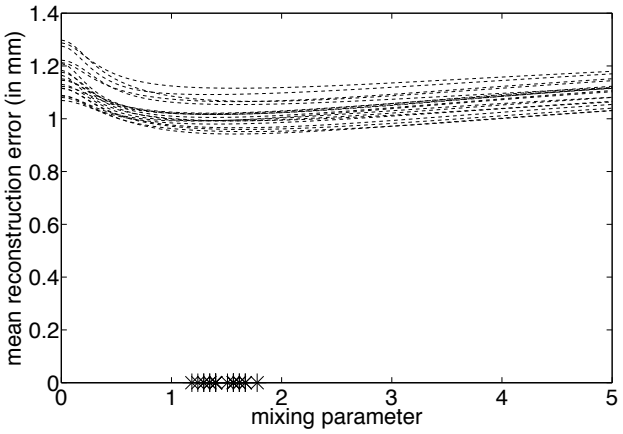


Figure 5.2: Example of error curves as a function of the MAP-UIW mixing parameter for the range [0,5]. Each curve is based on a leave-one-out error on a separate set, as described in the main text. On the x-axis the minimum error parameter values are also depicted as stars. For this experiments 10 vertebrae samples were used as a training set.

5.3.4 Mean value of the Gaussian prior

For the Gaussian prior described in Section 5.2, a crucial aspect is the design of the mean matrix $\boldsymbol{B}$ of the prior distribution of $\hat{\boldsymbol{\Sigma}}$ in (5.22) and (5.2.3). The design of this matrix must reflect some prior knowledge of the data it represents. For example neighboring points often covary for typical shapes. Therefore, we find it useful to define $\boldsymbol{B}$ as a function of the distances between points similarly to [107]:

$$\boldsymbol{A_{ij}} = \text{dist}\left(\boldsymbol{p}_i - \boldsymbol{p}_j\right), \tag{5.31}$$

where $\boldsymbol{p}_i$ is the i'th point of the mean of the aligned training examples. Subsequently, we reweigh the distances with an exponential decrease to obtain:

$$\boldsymbol{B} = \exp\left(-\frac{1}{\beta}\boldsymbol{A} \otimes \boldsymbol{I}_l\right), \tag{5.32}$$

where $\exp()$ is taken element-wise, $\otimes$ is the Kronecker product, and $\boldsymbol{I}_l$ is the $l \times l$ identity matrix, and β is a scale parameter. In this way, the resulting matrix $\boldsymbol{B}$ is of size $q \times q$, and $\boldsymbol{A}$ is of size $q/l \times q/l$. This definition of $\boldsymbol{B}$ focuses the regularized estimates to a region of covariance space, where shape point covariances are related by proximity.

In [107] geodesic distance were used, but in this work we prefer Euclidean. In general, whether one chooses the Euclidean or the geodesic distance is a modeling choice. If, for example, the shape change is due to friction, then it is fair to assume

that shape changes correlate on the surface with nearby points. On the other hand, shape changes for soft tissue may show correlations between surface points that are better modeled by their direct Euclidean distance than their geodesic surface distance. Regardless, considering only points on the surface of an object and for small values of β in comparison to the distance between sample points, the Euclidean and the geodesic distances yield similar values of $\boldsymbol{B}$ in (5.32). Further, when objects are flat sheet as is the case for the tibial knee cartilage sheets described in Section 5.4.2, the Euclidean distance approximates the geodesic distance well. Finally, the Euclidean distance is the easiest to calculate. We choose to model only correlation between nearby points and therefore small values of β and Euclidean distance.

In this section, we have given an example of the mean value matrix $\boldsymbol{B}$ based on the idea that proximity implies covariance. Many other, and perhaps better, choices are possible, but exhaustive investigation is beyond the scope of this paper.

5.4 Experiments

We now describe the results obtained using the validation protocol and setup described in Section 5.3. First, we report the matrix comparisons between ground truth and covariance estimates for the synthetic data. Secondly, we discuss the shape reconstruction for vertebrae and cartilage shapes.

5.4.1 Synthetic shapes, setup and data

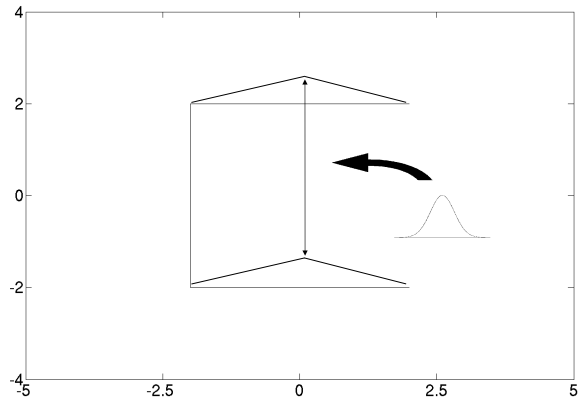
In this subsection, we describe the data setup of the experiments with synthetic data and the results of performing a matrix comparison ad described in 5.3.1. The synthetic data set is designed to explore two types of variability: linear shape change and independent landmark random variations. First, we design a basis shape resembling a vertebra. Second, the linear shape variation is created by covarying two points along the vertical direction by a normally distributed displacement, which is a typical shape variation seen in real data. Third, normally distributed noise is added to each landmark to simulate annotation noise. We impose 3 different covariance matrices:

$$\boldsymbol{C}_1 = \gamma \begin{bmatrix} \kappa_1 & -\kappa_2 \\ -\kappa_2 & \kappa_2 \end{bmatrix}, \tag{5.33a}$$

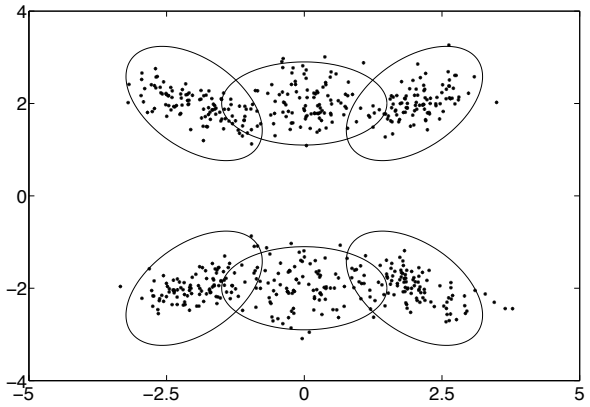
$$\boldsymbol{C}_2 = \gamma \begin{bmatrix} \kappa_1 & \kappa_2 \\ \kappa_2 & \kappa_2 \end{bmatrix}, \tag{5.33b}$$

$$\boldsymbol{C}_3 = \gamma \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix}. \tag{5.33c}$$

Figure 5.3(a) illustrates the linear shape variation and Figure 5.3(b) examples of noise. For the noise, we have used $\gamma = 2$, $\kappa_1 = 0.3$, and $\kappa_2 = 0.1$, and $\boldsymbol{C}_1$ for the top right and bottom left corners, $\boldsymbol{C}_2$ for the remaining corners, and $\boldsymbol{C}_3$ for the mid sections. Examples of resulting shapes are shown in Section 5.4. We estimate the ground truth covariance matrix by using ML on a very large and independent data set of 10.000



(a)



(b)

Figure 5.3: (a) Mean shape of the synthetic data and an example of linear shape variations obtained from a Gaussian distribution, where the two central points covary vertically. (b) Example of noise variation, where the ellipses denote the standard deviation. Samples of the resulting shapes are shown in Figures 5.4 and 5.5.

samples. Furthermore, we generate 5000 samples, from which the 25 crossfolds are generated as described in Section 5.3.3.

The experiments on the synthetic data previously described lead to the results depicted in Figures 5.4 and 5.5. Here, (5.25) is used to compare the ground truth and the estimated matrices on a small sample size using the different methods (ML, ML-T, MAP-G, MAP-IW, and MAP-UIW) for two noise levels. For each method, the variation for cross-validation is illustrated using standard error of the mean (SEM) error bars. Furthermore, significance of differences between the best method and Tikhonov regularization is illustrated by stars, where '*' denotes a t-test p-value below 0.05, '**' below 0.01, and '***' below 0.001. We observe that all the suggested methods outperform ML, and further that the MAP-IW and MAP-UIW priors are suitable choices when the shape variation is dominated by noise (large γ from (5.33), illustrated in Figure 5.5), while the MAP-G prior is a better choice when there is a dominant shape variation (illustrated in Figure 5.4). In addition, Figures 5.4 and 5.5 show that with more available data less regularization is required; and the differences between the methods are less pronounced. It is also important to bear in mind that except for MAP-G, all the methods share the same eigenvectors and differ only in the eigenvalues. This is likely reflected by the clustering of graphs not including MAP-G.

5.4.2 Increasing shape resolution, setup and real data

In this section, we explain the experimental setup and results for increasing shape resolution. These experiments are performed on two different data sets: vertebrae and cartilage shapes. For the aim of detecting osteoporotic fragility fractures, the use of high-resolution (full boundary) vertebral shapes may lead to more reliable results than an analysis with low-resolution shapes [23, 61]. For studies on osteoarthritis, more detailed 3D shape models of cartilage may also lead to more accurate results.

