2018

Multi-Objective Design Optimization of a Mobile-Bearing Total Disc Arthroplasty Considering Spinal Kinematics, Facet Joint Loads, and Metal-on-Polyethylene Contact Mechanics

Chaochao Zhou
Binghamton University--SUNY, czhou12@binghamton.edu

Recommended Citation

Zhou, Chaochao, "Multi-Objective Design Optimization of a Mobile-Bearing Total Disc Arthroplasty Considering Spinal Kinematics, Facet Joint Loads, and Metal-on-Polyethylene Contact Mechanics" (2018). *Graduate Dissertations and Theses*. 99.
https://orb.binghamton.edu/dissertation_and_theses/99

MULTIOBJECTIVE DESIGN OPTIMIZATION OF A MOBILE-BEARING TOTAL DISC ARTHROPLASTY CONSIDERING SPINAL KINEMATICS, FACET JOINT LOADS, AND METAL-ON-POLYETHYLENE CONTACT MECHANICS

BY

CHAOCHAO ZHOU

B.S., Harbin Institute of Technology, 2009
M.S., Harbin Institute of Technology, 2011

DISSERTATION

Submitted in partial fulfillment for the requirements for
the degree of Doctor of Philosophy in Mechanical Engineering
in the Graduate School of
Binghamton University
State University of New York
2018

Accepted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Mechanical Engineering
in the Graduate School of
Binghamton University
State University of New York
2018

November 16, 2018

Seungbae Park, Chair
Department of Mechanical Engineering, Binghamton University

Ryan Willing, Faculty Advisor
Department of Mechanical Engineering, Binghamton University

Frank Cardullo, Member
Department of Mechanical Engineering, Binghamton University

Amber Doiron, Member
Department of Biomedical Engineering, Binghamton University

Susan Lu, Outside Examiner
Department of Systems Science and Industrial Engineering, Binghamton University

Abstract

Total disc arthroplasty (TDA) is a motion-preserving surgical technique used to treat spinal disorders, when more conservative medical therapies fail. Unfortunately, a high incidence of revision surgery exists due to postoperative complications including abnormal kinematics, facet joint arthritis, and implant failures. However, TDA is still an attractive option, since an optimally designed artificial disc is expected to reproduce native segmental biomechanics. Correspondingly, it would mitigate the development of adjacent segment diseases (a major concern of spinal fusion) caused by altered segmental biomechanics.

Design optimization is a process of finding the best design parameters for a component/system to satisfy one/multiple design requirements using optimization algorithms. The shape of a candidate design is parametrized using computer-aided design, such that design parameters are manipulated to minimize one/multiple objective functions subject to performance constraints and design space bounds. Optimization algorithms typically require the gradients of the objective/constraint functions with respect to each design variable. In the traditional design optimization, due to the high computational cost to calculate the gradients by performing finite element analysis in each optimization iteration, it often results in a slow process to seek the optimal solution. To address the problem, an artificial neural network (ANN) was implemented to derive the analytical expressions of the objective/constraint function and their gradients. By incorporating analytical gradients, we successfully developed a multiobjective optimization (MOO) framework considering three performance metrics simultaneously.

Furthermore, a new mobile-bearing TDA design concept featuring a biconcave polyethylene (PE) core was proposed, to strengthen the PE rim, where a high risk of fracture exists. It was hypothesized that there is a trade-off relationship among postoperative performance metrics in terms of spinal kinematics, facet joint loading, and metal-on-polyethylene contact mechanics. We tested this hypothesis by refining the new TDA to match normal segmental biomechanics and alleviate PE core stress. After performing MOO, the best-trade-off TDA design was determined by the solved three-dimensional Pareto frontier. The novel MOO framework can be also used to improve existing TDA designs, as well as to push the cutting edge of surgical techniques for the treatment of spinal disorders.

Acknowledgements

First, I would like to acknowledge the guidance and support of my advisor, Dr. Ryan Willing during the six-year Ph.D. studies. You opened a door for me to the wonderful world of biomechanics, and offered me enough freedom to seek the research direction that I am truly interested in. Although it was painful, your strict training and requirements in research, writing, and presentation shaped me to be an independent scientific investigator, and will benefit to me for my future academic research.

I also like to express my gratitude to the faculty in Department of Mechanical Engineering and Department of Biomedical Engineering. Your education equipped me with knowledge and expertise in mechanics, mathematics, and biology. Since my Ph.D. research heavily relied on computer simulation, I am thankful for the stuff in Watson School Information Technology Department for providing access to high-performance computational resources. Although each member in Dr. Willing's lab worked on different research topics, I appreciate our discussions that greatly inspired me. Besides, I miss my friends who lived and studied with me at Binghamton, and I cannot forget your help and encouragement when I confronted a difficulty.

One of the biggest acknowledgements should be given to my family. My parents always required me to concentrate on my Ph.D. studies, and even rarely told me anything that distracted and concerned me. Without your unwavering support and constant encouragement, my Ph.D. work cannot be completed.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AF	Annulus fibrosus
ALL	Anterior longitudinal ligament
ANN	Artificial neural network
BOO	Bi-objective optimization
CL	Capsular ligament
CoCr	Cobalt-chrome
CS	Coordinate system
CTE	Coefficient of thermal expansion
EP	Endplate
FE	Finite element
FEA	Finite element analysis
FJ	Facet joint
FJF	Facet joint force
FSU	Functional spine unit
GS	Ground substance
ICR	Instantaneous center of rotation
IDP	Intradiscal pressure
ISL	Interspinous ligament
ITL	Intertransverse ligament
IV	Intervertebral
IVD	Intervertebral disc
L1-L5	Lumbar vertebra 1-5
LBP	Lower back pain
LF	Ligamentum flavum
MOO	Multiobjective optimization
MoP	Metal-on-polyethylene
MPC	Multipoint constraint

MRI	Magnetic resonance imaging
NBI	Normal-boundary interaction
NC	Normal constraint
NN	Neural network
NP	Nucleus pulposus
PCP	Polyethylene contact pressure
PE	Polyethylene
PLL	Posterior longitudinal ligament
PS	Principal stress
ROM	Range of motion
S1	Sacrum
SOO	Single-objective optimization
SSL	Supraspinous ligament
T1-T12	Thoracic vertebra 1-12
TDA	Total disc arthroplasty
TDR	Total disc replacement
TOO	Tri-objective optimization
UHMWPE	Ultra-high-molecular-weight polyethylene
VA	Vertebral arch
VB	Vertebral body

Chapter 1: Introduction

1.1. Motivation

Lower back pain (LBP) is a public health concern in western countries. In the United States, 80% of adults suffer from LBP in their lifetimes [1]. It is the most common cause of job-related disability and a leading contributor to lost work days. In an analysis of 30,074 case reports to the National Health Interview Survey in 1998, it was estimated that 4.6% LBP patients lost their jobs, and LBP resulted in 149 million lost work days annually [2]. Due to the high prevalence rate, not only the suffering from LBP does cause distress to their families, but also a substantial economic burden on the whole society [3].

A significant portion of LBP is associated with intervertebral disc (IVD) degeneration. The IVD is a soft tissue, which consists of a gel-like center, the nucleus pulposus (NP), surrounded by an outer fibrous ring, the annulus fibrosus (AF). Disc degeneration is commonly initiated by natural daily stresses and minor injuries in the annular lamellae [4, 5]. As discs damage and lose water, they start to collapse. The decrease in the IVD height can result in pressure on the nerves in the human spinal column and abnormal loads on other spinal tissues. Therefore, disc degeneration further induces other pathologic conditions (*e.g.*, spinal stenosis and joint arthritis) in the lumbar spine, causing pain and weakness.

When more conservative medical therapies fail, surgical procedures such as spinal fusion and total disc arthroplasty (TDA) are required to relieve pain. In spinal fusion, the diseased disc is replaced by implanting a high-stiffness metal or polymeric cage loaded with bone graft, which promotes the fusion of adjacent vertebral bodies. Although the intervertebral height and lordosis (inclination angle) are restored, the insertion of a high-stiffness implant compromises many aspects of the normal disc biomechanics. In contrast, TDA using an artificial disc featuring metal-on-polyethylene (MoP) articulations or a viscoelastic core can relieve pain and pressure, with fewer restrictions on postoperative activities.

In the literature, there has not been a consensus for a recommended surgical treatment that would decrease the incidence of complications. Inconsistent results regarding the biomechanical effects of different therapies such as spinal fusion and TDA on the instrumented spine have been reported previously. However, clinical studies have indicated that TDA causes complications (e.g., the development of adjacent segment diseases) which also occur in patients who receive spinal fusion, with additional concerns of spinal tissue overloading and abnormal motion patterns at the treated segment, as well as implant failures due to wear, fracture and creep. The more complex complications impede the application of TDA, despite the benefit of motion preserving. Therefore, spinal fusion is still the gold standard treatment for spinal disorders, due to the higher risk of revision surgery after TDA.

Theoretically, spinal fusion that creates a high-stiffness construct in the intervertebral space can completely restrict the motion of the treated spinal segment, unless pseudo-arthrodesis (the failure to achieve the complete bony union) occurs. Therefore, the segmental behavior after spinal fusion is less sensitive to implant design. On the contrary, TDA design may significantly influence the postoperative segmental kinematics, MoP contact mechanics, and spinal tissue load-sharing, according to the knowledge from the intense studies of total knee/hip arthroplasty in the literature. Previously reported biomechanical comparisons between spinal fusion and TDA were often conducted by considering a unique commercial available TDA design. To the best of our knowledge, design optimization of a TDA has not been performed in previously published studies. Due to the lack of an in-depth understanding of the relationship between TDA design and postoperative spinal responses, whether the motion-preserving surgical treatment using an optimally designed implant can translate into a better spinal function is unclear.

1.2. Objectives

The overarching goal of this work is to find the optimum TDA designs using multiobjective design optimization simultaneously considering postoperative segmental kinematics, MoP contact mechanics, and spinal tissue load-sharing. To perform TDA multiobjective optimization, several sub-objectives need to be met. They include:

1. Development of a numeric framework to model the material anisotropy and heterogeneity across the annulus pulposus, and simulate the responses of the native intervertebral disc in various physiologically relevant loading scenarios.

2. Development of a multi-segment lumbar spine finite element (FE) model in order to investigate the behaviors at different levels of the intact and TDA-treated lumbar spines.

3. Performance assessment of segmental kinematics, MoP contact mechanics, and spinal tissue load-sharing in lumbar segments treated by different commercially available TDAs, and their sensitivities to implant design and facet joint articulation.

4. Proposal of a new biconcave mobile-bearing TDA design concept which improves MoP contact mechanics, and parametric FE modeling of the new TDA design considering the important design parameters which influence the TDA performance.

5. Development of an artificial neural network based multiobjective optimization algorithm to generate the complete Pareto frontier in a high-dimensional objective space.

6. Multiobjective design optimization of the biconcave mobile-bearing TDA simultaneously considering segmental kinematics, MoP contact mechanics, and facet joint loading in the treated lumbar segment, to achieve the best tradeoff among these performance measures.

1.3. Hypotheses

1.3.1. Hypothesis in the Study of Intervertebral Disc Biomechanics

Typically, soft tissue observed in clinical images is prestressed by internal pressure and tissue hydration; the normal disc is pressurized with a physiological intradiscal pressure (IDP) due to nucleus swelling. Simulations based on the geometry and fiber orientation from clinical images without considering these residual stresses will cause a discrepancy with the physical reality. It was reported that mild disc degeneration due to the loss of the capability of nucleus swelling caused a substantial decrease in the resistance to intervertebral rotations. Therefore, it is necessary to incorporate nucleus swelling into an IVD computational model to better understand IVD biomechanics. We hypothesize that thermoelastic analysis can be repurposed to provide an accurate representation of the disc prestrain under nucleus swelling. The geometry and annulus fiber orientation of the prestrained disc under a physiological IDP are identical to those observed in clinical images. Moreover, IDP and fiber orientation regulate the disc mechanical behavior in different loading scenarios.

1.3.2. Hypothesis in a New Mobile-bearing Total Disc Arthroplasty Design Concept

Impingement in TDAs with MoP articulations is defined as unintended contact between two nonbearing surfaces. According to well-documented retrieval studies [6], it is highly associated with excessive wear and fracture of the polyethylene [7]. However, we hypothesize that impingement as a contact constraint can prevent excessive intervertebral rotations and spinal tissue overloading. Therefore, the author proposed a new TDA concept featuring a biconcave PE mobile core, with a smooth transition from the dome to the rim. It is further hypothesized that the biconcave-core TDA design results in a significant increase in the strength of the PE rim, which can serve an intended bearing.

1.3.3. Hypothesis in Multiobjective Design Optimization of Total Disc Arthroplasty

TDA treatment using the anterior surgical approach may cause postsurgical complications such as excessive segmental rotation [8], spinal tissue overloading [9] and polyethylene inlay wear/fracture failures [10]. We hypothesize that these complications can be mitigated and native disc biomechanics can be matched by optimally designing a TDA. In this study, spinal fusion is considered as a TDA design extremity, and a motion-sacrificing configuration can be realized when the complete congruence on the MoP articulations is designated. To quantify TDA complications, three performance metrics regarding the segmental range of motion (ROM), facet joint force (FJF) and polyethylene contact pressure (PCP) were introduced. It is further hypothesized that there is a tradeoff among these three performance metrics. Multiobjective design optimization will provide a Pareto curve which quantifies the tradeoff relationship. Furthermore, it will allow us to achieve the best tradeoff design by changing design variables of a TDA.

1.4. Contributions

1.4.1. Modeling of Material Anisotropy and Heterogeneity in Intervertebral Disc

In the study of intervertebral disc biomechanics, a new constitutive model was developed to describe the nonlinear tensile behavior of annular reinforcing collagen fibers. Based on the constitutive model, algorithms used to fit and interpolate the stress-stretch curves of collagen fibers were developed, to model the material anisotropy and heterogeneity across the annulus fibrosus.

1.4.2. Investigation of Changes in Disc Geometry and Mechanics by Nucleus Swelling

Thermoelastic analysis was introduced in an intervertebral disc FE model to simulate the disc prestrain in nucleus swelling, and inversely derive the configuration of the unloaded (degenerated) disc due to the lack of nucleus swelling using an iterative "pull-back" algorithm [11]. By comparing the resulting intradiscal pressure (IDP) in the nucleus and principal stresses in the annulus to those using a reported disc prestrain simulation method, the author demonstrates that the disc prestrain is mechanically equivalent to thermal expansion. However, different from the previous disc prestrain method, the thermal expansion based disc prestrain method enables the "pull-back" of the entire disc including the annulus and nucleus, such that the effect of disc degeneration (the loss of the physiological IDP) on the disc behavior can be investigated.

1.4.3. Comparison of Normal and Degenerated Disc Behaviors

The behaviors of the normal (prestrained) disc and degenerated (unloaded) disc were computationally investigated by applying the reported physiological ranges of motions (ROMs) in different loading scenarios. The author revealed that the disc stiffness is regulated by the IDP and fiber orientation; a decrease of the IDP and fiber angles causes a distinct deterioration in the resistance to intervertebral rotations. Therefore, calibration of the IDP and fiber orientation can improve the model prediction of disc behavior.

1.4.4. Proposal of a New Biconcave Total Disc Arthroplasty Concept

According to well-documented TDA retrieval studies, the author proposed a new mobile-bearing TDA design concept featuring a biconcave polyethylene core, which strengthens the polyethylene rim, where a high risk of fracture exists. Using kinematic analysis and contact simulation, the author has successfully demonstrated that the new TDA design is capable of alleviating complications which commonly occurs in the treatments using commercially available TDAs.

1.4.5. Multiobjective Optimization of Total Disc Arthroplasty

The proposed mobile-bearing TDA concept was refined using a neural network (NN) based multiobjective design optimization algorithm. Multiobjective design optimization considering two objectives has been widely used to optimally design implants [12, 13]. Here, the author derived the analytical gradient of the NN predictions with respect to the NN inputs, and incorporated it into an evolutionary multiobjective optimization algorithm. Thus, the computational efficiency was significantly boosted to allow multiobjective optimization considering more objectives in this study. Specifically, the best tradeoff of the TDA performance was achieved considering three objectives (segmental kinematics, contact mechanics, and spinal tissue load-sharing) simultaneously.

Chapter 2: Background

2.1. Lumbar Spine

As shown in Figure 2.1, the human spinal column consists of a series of vertebrae (segmented bones), which are separated by intervertebral discs. The cervical, thoracic and lumbar spines are three main spine regions from the cranial to the caudal defined according to different curvatures observed in the spine column. The lumbar spine region (Figure 2.1a) comprises five lumbar vertebrae (L1-L5), and sometimes also includes the caudal-most thoracic vertebrae (T12) and the sacrum (S1).

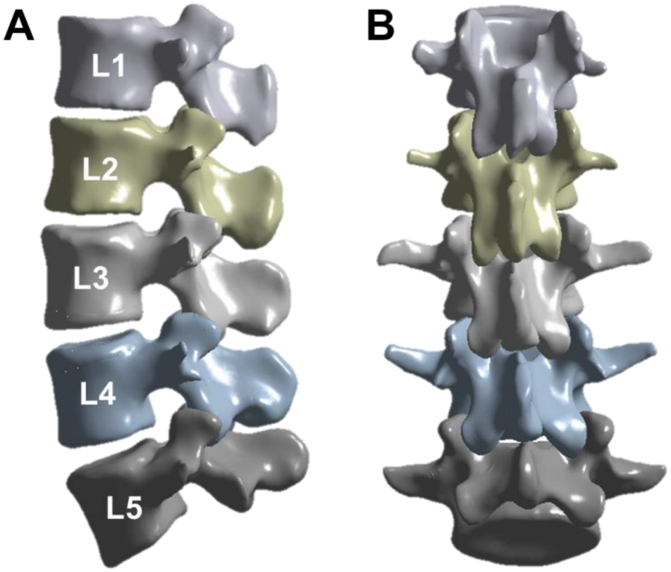


Figure 2.1: 3D model of the human lumbar spine created from CT images. (a) The sagittal view of the L1-L5 lumbar spine; (c) The posterior view of the lumbar spine.

Intervertebral rotations and translations in six degrees of freedom in three anatomic planes (Figure 2.2) are allowed in each motion segment or functional spine unit (FSU), which consists of two adjacent vertebrae, the intervertebral discs, and all intersegment ligaments. From L1 to L5 (Figure 2.1b), differences in anatomic dimensions of vertebrae [14] and orientations of the facet joints [15] can be observed, indicating that the load-sharing of spinal tissues varies from cranial to caudal in spinal motion [16].

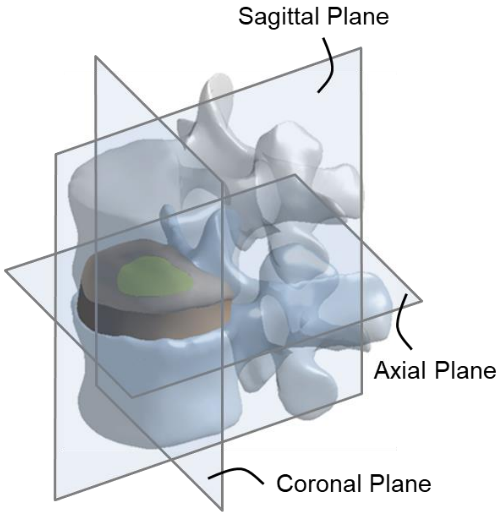


Figure 2.2: The anatomical planes of a lumbar functional spine unit, including the coronal (frontal) plane, sagittal plane and axial (transverse) plane.

2.2. Intervertebral Disc

The intervertebral disc (IVD) is a sophisticated soft tissue structure (Figure 2.3a) that consists of a gel-like center, the nucleus pulposus (NP), and an outer fibrous ring, the annulus fibrosus (AF). The NP contains type II collagen and polysaccharide, capable of trapping ions to generate an osmotic pressure for nucleus swelling [17]; therefore, the high water content is commonly observed in the NP of the healthy disc [18]. The AF ground substance (GS) surrounding the NP is reinforced by collagen fiber lamellae [19–22], with a crossing pattern of fiber bundles due to the alternating fiber orientations in adjacent lamellae (Figure 2.3b). The fiber angles with respect to the disc transverse plane vary radially and circumferentially across the AF [19, 20]. Meanwhile, the relative content of type I and II collagens gradually changes across the AF, as a consequence of the adaptation in tissue remodeling to loading history and physiological aging (Brickley-Parsons and Glimcher. 1984). Therefore, the different collagen content distribution and fiber orientation within the AF result in the material heterogeneity of the AF composite.

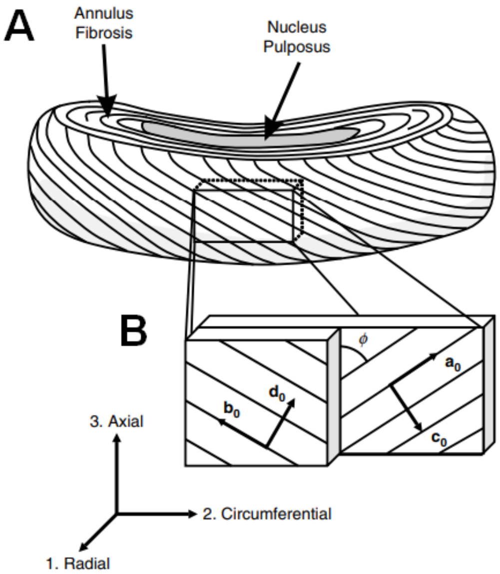


Figure 2.3: Anatomy of the human lumbar intervertebral disc [24], used with permission from Elsevier. (a) The NP and AF of a disc; (b) The AF lamellae reinforced by collagen fibers with a crossing pattern.

The disc mechanical functions in a spine motion segment follow the disc structure. The main function of the disc is the load-bearing of a dynamic compressive load, associated with the change in body positions in life activity [25]. The incompressible disc matrix resists the compressive load, and interstitial fluid plays a vital role in absorbing the impact [26, 27]. The fiber lamellae embedded in the AF constrain disc bulging, and further pressurize the disc matrix to avoid disc collapse. Besides, the disc serves to connect adjacent vertebrae to each other, and guide and restrict the intervertebral rotations, such as flexion, extension, lateral bending and axial torsion. *In-vitro* experiments have demonstrated that there exists a neutral zone in the hysteresis curve of the kinematic response, which describes the laxity of the motion segment [28], as shown in Figure 2.4. Beyond the neutral zone, pronounced tissue stiffening occurs in each loading scenario, so it determines the physiological ranges of motion (ROMs) and avoids excessive intervertebral rotations. Both the microscopic structural change and re-orientation of collagen fibers in large deformation contribute to the stiffening effect in the AF tissue. Polarized light microscopy demonstrates that initially crimped fiber bundles with a periodic wave pattern are gradually stretched and unfolded in uniaxial tensile testing [29]. In the meantime, the fibers with the crossing pattern are re-oriented along the loading direction [30]. Thus the increasing fiber forces stiffen the disc tissue.

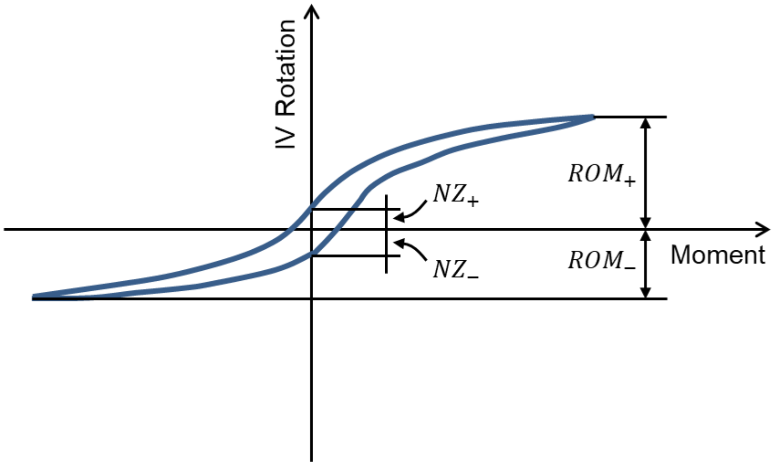


Figure 2.4: Graphic representation of the hysteresis curves of a lumbar segment applied a moment (NZ = neutral zone; ROM = range of motion; IV = intervertebral) [28], modified with permission from Elsevier.

The IVD behavior is strongly sensitive to the collagen fiber orientation [31–34], which varies radially and circumferentially across the IVD AF [19, 20]. The regional material variation may result from the tissue evolution in the adaptation to the loading history and physiological aging [23, 35]. To develop a computational biomechanical model with high fidelity, the experimental data available in the literature should be incorporated [36]. Holzapfel et al (2005) and Zhu et al. (2008) measured the stress-stretch curves of the fiber lamellae at different anatomic sites in the lumbar IVD specimens. A summary of reported material properties and *in-vitro* disc behavior using disc specimens at the upper lumbar levels [19, 37] and lower lumbar levels [20, 28] is listed in Table 2.1. These experimental data have been used to model the material heterogeneity of the AF in previous FE models [33, 34, 38–40].

Table 2.1: Summary of the IVD material properties and IVD mechanical behavior data in in-vitro experiments.

Single lamellar testing	References	Holzapfel et al (2005)	Zhu et al (2008)
	Testing specimens	L1-L2 discs: 11 Age: 57.9 ± 15.4 years Gender ratio: m:f = 8:3	L4-L5/L5-S1 discs: 5 Age: 28–45 years Gender ratio: m:f = 5:0
	Uniaxial tensile testing	Stress-stretch curves at 4 testing sites	Linear-fit tangent moduli at 8 polar sectors
	Fiber orientations	Linear-fit relation in the polar direction	Variation in both radial and polar direction
	Lamellar thicknesses	Available at 5 testing sites	Not available
In-vitro IVD behavior	References	Li et al (2009)	Heuer et al (2007a)
	Testing specimens	L2-L3 discs: 2 Age: 23–44 years	L4-L5 discs: 8 Age: 52 (38–59) years
	Loading condition in each loading scenario	The range of displacement /rotation were applied, the rest DOFs were allowed	Pure moment was applied

2.3. Ligaments and Facet Joints

As presented in Figure 2.5, there are seven ligaments in the lumbar spine, including anterior longitudinal ligament (ALL), posterior longitudinal ligament (PLL), ligamentum flavum (LF), intertransverse ligament (ITL), capsular ligament (CL), interspinous ligament (ISL) and supraspinous ligament (SSL). Similar to the annular lamellae, the behavior of spinal ligaments can be described as a tension-only, nonlinear stress-stretch or force-deflection curve, with an effect of tissue stiffening. Exceptionally, the SSL is able to resist compression, since Heuer et al. (2007a) reported that the ROMs of the L4-5 specimens significantly increased (Wilcoxon rank test) in extension after removing the SSL.

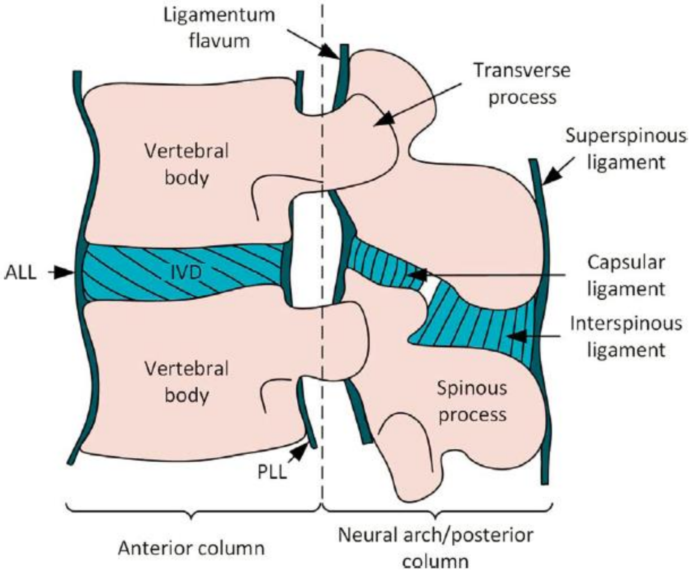


Figure 2.5: Illustration of seven lumbar spinal ligaments. Adopted from [41], under the terms of the Creative Commons user license (https://creativecommons.org/licenses/by/4.0/).

The facet joint surrounded by the CL (Figure 2.5) is a synovial plane joint between the articular processes of two adjacent vertebrae. It is made up of two major functional components, the articular cartilage and the capsular ligament. According to *in-vitro* experiments [28, 42] and simulations [16, 34], the dissection of the facet joints caused distinct changes in kinematic responses in both flexion and extension. Thus facet joints play an important role in the spinal function and stability only secondary to IVDs. Experiments on the isolated facet joints [43] have demonstrated that compression and tension are resisted by the articular cartilage and the capsular ligament, respectively.

2.4. Degenerative Disc Disease

Most spinal pathologies result from disc degeneration, in which the disc loses its capability for tissue hydration (hence the degenerated disc is like a flat tire). The underlying causes of disc degeneration include genetic inheritance and inadequate metabolite transport [4, 44–46]. Excessive loading initiates tissue damages in the IVD, which changes the cell environment; an aberrant, cell-mediated response to the environmental changes further accelerates tissue disruption. Due to extremely low cellularity, healing potential in adult IVDs is limited, so the progressive structural failure is almost irreversible. Adams et al. (2000) mimicked the disc degeneration process using *in-vitro* experiment. They observed that minor damages in the vertebral body endplates led to the progressive structural deterioration in the IVDs in dynamic loading. Similar to the *in-vivo* disc degeneration, the resulting tissue disruptions in the disc specimens include inwardly collapsing annulus, extreme outward bulging, complete radial fissures, and nucleus migration. Generally speaking, disc degeneration resembles failure in metals due to fatigue, which involves the crack initiation and propagation stages.

The extent of tissue deterioration in the degenerated discs is often evaluated using the Pfirrmann magnetic resonance imaging (MRI) classification system or Pfirrmann grades for short [47]. According to MRI images, the IVD can be graded as the healthy disc (Grade I and II), and the mildly (Grade III), moderately (Grade IV) and severely (Grade V) degenerated discs. The MRI features of the healthy and degenerated discs due to different MRI signal intensities are listed in Figure 2.6 and Table 2.2.

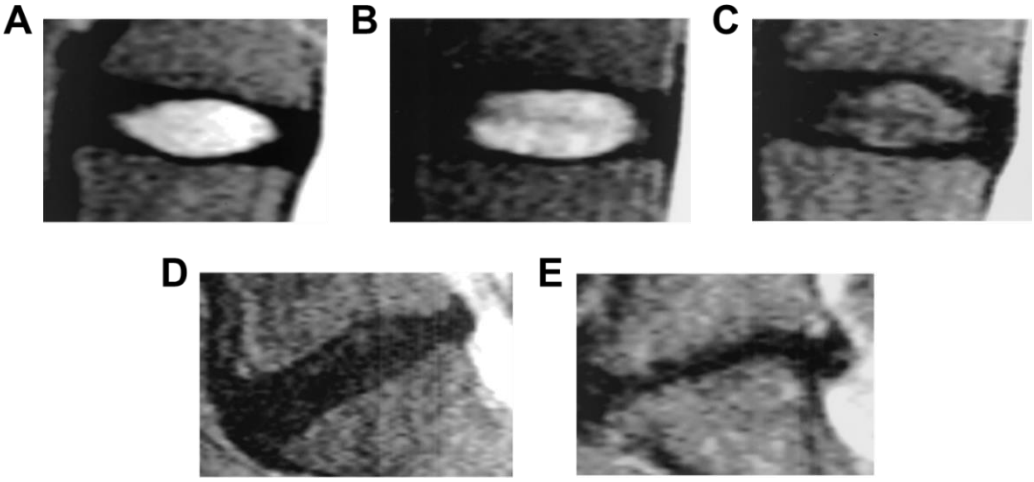


Figure 2.6: MRI images of degenerated discs with different grades [47], reproduced with permission from Wolters Kluwer Health, Inc. (a) Grade I; (b) Grade II; (c) Grade III; (d) Grade IV; (e) Grade V.

Table 2.2: The MRI classification of disc degeneration [47].

Grade	Structure	Distinction of NP and AF	Signal Intensity	Height of IVD
I	Homogeneous, bright white	Clear	Hyperintense, isointense to cerebrospinal fluid	Normal
II	Inhomogeneous with or without horizontal bands	Clear	Hyperintense, isointense to cerebrospinal fluid	Normal
III	Inhomogeneous, gray	Unclear	Intermediate	Normal to slightly decreased
IV	Inhomogeneous, gray to black	Lost	Intermediate to hypointense	Normal to moderately decreased
V	Inhomogeneous, black	Lost	Hypointense	Collapsed disc space

In general, the nucleus region of a healthier disc appears lighter in color (hyperintense) than the surrounding tissues, with a uniform signal intensity equivalent to that of the cerebrospinal fluid. However, the nucleus region of a more degenerated disc would be darker (hypointense). In terms of the disc size, the heights of mildly, moderately and severely degenerated discs are commonly 80%, 60% and 40% of the healthy disc height, respectively [48].

2.5. Lower Back Pain

The thoracolumbosacral region undergoes the most severe loading conditions and the largest range of motion. Hence, it is most susceptible to wear and tear, and spinal disorders commonly occur in the region. Although the etiology of lower back pain is not well known, spinal insatiability (referred to abnormally large intervertebral rotations) is considered as one of the important causes [49]. Through a comprehensive literature review, it is found that there are three major sources of lower back pain, including discogenic pain, pain due to the impingement of spinal cord and nerve roots, and pain triggered by facet joint arthritis. Three common pathways causing pain in the lumbar spine are summarized in Figure 2.7. It can be noted that all of them are associated with disc degeneration.

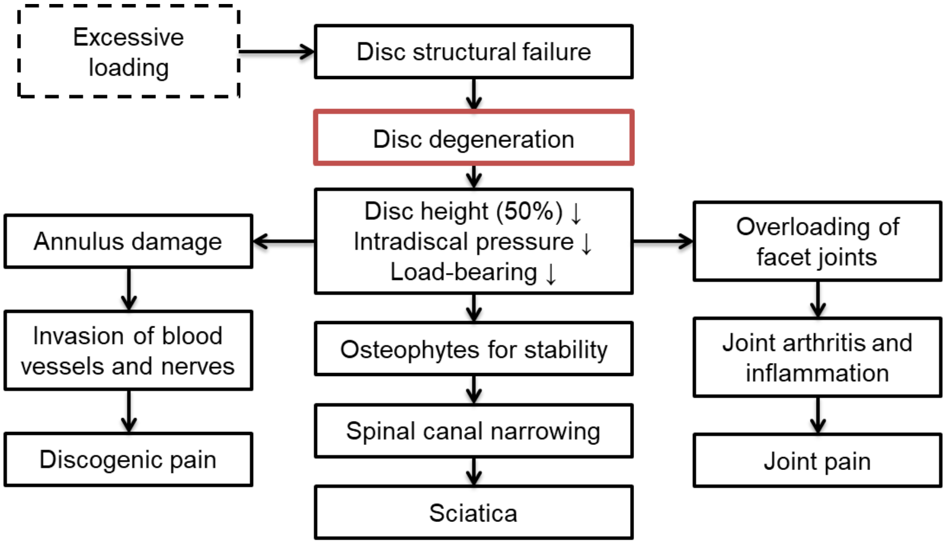


Figure 2.7: Three common pathways causing pain.

2.5.1. Discogenic pain

According to the sites where damages initiate, disc degeneration can be classified as either annulus-driven or endplate-driven [4, 44]. Annulus-driven disc degeneration (Figure 2.8a) commonly occurs in L4-S1 due to annulus damage. The upper lumbar and thoracic spines often suffer from endplate-driven disc degeneration (Figure 2.8b) due to endplate fracture [45]. Especially, fissures in the annular lamellae (Figure 2.8a) are mechanically and chemically conductive to revascularization and reinnervation, so cyclic loads and excessive stress in degenerated discs [5] may irritate invaded nerves, causing the discogenic pain [4].

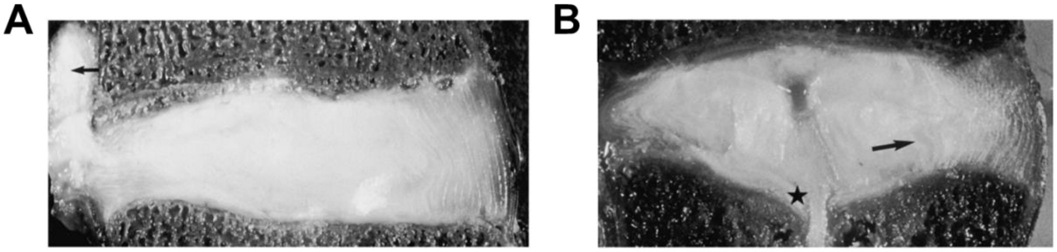


Figure 2.8: Examples regarding two phenotypes of disc degeneration [44], reproduced with permission from Wolters Kluwer Health, Inc. (a) Annulus-driven degeneration where a complete radial fissure was noted in the posterior anulus (arrow); (b) Endplate-driven degeneration that causes the inward bulging of the inner anulus lamellae (arrow) and damage to the vertebral body endplate (star).

2.5.1. Spinal Stenosis

Disc degeneration may cause a significant decrease in the disc height, disc herniation (extrusion of the gel-like nucleus material), and growth of bone spurs around the vertebral body margin to restore spinal stability [45]. These pathological conditions narrow the spinal canal and foramen, and thus compress the spinal cord and nerve roots, as illustrated in Figure 2.9. Furthermore, it can produce symptoms during gait, such as sciatica (tingling, weakness or numbness that radiates from the low back and into the buttocks and legs) [50]. People with spinal stenosis may describe the onset of leg pain or weakness with walking, but with relief of symptoms with sitting. Many will also describe increased tolerance to walking when flexed forward.

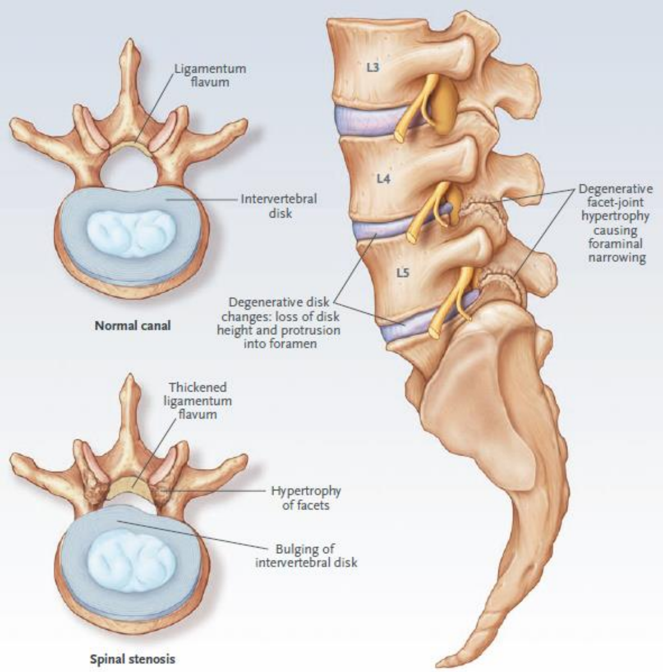


Figure 2.9: Spinal stenosis in the lumbar spine. Reproduced with permission from [51], Copyright Massachusetts Medical Society.

2.5.1. Facet Joint Arthritis

At last but not least, facet joint arthritis may be a consequence of disc degeneration [52], as presented in Figure 2.9. Loss of the intradiscal pressure (IDP) in the mildly degenerated disc significantly decreases the load-bearing capability and increases ROMs [48]. In the severely degenerated disc, the disc height substantially decreases, and the intervertebral space is obliterated, as the spine fuses [45, 53]. In both cases, the facet joints may carry an abnormal load on the spine in flexion and extension. Due to the excessive compression of articular cartilage and stretch of capsular ligament, facet joints may become inflamed, irritated or swollen, as well as nerves innervating the facet joint are impinged, causing pain and other uncomfortable symptoms [43, 46].

2.6. Surgical Procedures and Complications in the Treatment of Spinal Disorders

2.6.1. Spinal Fusion

Over the past 50 years, spinal fusion procedure (Figure 2.10) using implanted cages loaded with bone graft in combination with stabilization devices (*e.g.*, pedicle screw system) to fuse two or more vertebrae has become the gold standard for treatment of numerous pathologic conditions in the human spine [54]. Implant failures may occur at the treated level include pseudo-arthrosis (the failure to achieve the complete bony union), broken screws, device loosening, implant migration and subsidence [55].

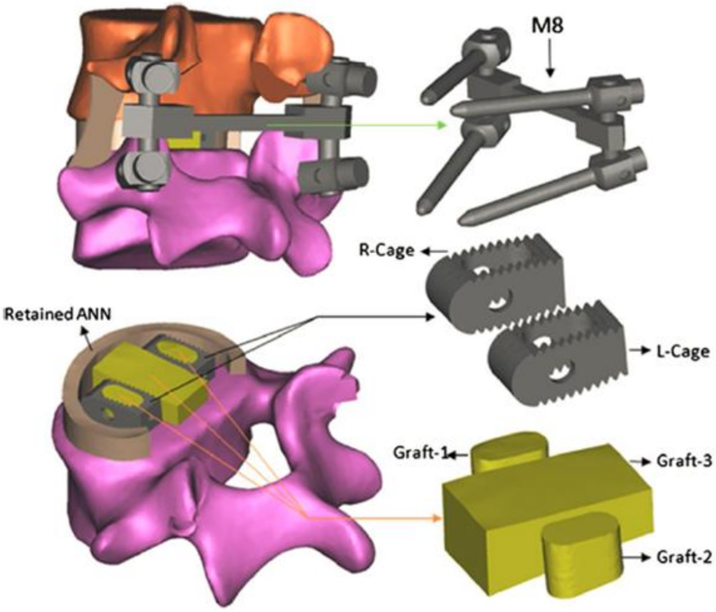


Figure 2.10: Schematic of posterior lumbar interbody fusion with pedicle screw stabilization devices. Adopted from [56], under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0).

Long-term follow-up studies of disc replacements have suggested that adjacent segment degeneration (disc degeneration at levels adjacent to a disc replacement) and adjacent segment diseases (development of new symptoms correlating with adjacent segment degeneration) are common problems [54]. It was reported that adjacent segment diseases were associated spinal fusion, which sacrifices spinal motion and causes different biomechanics from the native IVD [57–59]. For example, a cage plus a rigid stabilized instrument in spinal fusion (Figure 2.10) has a much higher stiffness than that of a native disc. As a result, stress concentration commonly occurs in the disc and facet joints at the level adjacent to the spinal fusion level, indicating a higher risk of disc degeneration and facet joint arthritis at the adjacent levels [60, 61].

To treat adjacent segment diseases, dynamic stabilization devices (*e.g.*, the Dynesys device which consists of polyethylene terephthalate cords and polycarbonate urethane spacers) have been developed to replace the rigid rods in the traditional pedicle screw system [62]. The viscoelastic material property of the polyurethane spacers in the dynamic stabilization device can allow micromotion and reduce stress in IVDs and facet joints at the adjacent levels. However, native disc biomechanics still cannot be completely restored, due to the high stiffness of the interbody fusion cage [63–65].

2.6.2. Total Disc Arthroplasty

Total disc arthroplasty (TDA) is an attractive option to treat lower back pain and restore the physiological ROM of the lumbar segment. The surgical treatment for the lumbar spine is conducted according to an anterior surgical procedure (Figure 2.11), in which the anterior longitudinal ligament, the anterior portion of the annulus and the entire nucleus are removed. Only the posterior and lateral portion of the annulus remains in place [9, 66].

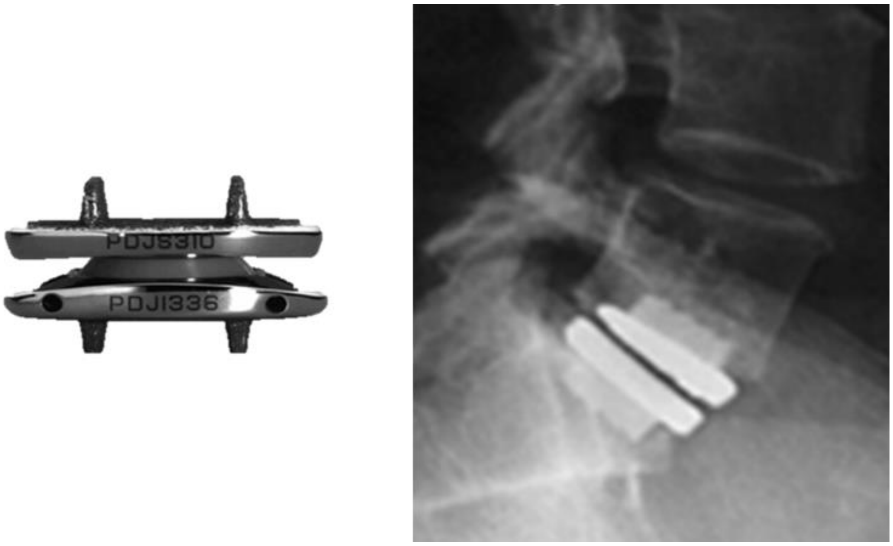


Figure 2.11: Lumbar total disc arthroplasty (Prodisc I implant) and a ten-year follow-up lateral radiograph after L5-S1 disc replacement [67]. Modified with permission from Wolters Kluwer Health, Inc.

Commercially available TDA designs can be classified into the un-/semi-constrained (ball-in-socket designs featuring metal-on-polyethylene articulations) TDA designs and constrained TDA designs (a viscoelastic core which restrained between two metallic endplates). The unconstrained artificial disc (Figure 2.12a) adopts an ultra-high-molecular-weight polyethylene mobile core which articulates both the superior and inferior metal endplates, and the semi-constrained artificial disc (Figure 2.12b,c) features a fixed-bearing ball-in-socket articulation. In both unconstrained and semi-constrained designs, commonly used material couplings are metal/polyethylene (Figure 2.12a,b) and metal/metal (Figure 2.12c) [9, 68].

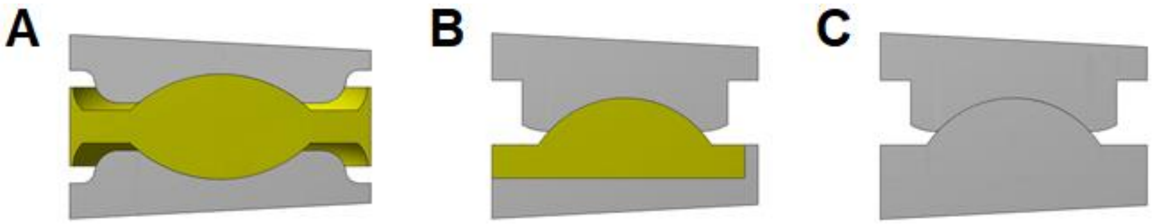


Figure 2.12: The cross-section views of unconstrained (mobile-bearing) and semi-constrained (fixed-bearing) lumbar TDAs. (a) SB Charité III; (b) ProDisc II; (c) Maverick.

The constrained TDA design (*e.g.*, AxioMed) commonly consists of titanium alloy retaining plates, with attached end caps, bonded to a viscoelastic, silicone polycarbonate urethane core [62]. Moreover, a novel TDA (M6-L) incorporates the artificial annulus fibers, so it creates a more physiological IVD structure compared to other TDA designs and can match native disc kinematics [9]. The constrained TDAs can avoid early loosening and long-term wear as using a non-elastic polyethylene core in unconstrained and semi-constrained TDAs, and effectively absorb shock [69, 70]. However, the long-term creep deformation of the viscoelastic core and the low bonding strength between the core and the metallic plates are limiting factors in their applications [71].

TDA provides a motion-preserving technique to treat disc degeneration diseases and thus theoretically mitigate adjacent segment degeneration. However, the treatment using the anterior surgical procedure substantially deteriorates the anterior spinal and disc tissues, compared to the intact segment that can provide significant resistance to intervertebral rotations [72]. Therefore, it may result in excessive segmental rotation [8] and facet joint overloading due to TDA hypermotion at both the treated level [55, 61, 66, 73, 74] and at the adjacent level [9, 75]. Implant failures such as subsidence and migration of metallic endplates are also a concern.

In this study, we focus on the unconstrained and semi-constrained TDAs with metal-on-polyethylene articulations. In these TDAs, damage and wear on the polyethylene components would occur, as observed in hip and knee replacements, with different extent and severity [68, 76]. Osteolysis caused by polyethylene wear debris is a common cause of aseptic loosening implant failure in MoP arthroplasty. Furthermore, well-documented retrieval studies for mobile-bearing (unconstrained) SB Charite TDAs have revealed that the polyethylene rim fracture and deformation (Figure 2.13) are a more noticeable problem rather than polyethylene wear [10, 76, 77]. In these retrieved Charite TDAs, the median unworn thickness of the polyethylene rim was only 2.9 mm (range: 0.9 to 5.0 mm), which could be regarded as the original design dimension [76]. Due to the unsymmetric motion of the superior and inferior MoP articulations *in vivo*, the rim was repeatedly bent and micro-cracks initiated in the rim areas. Hence, fractures and plastic deformations were commonly observed at the rims of these retrieved Charite TDAs, as presented in Figure 2.13.

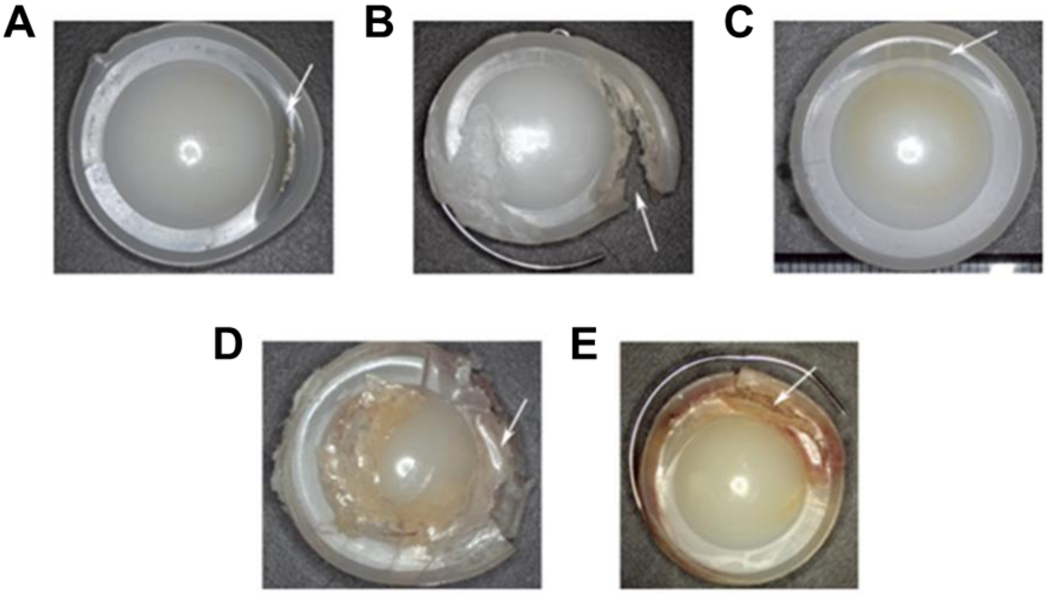


Figure 2.13: PE damage of retrieved Charite TDAs [76], reproduced with permission from Elsevier. (a) Burnishing, plastic deformation, and transverse crack (white arrow); (b) Transverse crack, rim fracture (white arrow); (c) Burnishing, plastic deformation (white arrow); (d) Rim fracture, transverse cracks (white arrow); (e) Burnishing, plastic deformation, and transverse crack (white arrow).

Aiming at these complications, Galbusera et al. (2008) suggested the key biomechanical parameters for TDA design: 1) restoration of a physiological kinematics and mobility, avoiding segmental instability; 2) restoration of a correct spinal alignment; 3) protection of the biological structures, such as the adjacent intervertebral discs, the facet joints and the ligaments, from overloading and resulting accelerated degeneration; 4) device stability and wear. This present research will focus on assessment of commercially available TDA designs using these key biomechanical parameters. However, the TDA device integrity and wear are not a focus of this study. We assume that commercially available artificial prostheses have been strictly tested according to ASTM F2326 (TDA static and dynamic tests) and F2423 (TDA functional, kinematic and wear assessment) protocols.

2.6. Finite Element Analysis

Cadaveric simulator and *in-vivo* imaging systems [37, 78] have been used to investigate spine kinematics, but sporadic experiments were designed to measure stress, strain, loads in various structures. For example, IVD stress profiles were measured using needle transduce [5] and facet joint deformation using strain gage [46]. Generally, measurements and experiments are impractical, when complicated internal loads and other clinically relevant parameters need to be evaluated. Therefore, finite element (FE) analysis is a powerful tool, which supplements the *in-vivo* and *in-vitro* studies.

2.6.1. Intervertebral Disc Prestrain FE Analysis

Typically, soft tissue such as ligament, cornea, vessel, and IVD observed in clinical images is a configuration prestressed by internal pressure or tissue hydration [11, 38, 79–81]. Therefore, simulation based on the image configuration without considering the residual stress may cause a significant discrepancy with the actual stress. In previous studies, different prestrain simulation methods [11, 38, 79–81] been developed to iteratively compute the unloaded soft tissue configuration (an assumed configuration without prestress) and simulate the prestrained soft tissue configuration.

For the IVD tissue, Marini et al. (2014; 2016) performed a prestrain simulation, in which an IDP of 0.1 MPa (representing the physiological IDP in the supine position) was applied on the internal space where the NP was removed from the IVD FE model. After the unloaded AF (losing the physiological IDP) was inversely computed from the image geometry using a “pull-back” algorithm, the NP was reinserted to the IVD FE model, and an adaptation simulation was performed by applying the initial strain distribution. Due to the incompressibility of the NP, the disc geometry was not influenced by the reinsertion of the NP into the IVD FE model. Therefore, a prestrained IVD identical to the image geometry was obtained, and the simulated responses of the prestrained IVD in different loading scenarios could compare favorably with *in-vitro* experimental measurements [28, 37]. However, it is noted that the NP was not considered when they conducted the “pull-back” computation. Therefore, the entire geometry of a degenerated disc due to the loss of the IDP cannot be obtained using their approach, and thus the effect of IDP on the disc behavior cannot be investigated.

Jacobs et al (2014) developed a viscoelastic lumbar IVD FE model to investigate the time-dependent disc response in compression. Homogeneous material properties were assigned to soft tissues (NP, AF, cartilaginous and bony endplates) in their IVD FE model. Prior to loading, the disc bulges of the NP and AF were simulated after tissue hydration in saline solution for 24 hours. However, they did not derive the configuration of the unloaded disc on which prestrain simulation should be based, so the prediction accuracy of the resulting stress distribution in physiological relevant loading scenarios might be compromised. In addition, the computational cost using viscoelastic analysis would also be a concern to implement the iterative “pull-back” algorithm.

2.6.2. Loading Protocols

Two testing protocols have been used to perform the biomechanical tests of the intact and treated lumbar spine in both *in-vitro* experiments and computational simulations. In the load-control protocol, the spine is applied an axial follower compressive load to simulate the upper body weight and the activation of spinal muscles [82], followed by a pure moment in physiologically relevant loading scenarios (flexion, extension, lateral bending and axial torsion). In real life, people bend or torque their spines within a similar, limited ROM regardless of whether their spines are healthy or subjected to surgical treatments. Hence, a hybrid loading protocol was proposed by Goel et al. (2005). It requires that the overall rotation of the treated spine in an axial compression and a pure moment (load-control) is equal to that of the intact spine subjected to the same loading scenario.

2.6.3. Intact Lumbar Spine FE Analysis

Most spinal tissues have complicated anisotropic and heterogeneous material properties. Previous studies [33, 34] have demonstrated that many different parameter combinations in FE spine models may yield simulation results in good agreement with *in-vitro* or *in-vivo* experimental measurements. However, the mechanical behavior of individual tissue is confounded, so it may cause poor prediction for the responses of a postoperative spine, in which implants are inserted by partially dissecting spinal tissues. Therefore, before the development of multi-segment spine models, it is recommended to calibrate material properties of spinal tissues in a spinal segment FE model using a procedure of stepwise reduction of functional tissues [28, 42]. As such, simulation results can match not only the overall response of a spine but also the responses of individual components.

FE analysis has been widely used to investigate the spinal behavior in different loading scenarios noninvasively. Dreischarf et al. (2014) reviewed eight well-established FE models of the lumbar L1-L5 spine incorporating all functional spinal tissues (vertebrae, IVDs, ligaments, and cartilages) in different research centers around the globe. Their simulation results can match well with *in-vitro* measurements of the nonlinear spine behavior using pure and combined loading modes, and the resulting functions of force-displacement and moment-rotation can be used as input data for musculoskeletal models [32].

2.6.4. Instrumented Lumbar Spine FE Analysis

FE spine models can be used to compare the mechanical responses of the healthy spine to those of the instrumented spine subjected to different surgical treatments [73, 83–86]. In general, the responses at the treated level and the adjacent levels are sensitive to the implant stiffness (TDA vs. interbody cage), although slight differences in reported simulation results exist. The main problems in a treated lumbar spine occur at the implanted level, compared to the spinal fusion which induces abnormal responses at the adjacent levels.

Using the load-control protocol, Chen et al. (2009) reported that the fixed-bearing TDA (ProDisc) led to instability at the treated level, which might accelerate degeneration at the highly stressed annulus and facet joint, but no instability at the adjacent levels. In contrast, the posterior interbody fusion procedure revealed possibly accelerative degeneration of the annulus and facet joint at both adjacent levels. However, Denozière and Ku (2006) showed that the same fixed-bearing TDA design caused a higher risk of instability and further degeneration than predicted for the fused model at both the treated and adjacent levels.

For the mobile-bearing TDA (Charite), a decrease in segmental rotational stiffness resulting from TDR and the removal of soft tissue structures caused excessive motion and increased loading in the facets at the implanted level [87]. Using the hybrid loading protocol, Goel et al. (2005) found that the mobile-bearing TDA slightly increases motion with an increase in facet loads at the implanted level, when compared to the adjacent segments where ROMs and loads decrease at the adjacent levels. In addition, impingement is a concern that is sensitive to disc height distraction, anterior-posterior position, implant lordosis and spinal sagittal orientation [7].

Through our survey of the reported simulations of TDA treatments, we can conclude that TDA design significantly influences TDA kinematics, contact mechanics, and spinal tissue load-sharing. Therefore, it is essential to employ design optimization to elucidate the association of TDA responses with TDA design.

2.7. Multiobjective Optimization

Multiobjective optimization (MOO) is used to consider more than one performance metric during optimization. As the number of objectives (performance metric) increases, the dimensions of the objective space also increases. By convention, an optimization problem is defined as a minimization problem (a maximization problem can be converted to a minimization problem). In single-objective optimization (SOO), the optimum is a point in the 1-D objective space (Figure 2.14a). As the number of objectives increases, the set of optimums in bi-objective and tri-objective optimizations is known as the Pareto frontier. Its typical shapes in 2-D and 3-D objective spaces are a curved line (Figure 2.14b) and a curved face (Figure 2.14c), respectively. Since the Pareto frontier is comprised of optimums, no point outperforms another in terms of all objective functions.

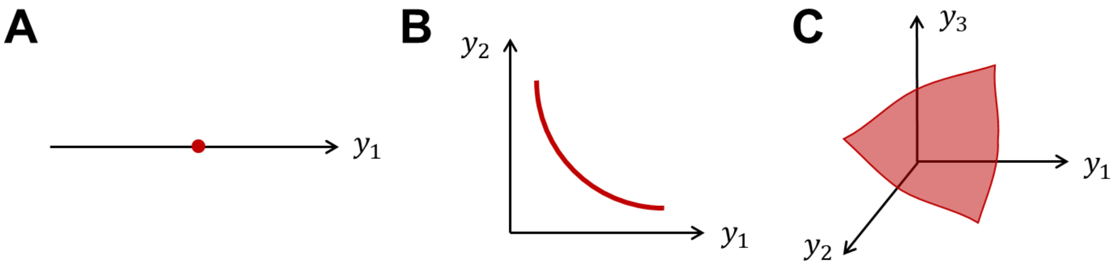


Figure 2.14: The optimums in 1-D (a), 2-D (b), 3-D (c) objective spaces.

Using MOO, we can describe the trade-off among objectives using the Pareto frontier. For the convenience of the visualization, we consider a Pareto curve (Figure 2.15) in a 2-D objective space, which is obtained using bi-objective optimization. It can be observed that as the objective y_1 is improved (decreased), the other objective y_2 is compromised (increased), as shown the green Pareto curve along the direction marked by an arrow in Figure 2.15. Hence, this Pareto curve allows the designer to explore how much be sacrificed in terms of either objective function in order to improve in the other, while still maintaining an optimum design.

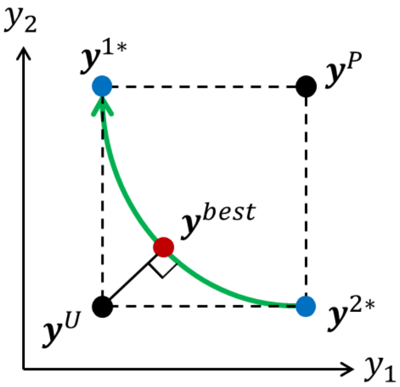


Figure 2.15. The Pareto curve (green) in 2-D objective space. Blue points: the anchor points ($\boldsymbol{y}^{1*}$ and $\boldsymbol{y}^{2*}$). Black points: the Utopian point ($\boldsymbol{y}^{U}$) and the pseudo Nadir point ($\boldsymbol{y}^{P}$), respectively. Red point: the best trade-off point ($\boldsymbol{y}^{best}$) on the Pareto curve in the minimum distance from the Utopian point ($\boldsymbol{y}^{U}$).

The end points of the Pareto curve are called the anchor points ($\boldsymbol{y}^{1*}$ and $\boldsymbol{y}^{2*}$), as indicated by the two blue points in Figure 2.15. These anchor points represent the performance of designs which was optimized (minimized) for just one objective, without considering the other; they are single-objective optimums solved by SOO. The anchor points can determine two special points, the Utopian point ($\boldsymbol{y}^{U}$) and the pseudo Nadir point ($\boldsymbol{y}^{P}$), as shown in Figure 2.15. The Utopian point ($\boldsymbol{y}^{U}$) is an ideal solution, only when the two anchor points (single-objective optimums) are coincident; therefore, the Utopian point is commonly impossible to be achieved. Correspondingly, the best trade-off point ($\boldsymbol{y}^{best}$, Figure 2.15) can be defined as a point on the Pareto curve in the minimum distance from the Utopian point ($\boldsymbol{y}^{U}$), representing the minimum sacrifices with respect to the Utopian point ($\boldsymbol{y}^{U}$).

The pseudo Nadir point ($\boldsymbol{y}^{P}$) represents a “bad” design point in the objective space, whose components was chosen as the worst performance measure of each single-objective optimum. Different from the pseudo Nadir point, the Nadir point ($\boldsymbol{y}^{N}$) is defined a point, whose components are the maxima that each objective can reach in the feasible objective space; therefore, the Nadir point is also an idea design point. To better understand the Utopia point, pseudo Nadir point, and the Nadir point, their relationship is described using an example in the 2-

D objective space, as shown in Figure 2.16. For further discussion of relevant concepts and formula in multiobjective optimization, please be referred to Appendix C.2.2.

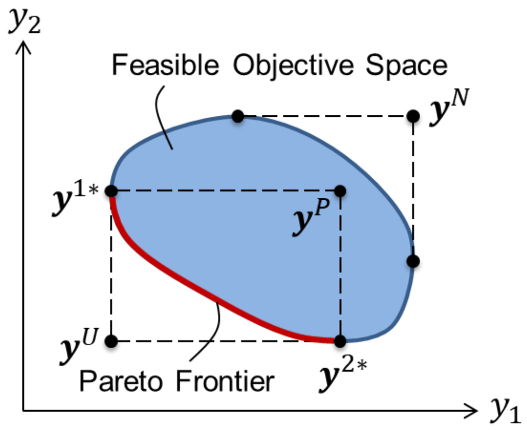


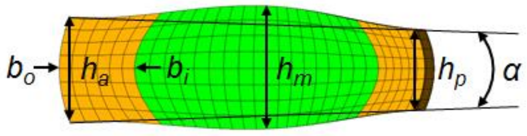
Figure 2.16: Illustration of the Utopia point ($\boldsymbol{y}^U$), pseudo Nadir point ($\boldsymbol{y}^P$), and the Nadir point ($\boldsymbol{y}^N$), modified from [88] with Dr. Achille Messac's permission.

Chapter 3: Investigation of the Lumbar Intervertebral Disc Biomechanics

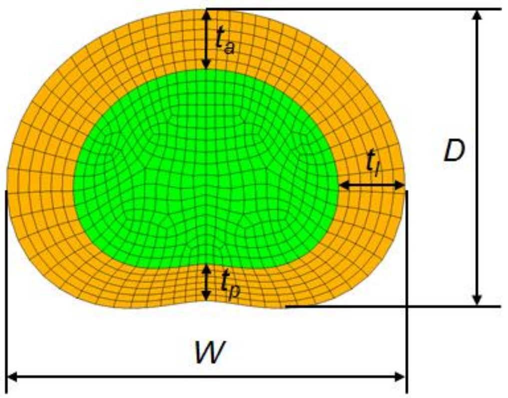
Adjacent vertebrae articulate through the intervertebral disc (IVD), which serves to guide and restrict intervertebral rotations. The intervertebral disc is prestressed by an intradiscal pressure (IDP) caused by nucleus swelling. The collagen fiber reinforced annulus fibrosus (AF) provides resistance to intervertebral rotations and pressurizes the nucleus pulposus (NP) to prevent disc collapse. Therefore, the objective is to investigate the changes in the disc geometry, fiber orientation and biomechanics by nucleus swelling. A sophisticated FE model of the disc was developed by incorporating non-homogeneous material properties across the AF, while nucleus swelling was simulated by repurposing thermoelastic analysis. The resulting stresses in the nucleus and annulus were compared to those using a reported simulation method of nucleus swelling, to demonstrate that nucleus swelling is mechanically equivalent to thermal expansion. The IVD mechanical responses were simulated in various physiologically relevant loading scenarios, by varying the intradiscal pressures (normal vs. degenerated discs) and fiber orientations.

3.1. Parametric Finite Element Modeling of Intervertebral Disc

Using a mesh morphing technique (Appendix A), a representative L2-L3 lumbar IVD with a transverse diameter of 50 mm and a sagittal diameter of 37 mm [39] was parametrically modeled, as shown in Figure 3.1 and Table 3.1. According to reported measurements [89], the disc heights ($h_a = 10.9$ mm, $h_m = 12.8$ mm and $h_p = 8.3$ mm) in the disc sagittal section were defined, considering the disc inclination angle (4°) in the sagittal plane [39], as well as the superior and inferior endplate bulge depths (1.9 and 1.4 mm, respectively) [90]. The mean disc height was defined as $(h_a + h_p)/2 = 9.6$ mm, which is equal to that when flat endplates were assumed [39].