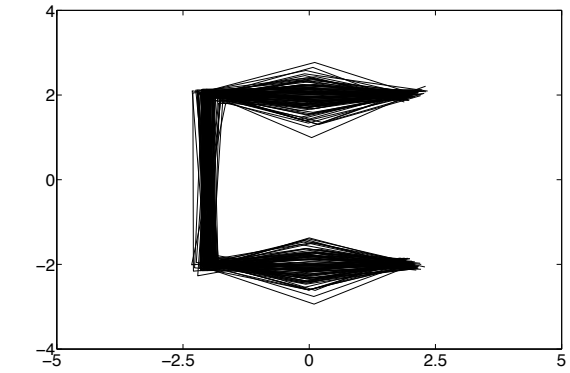
Using the shape reconstruction model (5.29), a high-resolution shape was reconstructed from a low-dimensional shape. The reconstruction error for a single shape is computed for the ζ points of the boundary as the root mean square error (RMSE),

$$E_{\text{full}} = \sqrt{\frac{1}{\zeta}\sum_{i=1}^{\zeta} \left\| \boldsymbol{p}_{i,reconst.} - \boldsymbol{p}_{i,orig.} \right\|_2^2}. \tag{5.34}$$

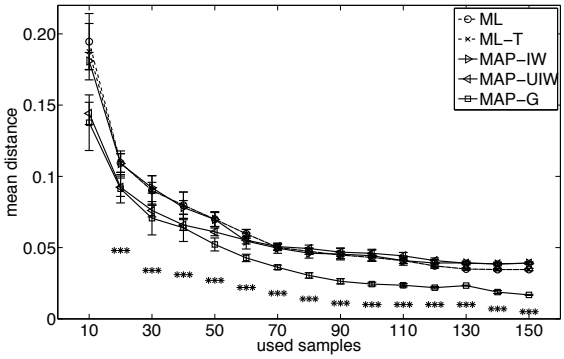
The performances of ML, ML-T, MAP-IW, and MAP-UIW methods are compared using mean E_{full} over all the test shapes for different number of principal eigenmodes.

Vertebra shapes

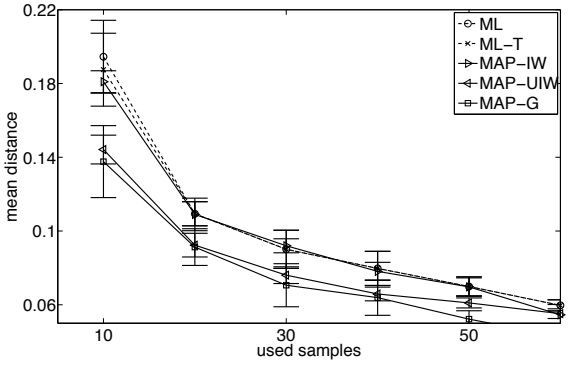
For general shape analysis, a detailed outline of the shapes are typically used (as illustrated in Figure 5.6(a)). However, for clinical studies on osteoporosis, height ratios are often used for fracture quantification [30, 83] and only the six vertebra contour points needed to compute these height ratios are routinely marked by expert radiologists. These six landmark points are shown as large stars in Figure 5.6(b).



(a)

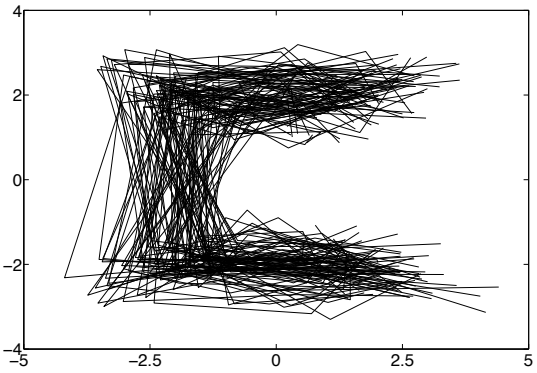


(b)

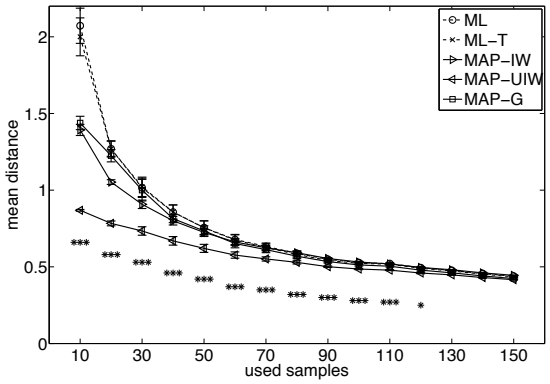


(c)

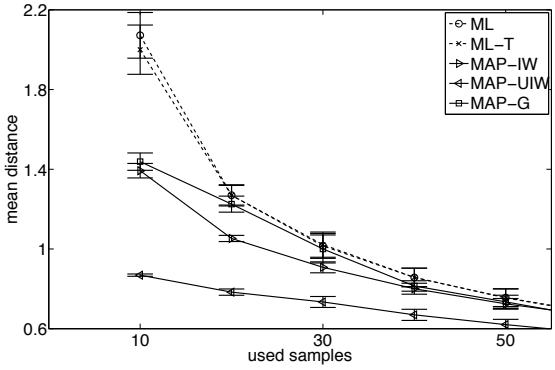
Figure 5.4: Frobenius norm of the difference between the ground truth covariance matrix and the estimation with ML, ML-T, MAP-IW, MAP-UIW, and MAP-G priors. All the methods are compared for different number of samples, all the parameters are determined by cross-validation as described in Section 5.3.3. Subfigure (a) shows the data obtained using (5.33) with $\gamma = 0.1$, (b) shows the difference as described in Section 5.4.1, while (c) shows a zoom of the graph in (b). The stars indicate statistical significance of the differences between MAP-G and ML-T (Tikhonov) where '*' denotes a t-test p-value below 0.05, '**' below 0.01, and '***' below 0.001.



(a)

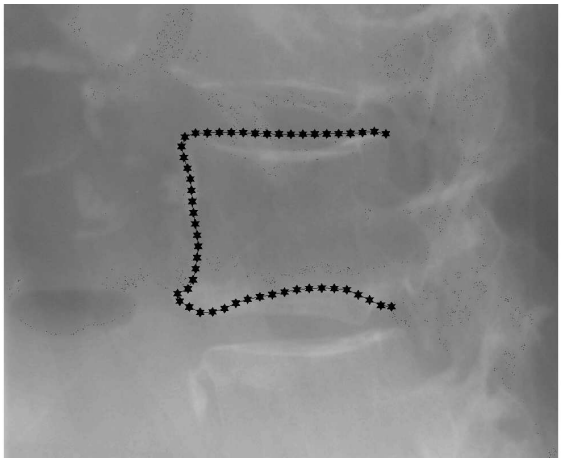


(b)

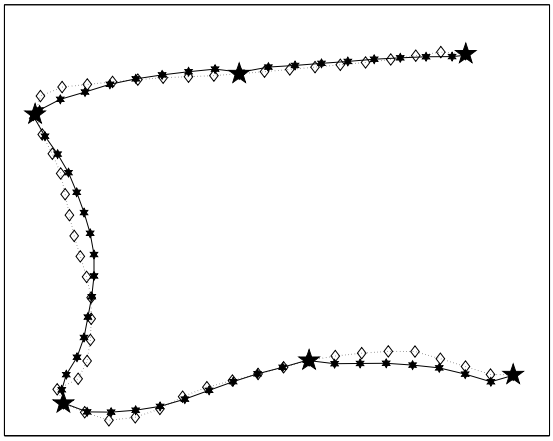


(c)

Figure 5.5: Frobenius norm of the difference between the ground truth covariance matrix and the estimation with ML, ML-T, MAP-IW, MAP-UIW, and MAP-G priors. All the methods are compared for different number of samples, all the parameters are determined by cross-validation as described in Section 5.3.3. Subfigure (a) shows the data obtained using (5.33) with $\gamma = 2$, (b) shows the difference as described in Section 5.4.1, while (c) shows a zoom of the graph in (b). The stars indicate statistical significance of the differences between MAP-UIW and ML-T (see Figure 5.4 for details).



(a)



(b)

Figure 5.6: (a) An image of a vertebra with the shape annotation. (b) An original (continuous line) and reconstructed (dashed line) shape annotation. The shape's 52 points are reconstructed using only the 6 points depicted as the big stars.

The aim is thus to interpolate the full contour (shown as full outline in Figure 5.6(b)) from these six points. During the experiments, the shape model is built from the full outline using 52 landmarks per shape. Hence, in our experiments $q = 104$ and $k = 12$.

We study the vertebra shape variation using images of 75 healthy elderly women,

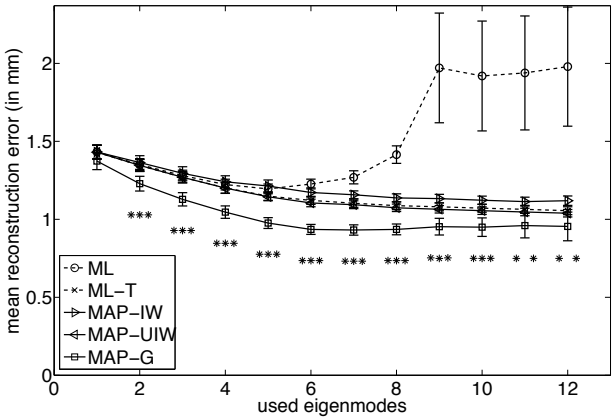
who maintain skeletal integrity over a 7 years observation period. The subjects had lateral X-ray of the lumbar spine taken twice, once in 1992-93 (baseline) and again in 2000-01 (follow-up). The used images were digitized using a Vidar DosimetryPro Advantage scanner at 45 microns (570 dpi) providing an image resolution of 9651×4008 pixels with 12-bit gray scale. The four lumbar vertebrae L1-L4 were annotated using an annotation tool developed in Matlab. Using both the baseline and follow-up images, we obtain 600 (75 patients × 2 visits × 4 vertebrae) vertebra shape annotations. To avoid biased results, we only use shapes from the follow-up set. Furthermore, we only use 1 of 4 possible vertebrae per patient selected at random from L1 to L4. Therefore, the total data set of 600 shapes is reduced to 75 shapes. From this we either train on 10 or 20 shapes, and always test on 35 shapes. Figure 5.7 shows the mean reconstruction errors for vertebrae shapes obtained using (5.34) for two different training sample sizes as a function of the number of eigenvalues t. The experiments demonstrate that all the MAP-based methods improve the reconstruction especially for small sample sizes. Where even these minor improvements in terms of vertebral shape reconstruction can actually affect positively the clinical interpretation in terms of, e.g., establishing first incident vertebral fracture risk [79]. Here, risk evaluation was based on pre-fracture vertebral deformities below established fracture thresholds. Again, the Gaussian prior appears the most suitable for dealing with shapes dominated by global shape variation rather than annotation noise. Figure 5.7(a) shows that the ML method has a pronounced error for 9 eigenmodes and more caused by the small training set (leave-one-out with only 10 samples). Similarly, the steep rise at 12 eigenvalues in Figure 5.7(b) reflects inherent co-linearity of the training set caused by the data set not spanning a subspace larger than 11 eigenvalues. Results not included for 40 training shapes show a similar but less pronounced behavior.

Cartilage shapes

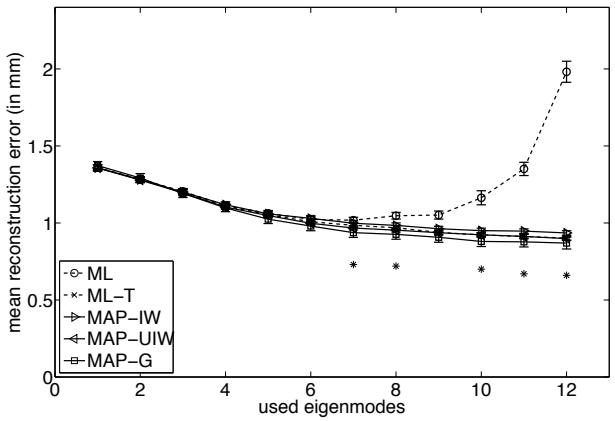
The cartilage data set was composed of 320 knee MRI scans from 80 subjects including both left and right knees and baseline and follow-up scans from a longitudinal 21-month study. In particular, we focus on the tibial cartilage sheets, which are approximately flat. Similarly to the vertebrae, the data set was reduced to 80 samples to ensure only one shape per patient. The MRI acquisition was done on an Esaote C-Scan low field 0.18T clinical scanner using 3D, T1-weighted sequences (flip angle 40°, TR 50ms, TE 16ms). The scans were sagittal with slice resolution of 0.7mm × 0.7mm and slice thickness of 0.8mm. The dimensions of the scans were 256 × 256 pixels with around 110 slices. The population includes healthy and diseased knees with varying degree of osteoarthritis from both men and women at ages from 21 to 78.

We represent the knee cartilage sheets using the m-rep model [67]. Here, a shape is defined by a spatially regular lattice of medial atoms as depicted in Figure 5.8(a). For each knee, we have a three-dimensional m-rep model of the medial tibial cartilage compartment resulting from a fully automatic segmentation [21]. In Figure 5.8(a) a cartilage shape model is illustrated for a cropped knee MRI.

For m-reps, each atom is representation by a tuple of position, orientation, and



(a)



(b)

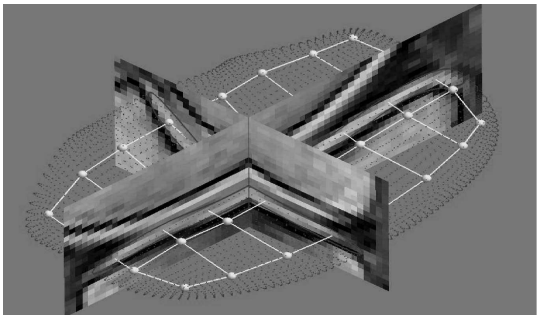
Figure 5.7: The vertebra reconstruction error for different number of training shapes: (a) 10 and (b) 20. The error measure is described in section 5.3.2 and it is defined in mm. The stars indicate statistical significance of the differences between MAP-G and ML-T (see Figure 5.4 for details).

radius parameters. A more detailed explanation of this model is given in [37]. For this paper, we only consider the position parameters to allow an analysis similar to the point distribution model used for the vertebra shapes.

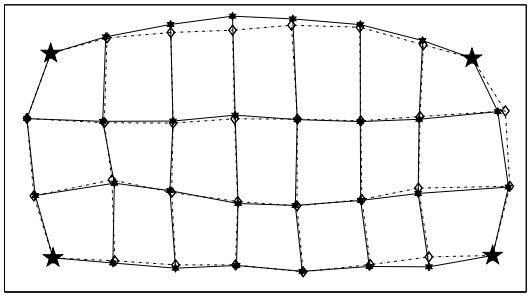
We evaluate the performance of our method for interpolating medial atoms in order to produce a higher resolution medial model of the cartilage sheets. We perform this evaluation by removing atoms from the models and by measuring how well the interpolation allows for reconstruction of the original model. We train on 10 or 20 shapes, and always test on 35 shapes from the remaining shapes. We produce a low-resolution lattice by removing atoms from the cartilage model, and calculate the mean reconstruction error between the original and the reconstructed 3D shapes using (5.34). To represent a medial model for a cartilage sheet, $\zeta = 32$ points are used, hence the dimensionality $q = 96$. During the experiments 24 points are removed leaving 4 points, such that the reduced dimensions is $k = 12$. A typical result of the reconstruction is shown in Figure 5.8. The mean reconstruction error using cartilage shapes is obtained using (5.34) and depicted in Figure 5.9. The cartilage shapes show an improvement using the Gaussian prior. Like for the vertebra, the MAP-G method is more appealing because the shapes are dominated more by global shape variations rather than noise. The observation that MAP-G achieves a given reconstruction error using fewer eigenmodes compared to the other methods indicate that MAP-G estimates a more compact representation of the data.

5.5 Conclusion

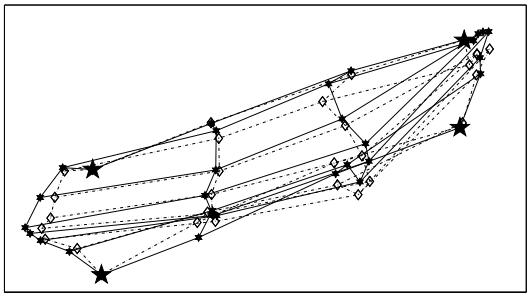
Efficient estimation of the covariance matrix is of high importance in statistical shape analysis. Often, the number of available training examples is limited and the estimation of a covariance matrix is a challenging task. We have studied a Bayesian approach to the problem as an alternative to the well-known methods of Maximum Likelihood (ML) and Tikhonov regularization (ML-T), and we have investigated three different priors: the Inverted Wishart (MAP-IW), the Uncommitted Inverted Wishart (MAP-UIW), and the Gaussian prior (MAP-G). Comparing with respect to the parameters, the ML method has no parameters, the ML-T has one parameter, the MAP-UIW prior has two, while the remaining two priors have $\frac{(q+1)\cdot q}{2} + 1$ parameters in their unconstrained form. Both the MAP-IW prior and the MAP-G prior require a full matrix of parameters, but we have suggested constrained forms, that can be deduced automatically. The MAP-IW method can reflect the noise variation of independent landmark building the $\boldsymbol{\Psi}$ matrix as a block diagonal matrix as described in Section 5.2.1. Contrarily, the MAP-G mean matrix $\boldsymbol{B}$ allows for inter-landmark interactions, like the suggested geometrical relationships in Section 5.3.4. The Wishart priors also depend on a nuisance parameter m, but we remind that for the reconstruction task the results do not seem to be very sensitive to its value. The presented methods perform as well as or better than the ML method especially in the case of small number of training examples. The choice of prior is naturally of great importance for the result. The MAP-IW and MAP-UIW priors assume statistically independent points, and therefore perform better for variations caused by uncorrelated noise. The



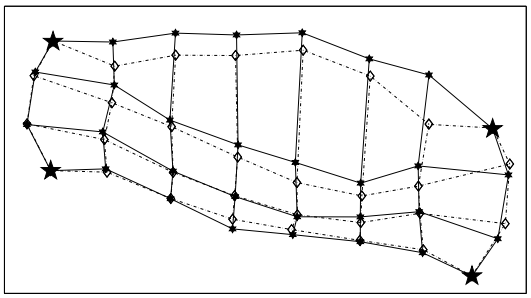
(a)



(b)

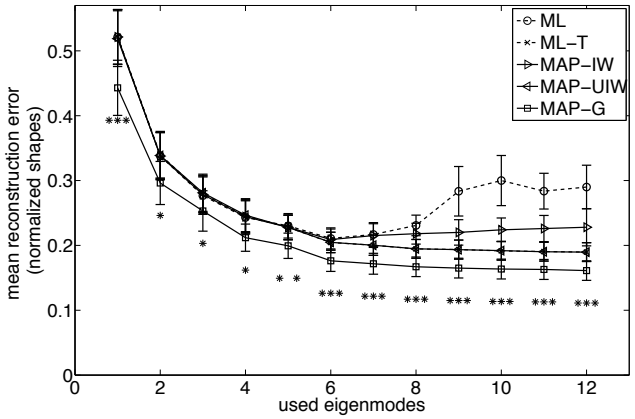


(c)

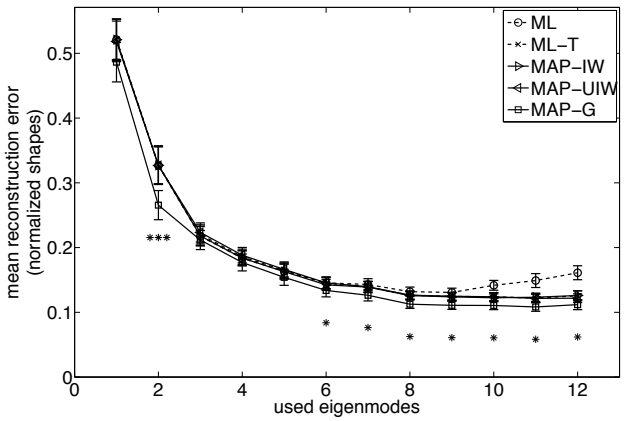


(d)

Figure 5.8: (a) Sagittal and axial MRI slice with the shape model for the medial tibial cartilage compartment. (b) Overlay of the original and the reconstructed cartilage shape viewed axially, (c) coronally, and (d) sagittally. The continuous line is the original shape and the dashed one is the reconstruction. Here the 32 points are reconstructed using only the 4 points depicted as big stars.