The sagittal section



The transverse section

Figure 3.1: The L2-L3 lumbar IVD FE model defined in the disc sagittal and transverse sections, according to reported measurements (Table 3.1). The green and yellow potions represent the NP and AF, respectively.

Table 3.1: The anatomic shape parameters of the L2-L3 lumbar IVD FE model.

Shape Parameters		References
W (mm)	52	Marini and Ferguson (2014)
D (mm)	39	
α (°)	4	
b_o, b_i (mm)	1, 2.5	
h_a, h_m, h_p (mm)	10.9, 12.8, 8.3	Van Der Houwen et al (2010)
t_a, t_l, t_p (mm)	7.8, 8.7, 4.8	
NP vol. ratio (%)	44	

The disc bulge under the 0.1 MPa IDP in the lying supine position [91] was assumed as 1 mm on the outer AF wall and 2.5 mm on the inner AF wall [38, 39]. The circumferentially varying AF thicknesses ($t_a = 7.8$ mm, $t_l = 8.7$ mm and $t_p = 4.8$ mm measured at the disc transverse section) were chosen such that the nucleus volume has a ratio of 44% to the total disc volume [33], with its center posterior to the disc centroid. As shown in Figure 3.2, the L2-L3 IVD finite element model using the reported anatomic dimensions includes main shape features, in good agreement with the mean shape of the L3-L4 disc in a statistical shape analysis [92].

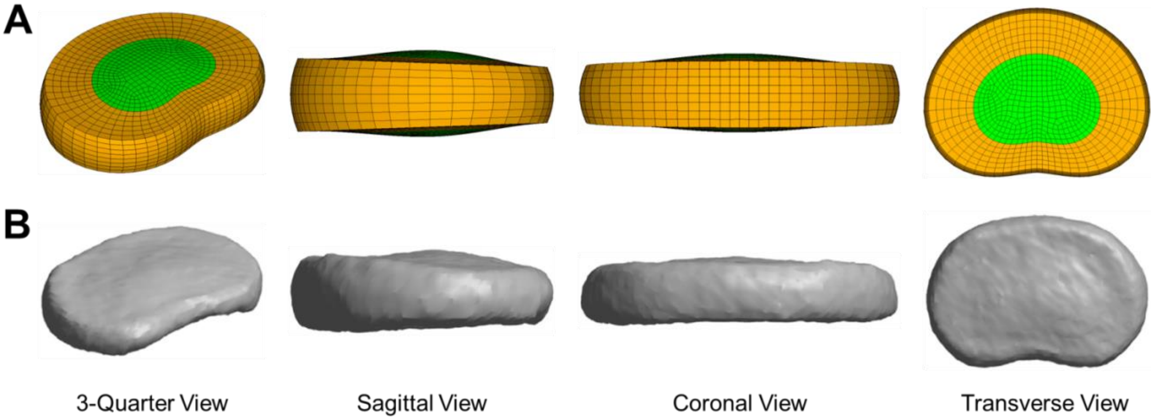


Figure 3.2: Comparison of the L2-L3 disc parametric FE model (a) with the mean shape (b) of the L3-L4 disc in a statistical shape analysis [92], reproduced with permission from Elsevier.

3.2. Constitutive Material Model

3.2.1. Nucleus Pulposus

The NP was assumed as nearly-incompressible, homogeneous and isotropic hyperelastic materials using a Neo-Hookean model [38, 39], as expressed below:

$$\Psi^{NP}(J, \bar{I}_1) = \Psi^{NP}_{vol}(J) + \Psi^{NP}_{iso}(\bar{I}_1) = \frac{\kappa^{NP}}{2}(J-1)^2 + c_1^{NP}(\bar{I}_1 - 3) \tag{3.1}$$

where $\Psi^{NP}(\bar{I}_1, J)$ is the strain energy functions of the NP. It is decoupled into a volumetric part, $\Psi^{NP}_{vol}(J)$, and an isochoric (volume-preserving) part, $\Psi^{NP}_{iso}(\bar{I}_1)$. $J = \det(\boldsymbol{F})$ is the determinant of the deformation gradient ($\boldsymbol{F}$), which is used to measure the volume ratio. To characterize the isochoric deformation, $\bar{I}_1 = \mathrm{tr}(\overline{\boldsymbol{C}})$ is defined as the first invariant of the modified right Cauchy-Green tensor, $\overline{\boldsymbol{C}} = \overline{\boldsymbol{F}}^T \overline{\boldsymbol{F}}$, in which $\overline{\boldsymbol{F}} = J^{-1/3}\,\boldsymbol{F}$ satisfying $\det(\overline{\boldsymbol{F}}) = 1$ in the isochoric deformation. κ^{NP} and c_1^{NP} are two material constants of the NP tissue. The Neo-Hookean model (Equation 3.1) provided by ANSYS was used to simulate the NP using material constants κ^{NP}= 67.163 MPa and c_1^{NP} = 3.353 kPa [38].

3.2.2. Annulus Fibrosus

A user material subroutine (ANSYS USERMAT subroutine) developed in Appendix B was used to simulate the fiber reinforced AF composite. The AF was assumed as the nearly-incompressible, heterogeneous and two-family fiber reinforced hyperelastic material. The strain energy was decoupled into a volumetric part, $\Psi_{vol}^{AF}(J)$, an isochoric part, $\Psi_{iso}^{AF}(\bar{I}_1)$, and two fiber-reinforcing parts of the two fiber families in different fiber directions, $\Psi_{fib}^{AF}(\lambda = \lambda_{\boldsymbol{a}})$ and $\Psi_{fib}^{AF}(\lambda = \lambda_{\boldsymbol{g}})$:

$$\Psi^{AF}(J, \bar{I}_1, \lambda_1, \lambda_2) = \Psi_{vol}^{AF}(J) + \Psi_{iso}^{AF}(\bar{I}_1) + \Psi_{fib}^{AF}(\lambda_{\boldsymbol{a}}) + \Psi_{fib}^{AF}(\lambda_{\boldsymbol{g}}) \tag{3.2}$$

The volumetric part $\Psi_{vol}^{AF}(J)$ and isochoric part $\Psi_{iso}^{AF}(\bar{I}_1)$ simulate the isotropic material responses of the AF ground substance. Two fiber reinforcing parts $\Psi_{fib}^{AF}(\lambda_{\boldsymbol{a}})$ and $\Psi_{fib}^{AF}(\lambda_{\boldsymbol{g}})$ were used to model the material transverse isotropy (the stiffness in a direction is much greater than that orthogonal to the direction) in the spatial fiber directions $\boldsymbol{a}$ and $\boldsymbol{g}$ in a crossing pattern. The strecth ratio ($\lambda_{\boldsymbol{a}}$ or $\lambda_{\boldsymbol{g}}$) was introduced to descibe the change in the fiber length along the spatial fiber direction ($\boldsymbol{a}$ or $\boldsymbol{g}$). It should be noted that the reinforcing fiber parts do not contribute to the compressibility (*i.e.*, $J \approx 1$ is held, so $\overline{\boldsymbol{C}} = \boldsymbol{C}$), thus two pseudo-invariants $\lambda_{\boldsymbol{a}} = \sqrt{\boldsymbol{a}_0 \cdot \boldsymbol{C}\boldsymbol{a}_0}$ and $\lambda_{\boldsymbol{g}} = \sqrt{\boldsymbol{g}_0 \cdot \boldsymbol{C}\boldsymbol{g}_0}$ [29, 93] are introduced in the material model. Here, $\boldsymbol{C}$ is the right Cauchy-Green tensor $\boldsymbol{C} = \boldsymbol{F}^T\boldsymbol{F}$. $\boldsymbol{a}_0$ and $\boldsymbol{g}_0$ are the fiber directions in the reference configuration, satisfying

$\boldsymbol{Fa}_0 = \lambda_{\boldsymbol{a}}\boldsymbol{a}$ and $\boldsymbol{Fg}_0 = \lambda_{\boldsymbol{g}}\boldsymbol{g}$, respectively. The two pseudo-invariants ($\lambda_{\boldsymbol{a}}$ and $\lambda_{\boldsymbol{g}}$) represent the stretch ratios in the spatial crossing patterned fiber directions, $\boldsymbol{a}$ and $\boldsymbol{g}$.

Same as the previous model [38, 39], the AF ground substance (GS) was represented by a modified Yeoh's model (combing Equations 3.3 and 3.4 together).

$$\Psi_{vol}^{AF}(J) = \frac{\kappa}{2}(J-1)^2 \tag{3.3}$$

$$\Psi_{iso}^{AF}(\bar{I}_1) = c_1(\bar{I}_1 - 3) + c_2(\bar{I}_1 - 3)^2 \tag{3.4}$$

where the material constants of the AF GS were chosen as $\kappa = 73.521$ MPa, c_1 = 38.793 kPa, and c_2 = 55.049 kPa [38].

Since the single lamellar mechanical behavior is tension-only and nonlinear [19, 20], a piecewise function was proposed to formulate the strain energy function $\Psi_{fib}^{AF}(\lambda)$ of the reinforcing fiber part as shown in Equation 3.5. The constative relationship of the reinforcing fibers intimately interacts with the algorithms for fitting and interpolation of the experimentally measured stress-stretch curves (described in the next section), in order to realize the modeling of the material heterogeneity across the annulus fibrosus.

$$\Psi_{fib}^{AF}(\lambda) = \begin{cases} 0, & \lambda < 1 \\ c_3\left\{\frac{1}{c_4}\left[e^{c_4(\lambda-1)} - 1\right] - (\lambda - 1)\right\}, & 1 \le \lambda < c_5 \\ \frac{E^*}{2}(\lambda - c_5)^2 + T^*(\lambda - c_5) + \Psi^*, & \lambda \ge c_5 \end{cases} \tag{3.5}$$

where $E^* = c_3 c_4 e^{c_4(c_5-1)}$, $T^* = c_3\left[e^{c_4(c_5-1)} - 1\right]$ and $\Psi^* = c_3\left\{\frac{1}{c_4}\left[e^{c_4(c_5-1)} - 1\right] - (c_5 - 1)\right\}$. The AF composite material model has 12 material constants including κ, c_1, c_2, c_3, c_4, c_5, three components of $\boldsymbol{a}_0$ and three components of $\boldsymbol{g}_0$. Especially, $\lambda^* = c_5$ is defined as the uncrimped stretch of the AF collagen fiber [94].

3.3. Modeling of Material Anisotropy and Heterogeneity in Annular Lamellae

3.3.1. Fitting of Fiber Lamellar Tensile Properties

It is assumed that adjacent lamellae have the same tensile properties, but alternating fiber orientations represented by $\boldsymbol{a}$ and $\boldsymbol{g}$. Therefore, only the fiber family in one fiber direction (*e.g.*, $\boldsymbol{a}$) needs to be considered, when fitting the lamellar tensile properties. According to previously reported uniaxial tensile test data of single lamellae extracted from L1-L2 disc specimens [19], a representative tension-only, nonlinear engineering stress (T) – stretch ratio (λ) curve of a single lamella along the fiber direction is shown in Figure 3.3.

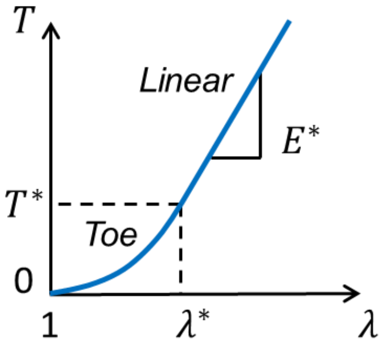


Figure 3.3: Illustration of engineering stress–stretch curves in the material model.

Using Equation 3.5, the Cauchy stress of an arbitrary fiber with a stretch ratio λ in the fiber direction $\boldsymbol{a}$ can be written as:

$$\boldsymbol{\sigma}_{fib}^{AF} = \frac{\lambda}{J}\frac{\partial \Psi_{fib}^{AF}(\lambda)}{\partial \lambda}\boldsymbol{a} \otimes \boldsymbol{a} = \frac{\lambda}{J}\Psi_{\lambda}^{fib}\boldsymbol{a} \otimes \boldsymbol{a} = \sigma_{fib}^{AF}\boldsymbol{a} \otimes \boldsymbol{a} \tag{3.6}$$

where the notation $\Psi_{\lambda}^{fib} = \frac{\partial \Psi_{fib}^{AF}(\lambda)}{\partial \lambda}$ have been employed. The scalar σ_{fib}^{AF} represents the Cauchy/true stress magnitude in the fiber direction $\boldsymbol{a}$. In a uniaxial test along the fiber direction $\boldsymbol{a}$, the fiber engineering stress $T(\lambda)$, typically used in experimental characterization, can be obtained according to the nearly-incompressible constraint ($J \approx 1$).

$$T(\lambda) = \frac{\sigma_{fib}^{AF}}{\lambda} = \Psi_{\lambda}^{fib}(\lambda) \tag{3.7}$$

The single lamellar engineering stress (T) – stretch ratio (λ) relation (Figure 3.7) can be expressed as a piecewise engineering stress-stretch function. Substituting Equation 3.5 into Equation 3.7 and differentiating $\Psi_{fib}^{AF}(\lambda)$ with respect to λ, $T(\lambda)$ is obtained,

$$T(\lambda) = \begin{cases} 0, & \lambda < 1 \qquad \text{(compression region)} \\ c_3\left[e^{c_4(\lambda-1)} - 1\right], & 1 \le \lambda < \lambda^* \quad \text{(exponential toe region)} \\ E^*(\lambda - \lambda^*) + T^*, & \lambda \ge \lambda^* \qquad \text{(linear region)} \end{cases} \tag{3.8}$$

where $\lambda^* = c_5$ is the uncrimped stretch ratio of collagen fiber, at which the tensile behavior transits from the exponential region to the linear region. T^* is defined as the uncrimped engineering stress, when the fibers are stretched to the uncrimped stretch ratio of λ^* and starts the linear tensile behavior.E^* is the tangent modulus in the linear region.

The formulation (Equation 3.8) is similar to the previously reported constitutive model of ligaments and tendons [29, 94, 95], where 4 material constants were used to describe the fiber tension. However, it should be noted that there are only 3 material constants (c_3, c_4 and $\lambda^* = c_5$) in Equation 3.8, since we have assumed that the $T(\lambda)$ function is continuously differentiable (C^1) at the fiber uncrimped stretch (λ^*). The assumption largely facilitates fitting and interpolation of fiber uniaxial tensile test data to simulate the gradual transition in the fiber stiffness distribution across the AF, due to the AF material heterogeneity.

According to the property of C^1 continuity at λ^*, they can be expressed in terms of c_3, c_4 and λ^*, using the function of the exponential toe region in Equation 3.8,

$$T^* = T(\lambda^{*-}) = c_3\left[e^{c_4(\lambda^*-1)} - 1\right] \tag{3.9}$$

$$E^* = \frac{\partial T(\lambda^{*-})}{\partial \lambda} = c_3 c_4 e^{c_4(\lambda^*-1)} \tag{3.10}$$

By combining Equations 3.9 and 3.10, c_3 and c_4 can be uniquely determined by T^* and E^*. Therefore, in the uniaxial fiber material model (Equations 3.5 and 3.8), the three material parameters, c_3, c_4 and λ^* are equivalent to T^*, E^* and λ^*. Although no explicit solutions to c_3 and c_4 in terms of T^* and E^* exist, they can be numerically solved in MATLAB R2014a. To fit experimental T–λ curves, it is more convenient to use T^*, E^*, and λ^* rather than using c_3, c_4, and λ^*, because their physical meanings are so clear that they can be directly measured from the fiber uniaxial tensile curve, as illustrated in Figure 3.3.

According to the experiment of Holzapfel et al. (2005), tensile tests were performed using single lamella specimens of the L1-L2 discs dissected at 4 anatomic sites (VLe = ventral-lateral external; De = dorsal internal; VLi = ventral-lateral internal; Di = dorsal internal, Figure 3.4a). The load axis was aligned with the fiber orientation of a single annular lamella specimen (Figure 3.4b). Using Equation 3.8, the fitted T–λ curves of the AF lamellae at 4 testing sites (Figure 3.4a) of the L1-L2 disc specimens [19] are shown in Figure 3.4c.

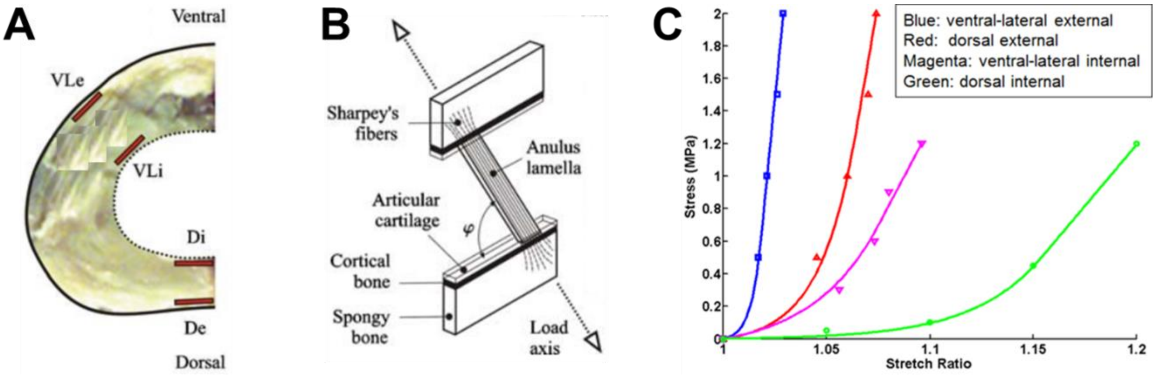


Figure 3.4: Description of previously reported single lamellar tensile test [19], modified with permission from Springer Nature. (a) Four test sites of the L1-L2 disc specimens where single lamella specimens were dissected. (b) Schematic diagram of the tensile test, in which the load axis was aligned with the fiber orientation of a single lamella specimen. (c) Fitting of the engineering stress (T) – stretch ratio (λ) curves to the measurements at the 4 testing sites using our formulation (Equation 3.8).

3.3.2. Interpolation of Fiber Lamellar Tensile Properties

The heterogeneous distribution of the AF lamellar tensile properties results from the gradually varying collagen content distribution across the AF [23, 39]. However, the lamellar tensile properties were only measured at a limited number of anatomic sites (Figure 3.4). Therefore, more T–λ tensile curves need to be interpolated to simulate the material heterogeneity due to the gradual change in collagen content distribution [39].

As illustrated in Figure 3.5a, the linear segments of Curves 1 and N necessarily intersect at a point denoted by (λ_C, T_C). Supposing that $N-2$ curves (Figure 3.5b) passing through the common intersection (λ_C, T_C) are interpolated between Curve 1 and Curve N, the linear segment of each curve can be defined using the coordinate (λ_C, T_C) of the intersection point,

$$T = E_i^*(\lambda - \lambda_c) + T_c, \quad i = 1,\ 2,\ \ldots,\ N \tag{3.11}$$

where E_i^* is the slope of the linear segment of each curve.

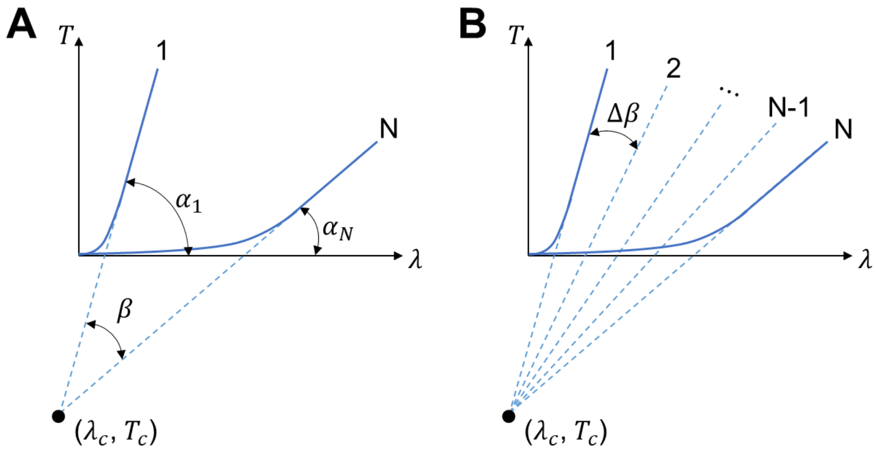


Figure 3.5: Illustration of the interpolation algorithm. (a) The linear segments of Curves 1 and N intersect at a point (λ_C, T_C); (b) $N-2$ lines passing through the intersection point are interpolated.

Combining the linear segments equations of Curves 1 and N in Equation 3.11,

$$\begin{cases} T = E_1^*(\lambda - \lambda_1^*) + T_1^* \\ T = E_N^*(\lambda - \lambda_N^*) + T_N^* \end{cases} \tag{3.12}$$

The intersection point (λ_C, T_C) can be solved,

$$\begin{cases} \lambda_c = \dfrac{T_1^* - T_N^* + E_N^*\lambda_N^* - E_1^*\lambda_1^*}{E_N^* - E_1^*} \\ T_c = \dfrac{E_N^*T_1^* - E_1^*T_N^* + E_1^*E_N^*(\lambda_N^* - \lambda_1^*)}{E_N^* - E_1^*} \end{cases} \tag{3.13}$$

It can be demonstrated that the intersection angle between Curves 1 and N is $\beta = \alpha_1 - \alpha_N$, where α_1 and α_N are the inclination angles regarding the abscissa axis λ (Figure 3.5a). To model the gradual change in the tensile properties, it is assumed that the angles between adjacent curves are identical, i.e., $\Delta\beta = \beta/(N-1)$, as shown in Figure 3.5b. Therefore, the inclination angle (α_i) and slope (E_i^*) of the linear segment each curve can be obtained,

$$\alpha_i = \alpha_1 - (i-1)\Delta\beta, \quad i = 1,\ 2,\ \ldots,\ N \tag{3.14}$$

$$E_i^* = \tan(\alpha_i)\,, \quad i = 1,\ 2,\ \ldots,\ N \tag{3.15}$$

Furthermore, it is assumed that the uncrimped stretches (λ_i^*) are equally spaced,

$$\lambda_i^* = \lambda_1^* + \frac{(i-1)(\lambda_N^* - \lambda_1^*)}{N-1}, \quad i = 1,\ 2,\ \ldots,\ N \tag{3.16}$$

According to the definition of the uncrimped stress (T^*), it is the stress that corresponds to the uncrimped stretch (λ^*), i.e., $T^* = T(\lambda^*)$. Thus, substituting Equation 3.16 into Equation 3.11, the uncrimped stress of each curve can be obtained,

$$T_i^* = E_i^*(\lambda_i^* - \lambda_c) + T_c, \quad i = 1,\ 2,\ \ldots,\ N \tag{3.17}$$

Combining Equations 3.15 and 3.17, the material constant (c_3^i and c_4^i) of each curve can be numerically solved using E_i^* and T_i^* according to Equations 3.9 and 3.10.

To characterize the circumferentially and radially varying AF lamellar tensile properties, the AF FE model was divided into 6 polar sectors and 6 layers in the radial direction according to the previous experiment [19], as shown in Figure 3.6. Using the interpolation algorithm, 36 T–λ curves (Figure 3.7) of the AF lamellae were interpolated using the 4 fitted T–λ curves (Figure 3.4c); the resulting tensile properties (c_3, c_4 and λ^*) corresponding to each T–λ curve (Figure 3.7) were assigned to the 36 AF regions (6 polar sectors × 6 layers, Figure 3.6). For the implementation of the ANSYS USERMAT subroutine, the engineering stress (T) can be trivially converted to the Cauchy’s tress using Equation 3.7.

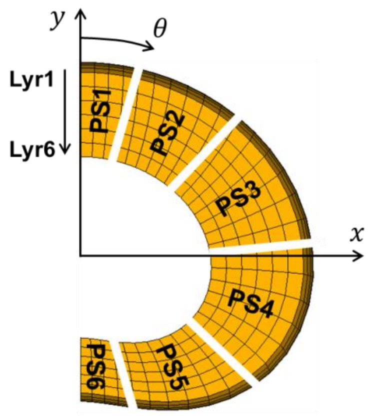


Figure 3.6: The divided polar sectors and layers of the AF lamellae.

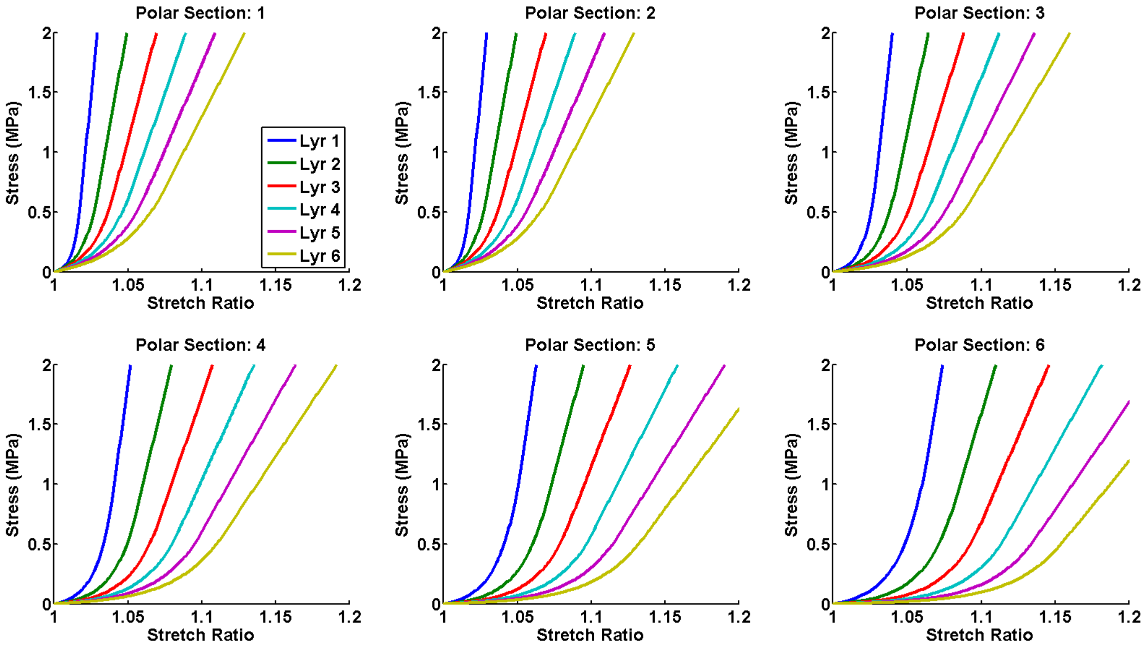


Figure 3.7: The T–λ curves in different regions of the AF lamellae using the proposed interpolation algorithm.

3.3.3. Assignment of Annulus Fibrosis Fiber Orientation

Collagen fibers are embedded within the AF in a crossing pattern. According to the measurement of L1-L2 lumbar disc specimens [19], the alternating fiber angles ($|\varphi_{exp}|$, measured with respect to the disc transverse plane) are linearly related the polar coordinate (θ, Figure 3.6),

$$\left|\varphi_{exp}\right| = 23.2° + 0.13\theta \tag{3.18}$$

To model the AF composite material, an algorithm was proposed for assigning the alternating fiber orientations to the AF FE elements. In this study, we only consider 8-node hexagonal linear solid elements (SOLID185 in ANSYS), which are used to mesh the disc geometry.

It is assumed that all the fibers in each AF element are aligned on the middle plane across the AF, which is determined using two vectors ($\boldsymbol{d}_1$ and $\boldsymbol{d}_2$), each passing through the centroids of two opposite element faces, as shown in Figure 3.8. Especially, $\boldsymbol{d}_1$ represents the circumferential direction across the AF. Since $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ are not necessarily orthogonal, so a new vector $\boldsymbol{d}_2'$ perpendicular to $\boldsymbol{d}_1$ within the middle plane is introduced,

$$\boldsymbol{d}_2' = (\boldsymbol{d}_1 \times \boldsymbol{d}_2) \times \boldsymbol{d}_1 \tag{3.19}$$

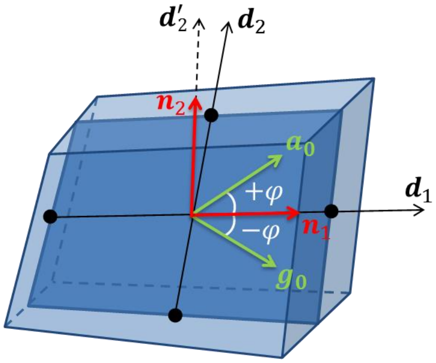


Figure 3.8: Illustration of assignment of fiber orientations to each AF element.

After normalization of $\boldsymbol{d}_1$ and $\boldsymbol{d}_2'$, a set of orthonormal basis vectors ($\boldsymbol{n}_1$ and $\boldsymbol{n}_2$) within the middle plane is obtained,

$$\boldsymbol{n}_1 = \frac{\boldsymbol{d}_1}{|\boldsymbol{d}_1|}, \quad \boldsymbol{n}_2 = \frac{\boldsymbol{d}_2'}{|\boldsymbol{d}_2'|} \tag{3.20}$$

Therefore, the reference directions ($\boldsymbol{a}_0$ and $\boldsymbol{g}_0$) of the two fiber families with alternating fiber angles ($\pm\varphi$) can be determined using the basis vectors ($\boldsymbol{n}_1$ and $\boldsymbol{n}_2$),

$$\boldsymbol{a}_0 = \cos\varphi\boldsymbol{n}_1 + \sin\varphi\boldsymbol{n}_2, \quad \boldsymbol{g}_0 = \cos\varphi\boldsymbol{n}_1 - \sin\varphi\boldsymbol{n}_2 \tag{3.21}$$

To match *in-vitro* measurements of the disc mechanical responses in various physiologically relevant loading scenarios [78], the fiber angles in the disc FE model were tuned to $|\varphi_{tune}| = |\varphi_{exp}| + 10°$, 10° larger than the experimentally measured fiber orientation ($|\varphi_{exp}|$), as shown in Figure 3.9.

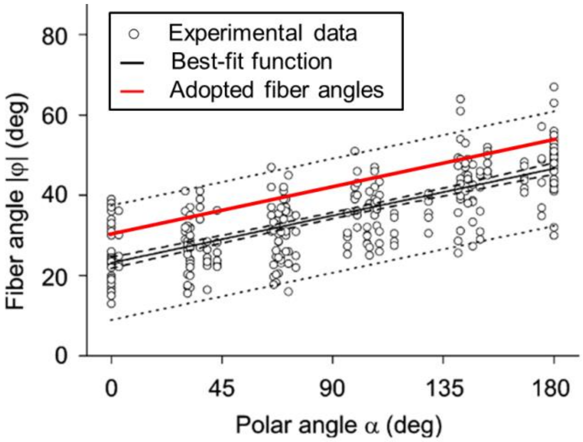


Figure 3.9: Comparison of the experimentally measured fiber orientation ($|\varphi_{exp}|$) with the model adopted fiber orientation ($|\varphi_{tune}|$). Modified from [19] with permission from Springer Nature.

3.4. Pull-back Computation and Disc Prestrain Simulation

Typically, soft tissues observed in clinical images are prestressed by internal pressure and tissue hydration. The intervertebral disc is a soft tissue, and the normal disc is pressurized with a physiological intradiscal pressure (IDP) due to nucleus swelling [91]. As a result of nucleus swelling, the resulting hydrostatic pressure in the nucleus is well known as the intradiscal pressure (IDP). Simulations based on the image configuration without considering these residual stresses may cause a significant discrepancy with the physical reality [80]. Therefore, we hypothesize that disc geometry, fiber orientation and mechanics are influenced by nucleus swelling. To accurately predict the normal disc behavior and internal stresses, it is essential to introduce the nucleus swelling simulation in an IVD element (FE) model, in order to consider the residual stress distribution.

As shown in the Figure 3.10, the disc geometry created using well-documented anatomic dimensions of the L2-L3 disc represents the image configuration ($\boldsymbol{X}_I$). However, the disc image configuration ($\boldsymbol{X}_I$) has been a consequence of disc prestrain caused by nucleus swelling. Thus there is a distribution of residual stress in the disc prestrained configuration ($\boldsymbol{X}_P$), although the disc geometry and fiber orientation of the prestrained configuration ($\boldsymbol{X}_P$) are completely identical to the image configuration ($\boldsymbol{X}_I$). To simulate the residual stress distribution under a physiological IDP, we need to inversely derive the fictitious unloaded configuration ($\boldsymbol{X}_U$), which is assumed to be configuration due to the loss of the IDP. The unloaded configuration ($\boldsymbol{X}_U$) is typically chosen as the reference configuration, such that the prestrained disc configuration ($\boldsymbol{X}_P$) can be obtained by applying the physiological IDP to the nucleus of the unloaded configuration ($\boldsymbol{X}_U$).

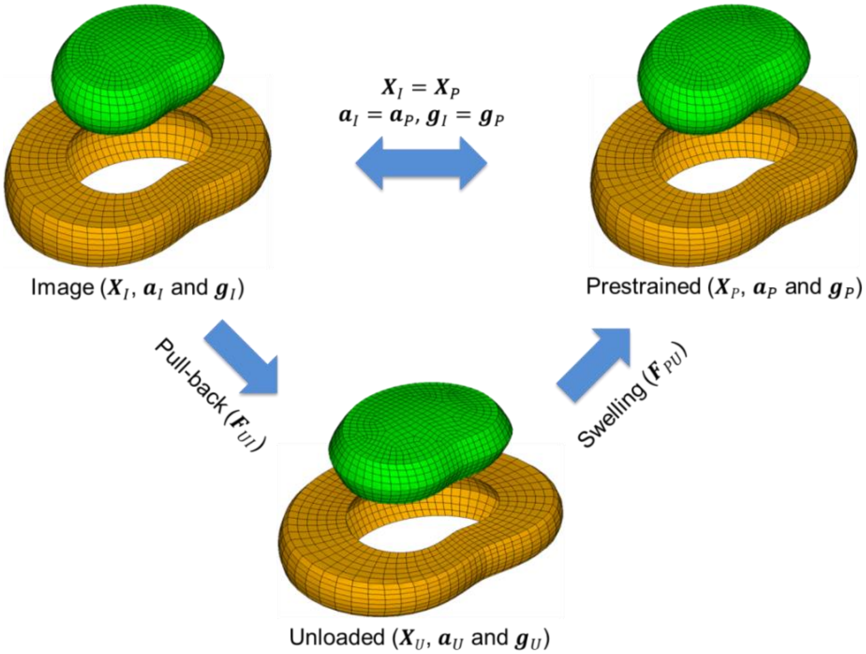


Figure 3.10: The relationship among the image configuration ($\boldsymbol{X}_I$), unloaded configuration ($\boldsymbol{X}_U$) and prestrained configuration ($\boldsymbol{X}_P$) of an intervertebral disc.

The relationship among the three configurations can be described as the following equation [38, 79]:

$$\boldsymbol{F}_{PI} = \boldsymbol{F}_{PU}\boldsymbol{F}_{UI} = \boldsymbol{I} \tag{3.22}$$

where $\boldsymbol{F}_{PI}$, $\boldsymbol{F}_{UI}$, and $\boldsymbol{F}_{PU}$ are the deformation gradients (see the definition in Appendix B.1) from $\boldsymbol{X}_I$ to $\boldsymbol{X}_P$, from $\boldsymbol{X}_I$ to $\boldsymbol{X}_U$, and from $\boldsymbol{X}_U$ to $\boldsymbol{X}_P$, respectively. It is noted that $\boldsymbol{F}_{PI} = \boldsymbol{I}$ is held, since $\boldsymbol{X}_I = \boldsymbol{X}_P$.

3.4.1. Methods

In this study, a pull-back algorithm developed by Riveros et al. (2013) was adopted to iteratively compute the initial configuration of intervertebral disc, as described in Figure 3.11. The computation was initialized by assuming the unloaded (zero-pressure) configuration ($\boldsymbol{X}_U$) equal to the image configuration ($\boldsymbol{X}_I$). In each iteration, a nucleus swelling simulation was performed using the unloaded configuration ($\boldsymbol{X}_U$). The resulting prestrained configuration ($\boldsymbol{X}_P$) equilibrating an IDP of 0.1 MPa [91] was compared with the image configuration ($\boldsymbol{X}_I$). If the geometric error of $\boldsymbol{X}_P$ from $\boldsymbol{X}_I$, defined by the infinity normal ($||\boldsymbol{X}_I - \boldsymbol{X}_P||_\infty$), was larger than a tolerance of 0.2 mm (< image resolution) or the number of iterations less than the preset maximum of 15, the pull-back computation would be repeated by superposing the difference ($\boldsymbol{X}_I - \boldsymbol{X}_P \leq \boldsymbol{0}$) onto the unloaded configuration ($\boldsymbol{X}_U$). Meanwhile, the alternating fiber orientations ($\boldsymbol{a}_I$ and $\boldsymbol{g}_I$) in the image configuration were pulled back to the unloaded configuration using the numerical deformation gradient ($\boldsymbol{F}^{-1}$), which is formulated in Appendix B.1.2. It is noted that the geometric tolerance and the maximum number of iterations were adopted from the previous algorithm [11]; further decreasing the geometric tolerance and increasing the maximum number of iterations caused convergence difficulty and additional computational cost, respectively. When the termination condition was satisfied, the iterative computation stopped, and both the unloaded configuration ($\boldsymbol{X}_U$, $\boldsymbol{a}_U$ and $\boldsymbol{g}_U$) and the prestrained configuration ($\boldsymbol{X}_P$, $\boldsymbol{a}_P$ and $\boldsymbol{g}_P$) identical to the image configuration ($\boldsymbol{X}_I$, $\boldsymbol{a}_I$ and $\boldsymbol{g}_I$) were obtained.

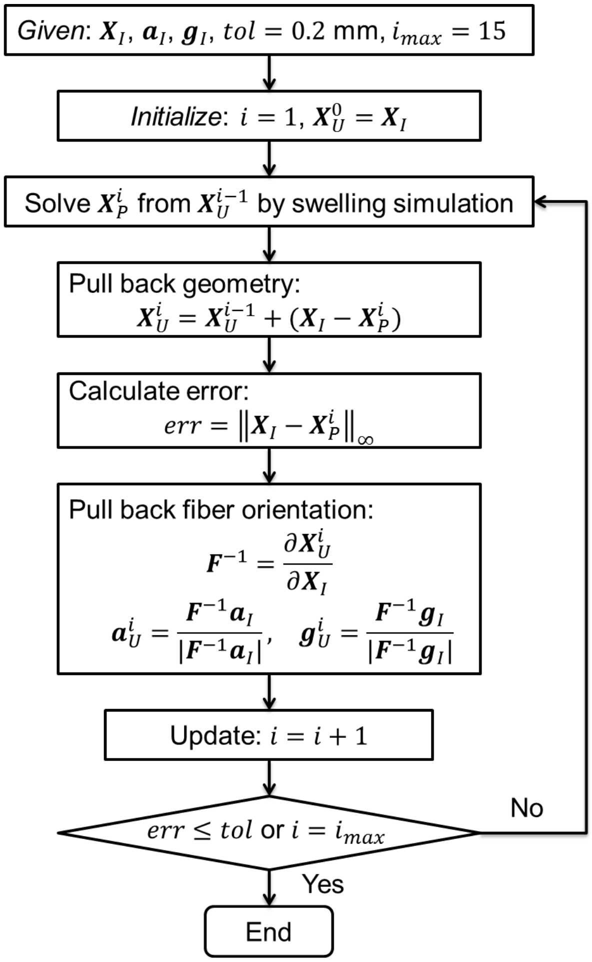


Figure 3.11: The flowchart of the pull-back algorithm.

Nucleus swelling in the pull-back computation was simulated by repurposing thermoelastic FE analysis in ANSYS. Two rigid endplates were attached to the top and bottom of the IVD. No constraints were applied on the superior endplate, but the inferior endplate was fully constrained. Thermal expansion in the NP was simulated in 1℃ temperature rise by assigning a coefficient of thermal expansion (CTE) to the NP. The NP CTE was calibrated to 0.038 /K, such that the simulated IDP in the NP is equal to 0.1 MPa [91]. It is found that there is an approximate linear relationship between the CTE and the IDP, but depending on the disc geometry and material properties. Therefore, the calibration of the CTE commonly requires that the pull-back computation is performed three times, if any change in the disc FE model is made.

In the previous report pull-back computation [38, 39], the physiological intradiscal pressure (IDP) of 0.1 MPa was directly applied on the internal surface of the AF and rigid endplates, so only the AF was pulled back to the unloaded configuration. The prestrain of the entire disc (AF and NP) was achieved by performing a one-increment adaptation simulation, in which the re-inserted NP was pressurized by applying the strain distribution caused by the IDP of 0.1 MPa. Since we introduced the nucleus swelling simulation in the iterative computation, both the AF and NP can be pulled back to the unloaded configuration simultaneously.

3.4.2. Results

After 12 iterations, the disc image configuration was pulled back to the unloaded configuration, and the shape error (represented by the infinite normal, $||\boldsymbol{X}_I - \boldsymbol{X}_P||_\infty$) between the prestrained configuration and the image configuration was 0.138 mm, within the preset tolerance of 0.2 mm. By calibration of the NP CTE to 0.038 /K, the NP of the prestrained disc is hydrostatically pressurized with an average IDP of 0.1 MPa (Figure 3.12), in agreement with the *in-vivo* measurement of the normal disc in the lying supine position [91].

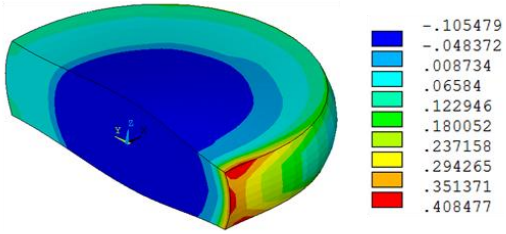


Figure 3.12: The simulated hydrostatic pressure distribution [MPa] in the XP disc.

Compared to the unloaded disc, the average disc height increases by 0.8 mm and the average disc bulge decreases by 0.4 mm after nucleus swelling (Figure 3.13). Moreover, the average criss-cross fiber angle within the outermost fiber lamellae in the image configuration (Figure 3.14a) is larger than that in the unloaded configuration (Figure 3.14b) by 1.7 °.

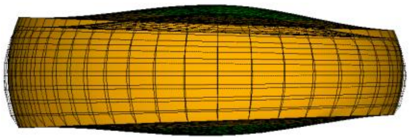


Figure 3.13: Comparison of the prestrained disc geometry (solid lines) with the unloaded disc geometry (dash lines).

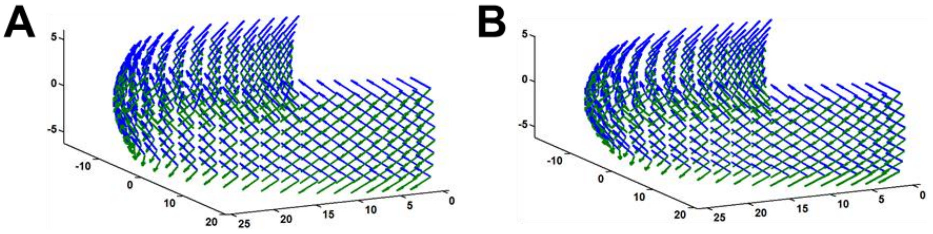


Figure 3.14: Changes in the fiber orientation in the outermost AF lamella due to nucleus swelling. (a) The image configuration; (b) The unloaded configuration.

To validate the proposed nucleus swelling simulation method using thermoelastic analysis (Figure 3.15a), an IDP of 0.1 MPa was directly applied on the AF and endplate internal surfaces after removing the NP (Figure 3.15b) according to a previously reported disc prestrain simulation [38, 39]. The same principle stress distributions occur in the AFs by either thermoelastic analysis (Figure 3.15a) or applying the IDP of 0.1 MPa (Figure 3.15b). Hence, it demonstrates that thermal expansion can be repurposed to provide an accurate representation of disc prestrain under nucleus swelling.

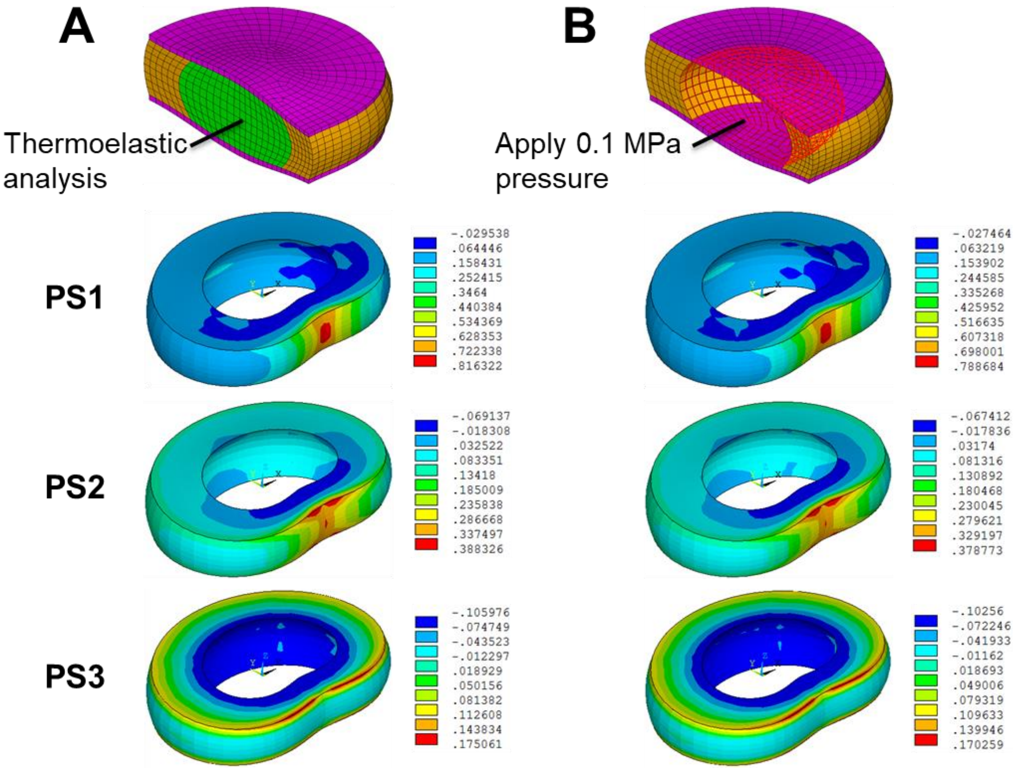


Figure 3.15: The simulated distribution of three principal stresses (PS1, PS2 and PS3, MPa). (a) Nucleus swelling was simulated using thermoelastic analysis; (b) A pressure of 0.1 MPa was directly applied on the AF and endplate internal surfaces after removing the NP.

3.4.3. Discussion and Conclusion

According to our simulation results, the disc geometry and fiber orientation do not change obviously after nucleus swelling. It matches the Pfirrmann grades, in which mildly disc degeneration causes the loss of the IDP and water content [47]. By comparing with the previously reported disc prestrain method, we demonstrate that thermal expansion can provide an accurate representation of disc prestrain in nucleus swelling. This ensures that the entire disc including the AF and NP can be pulled back to the unloaded disc which experiences mild disc degeneration. Therefore, it enables the investigation of the IDP effect on the disc behavior, as described in the subsequent section.

To develop the FE computational model, the model assumes that the annular fibers are at a stretch ratio of 1 when the disc is in its “pull-back” configuration which is considered as the reference configuration. However, the actual distribution of stretch ratios is not uniform, since some tissues could be slacker or more taught in the unloaded configuration. In addition, the computational cost of the iterative “pull-back” computation and calibration of the coefficient of thermal expansion to meet the physiological IDP results in a hurdle to implementing them in multi-segment spine FE models.

3.5. Mechanical Responses of Normal Disc and Degenerated Disc

3.5.1. Methods

The behaviors of the normal (prestrained) disc and degenerated (unloaded) disc were computationally investigated by applying the reported physiological ranges of motions (ROMs) in different loading scenarios [37, 78]. In both disc FE models, the inferior endplate was fully constrained, and the superior endplate was applied ±5 °in sagittal bending, ±3 °lateral bending and ±2.5 °axial torsion [37], as illustrated in Figure 3.16. In each loading scenario, the rest five displacement or rotation degrees of freedom of the superior endplate were allowed [78, 96]. The simulated moment reactions were compared with previously reported *in-vitro* measurements [78].

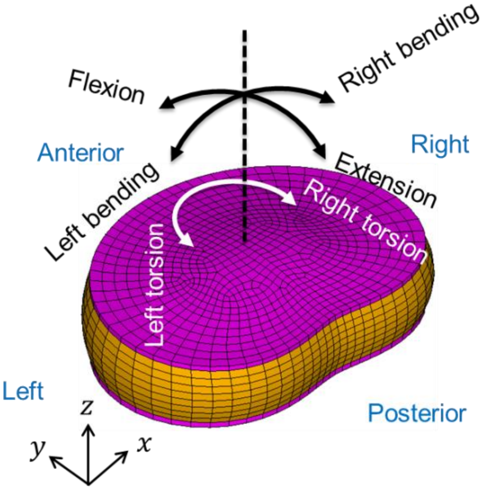


Figure 3.16: The loading scenarios in which the disc responses were simulated.

3.5.2. *Results*

As shown in the Figure 3.17, the simulated behaviors of the normal disc ($\boldsymbol{X_P}$) with the fiber orientation of $|\varphi_{tune}| = |\varphi_{exp}| + 10°$ were in good agreement with the *in-vitro* experimental measurements in different loading scenarios [78].

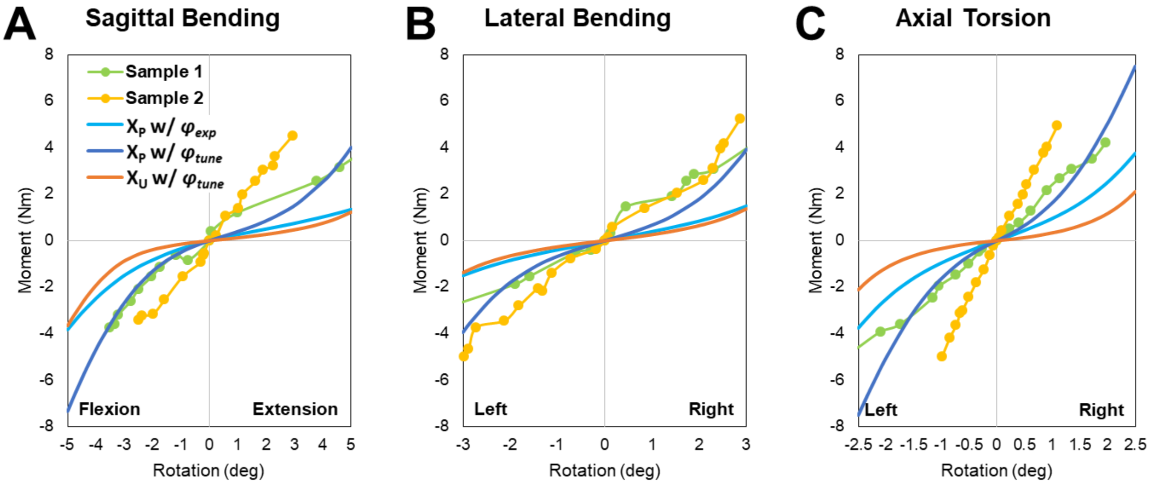


Figure 3.17: Comparison of the simulated rotation-moment curves of the normal vs. degenerated L2-L3 discs with the *in-vitro* measurements in different loading scenarios [78].

The resistances to intervertebral rotations are distinctly compromised in all loading scenarios, when the fiber angles are decreased to $|\varphi_{exp}|$. Disc degeneration ($\boldsymbol{X_U}$ with $|\varphi_{tune}|$) also resulted in a substantial decrease in the disc stiffness to resist intervertebral rotations (Figure 3.17a-c).

The maximum stresses of the normal disc ($\boldsymbol{X_P}$ with $|\varphi_{tune}|$) in flexion (11 MPa, Figure 3.18a) and torsion (4.94 MPa, Figure 3.18d) were located at the dorsal-outer AF lamellae, indicating the location with a high risk of AF injury. Compared with the stress distribution of the normal disc (Figure 3.18), the degenerated disc ($\boldsymbol{X_U}$ with $|\varphi_{tune}|$) was less stressed in all loading scenarios, as shown in Figure 3.19.

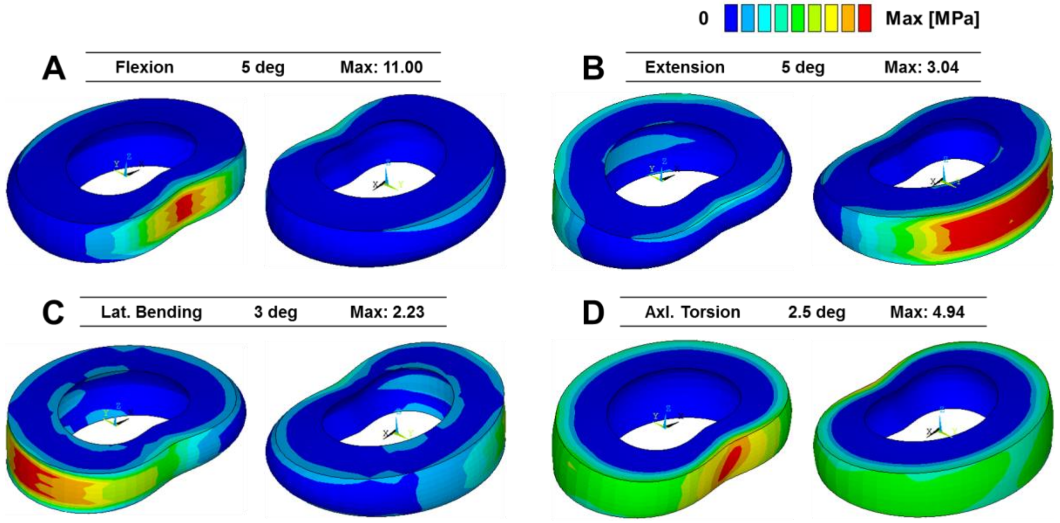


Figure 3.18: The simulated von Mises distributions in the AF of the normal disc in different loading scenarios, where the physiological ROMs were applied on the disc FE model.

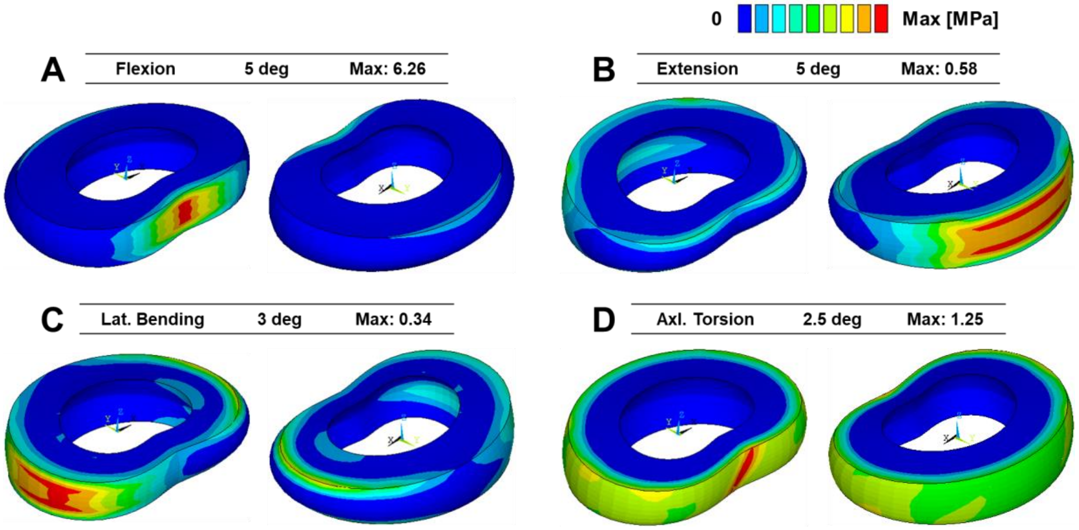


Figure 3.19: The simulated von Mises distributions in the AF of the degenerated disc in different loading scenarios, where the physiological ROMs were applied on the disc FE model.

3.5.3. Discussion and Conclusion

Using a well-established discrete modeling technique (Appendix B.4) for composite material [97] available in ANSYS 15.0, we validated the derived mathematical formulations used to develop the USERMAT subroutine. However, patch test results (Appendix B.4) show that fibers with angles less than 25 °(corresponding to $|\varphi_{exp}| = 23.2°$ at the AF anterior) cannot reinforce an isolated AF specimen in tension along the disc craniocaudal axis, regardless of the material properties (high vs. low fiber stiffness) and model techniques (discrete vs. continuum methods) used for the AF GS and fibers. It indicates that new mechanisms need to be introduced to the AF composite constitutive model to improve model fidelity. Here, we tuned fiber angles to $|\varphi_{tune}|$, 10 °larger than the experimentally measured fiber orientation ($|\varphi_{exp}|$), to match the disc behavior in experimental measurements [78].

Our simulation results show that the disc stiffness is regulated by the IDP and fiber orientation; a decrease of the IDP and fiber angles causes a distinct deterioration in the resistance to intervertebral rotations. Therefore, calibration of the IDP and fiber orientation can improve the model prediction of disc behavior. In the literature, the technique that calibrates the regional variation in annular fiber orientation to match the simulated responses of a disc FE model to *in-vitro* experimental measurements has been widely implemented [31–34].

Chapter 4: Development and Calibration of a Lumbar Spine Finite Element Model

The development and calibration of a representative multi-segment L1-L5 lumbar spine finite element (FE) model were described in this chapter. This model will be used to investigate TDA biomechanics while functional spinal structures are considered. Though a comprehensive literature survey, we have identified common complications in patients who received the treatment of TDA, including abnormal kinematics, polyethylene wear/fracture, and facet joint overloading. The previous continuum disc FE model and the implementation of the "pull-back" algorithm and nucleus swelling simulation enable providing a more accurate prediction for stress. However, the high computational cost and model complexity limit their applications in the multi-segment spine FE model. Instead, a simpler discrete disc model using reinforcement elements available in commercial software was adopted, since disc stress was not a focus in the studies related to TDA performance assessment. Besides, we still leveraged the fiber constitutive model (user material subroutine for reinforcing elements) to simulate the nonlinear tensile behavior, as well as the fitting and interpolation algorithms to model the heterogeneous material properties across the annulus fibrosus. Furthermore, material properties of all spinal soft tissues (disc and ligaments) were calibrated using a stepwise reduction procedure, and the effect of muscle co-activation was considered by implementing a well-established follower preload technique.

4.1. Geometric Modeling

In the study, a representative lumbar spine finite element (FE) model was required to simulate the normal spine mechanical behavior. Due to the relatively simple anatomy, a representative lumbar spine was parametrically modeled in PTC Creo 2.0, by incorporating well-documented average anatomical dimensions, as shown in Figure 4.1. Compared to previously lumbar spine parametric models [55, 98, 99], more features have been introduced to the current model. The modeling technique was briefly described here.

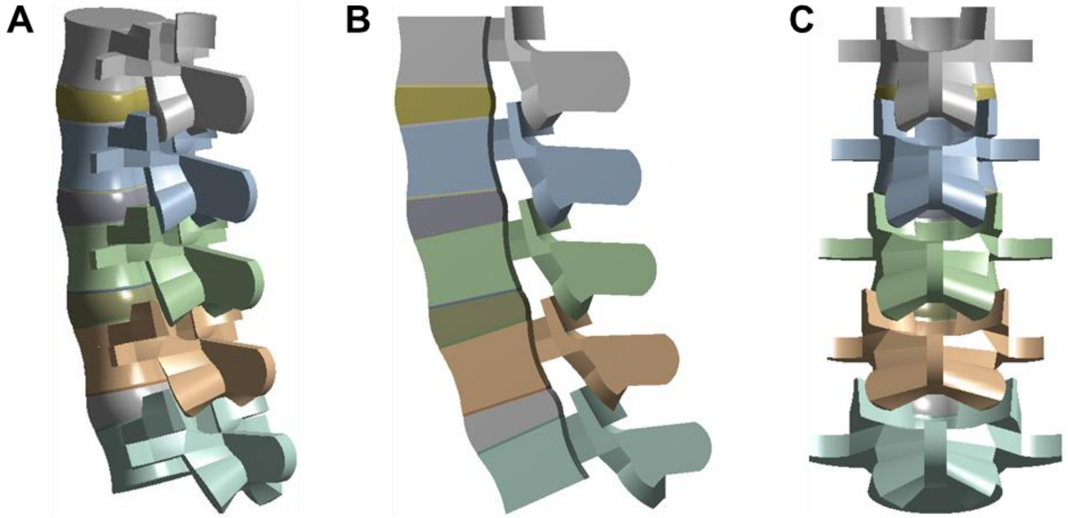


Figure 4.1: Parametric models of L1-L5 vertebrae. (a) 3D quarter view; (b) sagittal view; (c) posterior view.

The kidney-shaped profile of the vertebral body in the transverse plane was modeled according to parametric equations proposed by Little et al. (2007). As shown in Figure 4.2, the profile of the vertebral endplate consists of 3 elliptical arcs ($\widehat{aa}'$, $\widehat{ab}$ and $\widehat{a'b}'$) and 3 circular arcs ($\widehat{bc}$, $\widehat{b'c}'$ and $\widehat{c'c}'$). The periphery walls of vertebral bodies were modeled by sweeping a concave curve with a depth of 1mm across the middle transverse plane of each vertebral body. Previously reported geometric data, including vertebral anatomic dimensions [14] and articular facet dimensions and orientations [15], with a variation from L1 to L5 were incorporated, as listed in Tables 4.1 and 4.2, respectively. These dimensions together with nomenclatures were illustrated in Figure 4.3 and 4.4, respectively.

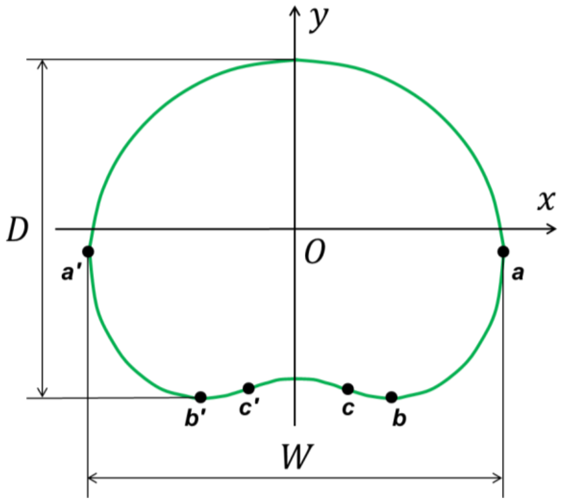


Figure 4.2: The profile of the vertebral endplate which was modeled according to parametric equations proposed by Little et al. (2007).

Table 4.1: The anatomic dimensions (mm) of L1-L5 lumbar vertebrae [14].

Endplate	L1	L2	L3	L4	L5
Endplate width upper (EPW_u)	41.2	42.6	44.1	46.6	47.3
Endplate depth upper (EPD_u)	34.1	34.6	35.2	35.5	34.7
Endplate width lower (EPW_l)	43.3	45.5	48.0	49.5	49.4
Endplate depth lower (EPD_l)	35.3	34.9	34.8	33.9	33.2
Vertebral body height posterior (VBH_p)	23.8	24.3	23.8	24.1	22.9
Spinal canal	L1	L2	L3	L4	L5
Spinal canal width (SCW)	23.7	23.8	24.3	25.4	27.1
Spinal canal depth (SCD)	19.0	18.2	17.5	18.6	19.7
Pedicle	L1	L2	L3	L4	L5
Pedicle width right (PDW_r)	8.0	7.8	10.2	13.4	18.0
Pedicle height right (PDH_r)	15.9	15.0	14.2	15.7	19.6
Pedicle inclination sagittal and right (PDI_{sr})	16.5	17.1	19.8	18.4	25.9
Process	L1	L2	L3	L4	L5
Spinous process length (SPL)	67.7	71.7	71.7	70.1	68.3
Transverse process width (TPW)	71.2	76.1	85.7	79.4	92.5

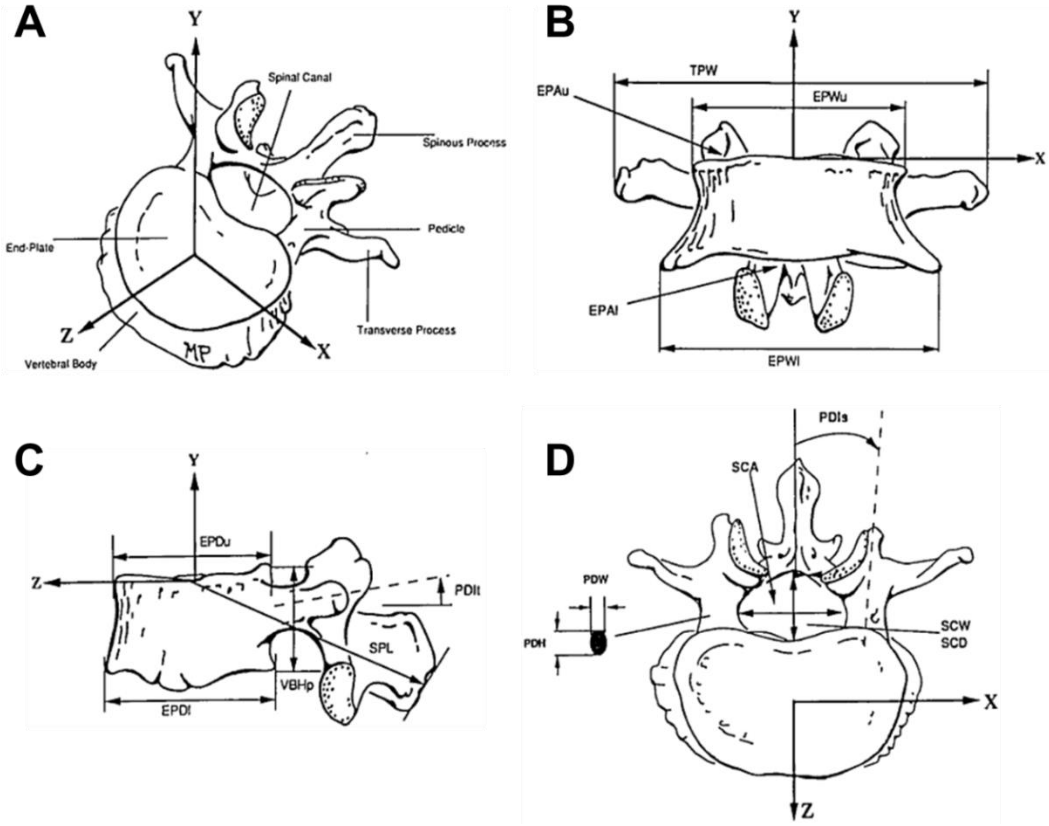


Figure 4.3: Illustration of vertebral anatomic dimensions [14], reproduced with permission from Wolters Kluwer Health, Inc. (a) Isometric view; (b) Frontal view; (c) Sagittal view; (d) Transverse view.