(a)



(b)

Figure 5.9: The cartilage reconstruction error for different number of training shapes: (a) 10 and (b) 20. The error measure is described in section 5.3.2 and it is defined in size normalized shapes. The stars indicate statistical significance of the differences between MAP-G and ML-T (see Figure 5.4 for details).

MAP-G prior is not limited to zero off-block-diagonal elements and may be used to steer the estimate towards preferred shape variations. Therefore, the choice of prior is related to various factors: the number of samples available, the complexity of the data, the noise variance in the data and the intended application context in which estimated covariance matrices will be used. In general, the prior and specifically for the MAP-G the $\boldsymbol{B}$ matrix has to be chosen using prior knowledge optimizing this interplay between prior, data, and application. The focus of this paper has been to introduce novel prior frameworks for covariance matrix estimation and to give simple examples of how they can be used to introduce actual prior knowledge in both artificial and real data sets. But even with the slightly naïve $\boldsymbol{B}$ matrix suggested for the MAP-G prior, the improvements over standard approaches are significant for all but one of the examples given. Finally, the paper presented one case with a strong or near-optimal link between prior and data is present. This was the case for the very noisy version of the artificial data, where as the expected the Uncommitted Inverted Wishart prior excelled. In summary, with a sufficient number of samples the standard ML estimation may be the right choice, while for low sample sizes regularization is required. In this case, we argue that our methods can be valid substitutes to Tikhonov regularization.

Appendix A: Derivation of Maximum Likelihood Estimator (ML)

For completeness the following gives the derivation of the log-likelihood estimates of the mean vector and covariance matrix using matrix differential calculus [81]. This formulation will be used as a reference point for deriving expressions for the MAP estimates.

Using (5.6), for estimating $\boldsymbol{\Sigma}$ and $\boldsymbol{x}$ from a set of samples, it is possible to seek the maximum point of $p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n|\boldsymbol{\Sigma}, \boldsymbol{\mu})$ as,

$$\frac{\partial \log p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n|\boldsymbol{\Sigma}, \boldsymbol{\mu})}{\partial \boldsymbol{\Sigma}} = 0, \tag{5.35a}$$

$$\frac{\partial \log p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n|\boldsymbol{\Sigma}, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = 0. \tag{5.35b}$$

For practical reasons it is possible to rewrite the sum under the exponential function of (5.6) as,

$$\sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma^{-1}} (\boldsymbol{x}_i - \boldsymbol{\mu}) = \mathrm{tr}\left(\boldsymbol{\Sigma^{-1}} \boldsymbol{X} \boldsymbol{X^T}\right), \tag{5.36}$$

and to rewrite the logarithm of (5.6) as

$$L(x_1, \ldots x_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = - \frac{nq}{2} \log(2\pi) - \frac{n}{2} \log(\boldsymbol{\Sigma}) - \frac{1}{2} \mathrm{tr}\left(\boldsymbol{\Sigma^{-1}} \boldsymbol{X} \boldsymbol{X^T}\right). \tag{5.37}$$

The differential of L, considering only $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ as variables, is found to be

$$dL = - \operatorname{tr}\left(\frac{n}{2}\boldsymbol{\Sigma^{-1}}d\boldsymbol{\Sigma} + \frac{1}{2}(d\boldsymbol{\Sigma^{-1}})\boldsymbol{X}\boldsymbol{X^T} + \frac{1}{2}\boldsymbol{\Sigma^{-1}}d(\boldsymbol{X}\boldsymbol{X^T})\right). \quad (5.38)$$

For minimizing the partial derivatives of dL for $\boldsymbol{\Sigma}$, it is possible to isolate the first two terms of (5.38) considering $d\boldsymbol{\Sigma^{-1}} = -\boldsymbol{\Sigma^{-1}}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma^{-1}}$:

$$0 = -\operatorname{tr}\left(\frac{n}{2}\boldsymbol{\Sigma^{-1}}d\boldsymbol{\Sigma} - \frac{1}{2}\boldsymbol{\Sigma^{-1}}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma^{-1}}\boldsymbol{X}\boldsymbol{X^T}\right) \quad (5.39a)$$

$$= -\operatorname{tr}\left(\boldsymbol{\Sigma^{-1}}(d\boldsymbol{\Sigma})(n\boldsymbol{I} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{X^T})\right), \quad (5.39b)$$

where a non trivial solution is $0 = n\boldsymbol{I} - \boldsymbol{\Sigma^{-1}}\boldsymbol{\Sigma^{-1}}\boldsymbol{X}\boldsymbol{X^T}$ that is

$$\boldsymbol{\Sigma} = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T, \quad (5.40)$$

equivalent to (5.7b). Alternatively, starting from the Wishart likelihood,

$$p(\boldsymbol{S}|\boldsymbol{\Sigma}, n) = \frac{|\boldsymbol{S}|^{(n-q-1)/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\right)}{2^{nq/2}\,|\boldsymbol{\Sigma}|^{n/2}\,\Gamma_q\left(\frac{n}{2}\right)}, \quad (5.41)$$

the point of Maximum Likelihood (ML) for varying $\boldsymbol{\Sigma}$ is found to be,

$$0 = d\log p(\boldsymbol{S}|\boldsymbol{\Sigma}, n) \quad (5.42a)$$

$$= -\frac{1}{2}d\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) - \frac{n}{2}d\log|\boldsymbol{\Sigma}| \quad (5.42b)$$

$$= \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - n\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}\right), \quad (5.42c)$$

which implies that

$$0 = \boldsymbol{\Sigma}^{-1}\boldsymbol{S} - n\boldsymbol{I}, \quad (5.43)$$

or therefore $\boldsymbol{\Sigma} = \boldsymbol{S}/n$, also equivalent to (5.7b).

Appendix B: Derivation of Maximum A Posteriori Inverted Wishart Estimator (MAP-IW)

In the following, we will derive (5.17). Starting from (5.15) and only considering varying $\boldsymbol{\Sigma}$, we calculate the differential log as,

$$d\log p(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, m) = -\frac{(m+q+1)}{2}d\log|\boldsymbol{\Sigma}| - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Psi}d\boldsymbol{\Sigma}^{-1}\right) \quad (5.44a)$$

$$= -\frac{(m+q+1)}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}\right) + \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right) \quad (5.44b)$$

$$= \operatorname{tr}\left(\left(\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1} - \frac{(m+q+1)}{2}\boldsymbol{\Sigma}^{-1}\right)d\boldsymbol{\Sigma}\right), \quad (5.44c)$$

The maximum a posteriori using i.i.d. Gaussian likelihood is found to be,

$$0 = d\log p(\boldsymbol{x}_1, ..., \boldsymbol{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) + d\log p(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, m) \tag{5.45a}$$

$$= \frac{1}{2}\text{tr}\left(\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1}\right)d\boldsymbol{\Sigma}\right) + \text{tr}\left(\left(\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1} - \frac{(m+q+1)}{2}\boldsymbol{\Sigma}^{-1}\right)d\boldsymbol{\Sigma}\right), \tag{5.45b}$$

which implies that

$$0 = \boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1} - (m+q+1)\boldsymbol{\Sigma}^{-1} \tag{5.46a}$$

$$= \boldsymbol{X}\boldsymbol{X}^T - n\boldsymbol{\Sigma} + \boldsymbol{\Psi} - (m+q+1)\boldsymbol{\Sigma}, \tag{5.46b}$$

and

$$\boldsymbol{\Sigma} = \frac{1}{m+n+q+1}\left(\boldsymbol{X}\boldsymbol{X}^T + \boldsymbol{\Psi}\right), \tag{5.47}$$

which is seen to be identical to (5.17).

Appendix C: Derivation of Maximum A Posteriori Uncommitted Inverted Wishart Estimate (MAP-UIW)

In the following, we will derive (5.19). For the moment assume that $\boldsymbol{\Sigma}$ is invertible, consider (5.14) and let $\boldsymbol{\Psi} = \alpha^2\boldsymbol{\Sigma}^{-1}$. We use the improper inverted Wishart prior $p(\boldsymbol{\Sigma}|\alpha^2\boldsymbol{\Sigma}^{-1}, m)$. Considering only varying $\boldsymbol{\Sigma}$, the differential log of p is,

$$d\log p(\boldsymbol{\Sigma}|\alpha^2\boldsymbol{\Sigma}^{-1}, m) = \frac{m}{2}d\log\left|\boldsymbol{\Sigma}^{-1}\right| - \frac{1}{2}d\text{tr}\left(\alpha^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\right) - \frac{(m+q+1)}{2}d\log|\boldsymbol{\Sigma}| \tag{5.48a}$$

$$= \frac{m}{2}\text{tr}\left(\boldsymbol{\Sigma}d\boldsymbol{\Sigma}^{-1}\right) - \text{tr}\left(\alpha^2\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}^{-1}\right) - \frac{(m+q+1)}{2}\text{tr}\left(\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}\right) \tag{5.48b}$$

$$= -\frac{m}{2}\text{tr}\left((d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right) + \alpha^2\text{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right) - \frac{(m+q+1)}{2}\text{tr}\left(\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Sigma}\right) \tag{5.48c}$$

$$= \text{tr}\left(\alpha^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1} - \frac{(2m+q+1)}{2}(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right). \tag{5.48d}$$

The maximum a posteriori using independent and identically distributed (i.i.d.) Gaussian likelihood is found to be,

$$0 = d\log p(\boldsymbol{x}_1, ..., \boldsymbol{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) + d\log p(\boldsymbol{\Sigma}|\alpha^2\boldsymbol{\Sigma}^{-1}, m) \tag{5.49a}$$

$$= \frac{1}{2}\mathrm{tr}\left(\left(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1} - n\boldsymbol{I}\right)(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right) + \mathrm{tr}\left(\left(\alpha^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1} - \frac{(2m+q+1)}{2}\boldsymbol{I}\right)(d\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right), \tag{5.49b}$$

which implies that

$$0 = \frac{1}{2}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1} + \alpha^2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1} - \frac{(n+2m+q+1)}{2}\boldsymbol{I} \tag{5.50a}$$

$$= \frac{1}{2}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma} + \alpha^2\boldsymbol{I} - \frac{(n+2m+q+1)}{2}\boldsymbol{\Sigma}^2. \tag{5.50b}$$

Notice, that the above equation is also valid for positive semi-definite matrices $\boldsymbol{\Sigma}$. Hence, we conclude,

$$\boldsymbol{\Sigma}^2 = \frac{1}{(n+2m+q+1)}\left(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma} + 2\alpha^2\boldsymbol{I}\right), \tag{5.51}$$

which is seen to be (5.19).