Table 4.2: The L1-L5 articular facet dimensions (mm) and orientations (°) [15].

Articular facet	L1	L2	L3	L4	L5
Facet width upper and right (FCW_{ur})	12.7	14.6	16.0	16.1	17.4
Facet height upper and right (FCH_{ur})	10.2	11.1	13.8	14.1	16.3
Facet width lower and right (FCW_{lr})	12.4	12.7	13.8	14.7	15.6
Facet height lower and right (FCH_{lr})	15.3	16.0	15.7	16.2	18.4
Interfacet width upper (IFW_u)	26.2	26.4	28.6	31.4	35.0
Interfacet width lower (IFW_l)	24.8	26.6	29.1	34.8	40.6
Interfacet height right (IFH_r)	32.5	33.0	32.4	28.5	26.1
Card angle with respect to the transverse plane upper and right (CAX_{ur})	82.9	85.7	81.9	81.2	86.0
Card angle with respect to the sagittal plane upper and right (CAY_{ur})	39.5	44.4	48.0	58.8	62.3
Card angle with respect to the transverse plane lower and right (CAX_{lr})	81.3	82.8	75.6	70.5	71.0
Card angle with respect to the sagittal plane lower and right (CAY_{lr})	26.6	36.9	47.8	65.8	49.8

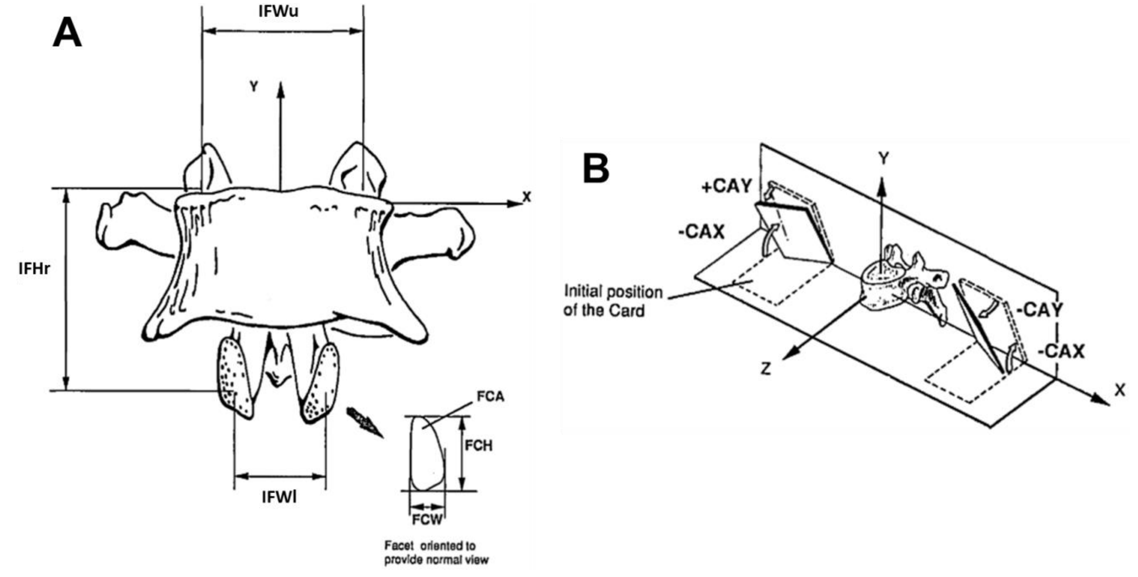


Figure 4.4: Illustration of articular facet linear and angular dimensions [15], reproduced with permission from Wolters Kluwer Health, Inc. (a) Linear dimensions; (b) The facet plane orientations defined by two card angles. It is equivalent to rotating a card initially lying in the transverse plane by an angle (CAX), followed by rotating the card by another angle (CAY).

The same parametric modeling technique [100] was implemented to model the disc profiles, which were required to coincide with the endplate profiles of adjacent vertebral bodies. The nucleus profiles were scaled based on the dimensions of the whole disc, such that the nucleus volume has a ratio of 44% to the total disc volume [33], with its center posterior 2 mm to the disc centroid. It was assumed that the annulus has an outer bulge of 1 mm and an inner bulge of 2 mm [39]. For simplicity, through the lumbar spine, the average disc heights and disc inclination angles at the sagittal plane were set as 13 mm and 6°, respectively [66].

4.2. Finite Element Model

4.2.1. Mesh Preprocessing

The geometric model of the L1-L5 lumbar spine was meshed in ANSYS 15.0 ICEM, where a multi-zone meshing technique was implemented. The mesh density was taken from a fully validated lumbar spine FE model [66]. As presented in Figure 4.5, the spine FE model consists of 9123 brick elements, 3341 shell elements and 11632 nodes. Subsequently, the FE model was further pre-processed when it was imported to ANSYS 15.0 Mechanical APDL.

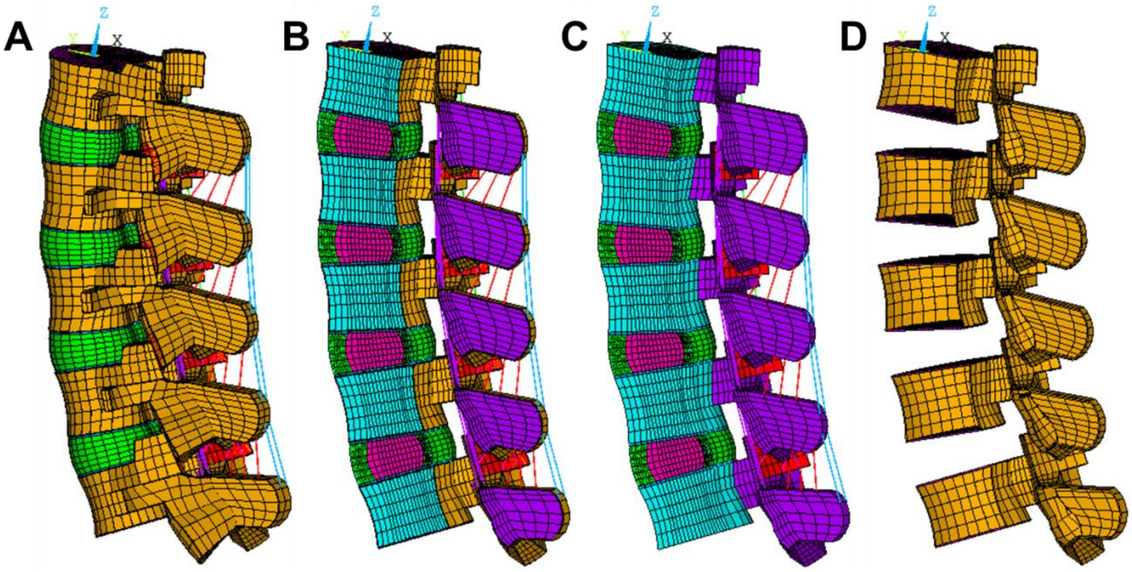


Figure 4.5: The ligamentous L1-L5 spine finite element model (a) and its right half model (b). The cancellous bone of the vertebral bodies and arches (c) is overlaid with shell elements (d), which represents cortical shells and bony endplates.

4.2.1. Modeling of Facet Joints

The articulating facet joint surfaces were modeled using surface-to-surface frictionless contact elements in combination with the penalty algorithm with normal contact stiffness of 200 N/mm [86, 101]. For the purpose of the validation, the augmented Lagrange contact algorithm was also employed to compare the simulated segmental ROMs, facet joint forces, and ligament forces with those using the penalty algorithm; no difference in these simulation results using either algorithm occur. The initial contact gap of the facet joint was adjusted to a uniform distance of 0.4 mm (Figure 4.6) [101], using the contact adjustment technique (Command: *CNCHECK, MORPH, …*) in ANSYS 17.1 MAPDL.

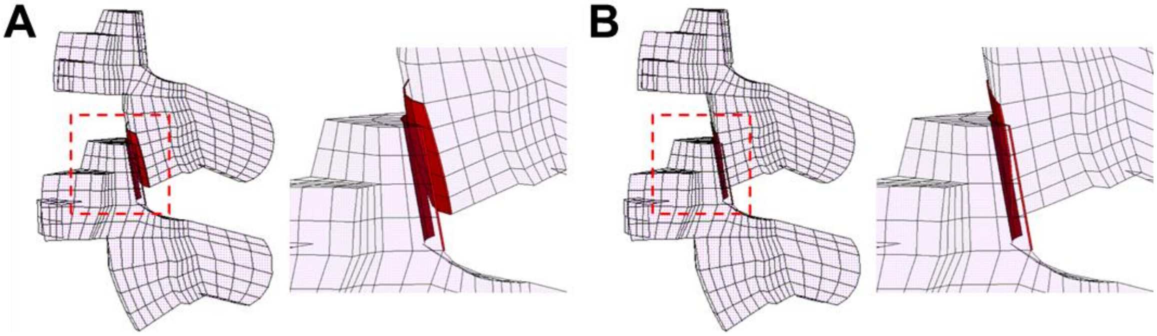


Figure 4.6: Adjustment of the initial contact gap of the facet joint. (a) The facet joint gap before adjustment; (b) The facet joint gap with a uniform distance of 0.4 mm after adjustment.

4.2.2. Modeling of Intervertebral Disc Annular Lamellae

As shown in Figure 4.7, six layers of fiber lamellae in a crossing pattern embedded within the annulus were modeled using reinforcing elements (REINF265) in ANSYS 15.0. It was assumed that adjacent layers were coincident, but with alternating fiber orientations. A new constitutive model was proposed to simulate the tension-only, nonlinear behavior of the fiber lamellae. The strain energy function of the fiber lamellae was formulated as:

$$\Psi_{fib}^{AF}(\lambda) = \begin{cases} 0, & \lambda < 1 \\ c_3\left\{\frac{1}{c_4}\left[e^{c_4(\lambda-1)} - 1\right] - (\lambda - 1)\right\}, & 1 \leq \lambda < c_5 \\ \frac{E^*}{2}(\lambda - c_5)^2 + T^*(\lambda - c_5) + \Psi^*, & \lambda \geq c_5 \end{cases} \tag{4.1}$$

where the strecth ratio (λ) was introduced to descibe the change in the fiber length along the spatial fiber direction. Especially, $\lambda^* = c_5$ is defined as the uncrimped stretch of the AF collagen fiber [94]. It should be noted that there are only 3 material constants (c_3, c_4 and $\lambda^* = c_5$) in Equation 4.1, different from previously reported constitutive models for collagen fibers [29, 94, 95]. The new model facilitates fitting and interpolation of fiber uniaxial tensile test data to simulate the material heterogeneity across the AF. An ANSYS USERMAT subroutine was developed to implement the new constitutive relationship for collagen fibers.

Moreover, annular lamellae were divided into 3 polar sections across the annulus, to which circumferentially varying lamellar thicknesses (*Sec* 1: 0.73 mm, *Sec* 2: 0.56 mm, *Sec* 3: 0.39 mm, Figure 4.7) were assigned, according to previously reported *in-vitro* measurements [19].

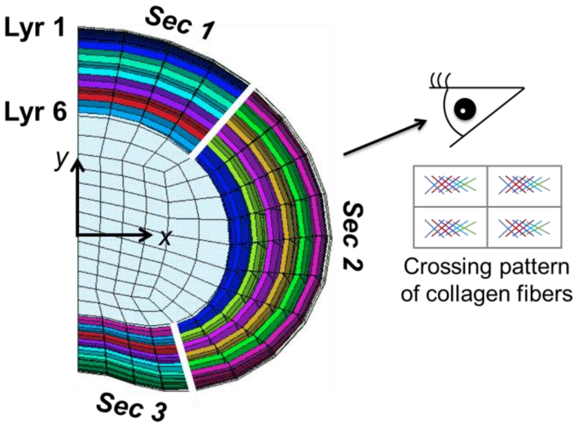


Figure 4.7: The reinforcing elements embedded within the brick elements of the annulus ground substance. They were divided into 3 polar sections, such that heterogeneous tensile properties were assigned. Different element colors indicate different tensile properties assigned. The inset shows the crossing pattern of collagen fibers viewed from the lateral.

The same fitting and interpolation techniques described in Chapter 3 were implemented to model radially and tangentially varying nonlinear tensile properties of the annular lamellae. The resulting fiber tensile properties (Figure 4.8) corresponding to different layers and polar sections (Figure 4.7) were assigned to the annular lamellae in the lumbar spine model. As shown in the inset of Figure 4.7, fiber orientations in adjacent lamellae alternate, causing a crossing pattern. The fiber angles within ventro-external (*Sec* 1, *Lyr* 1), ventro-internal (*Sec* 1, *Lyr* 6), dorsal-external (*Sec* 3, *Lyr* 1) and dorsal-internal (*Sec* 3, *Lyr* 6) annular lamellae were calibrated; fiber angles within other annular lamellae were linearly interpolated.

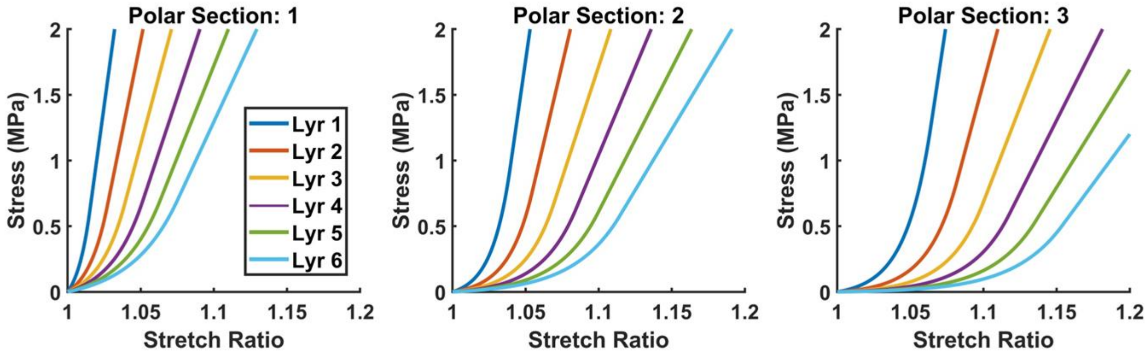


Figure 4.8: The radially and tangentially varying tensile properties of annular lamellae located at different layers and polar sections.

4.2.3. Modeling of Spinal Ligaments

As shown in Figure 4.9, there are 7 intersegmental ligaments in the lumbar spine, including ALL, PLL, CL, LF, ISL, SSL, and ITL. For most spinal ligaments, they are not capable of resisting compression. Thus, the ligament tensile force (F_{lig}) is assumed to be a tension-only, exponential function of the ligament deflection (D_{lig}) [48],

$$F_{lig} = \begin{cases} a\left(e^{bD_{lig}} - 1\right), & D \geq 0 \\ 0, & D < 0 \end{cases} \tag{4.2}$$

where a and b are two material constants which were determined by calibration. The resulting ligament force-deflection curve was assigned to nonlinear spring elements (COMBIN39) in ANSYS 15.0, which were used to be modeled each spinal ligament [33, 34].

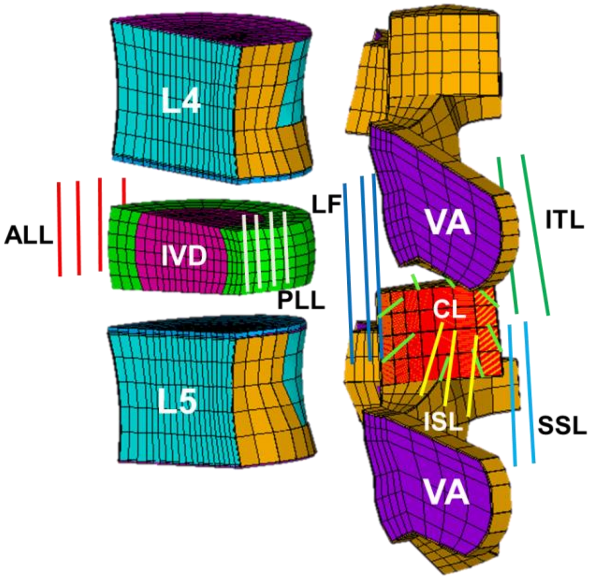


Figure 4.9: The FE model representations of seven intersegmental spinal ligaments.

According to the reported *in-vitro* experiment [28], the segmental rotation in extension significantly increased (Wilcoxon rank test) after removing SSL, indicating that it can resist compression. Therefore, it was assumed that the SSL can resist in compression by the same force-deflection curve as in tension [33, 34].

4.2.3. Material Properties

To model the material orthogonality of the cancellous bone in vertebral bodies, five local coordinate systems were established at the superior endplate centroids of each vertebral body. The z directions of local coordinate systems represent the craneo-caudal direction along the lordotic curve of the lumbar spine. The vertebral bodies and neural arches (Figure 4.5c) are covered by a cortical shell and bony endplates (Figure 4.5d), both of which were modeled by quadrilateral shell elements with the thicknesses of 0.6 mm [102] and 0.5 mm [103], respectively. The cartilaginous endplates and facet cartilages were explicitly modeled using brick elements, assuming that their thicknesses are 1 mm [102] and 0.25 mm [104], respectively. Cortical shells, bony endplates, cartilaginous endplates and facet cartilages were modeled using the isotropic elastic material properties. The disc nucleus pulposus and annulus fibrosus ground substance were simulated using Mooney-Rivlin hyperelastic material properties. The material properties [25] of each spinal tissue in the FE model were summarized in Table 4.3.

Table 4.3: A summary of the material properties adopted in the lumbar spine FE model.

Spinal Tissues	Material Properties
Vertebral Body Cancellous Bone	E_{xx} = E_{yy} = 140 MPa, E_{zz} = 200 MPa G_{xy} = G_{yz} = G_{xz} = 48.3 MPa υ_{xy} = 0.45, υ_{yz} = υ_{xz} = 0.315
Neural Arch Cancellous Bone	E = 3500 MPa, υ = 0.25
Cortical Bone	E = 12000 MPa, υ = 0.3
Bony Endplate	E = 10000 MPa, υ = 0.3
Cartilaginous Endplate	E = 24 MPa, υ = 0.4
Facet Cartilage	E = 35 MPa, υ = 0.4
Nucleus Pulposus	Mooney-Rivlin: c_1 = 0.12 MPa, c_2 = 0.03 MPa
Annulus Fibrosus Ground Substance	Mooney-Rivlin: c_1 = 0.18 MPa, c_2 = 0.045 MPa
Annular Lamellae	Tension-only, nonlinear stress-stretch relation (ANSYS USERMAT subroutine)
Spinal Ligaments	Tension-only, nonlinear force-deflection relation (ANSYS nonlinear spring elements)

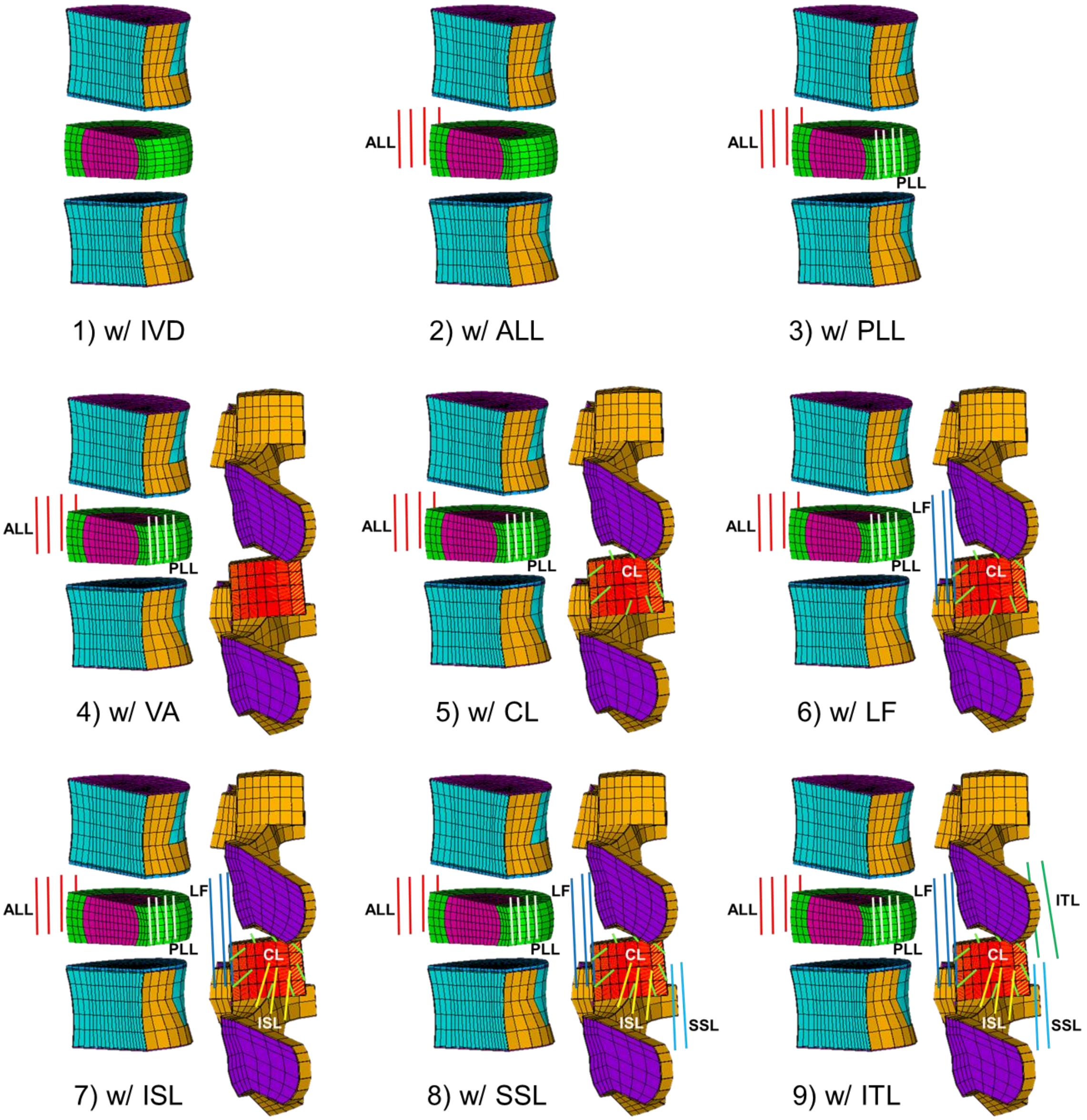
The spine mechanical behavior is significantly influenced by the material properties of soft tissues (intervertebral disc annular lamellae and spinal ligaments). Thus, they need to carefully modeled and calibrated, such that the resulting kinematic responses of either individual tissues or the entire lumbar spine closely matched published *in-vitro* measurements. The calibration procedure was reported in the following.

4.3. Calibration of Spinal Tissue Material Properties

4.3.1. Methods

Heuer et al. (2007a) quantified the mechanical function of each spinal tissue on lumbar segments in different loading scenarios and magnitudes, by step-wise reducing functional spine structures. In their *in-vitro* testing, eight specimens of lumbar motion segments (L4-5) with a median age of 52 years were employed. Spinal structures (ligaments, facet capsules, facet joints, vertebral arch and disc nucleus) were successively reduced. Specimens in intact and defect stages were mounted on a custom-built spine testing machine, and applied pure moments ramping up to 10 N in various physiologically relevant loading scenarios. Changes in the kinematic responses (range of motion and lordotic angle) of lumbar segments due to the step-wise reduction procedure were evaluated.

To calibrate material properties of each spinal tissue, the kinematic responses of the L4-L5 lumbar segment FE model in each reduced stages were simulated using a step-wise addition procedure of spinal structures, which was exactly opposite to the produce in the *in-vitro* experiment [28]. The step-wise addition procedure has been implemented to calibrate material properties of spinal tissues in previously reported computational models [16, 33, 34]. As shown in Figure 4.10, eight segment models in defect stages and one intact segment model (w/ ITL) were created as a result of the successive addition of spinal structures.



1) w/ IVD 2) w/ ALL 3) w/ PLL

4) w/ VA 5) w/ CL 6) w/ LF

7) w/ ISL 8) w/ SSL 9) w/ ITL

Figure 4.10: Defected/intact states of the L4-L5 lumbar segment model in the step-wise addition procedure (w/ = with).

The same loading conditions described in the *in-vitro* testing [28, 96] was used to simulate the kinematic responses of the defected/intact L4-L5 lumbar segment in flexion, extension, lateral bending and axial torsion, as illustrated in Figure 4.11. The L5 inferior endplate was fully constrained, and the L4 superior endplate was applied a pure moment of 10 Nm using a multi-point constraint technique in ANSYS 15.0. In each model (Figure 4.10), the material properties of only one spinal tissue, which was introduced compared to the previous stage, were calibrated using the optimization toolbox in MATALB R2014a, such that the simulated ranges of motion (ROMs) of the L4-5 segment matched in-vitro measurements in the 4 loading scenarios [28]. Therefore, the calibration order is the IVD, ALL, PLL, VA (where the initial gap of facet joint was tuned to 0.4 mm), CL, LF, ISL, SSL, and ITL (the ITL tensile properties were chosen from Rohlmann et al. 2006, due to the lack of experimental data in the *in-vitro* experiment of Heuer et al. 2007a).

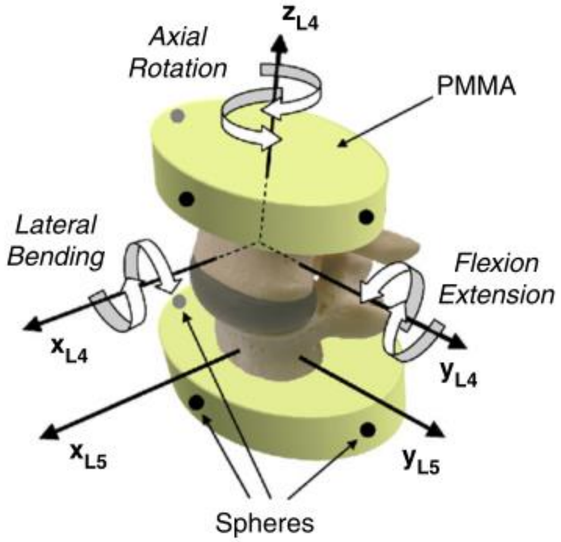


Figure 4.11: Illustration of the *in-vitro* loading conditions on a custom-built tester used to test the kinematic responses of the L4-L5 lumbar segment in different loading scenarios [42], reproduced with permission from Elsevier.

4.3.2. Results

After calibration of the annular fiber orientation and spinal ligament tensile properties, the simulated kinematic responses of the defected/intact L4-L5 segments are in good agreement with the *in-vitro* measurements [28] in the 4 physiologically relevant loading scenarios, as shown in Figure 4.12.

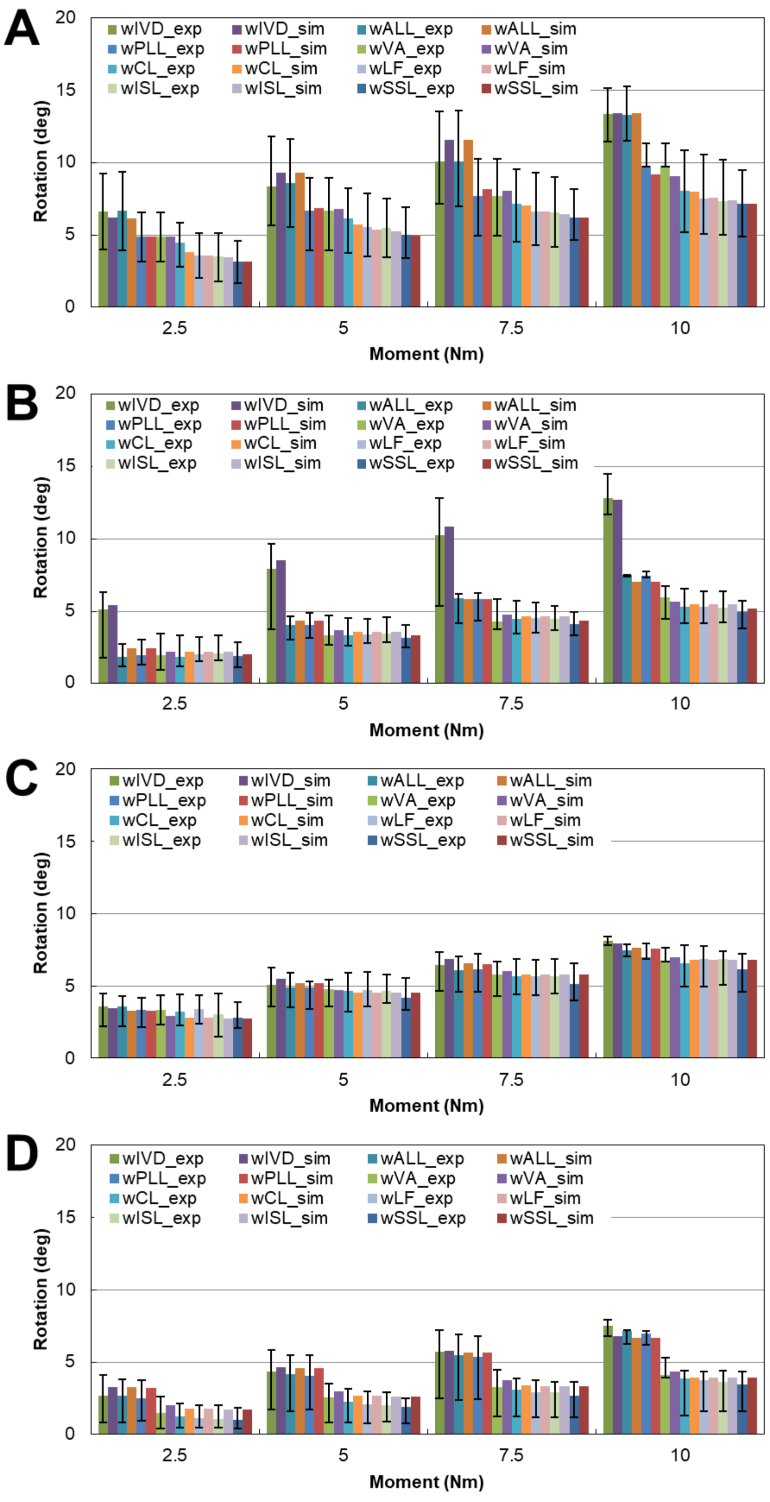
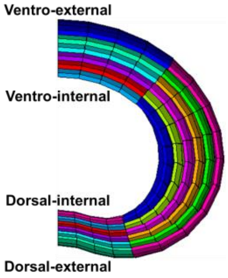


Figure 4.12: Comparison of the simulated kinematic responses of the defected/intact L4-L5 segment with those of *in-vitro* measurements [28] in flexion (a), extension (b), lateral bending (c) and axial torsion (d).

The calibrated fiber angles in different anatomic sites in the L4-L5 segment are listed in Table 4.4, where the *in-vitro* measurements of fiber angles reported in the literature are also presented.

Table 4.4: Comparion of the calibrated fiber angles at 4 anatomic sites with the reported *in-vitro* measurements.

	**Fiber Angle, φ (°)**			
	Calibrated	Tangentially varying	Radially varying	Fully anisotropic
	Heuer et al. (2007a)	Holzapfel et al. (2005)	Cassidy et al. (1989)	Zhu et al. (2008)
Ventro-external	30	24	28	25
Ventro-internal	58	24	45	54
Dorsal-external	44	46	28	90
Dorsal-internal	66	46	45	70
Donor age [years]	38 ~ 59	57.9 ± 15.4	31 ~ 80	28 ~ 45
IVD level	L4-5	L1-2	L1-2, L2-3, L4-5	L4-5, L5-S1

The calibrated tensile properties of spinal ligaments in the L4-L5 segment are presented in Figure 4.13. Generally, the stiffnesses of the anterior ligaments are larger than those of posterior ligaments.

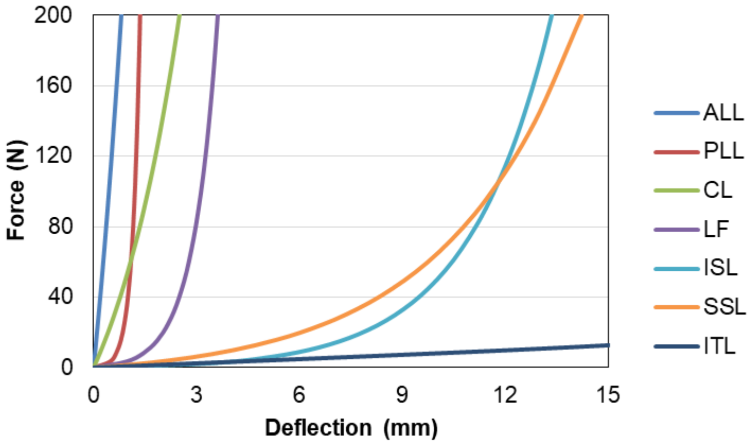


Figure 4.13: The calibrated tensile properties of spinal ligaments.