Appendix D: Derivation of Maximum A Posteriori Gaussian Estimate (MAP-G)

In the following, we will derive (5.2.3). Starting from (5.22) and only considering varying $\boldsymbol{\Sigma}$, we find the differential log as,

$$d\log p(\boldsymbol{\Sigma}) = -\frac{1}{2s^2}d\mathrm{tr}\left((\boldsymbol{\Sigma}-\boldsymbol{B})^T(\boldsymbol{\Sigma}-\boldsymbol{B})\right) \tag{5.52}$$

$$= -\frac{1}{s^2}\mathrm{tr}\left((\boldsymbol{\Sigma}-\boldsymbol{B})^T d\boldsymbol{\Sigma}\right) \tag{5.53}$$

The maximum a posteriori using i.i.d. Gaussian likelihood is found to be,

$$0 = d\log p(\boldsymbol{x}_1, ..., \boldsymbol{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) + d\log p(\boldsymbol{\Sigma}|\boldsymbol{B}, s) \tag{5.54a}$$

$$= \frac{1}{2}\mathrm{tr}\left(\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1}\right)d\boldsymbol{\Sigma}\right) - \frac{1}{s^2}\mathrm{tr}\left((\boldsymbol{\Sigma}-\boldsymbol{B})^T d\boldsymbol{\Sigma}\right), \tag{5.54b}$$

which implies that

$$0 = \boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\Sigma}^{-1} - \frac{2}{s^2}(\boldsymbol{\Sigma}-\boldsymbol{B})^T \tag{5.55a}$$

$$= \boldsymbol{X}\boldsymbol{X}^T - n\boldsymbol{\Sigma} - \frac{2}{s^2}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}-\boldsymbol{B})^T\boldsymbol{\Sigma}, \tag{5.55b}$$

and

$$\boldsymbol{\Sigma} = \frac{1}{n}\left(\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}} - \frac{2}{s^2}\boldsymbol{\Sigma}(\boldsymbol{\Sigma} - \boldsymbol{B})^T\boldsymbol{\Sigma}\right). \tag{5.56}$$

The equation above can be seen as the third degree matrix polynomial given in (5.2.3).

Chapter 6

Discussion

This thesis presented several techniques and algorithms regarding statistical shape models, covariance estimation, and their applications to osteoporotic vertebral fracture analysis. This chapter contains a summary of the thesis, conclusions, and a prospect of future works.

6.1 Summary

The different techniques employed in this thesis mainly comprised a combination of machine learning methods and shape representations. The study regarding vertebral fracture prediction, reported in Chapter 4, argued that vertebral fractures could not be exclusively considered as an abrupt change in the vertebrae. They can rather be considered as the result of a slowly deforming process involving bone quality decay and the bio-mechanics in the whole spine, which can be detected in advance. Therefore, a study involving a shape based approach was conducted for this purpose. Spine shapes were obtained from manual annotations on radiographs at baseline, when no fractures were present. Those shapes were than used as features for predicting the fracture risk. In particular, the conducted study was a longitudinal case-control investigation, where the women who sustained at least one incident lumbar fracture by follow-up represented the case group, and the osteoporosis risk factors matched women who maintained skeletal integrity by follow-up represented the control group. An expert radiologist graded the fracture risk of the patients' spines. The same radiologist and subsequently two technicians annotated the vertebrae independently. An LDA classifier was capable of assessing the fracture risk of the shapes giving an average AUC of 0.7 for the radiologist annotations and an average AUC of 0.6 for the technician annotations. Nevertheless, for all the sets of annotations the p-values of the comparisons between the scores and chance were smaller than 0.05. This method could be used in conjunction with the standard fracture risk factors, allowing a more reliable prediction. The remaining part of the thesis was devoted to improving shape based analysis, using a novel technique of covariance matrix estimation, called MAP-

PCA.

An accurate estimation of the covariance matrix is a crucial step in shape analysis and other statistical tasks. The number of available training examples is often small compared to the dimensionality of the data, and the consequent estimation of a covariance matrix often unreliable. In Chapter 5, a Bayesian approach to this problem, as an alternative to the well-known methods of Tikhonov regularization, was proposed. In particular, three different priors were investigated: the Inverted Wishart (MAP-IW), the Uncommitted Inverted Wishart (MAP-UIW), and the Gaussian prior (MAP-G). The MAP-IW and MAP-UIW methods can reflect the noise variation of independent landmark building a prior matrix as a block diagonal matrix. Differently, the MAP-G prior expresses inter-landmark interactions based on the geometrical relationships of landmarks. In the reported experiments, the Bayesian methods outperformed the traditional ML method and Tikhonov regularization, reducing to 1 mm the mean reconstruction error for the vertebral shapes, and to 20% mean error for the cartilage sheets, using only 10 samples for both experiments.

6.2 General discussion

Prediction of osteoporosis and vertebral osteoporotic fractures using shape model based analysis is not a trivial task. So far, the prediction of osteoporosis and related fractures has mainly been based on risk factors, such as, BMD, age, prior fragility fractures, and smoking [76]. In Chapter 4, a study based on a statistical shape based framework for predicting future fractures was reported. The shape based approach used in this study was based on manual annotations. Furthermore, the experiments were performed using two different approaches of shape alignment inside the framework: the well-known GPA method and the supervised DSA. The GPA method performed well only on the radiologist annotations, while the DSA method allowed also better performance for the technician annotations, since the DSA algorithm included class label information. However, the DSA algorithm was computationally more demanding.

Even using the DSA algorithm, the discriminative performances of the proposed algorithm on the two repeated annotations performed by the trained technicians, were not as high as for the original annotation but still statistically significant (p-values < 0.05). The discrepancy can reflect the fact that radiologist annotations are qualitatively more precise than technician annotations, probably due to more experience and acquired knowledge of the radiologist. This sensitivity to annotation quality has also been seen in another study using a different morphometric index [79], tested on almost the same data-set. With this in mind, the major conclusion is to advice the use of high quality, expert annotations for obtaining reliable features for fracture prediction. Nevertheless, the approach reported in Chapter 4 can be used in other prediction and classification contexts, where label information of the training data-set is available, since it is based on the general concept of shape, while the heights method is limited to vertebral applications.

Moreover, the few case patients available introduced the challenge of conducting

clinical trials, using small sample size compared to the data dimension. In Chapter 5, a novel methodology for estimating covariance matrices, when few samples are available using prior knowledge was proposed. The described methods perform as well as or better than the ML method especially in case of small number of training examples. The choice of prior is crucial for the result. The MAP-IW and MAP-UIW priors were designed assuming that the shapes are made of statistically independent points, and therefore they perform better on data with variations caused by uncorrelated noise of the landmarks. The MAP-G prior is not limited to zero off-block-diagonal elements and may be used in case of redefinition of eigenvectors. In experiments with real data reported in Chapter 5, the MAP-G prior outperformed the other methods. These results suggested that, in the used data, the overall shape information is stronger than the independent noise landmark information. Nevertheless, with a sufficient number of samples the standard ML estimation may be the right choice, while for low sample sizes regularization is required. Therefore, the choice of prior is related to various factors: the number of available samples, the complexity of the data, the noise variance in the data and the intended application context in which estimated covariance matrices will be used. Even so, the proposed methods can be valid substitutes to Tikhonov regularization. Figure 6.1 depicts how the space of the covariance estimation can be summarized. The figure remarks the prior choice issue. Depending on aspects, such as the statistical task in which the matrix will be used and the available data, a thoughtful choice of the prior must be carried out. A prior should include some knowledge available from the data. For instance, a proximity relationship between points for landmark based shapes or a particular curvature for some specific surfaces.

6.3 Future prospects

Regarding the shape based fracture prediction, the fracture status of each vertebra was assessed by follow-up by an expert radiologist using the Genant's SQ method, and re-evaluated by a computer algorithm using a modification of Genant's methodology with a strict measured threshold, where a vertebra was considered fractured if the ratios between any of the anterior, middle, and posterior heights was 0.8 or smaller. This additional step was deliberately used for filtering out subjects where there was borderline disagreement between the SQ and QM based fracture assessments, preventing the algorithm from being influenced by such a disagreement. It would be relevant to validate the described fracture-risk algorithm with respect to changing the "ground truth" fracture assessment methodology to other established methods, such as the QM methods of Melton and McCloskey [30, 83]. Moreover, the described method relies on shapes manually annotated from radiographs. Even if the framework can be extended with existing spine segmentation approaches [61, 106, 119]. By extending it with automatic segmentation of the spine [61], it will become independent from manual annotations. However, this can be a challenging task considering the discrepancy between the results obtained using radiologists or technicians annotations, and an automatic annotations would possibly be imprecise [98] as the technician an-

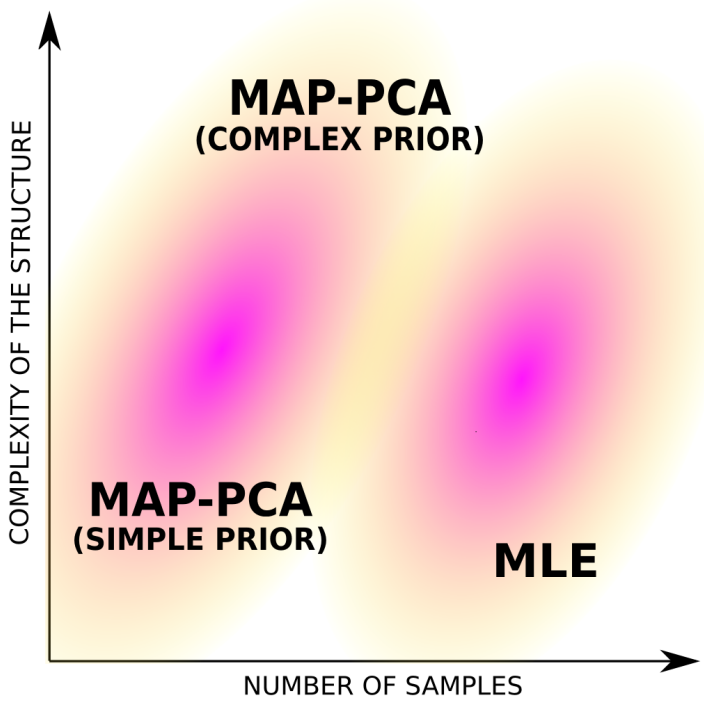


Figure 6.1: Representation of the covariance estimation problem space.

notations. Furthermore, an interesting extension of the prediction algorithm would be to merge biological data (such as CTX, NTX, etc.) with shape features during the prediction. Hopefully increasing the quality of the assessment, and disclosing further relationships between changes in shape and bio-markers. When other populations will be available, a comparison between non matched case-control group will also allow to include the entire set of risk factors of the Frax tool.