4.3.2. Discussion and Conclusion

Due to the different lumbar disc specimens used in *in-vitro* experiments (Table 4.4), it was observed that the phenotypes of *in-vitro* fiber orientation would be radially varied [22], circumferentially varied [19, 105], and both radially and circumferentially varied [20]. These experiment results revealed a high variability in the annular fiber orientation, which might result from the individual-specific adaptation in tissue remodeling to loading history and physiological aging (Brickley-Parsons and Glimcher. 1984). As we expected, the calibrated fiber orientation, which varies both radially and circumferentially, cannot exactly match any individual dataset of *in-vitro* experimental results listed in Table 4.4, since the *in-vitro* kinematic responses to which we calibrated were measured by Heuer et al. (2007a), who used eight spinal segments (L4-L5) with a median age of 52 years. Regrettably, they did not report the fiber orientation in their specimens, thus we cannot validate our calibrated fiber orientation and the fidelity in our FE model.

The trend in the variation of ligament stiffnesses match that reported by Schmidt et al. (2007a), who used the same *in-vitro* measurements of the segmental kinematic responses [28] to calibrate the ligament tensile properties. However, our calibrated ligament stiffnesses of the ISL and SSL are much smaller than their calibrated results. It should be noted that we calibrated the tissue material properties using 4 physiologically relevant loading scenarios, but they only simulated the kinematic responses of the defected/intact L4-L5 segment in flexion and extension.

4.4. Modeling of Compressive Follower Preload

The follower load is defined as a compressive load whose path approximates the tangent to the lordotic curve of the lumbar spine, passing through the instantaneous center of rotation of each segment. Patwardhan et al. (1999) have demonstrated that the follower load minimizes the segmental bending moments and shear forces, and significantly increases the load-carrying capacity. In contrast, the spine buckles under a vertical compressive load with a magnitude far below as seen *in vivo*. Therefore, the follower load is a consequence of the physiological co-activation of trunk muscles that alters the direction of the internal compressive force vector, in order to provide a greater margin of safety against both instability and tissue injury. In this work, we modeled the follower preload according to a previously reported experimental setup [82, 106], and the preload path was optimized to minimize the bending moments in the sagittal plane, as recommended by Dreischarf et al. (2010).

4.4.1. Methods

After performing the stepwise calibration procedure, the resulting material properties were assigned to the corresponding spinal tissues at the L1-2, L2-3 and L3-4 levels of the multi-segment lumbar spine model. According to the *in-vitro* experiment by Patwardhan et al. (1999), a compressive follower load was applied bilaterally by cables and dead weights, as shown in Figure 4.14a. The loading cables were firmly anchored to the cup holding the L1 vertebra and passed freely through cable guides attached to the vertebral bodies from L2 to L5. Correspondingly, in the L1-L5 spine FE model (Figure 4.14b), tension-only truss elements (LINK180) were used to simulate the load path which was tangent to the spine curve. Each truss element was connected to a pilot node which guides the motion of a vertebra. The nodal kinematic constraints of truss elements at the L2-L5 were defined using a cylindrical joint element (MPC184), which only allows a translation and a rotation degree of freedom along the joint axle.

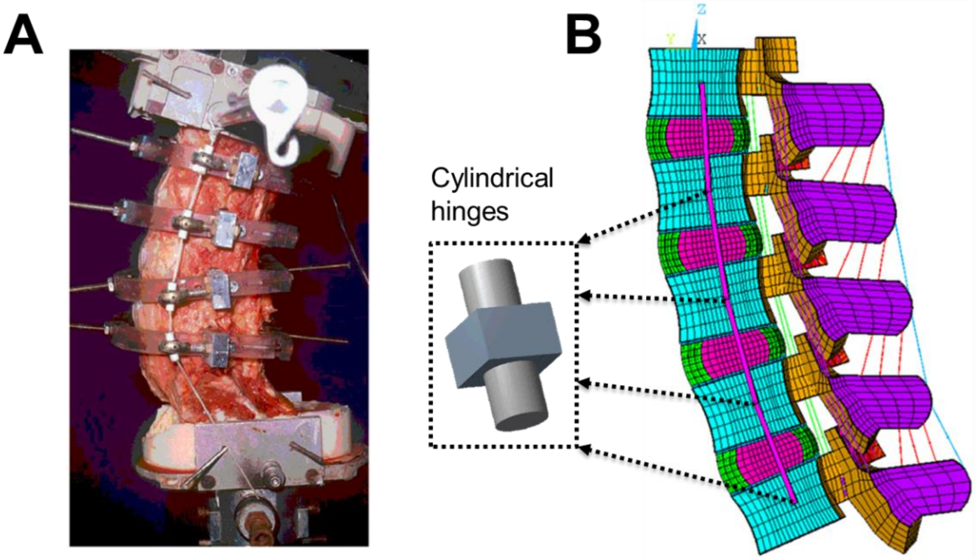


Figure 4.14: Modeling of the compressive follower preload according to the previously reported experimental setup. (a) The experimental setup in which a human cadaveric lumbar spine subjected to a compressive follower load [106], reproduced with permission from Elsevier; (b) The FE model of the lumbar spine undergoing a follower preload which was simulated by tension-only truss elements connected with cylindrical joint kinematic constraints.

For the L1-5 spine FE model, the L5 inferior endplate was fixed, and a symmetric boundary was applied to the sagittal plane. A compressive follower preload of 250 N (500 N for the complete model), which led to the simulated intervertebral rotations and intradiscal pressures that agreed well with *in-vivo* data [108], was applied to the spine FE model using the technique described above. As suggested by Dreischarf et al. (2010), a non-optimized follower preload path may cause change in the lordosis of the lumbar spine. Therefore, the preload path was optimized using the optimization toolbox in MATALB R2014a, such that no intervertebral rotation occurred at each level of the intact spine model [107].

4.4.2. Results

After optimization, the nodal positions of truss elements, which define the preload path, are 1.68 mm, 1.78 mm, 1.91 mm, 2.19 mm and 2.26 mm posterior to the centers of L1-L5 vertebral bodies, respectively. The optimal preload path is consistent with optimized preload paths reported in previous publications [107, 109].

4.4.3. Discussion and Conclusion

In the previous investigation of intervertebral disc biomechanics, the degenerated (unloaded) disc due to the loss of intradiscal pressure (IDP) was derived from the disc geometry and fiber orientation observed in clinic images, and disc prestrain under nucleus swelling was achieved by applying thermal expansion in the nucleus of the degenerated disc. According to our simulation results in various loading scenarios, the normal (prestrained) disc pressurized by a hydrostatic pressure in the nucleus provided a more considerable resistance to intervertebral rotations. However, the normal disc was modeled using an iterative algorithm, so the high computational cost limited its application to a multi-segment spine FE model.

In this study, the normal disc was modeled using stiffer material properties and the calibrated fiber orientation. It was demonstrated that the simulated kinematic responses of the normal disc could closely match in-vitro experimental measurements [28], although the resulting stress distribution might not very accurate. Moreover, material properties of all spinal ligaments in the intact multi-segment spine FE model have been calibrated using a step-wise addition procedure, and the well-established follower preload technique was used to consider the effects of muscle co-activation and body weight. Therefore, the spine FE model has achieved sufficiently high fidelity, as in the next chapter we will show that the simulated segmental ROMs and facet joint forces in each level are consistent with published simulation results. This meets our need to evaluate TDA performance metrics associated with postoperative complications, such as abnormal segmental kinematics and facet joint overloading.

Chapter 5: Performance Assessment of Total Disc Arthroplasty

The intact multi-segment lumbar L1-L5 spine finite element (FE) model was modified to simulate the behavior of a treated spine that was implanted different total disc arthroplasty (TDA) designs. First, we demonstrated that TDA led to substantial changes in the range of motion (ROM) and spinal tissue loads in the treated segment, but no distinct changes were observed at the adjacent levels regardless of different TDA designs. Therefore, the multi-segment spine FE model was further reduced to a single lumbar segment L3-L4 FE model, to investigate the effects of implant design and facet joint (FJ) articulation on the motion patterns and metal-on-polyethylene (MoP) contact mechanics. From the simulation results of commercially available TDA designs, we learned that FJ articulation could change the designed TDA motion patterns, and cause lift-off, impingement and unsymmetric motion in a mobile-bearing TDA design with a biconvex polyethylene (PE) core. The abnormal motion in the biconvex-core TDA further caused a downward bending of the rim with a median thickness of only 2.9 mm, and thus a higher risk of PE fracture existed. These results were in excellent agreement with well-documented retrieval studies in the literature. Therefore, the author proposed a new mobile-bearing TDA design featuring a biconcave PE core, which strengthens the PE rim in order to alleviate stress concentration and permanent deformation; they cause PE fracture due to the repetitive impingement in life activities. Using kinematic analysis and contact simulation, we successfully demonstrate that the new TDA design can improve the MoP contact mechanics and spinal kinemics in the treated lumbar segment.

5.1. Mechanical Behavior of a TDA in Multi-segment Lumbar Spine

Intervertebral disc degeneration induces spinal stenosis and facet joint arthritis, and is highly associated with lower back pain. When more conservative medical therapy fails, surgical procedures need to be performed to replace the diseased discs. Spinal fusion commonly using high-stiffness implants significantly alters native disc biomechanics, causing adjacent segment diseases. Total disc arthroplasty (TDA) is an attractive alternative to spinal fusion for the treatment of spinal pathologies, while the disc height and segmental motion are restored. However, in the literature, there has been no consensus that TDA can treat the implanted level without compromising the adjacent segments. Therefore, finite element (FE) simulations which supplement *in-vivo* and *in-vitro* studies were employed to investigate the mechanical responses of the lumbar spine after single-level TDA with a widely used socket-in-ball artificial disc. We hypothesize that different TDA design parameters influence the segmental range of motion (ROM) and the load-sharing of spinal tissues at both the treated and adjacent levels throughout common lumbar spine motions, extension and flexion.

5.1.1. Methods

The development and calibration procedures of an intact multi-segment L1-L5 lumbar spine FE model have been described in Chapter 4. The intact lumbar spine FE model (Figure 5.1a) was modified to simulate the mechanical responses of the treated lumbar spine (Figure 5.1b) subjected to an anterior surgical procedure [9, 66]. At the L3-L4 level, the anterior longitudinal ligament, the anterior portion of the annulus and the entire nucleus were removed. Only the posterior and lateral portion of the annulus remained in place. A widely used fixed-bearing TDA was inserted into the intervertebral space of the L3-L4 level.

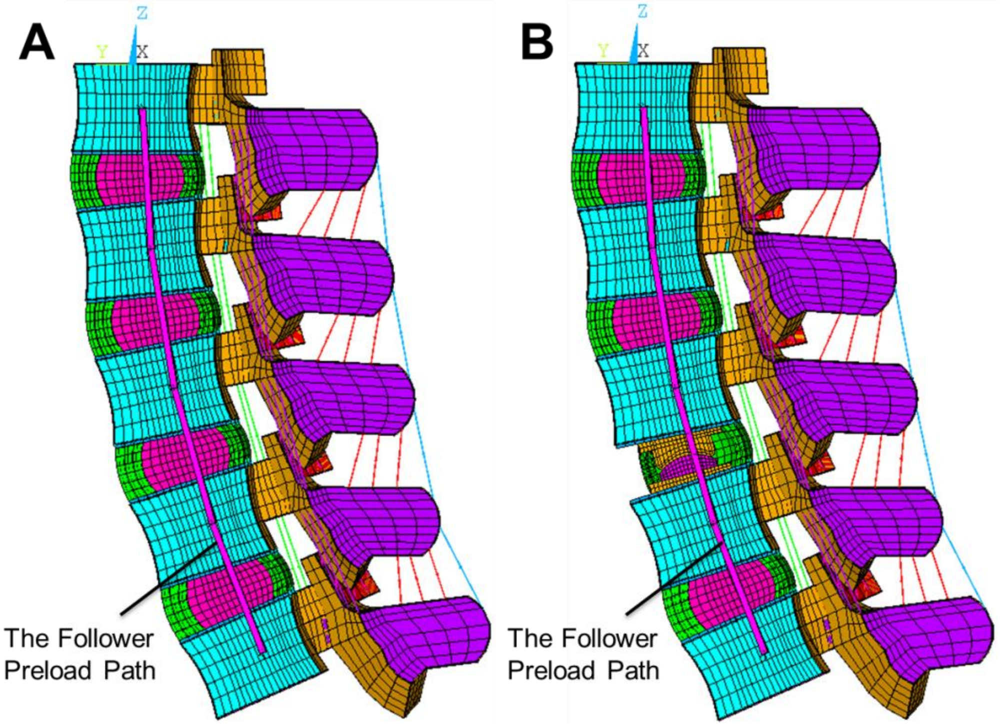


Figure 5.1: Finite element models of the intact L1-L5 lumbar spine (a) and the TDA-treated L1-L5 lumbar spine according to an anterior surgical procedure.

It was assumed that the TDA consists of two cobalt-chrome (CoCr) alloy endplates and an ultra-high-molecular-weight polyethylene (UHMWPE) inlay [55]. The UHMWPE inlay is bonded to the inferior CoCr endplate, and articulates with the superior CoCr endplate through a convex-up bearing surface. As illustrated in Figure 5.2, the TDA design was parametrically modeled considering four design parameters. g is the gap between CoCr-UHMWPE contact interfaces, used to simulate different TDA heights; d is the anteroposterior position of the TDA inferior endplate surface center in the intervertebral space with respect to the geometric center of the L4 superior endplate surface; h and r are the height and radius of the bearing surface of the UHMWPE inlay, respectively. The TDA design with g = 0, d = 0, h = 3 and r = 8 mm was chosen as the baseline design (Figures 5.2 and 5.3).

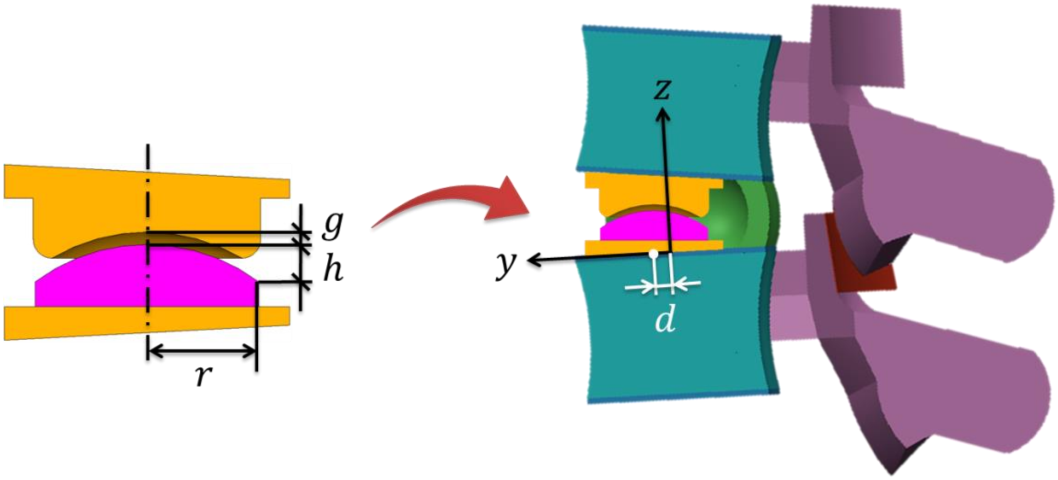


Figure 5.2: The parametric TDA model with four TDA design parameters (g, d, h and r), which was implanted into the L3-L4 level.

In parametric analysis, each design parameter was changed, while other three design parameters were kept the same as the baseline design. It resulted in six additional TDA designs different from the baseline design, as shown in Figure 5.3. For example, for the TDA design denoted by g = 1 mm, it implies other design parameters with d = 0, h = 3 and r = 8 mm. The baseline TDA design and other TDA designs with the change in only one of the design parameters was implanted into the L3-4 level (Figure 5.2). A coefficient of friction of 0.02 [101] was chosen for the material coupling of CoCr-UHMWPE. The Young's modulus and Poisson's ratio of CoCr was assumed to be 300 GPa and 0.27 [101], respectively. The UHMWPE inlay was assigned a Young's modulus of 2 GPa and a Poisson's ratio of 0.3 [101].

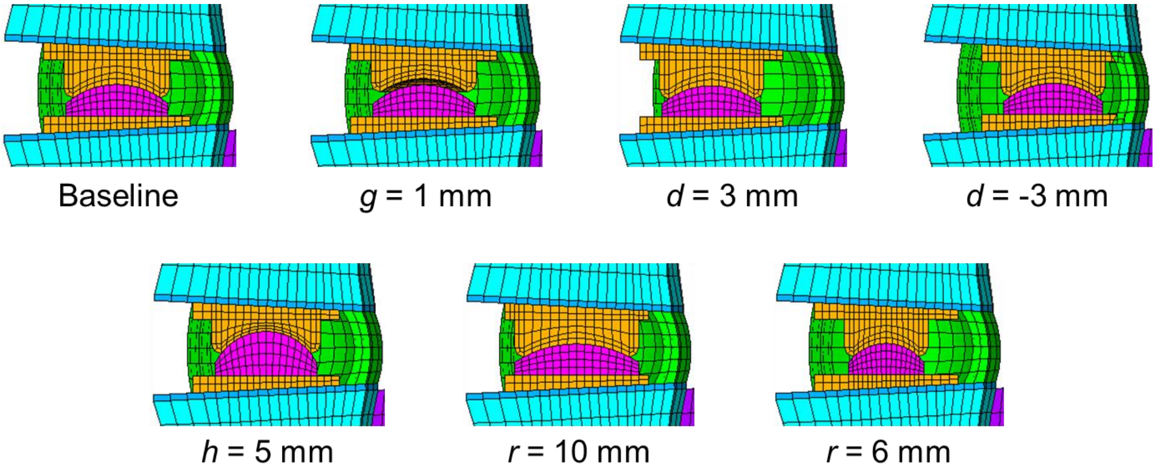


Figure 5.3: Different TDA designs from the baseline design in parametric analysis. The baseline TDA design was assign g = 0, d = 0, h = 3 and r = 8 mm. For each design, only one design parameter was changed, while other design parameters were kept the same as the baseline design.

According to the model symmetry in extension and flexion, the right half of the L1-L5 intact/treated lumbar spine (Figure 5.1a,b) was modeled to investigate mechanical responses of each level. The L5 inferior endplate was fixed, and a symmetry boundary was applied at the sagittal plane (Figure 5.4) to constrain the normal displacement and the out-of-plane rotations. A compressive follower preload of 250 N (500 N for the complete model) was applied to the intact/instrumented spine models using tension-only truss elements with cylindrical joint kinematic constraints, according to the experimental setup of Patwardhan et al. (1999). The preload path was optimized such that no intervertebral rotation occurred at each level of the intact spine model [107]. Subsequently, a moment ramping up to 3.75 Nm (7.5 Nm for the complete model) was applied to the L1 superior endplate of the intact/treated spine model in extension and flexion, respectively. The simulated ROMs, facet joint forces and ligament forces using different TDAs in extension and flexion at 3.75 Nm were compared.

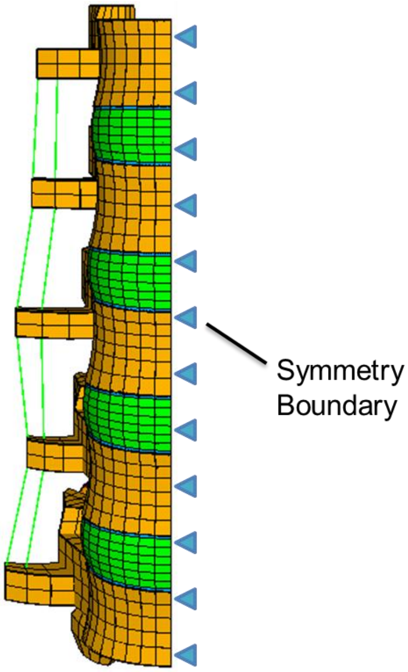


Figure 5.4: A symmetry boundary which is applied at the sagittal plane. It constrains the displacement normal to the sagittal plane, and the rotations out of the sagittal plane.

5.1.2. Results

When a follower preload of 250 N and a moment of 3.75 Nm are applied to the intact spine FE model, the average ROMs of each level are 4.3 ° in extension (Figure 5.5a-*left*) and 6.6 ° in flexion (Figure 5.5b-*left*). The facet joint forces in extension (Figure 5.5a-*middle*) at the L2-3, L3-4 and L4-5 levels are 36.2, 28.7 and 34.5 [N], respectively. However, no contact of facet joints occurs in flexion (Figure 5.5b-*middle*). These results closely match those reported by Schmidt et al. (2012).

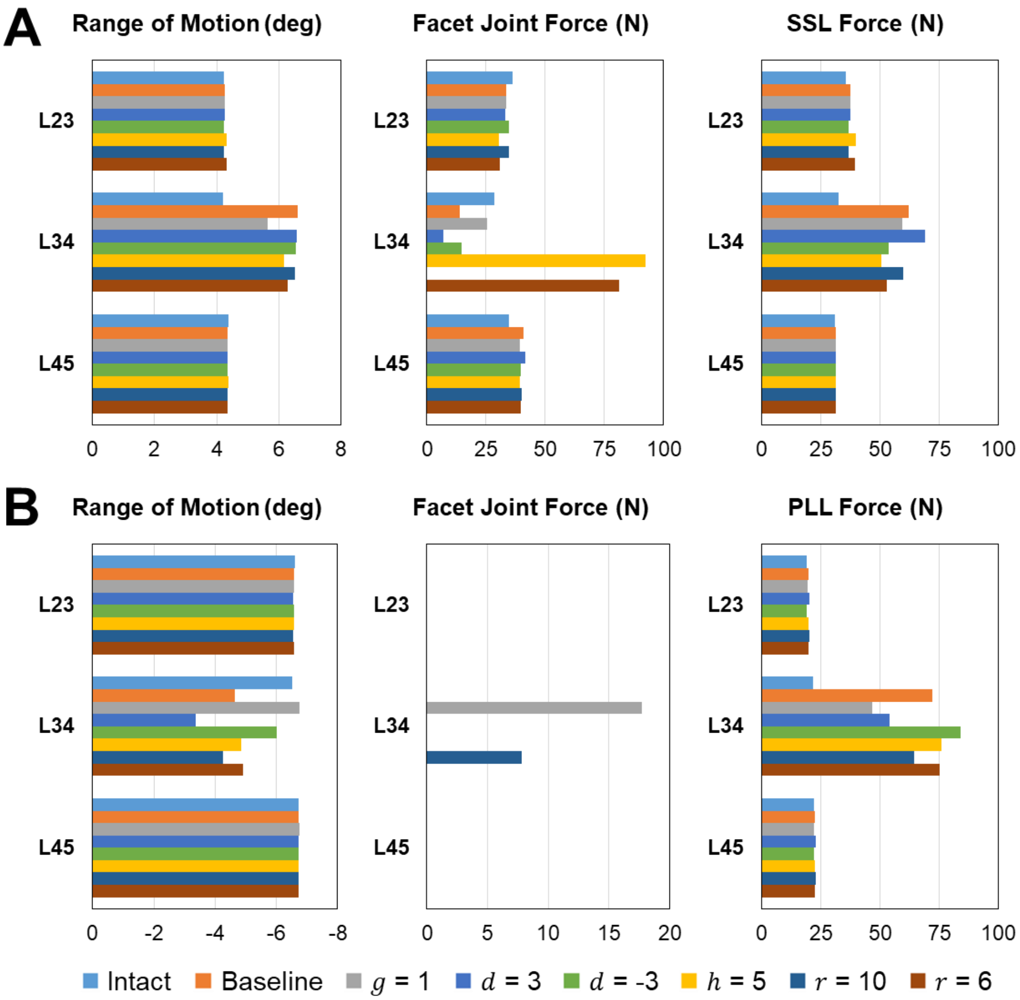


Figure 5.5: The simulated ROMs (left), facet joint forces (middle) and ligament forces (right) of the L1-5 intact and TDA-treated lumbar spines, each of which was modeled according to the symmetry with respect to the sagittal plane. Both models were applied a follower preload of 250 N and a moment of 3.75 Nm in extension (a) and flexion (b), respectively.

For TDA-treated lumbar spines, an obvious realignment in the neutral rotation occurs at the treated level (L3-L4) after applying the follower preload for different TDAs (1.2 °~ 3.6 °in extension angle), except for the TDA with d = -3 mm (0.2 °in flexion angle), which was positioned 3 mm posterior to the endplate center, as recommended by Schmidt et al. (2012). The TDA realignment after preloading influences the ROMs of the treated level (L3-L4) in both extension and flexion. Compared to the responses at the L3-L4 level of the intact lumbar spine, there is a same trend in the simulated ROM at the treated level of each treated lumbar spine that increases in extension (Figure 5.5a-*left*) and decreases in flexion (Figure 5.5b-*left*), although it is sensitive to different TDA designs.

In agreement with the study of Huang et al. (2003), distinct increases in facet joint forces at the treated level in extension (Figure 5.5a-*middle*) are induced by only two TDAs with h = 5 mm (91.7 N) and r = 6 mm (80.5 N), both of which cause a smaller radius of curvature on the CoCr-UHMWPE articulation. In contrast, facet joints at the treated level remain separated during extension, when the TDA with r = 10 mm is used. In flexion (Figure 5.5b-*middle*), facet joint loads only occur in two treated spines using the TDAs with g = 1 mm (18.0 N) and r = 10 mm (8.0 N), respectively. In terms of the load-carrying in ligaments, TDA leads to substantial increases in the SSL compressive force in extension (Figure 5.5a-*right*) and the tensile force of the posterior longitudinal ligament (PLL) in flexion (Figure 5.5b-*right*). It is worth noting that only slight changes in ROMs, facet joint forces, and ligament forces are observed at the adjacent levels (Figure 5.5a,b).

5.1.3. Discussion and Conclusion

Compared to the behavior of the intact lumbar spine, TDA using the anterior surgical procedure which removes spinal tissues causes a larger ROM in extension and changes the load-sharing of spinal tissues at the treated level where a TDA is implanted. Although we only focused on a convex-up fixed-bearing TDA design, it revealed that the ROM and the load-sharing of spinal tissues at the treated level in extension and flexion were very sensitive to TDA design parameters, including the disc height distraction, implant position, and radius of curvature of the metal-on-polyethylene bearing. However, it is noticed that the implantation of a TDA does not influence the mechanical responses of adjacent levels regardless of different TDA design parameters, indicating that the same load can be transferred to each level through a treated lumbar spine. Due to the slight change in the responses of adjacent levels, the multi-segment lumbar spine FE model can be simplified to a single motion segment FE model, to assess different TDA designs in terms of TDA kinematics, spinal tissue loading, and contact mechanics.

Not only the overall response of a spine but also the responses of individual components are required to match experimental measurements, in order to accurately predict spinal tissue load-sharing. Therefore, the well-established step-wise addition procedure was adopted to calibrate material properties of discs and ligaments. Especially, the supraspinous ligament (SSL) was assumed to be able to resist compression in the spine FE model, since Heuer et al. (2007a) reported that the ROMs of the L4-L5 lumbar specimens in extension significantly increased after removing the SSL. If the SSL compression resistance is neglected in our FE model, the simulated ROM in extension will increase beyond the standard deviation of the experimental data. In contrast, García Vacas et al. (2014) predicted that TDA resulted in a larger facet joint force at the treated level, and the capsular ligament was the only load-carrying ligament in extension, since they did not consider the resistance of the SSL to compression in their FE model. Further experimental characterization is essential to investigate whether buckling would occur in the SSL due to excessive compression.

5.2. Influence of Implant Design and Facet Joint Gap on TDA Kinematics and Contact Mechanics

To achieve natural segmental motion, it is desirable to match the range of motion (ROM) and the instantaneous center of rotation (ICR) of a total disc arthroplasty (TDA) to that of the intact segment; they are two important parameters of prosthesis design related to the quantity and quality of motion, respectively [112]. The calculated centrode (the path of ICR) is very sensitive to the nodal displacement of the segment in a computational FE model [42, 113], so adverse TDA motion will significantly change the desired kinematics [101]. For intact lumbar segments, the ICRs are located slightly below the superior endplate of the inferior vertebra, except for the L5-S1 segment whose ICR lies within the disc space [110]. The theoretical ICRs of the commercially available fixed-bearing (ProDisc) and mobile-bearing (Charite) TDAs approximately match those of the intact lumbar and lumbosacral segments, respectively. Unfortunately, it has been reported that the lift-off phenomenon (defined as the separation of TDA bearing surfaces) and unsymmetrical/one-sided motion (defined as the rotation on one articulation larger than the other in a mobile-bearing TDA) may occur [114], necessarily causing a change of designed TDA motion and an increased risk of polyethylene (PE) inlay wear/fracture [76]. Therefore, it is essential to understand the dependence of TDA kinematics and contact mechanics on TDA design and spinal anatomy. In this study, we attempt to elucidate the mechanisms of lift-off and/or unsymmetrical motion in commercially available TDAs, when they are implanted into a lumbar motion segment. It is hypothesized that both implant design and facet joint (FJ) articulation influence TDA motion and PE inlay stress during the metal-on-polyethylene (MoP) articulation.

5.2.1. Methods

The intact L1-L5 multi-segment lumbar spine finite element (FE) model developed in Chapter 4 was reduced to an intact L3-4 lumbar segment model by removing vertebrae and spinal tissues of other levels, as presented in Figure 5.6a. FE models of three commercially available MoP TDA designs, including the convex-up design (with a PE radius of curvature of 8 mm, Figure 5.7a), convex-down design (a reversed implantation of the convex-up TDA FE model, Figure 5.7b) and biconvex design (with the same radii of curvature of 11 mm on both MoP articulations, and the PE rim thickness of 2.5 mm, Figure 5.7c), were inserted into the L3-4 segment model (Figure 5.6a) according to an anterior surgical procedure [101]. They represent TDA designs with theoretical ICRs located inferiorly, superiorly and at the PE core center, respectively. The resulting TDA-treated lumbar segment model was presented in Figure 5.6b.

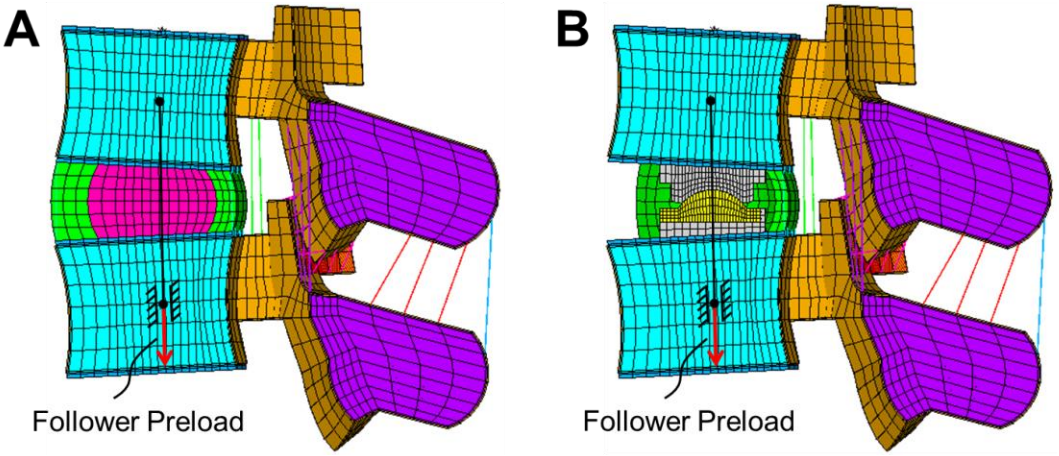


Figure 5.6: The intact L3-L4 lumbar motion segment FE model (a) and the treated segment FE model (b) which was created by inserting a TDA according to the anterior surgical procedure.

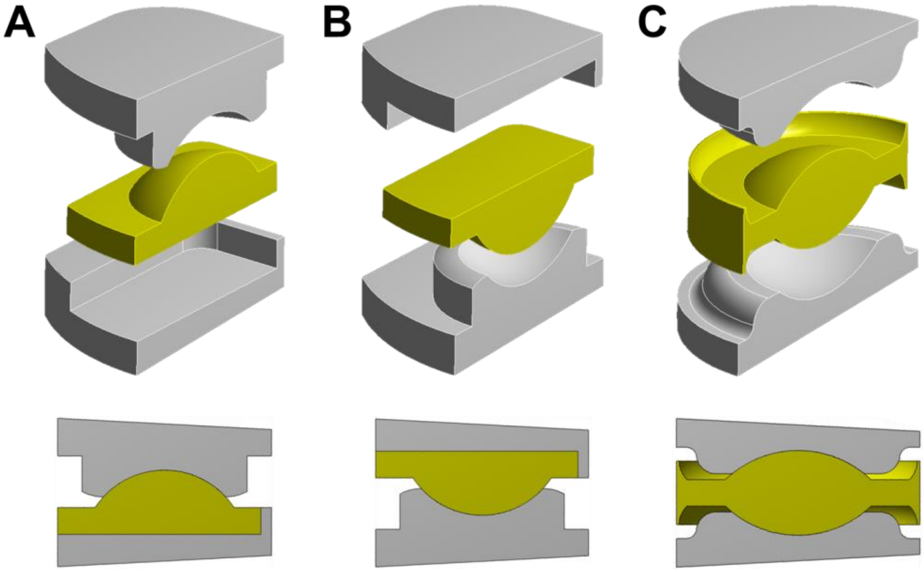


Figure 5.7: The FE models of the convex-up (a), convex-down (b) and biconvex (c) TDA designs. Gray elements represent the metallic endplates, and yellow elements represent the PE components.