Chapter 5 discussed a novel approach for covariance matrix estimation and gave simple examples of how they can be used to introduce prior knowledge. In particular, we considered some cases of artificial and real data, where either the global shape information or the independent noise information were predominant. In each case, one prior method outperformed the others, and this pointed out the issue of the choice of prior. As a future work, a further investigation of other priors related to shape based analysis, can be carried out. For example, different measurement of covariations between the landmarks, or the distance between two vertebral endplates given by the vertical heights.

Beyond the described improvements related to the single algorithms, all the works

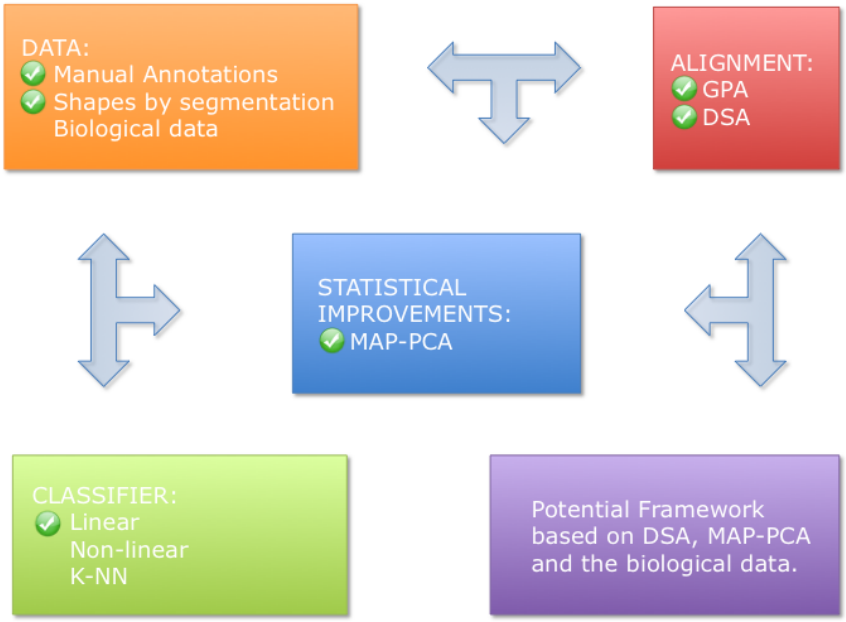


Figure 6.2: Prospect of present and possible future aspects of research on shape based fracture prediction. The already investigated aspects are marked by a ”checked“ sign.

presented in this thesis can be summarized in term of present and future investigations in Figure 6.2, where a combination of the aforementioned algorithms can be taken into account. Here, the already investigated aspects are marked by a ”checked“ sign. Other combinations can be also considered. The integration of a semi-supervised method for clustering can improve the results of data aligned with the DSA alignment or using a covariance matrix estimated with the MAPPCA algorithm. In particular, it is important to bear in mind that almost all the methods could benefit from a better regularization through MAP priors. Although the performance enhancement due to the combination of these elements is at the moment only theoretical.

Bibliography

[1] S. Allassonniere, Y. Amit, and A. Trouvé. Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society*, 69(1):3–29, 2007.

[2] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, 3rd edition, 2003.

[3] M. Aouache, A. Hussain, S.A. Samad, A.H. Hamzaini, and A.K. Ariffin. Active shape modeling of medical images for vertebral fracture computer assisted assessment system. In *Research and Development, 2007. SCOReD 2007.*, 2007.

[4] Y. Z. Bagger, L. B. Tanko, P. Alexandersen, H. B. Hansen, G. Qin, and C. Christiansen. The long-term predictive value of bone mineral density measurements for fracture risk is independent of the site of measurement and the age at diagnosis: results from the prospective epidemiological risk factors study. *Osteoporosis International*, 17:471–477, 2006.

[5] J.C. Baker-LePain, K.R. Luker, J.A. Lynch, N. Parimi, M.C. Nevitt, and N.E. Lane. Active shape modeling of the hip in the prediction of incident hip fracture. *Journal of Bone and Mineral Research*, 26(3):468–474, 2011.

[6] R. E. Bellman. *Dynamic Programming*. Courier Dover Publications, 2003.

[7] D. M. Black, N. K. Arden, and L. Palermo. Prevalent vertebral deformities predict hip fractures and new vertebral deformities but not wrist fractures. *Journal on Bone and Mineral Research*, 14:821–828, 1999.

[8] D.M. Black, D. E. Thompson, D. C. Bauer, and K. Ensrud. Fracture risk reduction with alendronate in women with osteoporosis: the fracture intervention trial. *Journal of Clinical Endocrinology Metabolism*, 85:4118–4124, 2000.

[9] V. Blanz and T. Vetter. Reconstructing the complete 3D shape of faces from partial information. Technical Report report of Computer Graphics no. 1, University of Freiburg, 2001.

[10] F. L. Bookstein. Landmark methods for forms without landmarks: Localizing group differences in outline shape. *Medical Image Analysis*, 1(1):225–243, 1997.

[11] F. L. Bookstein. Shape and the information in medical images: A decade of morphometric synthesis. *Computer Vision and Image Understanding*, 66(2):97–118, 1997.

[12] F.L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge University Press, 1997.

[13] M.L. Bouxsein and D. Karasik. Bone geometry and skeletal fragility. *Current Osteoporosis Reports*, 4:49–56, 2006.

[14] R. Burge, B. Dawson-Hughes, D.H. Solomon, J.B. Wong, A.King, and A. Tosteson. Incidence and economic burden of osteoporosis-related fractures in the united states. *Journal of Bone and Mineral Research*, 22(3):465–475, 2007.

[15] J.R. Center, T.V. Nguyen, N.A. Pocock, and J.A. Eisman. Volumetric bone density at the femoral neck as a common measure of hip fracture risk for men and women. *Journal of Clinical Endocrinology and Metabolism*, 89(6):2776–2782, 2004.

[16] R. D. Chapurlat, P. Garnero, G. Breart, and P. J. Meunier P. D. Delmas. Serum type i collagen breakdown product (serum ctx) predicts hip fracture risk in elderly women: the epidos study. *Bone*, 27:283–286, 2000.

[17] D. H. Cooper, T. F. Cootes, C. J. Taylor, and J. Graham. Active shape models, their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995.

[18] T. F. Cootes, A. Hill, C. J. Taylor, and J Haslam. Use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12:276–285, 1994.

[19] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Journal of Computer Vision and Image Understanding*, 61:38–59, 1995.

[20] T.F. Cootes, G.J. Edwards, and C.J.Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[21] E. B. Dam, P. T. Fletcher, and S. M. Pizer. Automatic shape model building based on principal geodesic analysis bootstrapping. *Medical Image Analysis*, 12(2):136–151, 2008.

[22] M. de Bruijne, M. T. Lund, L. B. Tanko, P. C. Pettersen, and M. Nielsen. Quantitative vertebral morphometry using neighbor-conditional shape models. In *Medical Image Computing and Computer-Assisted Intervention, R. Larsen, M. Nielsen and J. Sporring, eds., Lecture Notes in Computer Science*, pages 1–8, 2006.

[23] M. de Bruijne, M. T. Lund, L. B. Tanko, P. C. Pettersen, and M. Nielsen. Quantitative vertebral morphometry using neighbor-conditional shape models. *Medical Image Analysis*, 11:503–512, 2007.

[24] M. de Bruijne, P. Pettersen, L. Tanko, and M. Nielsen. Vertebral fracture classification. In *Proc. of SPIE Medical Imaging*, 2007.

[25] E.R. DeLong, D.M. DeLong, and D.L. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44:837–845, 1988.

[26] J. Dequeker, K. Gautama, and Y. S. Roh. Femoral trabecular patterns in asymptomatic spinal osteoporosis and femoral neck fracture. *Journal of Clinical Radiology*, 25:243–246, 1974.

[27] D. Diacinti, D. Pisani, F. Barone-Adesi, R. Del Fiacco, S. Minisola, V. Davide, G. Aliberti, and G. F. Mazzuoli. A new predictive index for vertebral fractures: The sum of the anterior vertebral body heights. *Bone*, 46:768–773, 2010.

[28] I.L. Dryden and K.V. Mardia. *Statistical Shape Analysis.* Wiley, 1998.

[29] R.O. Duda and P.E. Hart. *Pattern Recognition and Scene Analysis.* Wiley, 1973.

[30] R. Eastell, S. L. Cedel, H. W. Wahner, B. L. Riggs, and L. J. Melton. Classification of vertebral fractures. *Journal of Bone and Mineral Research*, 4(2):138–147, 1994.

[31] B. Efron and C. Morris. Multivariate empirical bayes and estimation of covariance matrices. *Annals of Statistics*, 4:22–32, 1976.

[32] J. D. Evans, D. Prudham, and I. Wandles. A prospective study of fractured proximal femur: Incidence and outcome. *Journal of Public Health*, 93:235–241, 1979.

[33] P.F. Felzenszwalb and D.P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61:2005–2046, 2003.

[34] P. Fillard, X. Pennec, V. Arsigny, and N. Ayache. Clinical DT-MRI estimation, smoothing, and fiber tracking with Log-Euclidean metrics. *IEEE Transactions on Medical Imaging*, 26(11):1472–1482, 2008.

[35] M. Fischler and R. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22:67–92, 1973.

[36] J.M. Fitzpatrick and M. Sonka. *Handbook of Medical Imaging.* SPIE Pubblications, 2009.

[37] P. T. Fletcher, S. Joshi, C. Lu, and S. Pizer. Gaussian distributions on lie groups and their application to statistical shape analysis. In C. Taylor and J. A. Noble, editors, *Information Processing in Medical Imaging*, pages 450–462, 2003.

[38] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 2004.

[39] Frax. *web-service and printable paper charts, http://www.shef.ac.uk/frax*. 2011.

[40] J. H. Friedman. Regularized discriminant analysis. *Journal of American Statistics Association*, 84:165–175, 1989.

[41] D. Fritsch, S. Pizer, L. Yu, V. Johnson, and E. Chaney. Segmentation of medical image objects using deformable shape loci. In *Proc. Information Processing in Medical Imaging*, pages 127–140, 1997.

[42] P. Garnero, O. Borel, and P. D. Delmas. Evaluation of a fully automated serum assay for c-terminal cross-linking telopeptide of type i collagen in osteoporosis. *Clinical Chemistry*, 47:694–702, 2001.

[43] S. Geisser. Discrimination, allocatory and separatory, linear aspects. *Classification and Clustering*, pages 301–330, 1977.

[44] H. K. Genant, C.Y.Wu, C.van Kuijk, and MC Nevitt. Vertebral fracture assessment using a semi-quantitative technique. *Journal on Bone and Mineral Research*, 8:1137–1148, 1993.

[45] H. K. Genant, M. Jergas, L. Palermo, M. Nevitt, R. S. Valentin, D. Black, and S. R. Cummings. Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures. *Journal on Bone and Mineral Research*, 11:984–996, 1996.

[46] H. K. Genant, C.Y. Wu, C. van Kuijk, and M.C. Nevitt. Vertebral fracture assessment using a semiquantitative technique. *Journal of Bone and Mineral Research*, 8:1137–1148, 1993.

[47] S. Goh, R.I. Price, S.Song, S.Davis, and K.P. Singer. Magnetic resonance-based vertebral morphometry of the thoracic spine: age, gender and level-specific influences. *Clinical Biomechanics*, 15:417–425, 2000.

[48] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall International, 2002.

[49] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society*, 53(2):285–339, 1991.

[50] A.A. Green, M. Berman, P. Switzer, and M.D. Craig. A transformation for ordering multispectral data in terms of image quality with implication for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65–74, 1988.

[51] J.S. Gregory, A. Stewart, P.E Undrill, D.M. Reid, and R.M. Aspden. Bone shape, structure and density as determinants of osteoporotic hip fracture: A pilot study investigating the combination of risk factors. *Investigative Radiology*, 40:591–597, 2005.

[52] J.F. Griffith, K. Engelke, and H.K. Genant. Looking beyond bone mineral density. imaging assessment of bone quality. *Skeletal Biology and Medicine*, 1192:45–56, 2010.

[53] L. R. Haff. Empirical bayes estimation of the multivariate normal covariance matrix. *Annals of statistics*, 8:586–597, 1980.

[54] T. Hastie, A. Buja, and R. Tibshirani. Penalized discriminant analysis. *Annals of Statistics*, 23(1):73–102, 1995.

[55] T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, 2003.

[56] T. Hildebrand and P. Ruegsegger. A new method for the model-independent assessment of thickness in three-dimensional images. *Journal of Microscopy*, 185:67–75, 1996.

[57] A. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 42(1):80–86, 2000.

[58] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational psychology*, 24:417–441, 1936.

[59] S.L. Hui, C.W.Slemanda, and C.C.Johnson. Age and bone mass as predictors of fracture in a prospective study. *Journal of Clinical Investigation*, 81:1804–1809, 1988.

[60] L. M. Hurxthal. Measurement of vertebral heights. *American Journal of Roentgenology*, 103:635–44, 1968.

[61] J.E. Iglesias and M. de Bruijne. Semiautomatic segmentation of vertebrae in lateral x-rays using a conditional shape model. *Academic Radiology*, 14(10):1156–1165, 2007.

[62] M. Ito, M. Ohki, K. Hayashi, M. Yamada, M. Uetani, and T. Nakamura. Relationship of spinal fracture to bone density, textural, and anthropometric parameters. *Calcification Tissue International*, 60(3):240–244, 1997.

[63] J. Iwamoto, T. Takeda, and S. Ichimura. Urinary cross-linked n-telopeptides of type i collagen levels in patients with rheumatoid arthritis. *Calcified Tissue Internatilnal*, 72(4):491–497, 2000.

[64] W. James and C. Stein. Estimation with quadratic loss. In *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 361–379, 1961.

[65] G. Jiang, R. Eastell, N. A. Barrington, and L. Ferrar. Comparison of methods for the visual identification of prevalent vertebral fracture in osteoporosis. *Osteoporosis International*, 15(11):887–896, 2004.

[66] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer Verlag, 2002.

[67] S. Joshi, S. M. Pizer, P. T. Fletcher, A. Thall, and G. Tracton. Multi-Scale 3-D Deformable Model Segmentation Based on Medical Description. In *Information Processing in Medical Imaging*, pages 64–77, 2001.

[68] J.A. Kanis, F. Borgstrom, C. De Laet, H. Johansson, O. Johnell, B. Jonsson, A. Oden, N. Zethraeus, B. Pfleger, and N. Khaltaev. Assessment of fracture risk. *Osteoporosis International*, 16:581–589, 2005.

[69] J.A. Kanis, LJ Melton, C.Christiansen, and C.C.Johnston. The diagnosis of osteoporosis. *Journal of bone and mineral research*, 9:1137–1141, 1994.

[70] K. Karhunen. Ueber lineare methoden in der wahrscheinlichkeitsrechnung. *Annals of Academic Science Fennicae*, 37:1–79, 1947.

[71] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[72] R.W. Keen. Burden of osteoporosis and fractures. *Current Osteoporosis reports*, 1(2):66–70, 2003.

[73] D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

[74] E. Klassen, A. Srivastava, W. Mio, and S. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383, 2004.

[75] P. A. Lachenbruch. *Discriminant Analysis*. Hafner press, 1975.

[76] C. De Laet, A. Oden, O. Johnell, B. Jonsson, and J.A. Kanis. The impact of the use of multiple risk factors on case finding strategies: a mathematical framework. *steoporosis International*, 16(1):313–318, 2005.

[77] R. Larsen and K. Hilger. Statistical shape analysis using non-euclidean metrics. *Medical Image Analysis*, 7(4):417–423, 2003.

[78] E.S. Leib, E.M. Lewiecki, N. Binkley, and R.C. Hamdy. International society for clinical densitometry. official positions of the international society for clinical densitometry. *Journal of Clinical Densitometry*, 7(1):1–6, 2004.

[79] M. Lillholm, A. Ghosh, M. de Bruijne, P. Pettersen, M. A. Karsdal, D. J. Leeming, C. Christiansen, and M. Nielsen. Vertebral fracture risk (vfr) score for fracture prediction in postmenopausal women. *Osteoporosis International*, 22:2119–2128, 2011.

[80] M. Loog and M. Bruijne. Discriminative shape alignment. In *Proc. of the 21st International Conference on Information Processing in Medical Imaging*, pages 459–466, 2009.

[81] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, 2nd edition, 1999.

[82] L. McCance and S. E. Huether. *Pathophysiology: The Biologic Basis for Disease in Adults and Children*. Mosby, 2009.

[83] E. V. McCloskey, T. D. Spector, K. S. Eyres, E. D. Fern, N. O'Rourke, S. Vasikaran, and J. A. Kanis. The assessment of vertebral deformity: a method for use in population studies and clinical trials. *Osteoporosis International*, 4(2):138–147, 1994.

[84] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis:a survey. *Medical Image Analysis*, 1(2):1–26, 1996.

[85] L. J. Melton. Epidemiology of vertebral fractures. *In: Christensen C, Johansen JS, Riis B (eds.) Osteoporosis, Copenhagen*, 1987.

[86] L. J. Melton, E.J.Atkinson, and C. Cooper. Vertebral fractures predict subsequent fractures. *Osteoporosis International*, 10:214–221, 1999.

[87] C. W. Miller. Survival and ambulation following hip fractures. *Journal of Bone Joint Surgery*, 60:930–934, 1978.

[88] J. B. Nase and J. B. Suzuki. Osteonecrosis of the jaw and oral bisphosphonate treatment. *Journal of American Dental Association*, 137(8):1115–1119, 2006.

[89] U. S. Department of Health-Human Services. *Bone Health and Osteoporosis: A Report of the Surgeon General*. 2004.

[90] World Health Organization. *Prevention and Management of Osteoporosis: Report of a WHO Scientific Group*. World Health Organization;, 2003.

[91] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[92] K. Overgaard, M. A. Hansen, S. B. Jensen, and C. Christiansen. Effect of calcatonin given intranasally on bone mass and fracture rates in established osteoporosis: a dose-response study. *British Medical Journal*, 305:556–561, 1992.

[93] B. Koller P. Ruegsegger and R. Muller. A tomographic system for the nondestructive evaluation of bone architecture. *Calcified Tissue International*, 58:24–29, 1996.

[94] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.

[95] P. C. Pettersen, M. de Bruijne, J. Chen, Q. He, C. Christiansen, and L. B. Tanko. A computer-based measure of irregularity in vertebral alignment is a bmd-independent predictor of fracture risk in postmenopausal women. *Osteoporosis International*, 18:1525–1530, 2007.