The Young's modulus of 215 GPa and Poisson's ratio of 0.3 for the cobalt chrome (CoCr) alloy were assigned to the TDA metallic endplates [7], which articulated PE components with a coefficient of friction of 0.02. In this study, we adopted material properties of the Charite TDA polyethylene inlay (a lightly cross-linked UHMWPE gamma-radiation sterilized in nitrogen with a standard dose of 30 kGy) to model all PE components, with a Young's modulus of 940 MPa and a Poisson's ratio of 0.46, respectively [7, 115]. The rate-independent multilinear isotropic hardening plasticity, the so-called J_2-plasticity model [7, 115], was used to simulate PE plastic deformation. It is noted that the J_2-plasticity model cannot fit well the stress-strain curve in very large plastic deformation, but the capability of predicting relatively small plastic deformation is acceptable [115]. The PE's J_2-plasticity model material parameters were listed in Table 5.1.

Table 5.1: The J_2-plasticity model material parameters of the γ-N_2 30 kGy UHMWPE [115].

Plastic strain (ε^P)	0	0.03	0.11	0.55	1.20	1.40	1.70
Yielding stress (σ_Y, MPa)	13.1	24.4	29.3	40.0	112.0	216.3	341.8

The facet joint articulation was simulated using frictionless contact. To investigate the effect of FJ gap on TDA motions, mesh morphing was used to adjust the initial gap to 0.0, 0.4, 0.8 and 1.2 mm (Figure 5.8), respectively. Especially, 0.4 mm represents the normal gap [113, 116]. For all intact/TDA-treated segment FE models, a compressive follower preload of 500 N was applied to the superior endplate of the L3, followed by a moment ramping up to 7.5 Nm in extension. The preload path has been optimized as described in Chapter 4, such that no initial rotation in the sagittal plane occurred after preloading [107]. Furthermore, all TDAs were placed 2 mm posterior to the disc centroid, such that the preload path passed through their theoretical ICRs (Figure 5.6b). Using the Reuleaux method [101], the centrode (the ICR path) of the intact segment with the normal FJ gap (0.4 mm) in extension was calculated.

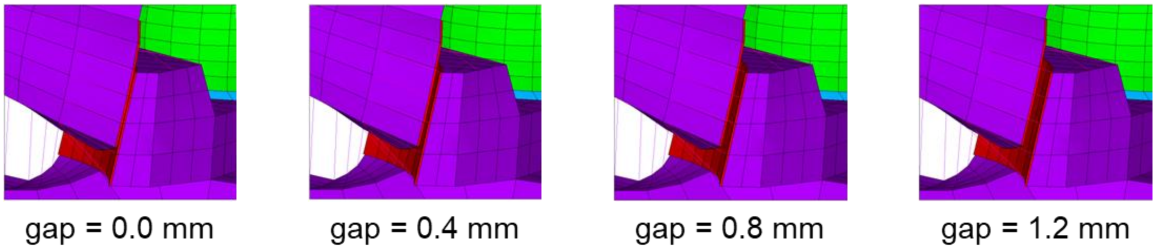

gap = 0.0 mm
gap = 0.4 mm
gap = 0.8 mm
gap = 1.2 mm

Figure 5.8: The facet joint FE models with initial gaps of 0.0, 0.4 (normal), 0.8 and 1.2 mm. Red elements represent the facet joint cartilages.

5.2.2. Results

In extension, the ICR of the intact L3-4 segment with the normal FJ gap starts at the disc center, and rapidly moves to the superior endplate of the L4 vertebra after 2.5 Nm (Figure 5.9), at which facet joints have engaged to provide an obvious resistance to the intervertebral rotation. The simulated ICR in extension is consistent with the description of the lumbar segmental ICRs [110]. As a result of tissue stiffening, the rotations of the intact L3-L4 segment during extension are 0.0 °after preloading, 2.6 °at 2.5 Nm, 3.6 °at 5 Nm and 4.2 °at 7.5 Nm, as shown in Figure 5.10.

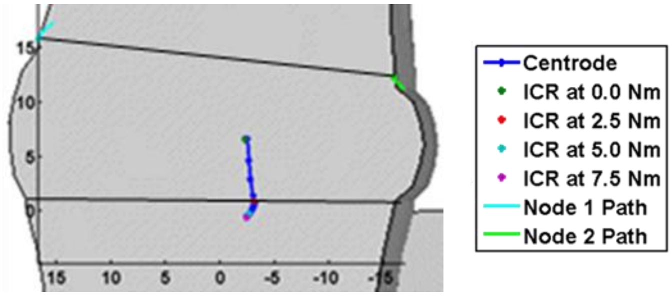


Figure 5.9: The centrode (ICR path) of the L3-L4 segment with the normal FJ gap in extension.

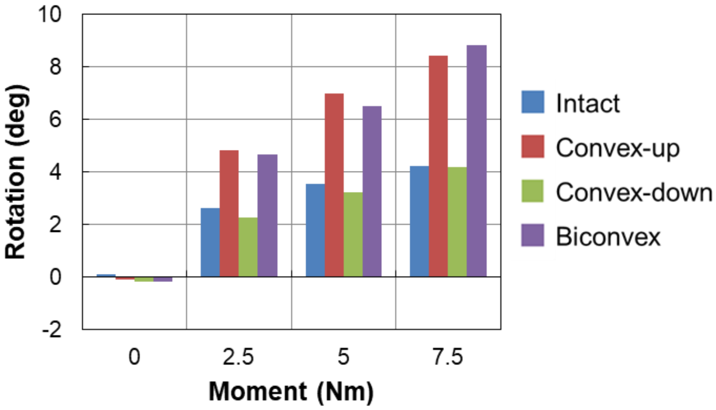


Figure 5.10: The extension ROMs of the intact segment and treated segments using convex-up, convex-down, and biconvex TDAs, when the normal FJ gaps were assigned. “0 Nm” indicates the ROMs after preloading.

Using the normal FJ gap, the ROM (4.2 °) of the segment treated by the convex-down TDA is in good agreement with that of the intact segment, as shown in Figure 5.10. However, the rotations of convex-up (8.4 °) and biconvex (8.8 °) TDA-treated segments are almost double those of the intact and convex-down TDA-treated segments in extension at 7.5 Nm. For the convex-up TDA design (Figure 5.11-top), the superior CoCr endplate lifts off from the PE component for FJ gaps of 0.0 and 0.4 mm, and posterior impingement occurs regardless of FJ gap size. For the convex-down TDA design (Figure 5.11-middle), only a slight anterior lift-off is observed when no initial FJ gap is simulated, and impingement never occurs.

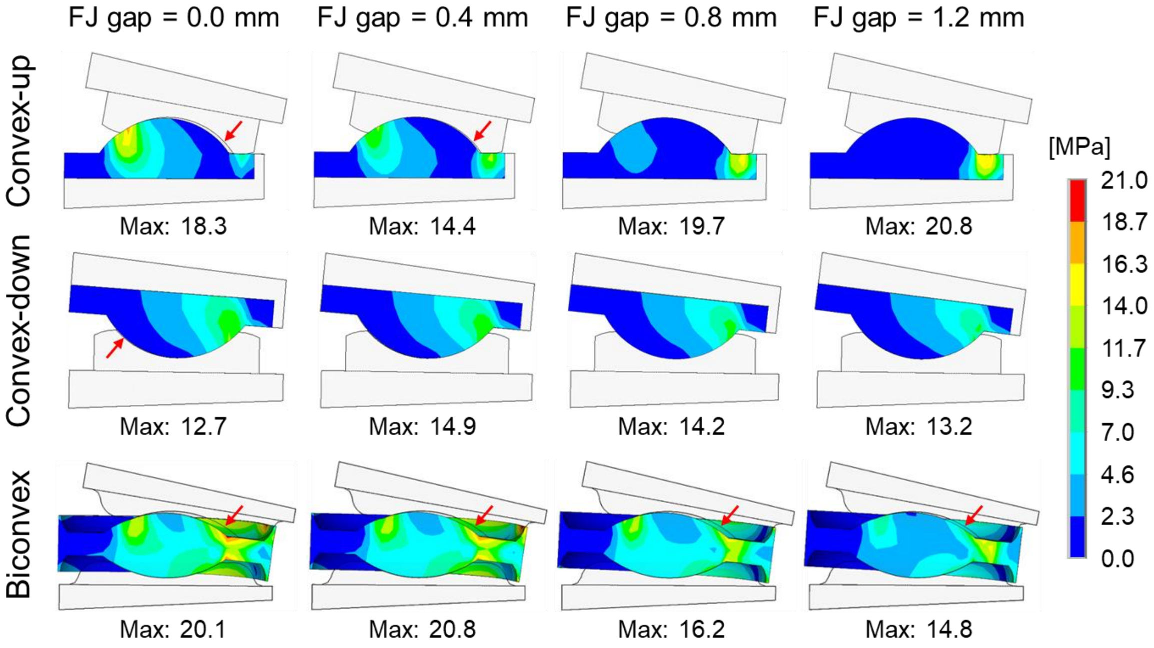


Figure 5.11. The TDA motion patterns and PE von Mises stresses of the convex-up (top), convex-down (middle) and biconvex (bottom) designs in extension at 7.5 Nm. The columns from left to right correspond to the treated segments with the FJ gaps of 0.0, 0.4, 0.8 and 1.2 mm. The red arrows indicate the locations of lift-off.

For the biconvex TDA design (Figure 5.11-bottom), posterior lift-off at the superior MoP articulation occurs in all cases, and superior MoP impingement causes a distinct downward bending of the PE rim. In all cases, the rotation of the superior MoP articulation is always larger than the inferior MoP articulation. This unsymmetric motion causes a downward bending of the PE rim. It is noted that the yield strength of the PE is 13 MPa, thus the PE rim is potentially plastically deformed. As the FJ gap increases, lift-off at the superior MoP articulation decreases, and inferior MoP impingement occurs in the segments with FJ gaps of 0.8 and 1.2 mm. As such, TDA motion becomes more symmetrical, and PE stress concentrations are mitigated.

5.2.3. Discussion and Conclusion

The extension ROM (4.2 °) of the single L3-L4 segment accurately matches that of the L3-L4 segment in the multi-segment L1-L5 lumbar spine (Figure 5.5a-*left*). It further demonstrates that the applied loads are evenly transferred through each level in the multi-segment lumbar spine, as a motion segment which is the smallest physiological motion unit of the spine exhibits biomechanical characteristics similar to each level of the entire spine. The segmental ROM of the convex-down TDA design is in excellent agreement with that of the intact segment (Figure 5.10), and a good congruence on the MoP remains during extension regardless of different facet joint gaps. Although the theoretical ICR of the convex-up TDA, which lies inferiorly, is closer to the ICR of the intact segment (Figure 5.9), it cannot be anticipated that the actual ICR matches the theoretical ICR due to adverse TDA motion (Figure 5.11-top).

Lift-off causes stress concentration and a higher wear/fracture risk for the PE component, due to a decrease in the contact area of the bearing surfaces. As the FJ gap increases, lift-off and/or unsymmetrical motion are alleviated in all TDA-treated segments. It clearly demonstrates that FJ articulation guides motion and impedes the MoP articulation. For the convex-up design (Figure 5.12a), the FJ tends to be separated by a coupled posterior translation of the superior vertebra in extension; the exact opposite occurs for the convex-down design (Figure 5.12b), causing FJ contact. Therefore, the convex-up design leads to a larger extension rotation and a higher incidence of impingement compared to the convex-down design. Combining both convex-up and convex-down TDA motions can explain the unsymmetrical motion in the biconvex design.

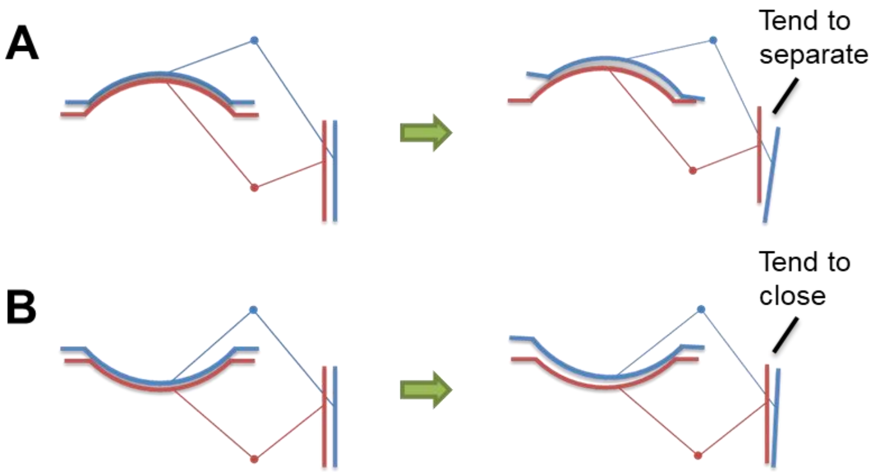


Figure 5.12: Illustration of the motion patterns of the convex-up (a) and convex-down (lower) TDAs in extension. (blue/red: the superior/inferior TDA component and the corresponding L3-inferior/L4-superior facet, both of which are assumed to be rigidly connected together)

Compared to the intact segment, TDA-treated segments lack the resistance to extension due to the dissection of the anterior longitudinal ligament, the anterior annulus, and the entire nucleus. Therefore, it is difficult to prevent TDA hypermotion and abnormal rotation (*e.g.*, lift-off and unsymmetric motion) in a lumbar segment subjected to anterior surgery. A lateral approaching surgical procedure [8] may be beneficial to restoring the normal segmental kinematics. In conclusion, facet joint articulation may change the designed motion patterns for both fixed- and mobile-bearing TDAs, resulting in lift-off and/or unsymmetric motion phenomena. These abnormal motion characteristics may not be identified using a standalone TDA; thus we suggest that TDA designs should be tested while incorporating spinal structures.

5.3. Design Optimization of a New Biconcave Mobile-bearing TDA Concept

The ball-in-socket total disc arthroplasty (TDA) designs, taken from total knee and hip arthroplasty (TKA/THA), provide a motion-preserving technique to treat disc degeneration diseases, while mitigate adjacent segment degeneration by maintaining spinal motion. However, the native disc is significantly different from cartilaginous interface joints such as knee and hip

[72]. Although impressive improvement in the joint functional capability after THA and TKA has been achieved, TDA treatment using anterior surgical procedure which removes intersegment tissues may cause abnormal biomechanics, and thus more complex complications such as excessive segmental rotation [8], spinal tissue overloading [9] and polyethylene (PE) inlay wear/fracture failures [10].

Impingement, which was observed in both mobile- and fixed-bearing TDAs, is defined as unintended contact between two nonbearing surfaces, and it is highly associated with excessive wear and fracture of the polyethylene [7, 87]. However, we hypothesize that impingement as a contact constraint can prevent excessive intervertebral rotations and spinal tissue overloading. Therefore, the author proposed a new TDA concept featuring a biconcave PE mobile core, with a smooth transition from the dome to the rim. The biconcave-core TDA was then optimized for natural range of motion (ROM) and minimal polyethylene contact pressure (PCP), respectively. The kinematics and contact mechanics of the ROM-optimal and PCP-optimal biconcave-core TDA designs was compared with those of the intact lumbar segment and the treated lumbar segment using a commercially available mobile-bearing TDA design with a biconvex PE core.

5.3.1. Methods

To perform design optimization, the proposed biconcave TDA concept (Figure 5.13a) was parametrically modeled considering seven design variables ($\boldsymbol{x}$), including the gaps (g_a, g_l, g_p) between the dome radii of the metallic endplate and PE core at the anterior, lateral and posterior, the superior and inferior radii of curvature (r_s, r_i) of the biconcave PE core, and the superior and inferior offset distances (d_s, d_i), as described in Figure 5.14. It is ensured that the curved rim faces of both the PE core and metallic EPs are tangent to the spherical dome bearings (with a radius of curvature of r_s/r_i) on both the superior and inferior MoP articulations. For each articulation, the circular PE edge and elliptical EP edge (represented by the black points in Figure

5.14a) are maintained on a same spherical face (indicated by the dash line), which is offset from the superior/inferior dome bearing (with a radius of curvature of r_s/r_i) by a distance of d_s/d_i. Table 5.2 lists the upper and lower boundaries of each design variable (g_a, g_l, g_p, r_s, r_i, d_s, d_i) in design optimization. For the purpose of comparison, a biconvex-core TDA with a dome radius of curvature of 10 mm and a rim thickness of 2.5 mm [117] was also modeled (Figure 5.13b), and subsequently implanted into an intact lumbar segment to simulate the resulting responses in different loading scenarios.

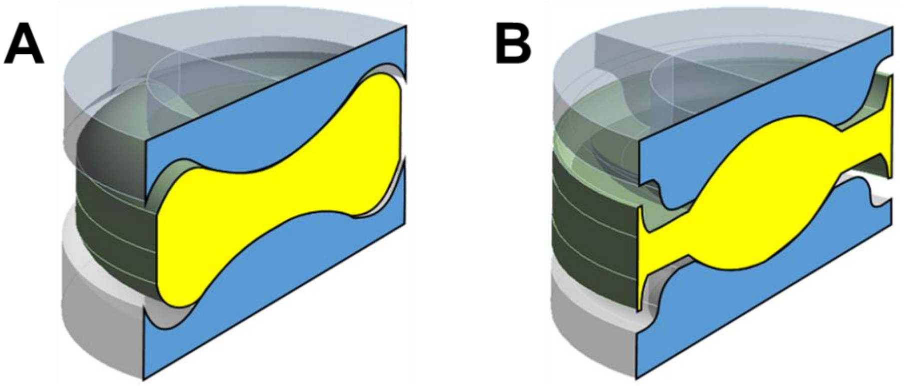


Figure 5.13: The 3-D oblique view of the biconcave (a) and biconvex (b) mobile-bearing TDA designs, both of which comprises a biconcave PE core and two metallic EPs.

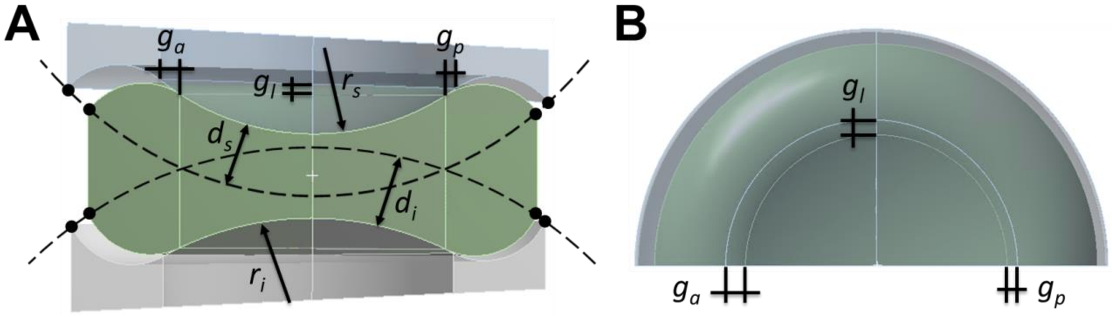


Figure 5.14: Design variables of the parametric biconcave mobile-bearing TDA model. (a) The sagittal view of the TDA; (b) The top view of the TDA.

Table 5.2: The lower and upper boundaries of the TDA design variables. Note that the unit of all design variables is mm.

	Lower Boundary ($\boldsymbol{x}_{lb}$)	Upper Boundary ($\boldsymbol{x}_{ub}$)
Dome Radii of Curvature (r_s and r_i)	10.00	20.00
Offset Distances (d_s and d_i)	3.00	4.20
Anterior Gaps (g_a)	0.50	1.50
Lateral Gaps (g_l)	0.00	1.00
Posterior Gaps (g_p)	0.00	1.00

To save computational cost in design optimization, the L3-4 intact and treated lumbar segment FE models developed previously (Figure 5.6) were simplified by modeling L3/L4 vertebrae and TDA metallic endplates as rigid bodies (Figure 5.15), using the multipoint constraint technique in ANSYS 15.0. A detailed comparison of the effects of deformable and rigid segment models on simulation results can be found in Appendix D.2. The PE cores were meshed using 20-node quadratic hexahedral elements (SOLID187). J_2-plasticity [7, 115] was used to model the material properties of TDA PE components (E = 940 MPa, υ = 0.46, σ_Y = 13 MPa) as described before. The biconcave / biconvex (Figure 5.15) mobile-bearing TDAs were inserted into the L3-4 segment according to an anterior surgical procedure [101]. Facet joint articulations and TDA articulations were modeled as frictionless contact and frictional contact with a coefficient of friction of 0.02, respectively.

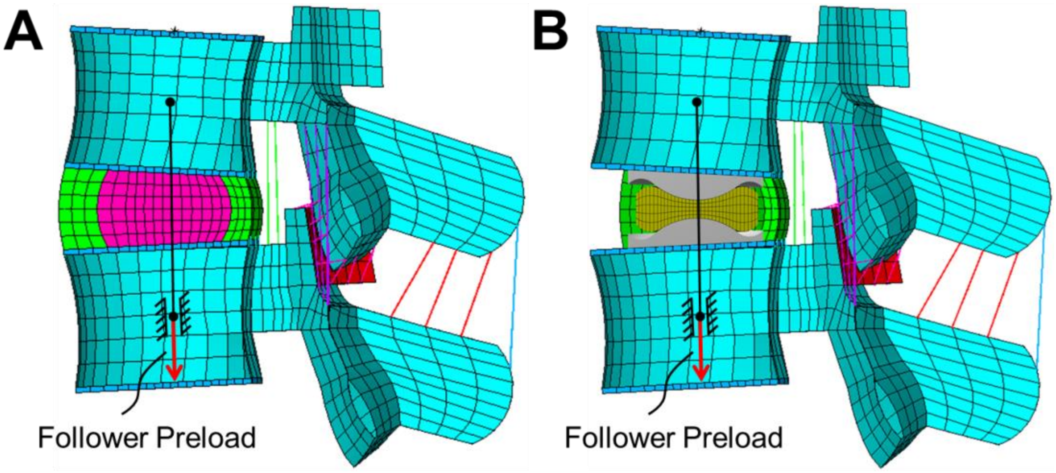


Figure 5.15: The intact (a) and TDA-treated (b) segment FE models, in which vertebrae and TDA metallic endplates were modeled as rigid bodies. Here, only the rigid treated model (b) implanted by a biconcave-core TDA is presented. The same treated segment model was also created for the biconvex-core TDA.

All intact/TDA-treated segment FE models were applied a compressive follower preload of 500 N, followed by a moment up to 7.5 Nm on the superior endplate of the L3 in 4 loading scenarios (flexion, extension, lateral bending, and axial torsion). The preload path has been optimized in terms of the intact multi-segment L1-L5 lumbar spine model (Chapter 4), such that no sagittal rotation occurred after preloading [107]. Furthermore, all TDAs were placed 2 mm posterior to the disc centroid, such that the preload path passed through their PE core centers. According to a mesh convergence study, the bearing surfaces of the biconcave and biconvex PE cores were meshed using an element size of 0.8 mm and 0.6 mm, respectively (Appendix D.1).

In each optimization iteration, the biconcave TDA-treated segment model was simulated in 4 loading scenarios simultaneously. Two single-objective design optimizations were performed in MATLAB R2014a, by defining two optimization objectives in terms of the segmental range of motion (ROM) and the polyethylene contact pressure (PCP), respectively:

$$J_{ROM}(\boldsymbol{x}) = \sum_{i} \sum_{j} \left| ROM_{ij}^{TDA} - ROM_{ij}^{INT} \right| \tag{5.1}$$

$$J_{PCP}(\boldsymbol{x}) = \underset{i}{\text{mean}}\left(\underset{j}{\max}(PCP_{ij})\right) \tag{5.2}$$

where i represents the 4 loading scenarios of flexion, extension, lateral bending, and axial torsion. j represents the 4 moments of 0 (preload), 2.5, 5, and 7.5 Nm in each loading scenario. The components of ROM_{ij}^{INT} are the normal segmental ROMs simulated using the intact segment FE model. Therefore, $J_{ROM}(\boldsymbol{x})$ and $J_{PCP}(\boldsymbol{x})$ represent the ROM metric (the ROM absolute error sum of the TDA-treated segment with respect to the intact segment) and the PCP metric (the mean maximum polyethylene contact pressure in four loading scenarios), respectively.

5.3.2. Results

After design optimization, the ROM-optimal shape and PCP-optimal shape of the biconcave TDA design are presented in Figure 5.16. Values of design variables (g_a, g_l, g_p, r_s, r_i, d_s, d_i) are tabulated for both optimal TDA designs in Table 5.3. Compared to the ROM-optimal design (Figure 5.16a), a complete congruence (g_p = 0.00 mm) occurs at the posterior rim articulation of the PCP-optimal design (Figure 5.16b). As listed in Table 5.4, the ROM metric (J_{ROM}) of the ROM-optimal TDA-treated segment is 7.04°, compared to 11.02° and 21.72° for the PCP-optimal and biconvex-core TDA-treated segments. The PCP metric (J_{PCP}) for the PCP-optimal TDA-treated segment is 20.73 MPa, compared to 26.95 MPa and 58.72 MPa for the ROM-optimal and biconvex TDA-treated segments. It is worth noting that the worst ROM and PCP metrics of ROM-optimal and PCP-optimal biconcave-core TDAs are 11.02 deg and 21.46 MPa, respectively, but both the metrics (21.72 deg and 58.72 MPa) of the biconvex TDA are even worse.

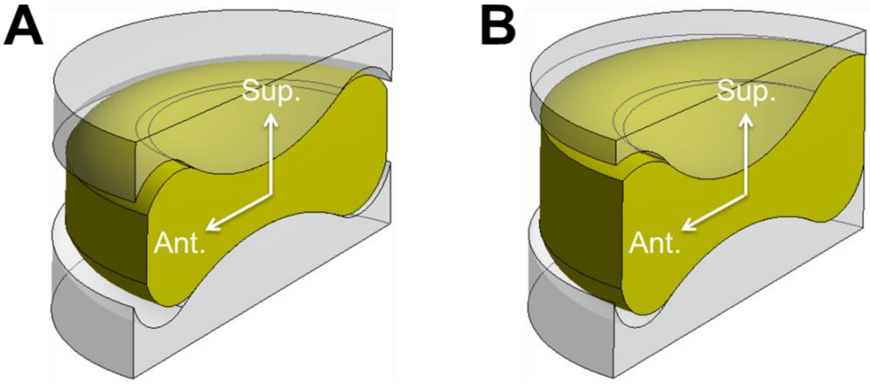


Figure 5.16: The ROM-optimal (a) and PCP-optimal (b) biconcave-core TDA designs.

Table 5.3: The design variables of the ROM-optimal and PCP-optimal biconcave TDA designs. The unit is mm.

	r_s	r_i	d_s	d_i	g_a	g_l	g_p
ROM-optimal	12.62	13.93	3.70	3.98	1.16	0.30	0.41
PCP-optimal	10.02	12.77	3.00	4.20	0.65	0.38	0.00

Table 5.4: The performance metrics of ROM-optimal, PCP-optimal and biconvex-core TDA designs in term of TDA kinematics and contact mechanics.

	J_{ROM} ($^\circ$)	J_{PCP} (MPa)
ROM-optimal	7.04	29.52
PCP-optimal	11.02	21.46
Biconvex-core	21.72	58.72

The flexion ROMs of all TDA-treated segments are in good agreement with that of the intact segment (Figure 5.17a). The biconvex TDA causes distinct hypermotion in other loading scenarios (Figure 5.17b-d). Since there is no posterior gap between the rims of PE component and metallic endplates in the PCP-optimal TDA (Figure 5.16b), it results in a substantial resistance to extension (Figure 5.17b).

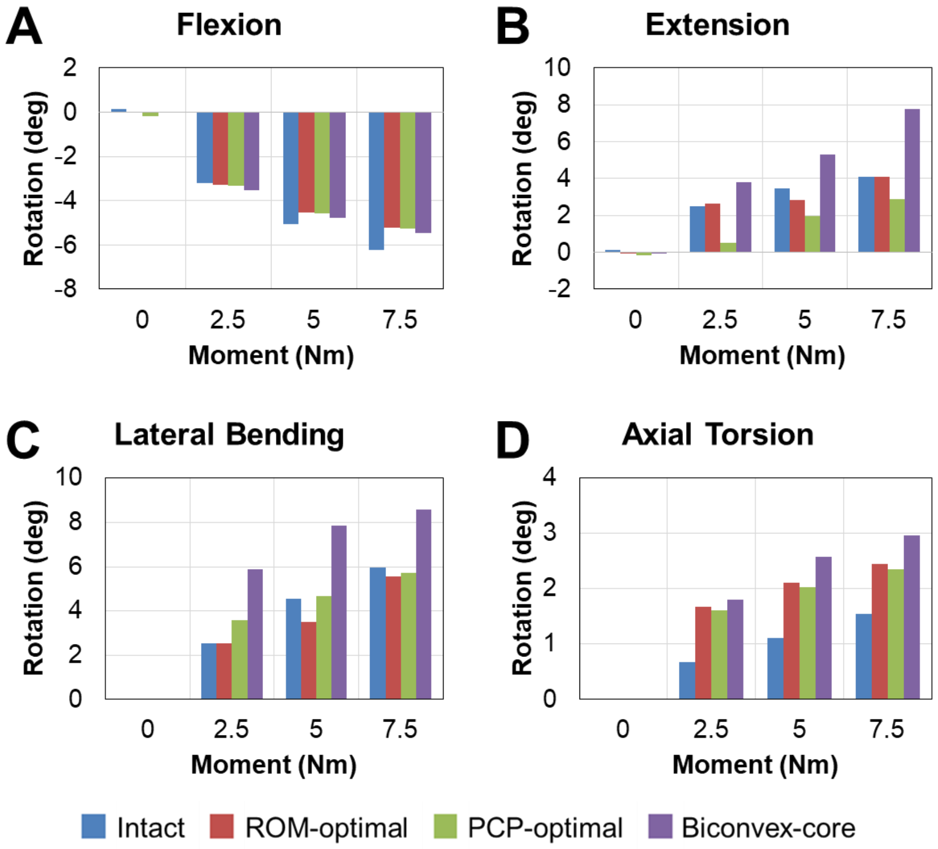


Figure 5.17. The ROMs of the intact and TDA-treated segments in different loading scenarios ("0 Nm" indicates the rotation after preloading).

In flexion at 7.5 Nm (Figure 5.18), contact only occurs on the PE core domes for all TDAs with a relatively uniform contact pressure distribution. However, the PE rims are the main bearing surfaces for all TDAs in other loading scenarios at 7.5 Nm (Figure 5.18). Especially, extension or lateral bending are the most severe loading scenarios which cause much larger contact pressures for the biconvex-core TDA design (Figure 5.18-bottom).

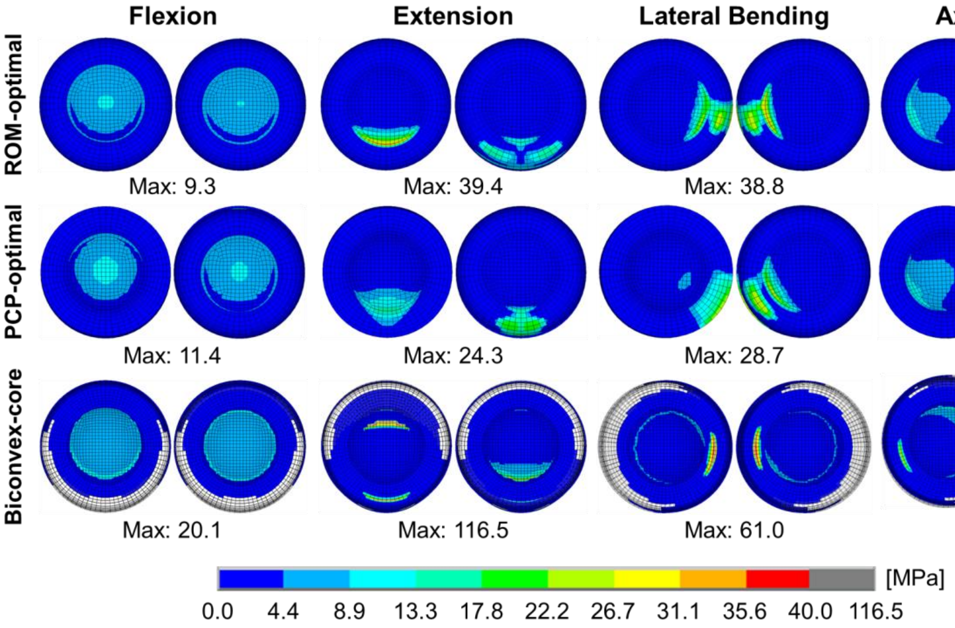


Figure 5.18: The contact pressures on the superior (left) and inferior (right) articulation surfaces of the PE core in different TDAs in different loading scenarios at 7.5 Nm. The transparent contact elements on the biconvex PE core indicate where the contact gap is beyond a preset threshold.

As shown in Figure 5.19a, lift-off and unsymmetric motion occur in the biconvex-core TDA in extension at 7.5 Nm, causing a distinct downward bending and plastic deformation ($\sigma_Y =$ 13 MPa) of the posterior PE rim. In contrast, stress concentrations are well relieved in both the ROM-optimal and PCP-optimal biconcave-core TDAs (Figure 5.19b,c) owing to more material at the PE rims and the smooth transition from the PE domes to the rims, whereas marginal lift-off of the PE core from the inferior metallic EP can still be noticed in each optimal biconcave-core TDA.