[96] S. Raganath. Contour extraction from cardiac mri studies using snakes. *IEEE Transactions on Medical Imaging*, 14:328–338, 1995.

[97] M. Roberts, T. F. Cootes, E. M. Pacheco, and J. E. Adams. Quantitative vertebral fracture detection on dxa images using shape and appearance models. *Academic Radiology*, 14:1166–1178, 2007.

[98] M. Roberts, E. Pacheco, R. Mohamkumar, T. Cootes, and J.E. Adams. Detection of vertebral fractures in dxa vfa images using statistical models of appearance and a semi-automatic segmentation. *Osteoporosis International*, 21(12):2037–2046, 2010.

[99] H. N. Rosen, A. C. Moses, J. Garber, I. D. Iloputaife D. S. Ross S. L. Lee, and S. L. Greenspan. Serum ctx: a new marker of bone resorption that shows treatment effect more often than other markers because of low coefficient of variability and large changes with bisphosphonate therapy. *Calcified Tissue Internatilnal*, 66:100–103, 2000.

[100] R. K. Ross, A. Paganini-Hill, P. C. Wan, and M. C. Pike. Effect of hormone replacement therapy on breast cancer risk: Estrogen versus estrogen plus progestin. *Journal of the National Cancer Institute*, 92(4):328–332, 1999.

[101] K. Rothman. *Epidemiology. An Introduction.* Oxford University Press, 2002.

[102] A. Ruyssen-Witrand, L.Gossec, S.Kolta, M.Dougados, and C.Roux. Vertebral dimensions as risk factor of vertebral fracture in osteoporotic patients: a systematic literature review. *Osteoporosis International*, 18:1271–1278, 2007.

[103] J. Schaefer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Journal of Statistical Applications in Genetics and Molecular Biology*, 4:online article, 2005.

[104] S. L. Silverman, C. Christiansen, H. K. Genant, S. Vukicevic, J. R. Zanchetta, T. J. de Villiers, G. D. Constantine, and A. A. Chines. Efficacy of bazedoxifene in reducing new vertebral fracture risk in postmenopausal women with osteoporosis: results from a 3-year, randomized, placebo-, and active-controlled clinical trial. *Journal of Bone and Mineral Research*, 23:1923–1934, 2008.

[105] C. G. Small. *The Statistical Theory of Shape*. Springer Series in Statistics, 1996.

[106] P. P. Smyth, C. J. Taylor, and J. E. Adams. Vertebral shape: Automatic measurement with active shape models. *Radiology*, 211(1):571–578, 1999.

[107] J. Sporring and K. Hommelhoff Jensen. Bayes reconstruction of missing teeth. *Journal of Mathematical Imaging and Vision*, 31:245–254, 2008.

[108] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate distribution. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, volume 1, pages 197–206, 1956.

[109] P. Switzer. Min/max autocorrelation factors for multivariate spatial imagery. *Computer Science and Statistics*, 1:13–16, 1985.

[110] P. Szulc and P.D. Delmas. Vertebral fracture initiative resource document. In *International Osteoporosis Foundation Report*, 2005.

[111] P. Taxel, P. Fall, K.M. Prestwood, L. Dulipsingh, D. Dauser, C. Ohannessian, and L.G. Raisz. Changes in urinary excretion of helical peptide during therapy for osteoporosis in older adults. *Journal of Clinical Chemistry*, 50:747–750, 2004.

[112] L. L. Thurstone. Multiple factor analysis. *Psychological Review*, 38(5):406–427, 1931.

[113] A. N. Tikhonov and A. N. Arsenin. *Solution of Ill-posed Problems*. Winston & Sons, Washington, 1977.

[114] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of Royal Statistical Society*, 21(3):611–622, 1999.

[115] T. Tomomitsu, K. Murase, T. Sone, and M. Fukunaga. Comparison of vertebral morphometry in the lumbar vertebrae by t1-weighted sagittal mri and radiograph. *European Journal of Radiology*, 56:102–106, 2005.

[116] D. J. Torgerson and S. E. M. Bell-Syer. Efficacy of bazedoxifene in reducing new vertebral fracture risk in postmenopausal women with osteoporosis: results from a 3-year, randomized, placebo-, and active-controlled clinical trial. *Journal of the American Medical Association*, 285:2891–2897, 2001.

[117] K. Uusi-Rasi, L. M. Semanick, J. R. Zanchetta, C. E. Bogado, E. F. Eriksen, M. Sato, and T. J. Beck. Effects of teriparatide [rhpth (1-34)] treatment on structural geometry of the proximal femur in elderly osteoporotic women. *Bone*, 36:948–958, 2005.

[118] A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum. A geometric snake model for segmentation of medical imagery. *IEEE Transactions on Medical Imaging*, 16(2):199–209, 1997.

[119] G. Zamora, H. Sari-Sarraf, and L. R. Long. Hierarchical segmentation of vertebrae from x-ray images. In *Proc. of SPIE Medical Imaging*, pages 5032–5044, 2003.

[120] R. M. D. Zebaze, G. Maalouf, and N. Maalouf E. Seeman. Loss of regularity in the curvature of the thoracolumbar spine: a measure of structural failure. *Journal of bone and mineral research*, 19:1099–1104, 2004.

[121] D. Zhang and G.Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.

List of Publications

Papers in international journals

- *"Maximum A Posteriori Bayes Estimation of Linear Shape Variation with Application to Vertebra and Cartilage Modeling."* Alessandro Crimi, Martin Lillholm, Anarta Ghosh, Marleen de Bruijne, Erik Dam, Mads Nielsen, Jon Sporring. IEEE Transaction on Medical Imaging, vol.30(8) pages 1514-1526, August 2011.
- *"Shape Based Assesment of Vertebral Fracture Risk using Discriminative Shape Alignment in Postmenopausal Women."* Alessandro Crimi, Marco Loog, Marleen de Bruijne, Mads Nielsen, Martin Lillholm. Academic Radiology, in press.
- *"A Bayesian Framework For The Automated Prediction Of Cardiovascular Disease From Lumbar Standard Radiographs."* Kersten Petersen, Melanie Ganz, Peter Mysling, Lene Lillemark, Alessandro Crimi, Mads Nielsen, Sami Brandt. IEEE Transaction on Medical Imaging, in press.

Papers in conference proceedings

- *"Bayes estimation of shape model with application to vertebrae boundaries."* Alessandro Crimi, Anarta Ghosh, Jon Sporring, and Mads Nielsen. In Proceedings of the SPIE: Medical Imaging 2009, Orlando, Florida, USA, 2009.
- *"Prior knowledge regularization in statistical medical image tasks."* Alessandro Crimi, Jon Sporring, Marleen de Bruijne, Martin Lillholm, Mads Nielsen. PMMIA workshop in conjunction with MICCAI 2009, Imperial college London, UK, 2009.

Published abstracts

- *"Morphometric Assessment of Vertebral Fracture Risk in Postmenopausal Women using Vertebral Shape Models and Discriminative Shape Alignment."* Alessandro Crimi, Marco Loog, Marleen de Bruijne, Mads Nielsen, Martin Lillholm.

Abstract annual meeting American Society for Bone and Mineral Research 2011 San Diego, California, USA 2011.

- *"Shape Based Analysis of Vertebra Fracture Risk in Post-menopausal Women."* Alessandro Crimi, Martin Lillholm, Paola Pettersen, Mads Nielsen. Abstract annual meeting American Society for Bone and Mineral Research 2010 Toronto, Canada 2010.

- *"Quantification of vertebra fractures in Post-menopausal women."* Alessandro Crimi, Mads Nielsen. Abstract annual meeting American Society for Bone and Mineral Research 2009 Denver, Colorado, USA 2009.

Acknowledgments

First and foremost, I would like to thank my supervisors Prof. Mads Nielsen and Dr. Martin Lillholm for their guidance, support and enthusiasm during my research. This work is mainly the result of the influence and opportunities they provided. I also thank Dr. Marleen de Brujine, Dr. Erik Dam, Dr. Anarta Ghosh, Dr. Marco Loog and Dr. Jon Sporring for their indispensable collaboration. Their useful advices helped me several times. Moreover, I have to mention the committee of my Ph.D. defence Dr. Francois Lauze, Dr. Martin Roberts, Dr. Wiro Niessen, their useful comments improved this work.

I would like to acknowledge the Center for Clinical and Basic Research, Ballerup, Denmark, for providing the annotated vertebrae radiographs, and Paola Pettersen, Jane Petersen and Annette Olsen for performing the annotations. I thank the Danish Research Foundation (den danske forskningfond) for providing necessary funding.

I cannot avoid being grateful to my colleagues at DIKU and Nordic-Bioscience, who created a friendly environment and not just a working environment. To be more precise, the "OP-gang" (Dan Jørgensen and Peter Mysling) who endured me for three years; the "neighbors" Dr. Melanie Ganz, Gopal Karemore, Lene Lillemark, Joselene Marques, Kersten Pettersen, Dr. Arish Qazi, Sudhakar Tummala, who shared tons of unhealthy cakes with me; my colleagues from DIKU: Chen Chen, Konstantin Chernoff, Dr. Sune Darkner, Dr. Aasa Feragen, Dr. Vladlena Gorbunova, Rabia Granlund, Dr. Christian Igel, Lars Lau, Dr. Pechin Lo, Adish Prasson, Dr. Lauge Sørensen, Stefan Sommer, Dr. Aditya Tatu and the always helpful and smiling Dina Johannessen and Camilla Jørgensen, and the others who I forgot.

Finally, I thank my family which supported me remotely during my studies, and Veronica who was always present even when I was not due to this work.

Thank to all of you, this learning "journey" did not produce just the results reported in this manuscript, but it was a real enriching experience of personal development. I learned that it is not possible to easily change the world, but instead it is enough to give a contribution that other people can reproduce or use as a starting point for their research. I also learned that it is very easy to draw erroneous conclusions, and the importance of being self-critic about own work as well as being humble. The most important lesson was that despite all the odds, through working hard, if there is a star, it will shine.