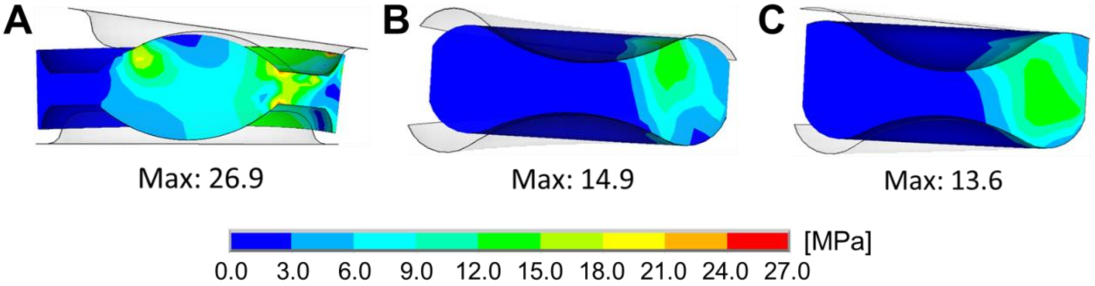


Figure 5.19: The motion patterns and von Mises stresses of the PE cores in the biconvex-core (a), ROM-optimal (b), and PCP-optimal (c) TDAs in extension at 7.5 Nm.

5.3.3. Discussion and Conclusion

Previous TDA retrieval studies [6] revealed that excessive wear and fracture of the PE rim cause biconvex-core TDA failures, due to the small PE rim thickness (median: 0.29 mm) of biconvex TDA. By changing the biconvex mobile core to a biconcave mobile core, the PE rim in the new biconcave TDA was significantly strengthened, whereby the PE rim can serve as a secondary bearing. Therefore, it is expected to considerably prolong implant life *in vivo*, compared to the commercially available biconvex-core TDA design.

Typically, there is a trade-off between the performance metrics of kinematics and contact mechanics [12]; as one metric improves, the other deteriorates. The trade-off relationship between these two metrics can be described by a Pareto front curve (Figure 5.20). The pseudo Nadir point can be defined by the two single-objective optimums (blue); each coordinate of the pseudo Nadir point were chosen as the worst performance metric of the two single-objective optimums. According to Table 5.4, it indicates that the biconvex-core TDA design is even worse than the pseudo Nadir point of the biconcave-core TDA design in terms of kinematics and contact mechanics.

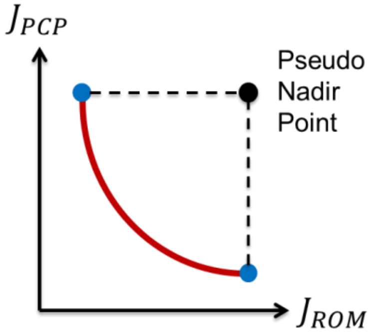


Figure 5.20: Illustration of a representative Pareto front curve in terms of two objectives, where two single-objective optimums (blue points) determine the pseudo Nadir point.

We will further expand the design variable space of the biconcave-core TDA and implement multi-objective optimization to consider both performance metrics simultaneously. To our best knowledge, it is the first time to optimally design a total disc arthroplasty, using a finite element model incorporating spinal structures. The new biconcave mobile-core total disc arthroplasty is expected to alleviate hypermotion and spinal tissue overloading, as well as decrease the risk of polyethylene wear and fracture failures.

Chapter 6: Multiobjective Optimization of Total Disc Arthroplasty

It remains controversial whether spinal fusion or total disc arthroplasty (TDA) is more beneficial for the treatment of disc degeneration and spinal disorders. Generally, spinal fusion creates a high-stiffness bone union within the intervertebral space, using an interbody cage loaded with bone graft. The motion-sacrificing technique completely restricts the intervertebral rotation, unless pseudo-arthrodesis slightly restores spinal motion. Therefore, the postoperative segmental response after spinal fusion is less sensitive to the interbody cage design. The small sensitivity to cage design limits the potential to match the native intervertebral disc biomechanics, despite the lower incidence and fewer types of complications after spinal fusion. In contrast, TDA design significantly influences the postoperative spine behavior. Here, it is hypothesized that an optimally designed TDA can mitigate postoperative complications and better mimic the native intervertebral disc. To that end, we leverage multiobjective design optimization to maximize benefits that TDA can bring.

In the previous chapter, we have proposed a new mobile-bearing TDA with a biconcave polyethylene core. It was demonstrated that the new TDA design could improve the treated lumbar segmental kinematics and metal-on-polyethylene (MoP) contact mechanics, according to our simulation results. In the study, we further investigated the performance of the new TDA in terms of facet joint loading, which was considered as an indicator of facet joint arthritis. To measure each performance metric, a computational FE model was created to simulate the responses of the TDA-treated lumbar segment. The new TDA design was refined in term of TDA kinematics, contact mechanics and facet joint loading, using a neural network based multiobjective design optimization algorithm. To our best knowledge, this is the first time that optimizes an implant design considering three performance metrics simultaneously. It should be noted that we optimally design the new TDA, in order to demonstrate that TDA kinematics, contact mechanics, and spinal tissue load-sharing are sensitive to TDA design, rather than competing with commercially available TDA designs.

6.1. Parametric Finite Element Modeling of Biconcave Mobile-bearing TDA

According to the study described in the previous chapter, the motion of a mobile-bearing TDA can be considered as a consequence of the superposition of individual convex-down and convex-up metal-on-polyethylene (MoP) articulations. Using ANSYS 15.0 Workbench DesignModeler, we developed a more sophisticated parametric model of the biconcave mobile-bearing TDA (Figure 6.1a), which consists of a polyethylene (PE) core and two metallic endplates (EPs).

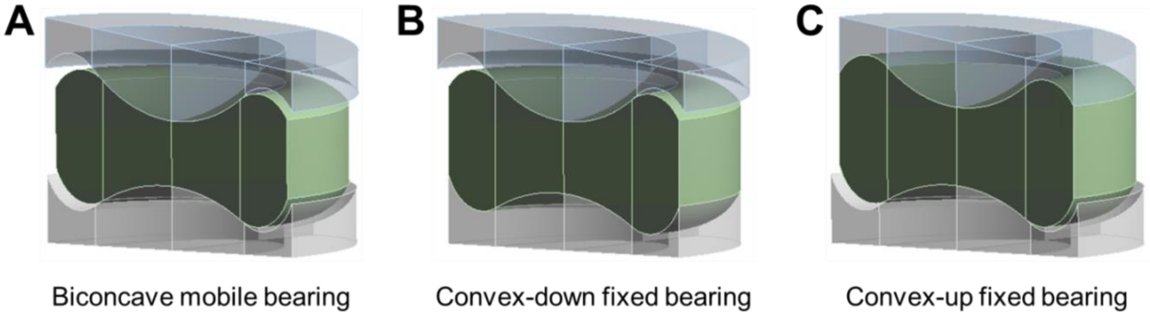


Figure 6.1: The parametric biconcave mobile-bearing TDA design (a) with the capability of changing itself to convex-down (b) and convex-up (c) fixed-bearing configurations.

The parametric TDA model enables itself to reduce to the two design extremities, a convex-down fixed-bearing TDA (Figure 6.1b) or a convex-up fixed-bearing TDA (Figure 6.1c), by assigning a complete congruence on either the superior or inferior rim contact interface. As such, multiobjective design optimization will determine whether a fixed-bearing design or a mobile-bearing design (which lies between the convex-down and convex-up fixed-bearing TDA configurations) is the optimal design in terms of different performance metrics.

Compared with the previous parametric TDA design described in Chapter 5, the number of design variables has been expanded to ten in the parametric TDA model. All design variables were chosen as the mating dimensions between the PE core and metallic EPs in the TDA. As presented in Figure 6.2a, r_s represents the radius of curvature of the superior dome bearing interface where the PE and metallic EP dome faces are coincident. It is ensured that the curved rim faces of the PE core and metallic EPs are tangent to the spherical dome bearings on both the superior and inferior MoP articulations. For the superior articulation, the circular PE edge and elliptical EP edge (represented by the black points in Figure 6.2a) are maintained on a same spherical face (indicated by the dash line), which is offset from the superior dome bearing interface (with a radius of curvature of r_s) by a distance of d_s.

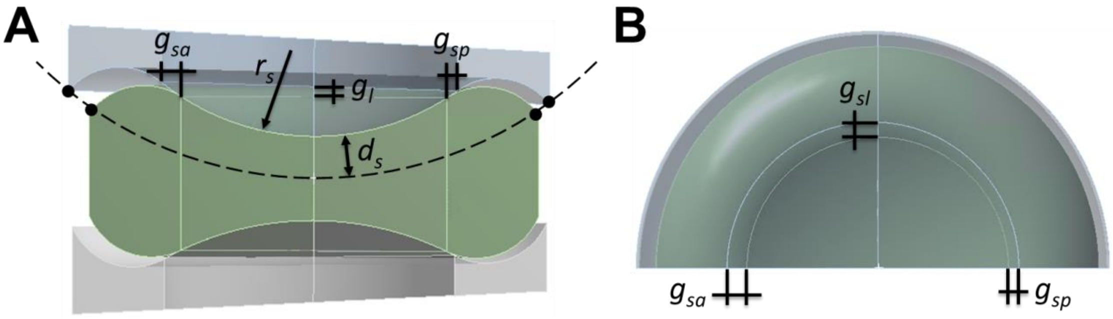


Figure 6.2: Design variables of the parametric biconcave mobile-bearing TDA model. Only the design variables (r_s, d_s, g_{sa}, g_{sl}, and g_{sp}) on the superior articulation were shown. (a) The sagittal view of the TDA; (b) The top view of the TDA.

The congruence between the PE and EP rims on the superior articulation were adjusted by the gaps (g_{sa}, g_{sl}, and g_{sp} in the transverse plane, Figure 6.2b) due to the differences between the PE dome radius and the EP dome radii at the anterior, lateral, and posterior. Similarly, there are the corresponding design variables (r_i, d_i, g_{ia}, g_{il}, and g_{ip}) for the inferior TDA articulation. Table 6.1 lists the ranges of each design variable. By randomly initializing the ten design variables within the ranges (Table 6.1), eight random TDA designs are generated as presented in Figure 6.3. It shows that the parametric TDA model can yield plentiful designs in the ten-dimensional design space.

Table 6.1: The lower and upper boundaries of the TDA design variables. Note that the unit of all design variables is mm.

	Lower Boundary ($\boldsymbol{x}_{lb}$)	Upper Boundary ($\boldsymbol{x}_{ub}$)
Dome Radii of Curvature (r_s and r_i)	11.00	20.00
Offset Distances (d_s and d_i)	3.10	4.10
Anterior Gaps (g_{sa} and g_{ia})	0.01	1.20
Lateral Gaps (g_{sl} and g_{il})	0.01	1.00
Posterior Gaps (g_{sp} and g_{ip})	0.01	0.80

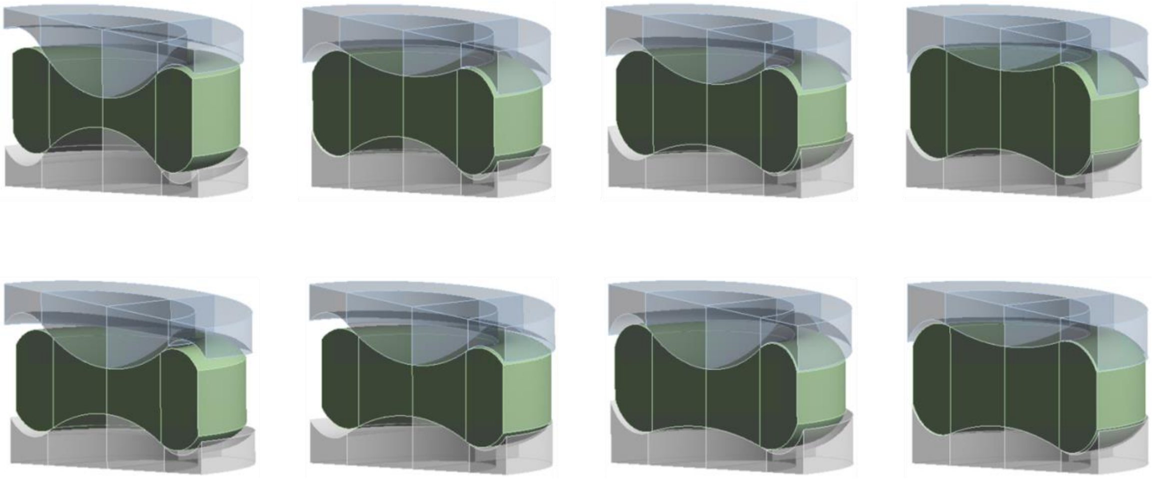


Figure 6.3: Eight random TDA designs by randomly initializing 10 design variables.

The parametric TDA design was meshed using 20-node quadratic hexahedral elements (SOLID187 in ANSYS), as shown in Figure 6.4. The mesh size of PE articulation surfaces was chosen as 0.75 mm, according to a mesh convergence study of the contact pressures on both MoP articulations. Within the ranges of design variables (Table 6.1), no failures occur when creating the TDA FE model (Figure 6.4), and the minimum element quality is greater than 0.35, according to a design exploration for 120 random designs using ANSYS Workbench 15.0.

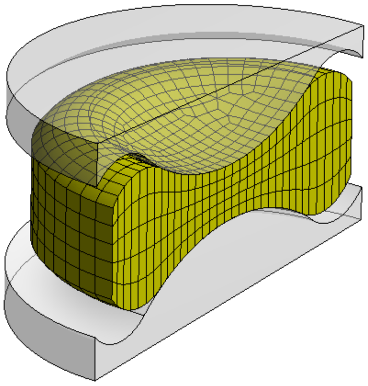


Figure 6.4: The FE mesh of the TDA parametric model, whose design variables were randomly chosen within the design space limitations (Table 6.1).

As shown in Figure 6.5, the TDA FE model was inserted into an intact L3-L4 segment FE model according to an anterior surgical procedure, in which the anterior longitudinal ligament, the anterior annulus, and the entire nucleus were dissected. The material properties of spinal tissues have been calibrated using a step-wise addition procedure [33]. To save computational cost in multiobjective design optimization, the vertebrae and TDA metallic endplates were modeled as rigid bodies (Figure 6.5a,b), using the multipoint constraint technique in ANSYS 15.0 MAPDL. J_2-plasticity [7, 115] was used to model the material properties of TDA PE components (E = 940 MPa, υ = 0.46, σ_Y = 13 MPa). Facet joint articulations and TDA articulations were modeled as frictionless contact and frictional contact with a coefficient of friction of 0.02, respectively.

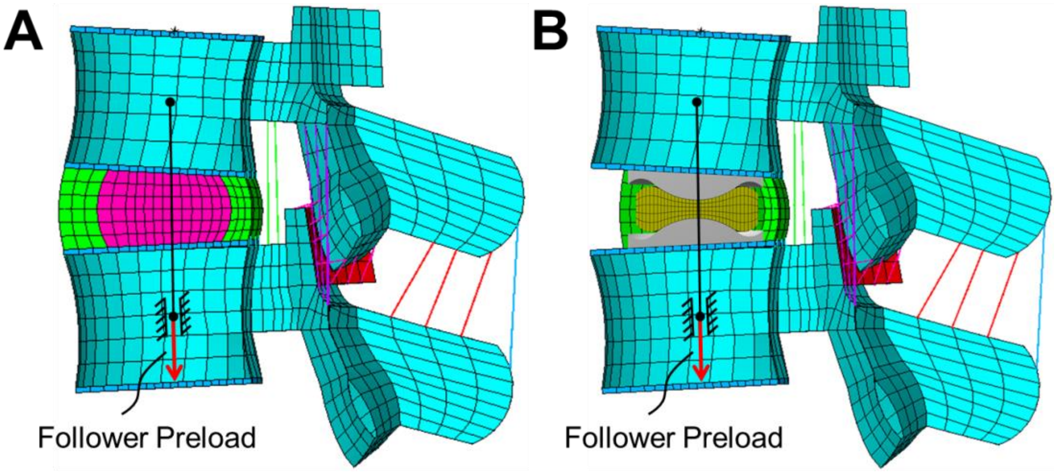


Figure 6.5: The FE models of the intact (a) and TDA-treated (b) L3-L4 segments.

The TDA-treated segment FE models (Figure 6.5b) were applied a compressive follower preload of 500 N, followed by a moment up to 7.5 Nm on the superior endplate of the L3 in four physiologically relevant loading scenarios (*i.e.*, flexion, extension, lateral bending, and axial torsion) [66], which were simulated in parallel. The kinematical responses of the intact segment FE model (Figure 6.5a) were also simulated using the same loading protocol [113], in order to provide the "benchmark" segmental ranges of motion for the assessment of the treated-segment kinematics. The path of the compressive follower preload was modeled using a tension-only cable element, in which the inferior node was applied a cylindrical kinematic constraint (Figure 6.5a,b) [82]. The preload path has been optimized during the development of the intact multi-segment L1-L5 lumbar spine model (Chapter 4), to minimize the sagittal rotation (flexion and extension) after preloading [107]. For the treated segment model (Figure 6.5b), the TDA was placed 2 mm posterior to the disc centroid, such that the preload path passed through the PE core center.

6.2. Implementation of Artificial Neural Network Based Multiobjective Optimization Algorithm

6.2.1. Performance Metrics of TDA Postoperative Complications

We hypothesize that TDA design can be optimized to alleviate postoperative complications, such as TDA abnormal motion, polyethylene wear/fracture, and facet joint overloading. Therefore, three TDA performance metrics (objectives) were defined to quantify the incidences of these complications; if a smaller value of a performance metric is measured from the response of a TDA design, it indicates that the TDA design is associated with a lower incidence of the corresponding complication. These performance metrics were predicted from the simulation results of the TDA-treated lumbar segment FE model.

In each loading scenario (flexion, extension, lateral bending, and axial torsion), we simulated the ranges of motion of the TDA-treated segment, the ipsilateral facet joint forces (chosen to be the average forces of the left and right facet joints in flexion and extension), and the maximum contact pressures on the superior and inferior polyethylene contact surfaces. Simulation results were output, when the moment was applied at 0.0, 2.5, 5.0, and 7.5 Nm, respectively (0.0 Nm represents the compressive follower preload). Three performance metrics regarding the range of motion (ROM), facet joint force (FJF), and polyethylene contact pressure (PCP) were formulated below,

$$t_1 = \left[\sum_i \sum_j \left| ROM_{ij} - ROM_{ij}^{INT} \right| \right]^{1.35} \tag{6.1}$$

$$t_2 = \operatorname*{mean}_i \left(\max_j \left(FJF_{ij} \right) \right) \tag{6.2}$$

$$t_3 = \left[\sum_i \sum_j PCP_{ij} \right]^{0.75} \tag{6.3}$$

where the index i represents flexion, extension, lateral bending, and axial torsion. The index j represents the applied moments at 0.0, 2.5, 5.0, and 7.5 Nm in each loading scenario. The components of ROM_{ij}^{INT} are the normal segmental ROMs simulated using the intact segment FE model (Figure 6.5a). The exponents in the ROM metric (t_1) and the PCP metric (t_2) were introduced such that all performance metrics have an approximately identical scale for a better visualization of the feasible objective space. The relationship between the FE predicted metrics $\boldsymbol{t}(\boldsymbol{x}) = [t_1(\boldsymbol{x}), t_2(\boldsymbol{x}), t_3(\boldsymbol{x})]^{\mathrm{T}}$ and the design variables $(\boldsymbol{x})$ is described in Figure 6.6. The design variables $(\boldsymbol{x})$ in the design space are mapped to the performance metrics $(\boldsymbol{t})$ in the objective space, through an implicit response function simulated by finite element analysis.

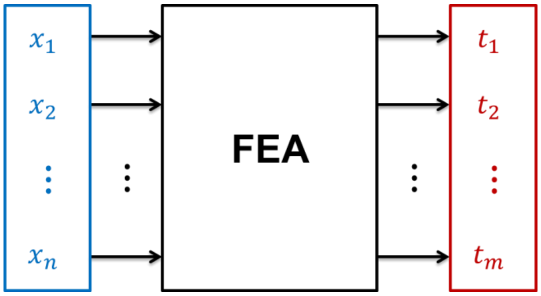


Figure 6.6: Illustrate the relationship between design variable and FE predicted performance metrics using finite element analysis (FEA).

6.2.2. Fitting of Performance Metrics Using Artificial Neural Network

In a traditional design optimization algorithm, finite element (FE) analysis is performed in each optimization iteration to provide a black-box cost function. Generally, evolutionary optimization algorithms based on gradient descent require the gradient of the cost function with respect to each design variable. Although it can be approximated using the finite difference method in an optimization algorithm, the implementation of the numerical gradient typically causes a computational cost multi-fold higher than that using the analytical gradient.

To improve computational efficiency in optimization, a three-layer multi-output feed-forward neural network (NN) was implemented to fit the response function of FE analysis, as shown in Figure 6.7. The NN is comprised of three layers, including an input layer, a hidden (middle) layer, and an output layer. The numbers of neurons at the input layer and the output layer match the number of design variables, $n = 10$, and the number of performance metrics, $m = 3$, respectively. It is noted that three performance metrics are predicted simultaneously using one multi-output NN, based on an assumption that there is a correlation among these performance metrics.

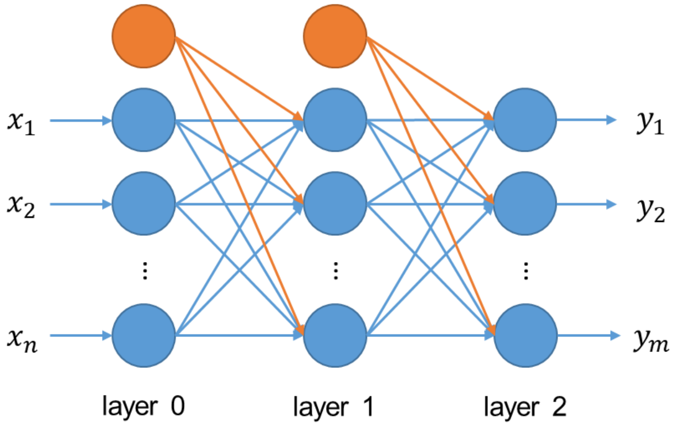


Figure 6.7: The architecture of a three-layer multi-output feed-forward NN. Especially, the orange neurons represent the bias neurons, whose outputs are always equal to 1.

The outputs ($\boldsymbol{y}$) of the NN (Figure 6.7) as a function of the inputs ($\boldsymbol{x}$) can be analytically expressed in the matriculated format,

$$\boldsymbol{y} = \boldsymbol{b}^{(2)} + \boldsymbol{w}^{(2)} a\left(\boldsymbol{b}^{(1)} + \boldsymbol{w}^{(1)} \boldsymbol{x}\right) \tag{6.4}$$

where $\boldsymbol{w}^{(1)}$ and $\boldsymbol{b}^{(1)}$ are the weights and bias terms of the NN from the input layer to the hidden layer. $\boldsymbol{w}^{(2)}$ and $\boldsymbol{b}^{(2)}$ are the weights and bias terms of the NN from the hidden layer to the output layer. $a(\bullet)$ is an element-wise activation function of the hidden layer. No activation exists at the input layer and the output layer.

Based on the continuous differentiability of Equation 6.4, we further formulate the gradient of the NN as shown in Equation 6.5, which is subsequently incorporated into a multiobjective optimization algorithm.

$$\operatorname{grad} \boldsymbol{y}(\boldsymbol{x}) = \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{w}^{(2)} \left[a'\left(\boldsymbol{b}^{(1)} + \boldsymbol{w}^{(1)} \boldsymbol{x}\right) \otimes \boldsymbol{e}_{r_h} \circ \boldsymbol{w}^{(1)} \right] \tag{6.5}$$

where $\otimes$ denotes the outer product. $\circ$ denotes the component-wise product (no summation contraction). $\boldsymbol{e}_{r_h}$ is a vector of all ones with the dimension of r_h, and r_h is the number of neurons at the hidden layer. $a'(\bullet)$ is the derivative of the activation function at the hidden layer.

A detailed derivation of Equations 6.4 and 6.5 can found in the Appendix C.1. In essence, NN prediction is performed by fitting an analytical function, $\boldsymbol{y}(\boldsymbol{x})$, and the prediction accuracy after training can be improved as the size of the sample $\boldsymbol{t}(\boldsymbol{x})$ which is simulated by FE analysis increases. Here, an individual example of $\boldsymbol{t}(\boldsymbol{x})$ is termed as a target or label, to which the NN prediction, $\boldsymbol{y}(\boldsymbol{x})$, attempts to match. Compared with Figure 6.6, the implicit response function of FE analysis, $\boldsymbol{t}(\boldsymbol{x})$, has been converted to a continuously differentiable function, $\boldsymbol{y}(\boldsymbol{x})$, using the NN, as shown in Figure 6.8. It is obvious that optimization of $\boldsymbol{y}(\boldsymbol{x})$ is computationally cheaper than optimization of $\boldsymbol{t}(\boldsymbol{x})$, as performed in traditional design optimization methods.

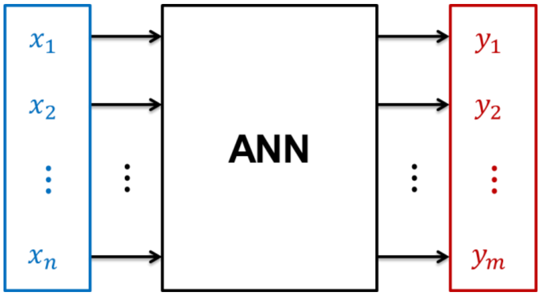


Figure 6.8: Illustrate the relationship between design variable and NN predicted performance metrics using artificial neural network (ANN).

6.2.3. Formulation of Tri-objective TDA Design Optimization Problem

TDA treatment using anterior surgical procedure may cause postsurgical complications such as excessive segmental rotation [8], spinal tissue overloading [9] and polyethylene inlay wear/fracture failures [10]. Aiming at these complications, the author proposed quantitative performance metrics (Equations 6.1, 6.2, and 6.3), used as objective functions in multiobjective optimization (MOO). Therefore, the problem of multiobjective TDA design optimization can be formulated as follows,

$$\overline{\min_{\boldsymbol{x}}}\, \boldsymbol{y}\,(\boldsymbol{x}) \tag{6.6}$$

subject to

$$\boldsymbol{x}_{lb} \leq \boldsymbol{x} \leq \boldsymbol{x}_{ub} \tag{6.7}$$

where $\overline{\min}$ denotes multiobjective optimization, instead of simultaneously / individually minimizing each objective. $\boldsymbol{y}$ is an m-dimensional (m-D) vector of design objectives (cost functions) in the objective space ($\mathbb{R}_{\boldsymbol{y}}^{m}$); here, a 3-D vector of design objectives, $\boldsymbol{y} = [y_1, y_2, y_3]^{\mathrm{T}}$, corresponding to three performance metrics of postoperative complications. $\boldsymbol{x}$ is an n-dimensional (n-D) vector of design variables in the design space ($\mathbb{R}_{\boldsymbol{x}}^{n}$), with a lower boundary ($\boldsymbol{x}_{lb}$) and upper boundary ($\boldsymbol{x}_{ub}$), respectively. For the TDA parametric design, the design space bounds within which no failures occur in model creation and meshing has been shown in Table 6.1. In this study, we avoided introducing inequality and equality constraints, since their gradients were also required in a constrained nonlinear optimization algorithm. In MOO, the cost functions, $\boldsymbol{y}(\boldsymbol{x}) = [y_1(\boldsymbol{x}), y_2(\boldsymbol{x}), y_3(\boldsymbol{x})]^{\mathrm{T}}$, were provided by the predictions of the artificial neural network described before.

6.2.4. Complete Pareto Frontier Multiobjective Optimization Algorithm

Using single-objective optimization (SOO), the single-objective optimum ($\boldsymbol{x}^{j*}$) of the design variables in the feasible design space ($\Omega_{\boldsymbol{x}}$) concerning each objective (y_j) can be obtained. The corresponding points in the feasible objective space ($\Omega_{\boldsymbol{y}}$) are defined as the anchor points ($\boldsymbol{y}^{j*}$),

$$\boldsymbol{y}^{j*} = \boldsymbol{y}(\boldsymbol{x}^{j*}) = [y_1(\boldsymbol{x}^{j*}) \quad y_2(\boldsymbol{x}^{j*}) \quad \cdots \quad y_m(\boldsymbol{x}^{j*})]^T \quad , \quad j = 1, 2, \cdots, m \tag{6.8}$$

By sequentially connecting these anchor points ($\boldsymbol{y}^{j*}$), the resulting plane is termed as the Utopian plane. In a 2-D objective space, the Utopian plane is a line segment, since there are only two anchor points. In a 3-D objective space, the Utopian plane is determined by a triangle consisting of the three anchor points.

Using the well-established normal-boundary interaction (NBI) method [118], the boundary ($\partial\Omega_{\boldsymbol{y}}$) of the feasible objective space where the Pareto frontier is located can be obtained by projecting the points on the Utopian plane along the plane normal vector. The NBI method is graphically described in Figure 6.9. The Pareto frontier is defined as the non-dominated boundary region, where no point outperforms another. However, it should be noted that the NBI method cannot ensure that all boundary points are non-dominated. To filter out the dominated points, a Pareto filtering algorithm [88, 119] needs to be used.

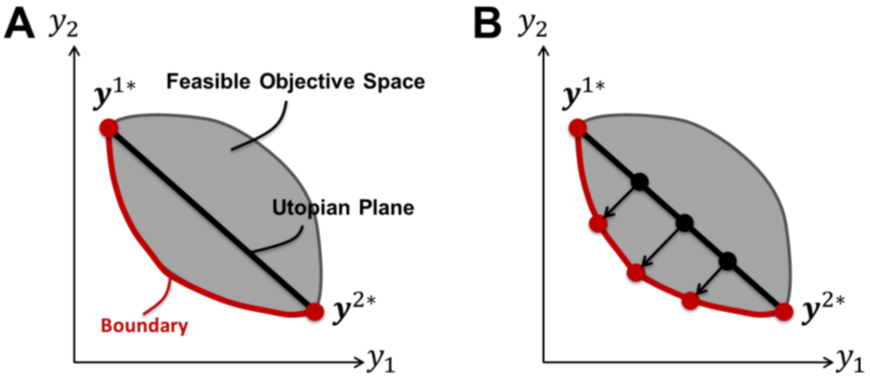


Figure 6.9: Description of the NBI method which is implemented considering two objectives. (a) A Utopian plane (line in the 2-D objective space) can be determined by anchor points ($\boldsymbol{y}^{1*}$ and $\boldsymbol{y}^{2*}$). (b) The boundary under the Utopian plane can be obtained by projecting the points on the Utopian plane along the plane normal vector.

The NBI can be stated as an optimization problem, in which the solution is the design variable vector ($\boldsymbol{x}$) corresponding to points on the boundary ($\partial\Omega_{\boldsymbol{y}}$) of the feasible objective space ($\Omega_{\boldsymbol{y}}$):

$$\max_{\boldsymbol{x},\rho} \rho \tag{6.9}$$

subject to

$$\boldsymbol{y}^{utp} + \rho\hat{\boldsymbol{n}} = \boldsymbol{y}(\boldsymbol{x}) \tag{6.10}$$

and Equation 6.7.

where an auxiliary scalar variable ($\rho \in [-\infty, \infty]$) is introduced. $\boldsymbol{y}^{utp}$ represents a point on the Utopian plane. $\hat{\boldsymbol{n}}$ is the unit projection vector in a direction towards the boundary ($\partial\Omega_{\boldsymbol{y}}$). $\boldsymbol{y}(\boldsymbol{x})$ is a point within the feasible design space ($\Omega_{\boldsymbol{y}}$), which is predicted using NN (Equation 6.4).

Therefore, the physical meaning of the NBI is that each point ($\boldsymbol{y}^{utp}$) on the Utopian plane is moved along the direction of $\hat{\boldsymbol{n}}$ as far as possible within the feasible objective space ($\Omega_{\boldsymbol{y}}$); the farthest location where $\boldsymbol{y}^{utp}$ can arrive represent a point on the boundary ($\partial\Omega_{\boldsymbol{y}}$).

In a high-dimensional ($m \geq 3$) objective space, the entire set of Pareto solutions may not be completely covered by the normal projection of the Utopian plane consisting of anchor points [88, 118, 119], as illustrated by an example in the 3-D objective space in Figure 6.10. Therefore, the Utopian plane needs to be enlarged such that a sufficiently large Utopian plane can cover the complete Pareto frontier.

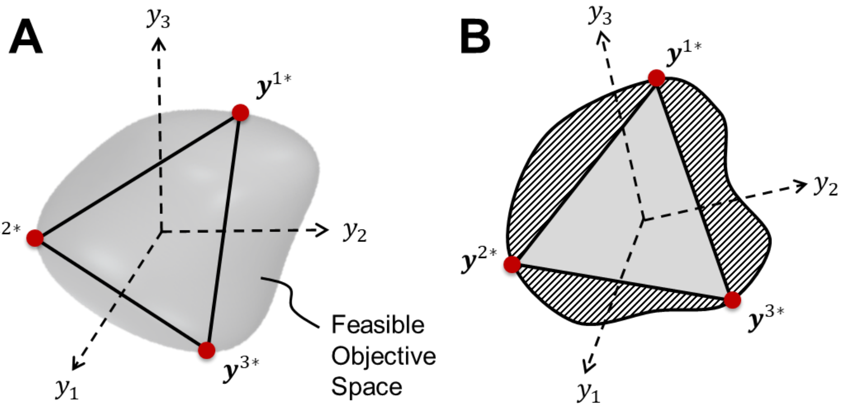


Figure 6.10: Illustration of the Utopian polygon in the 3-D feasible objective space, modified from [88] with Dr. Achille Messac's permission. (a) The Utopian plane defined by the triangle section consisting of three anchor points in the 3-D objective space; (b) Unobtainable Pareto solutions (hatched regions) which are located outside the normal projection of the Utopian polygon. The view direction has been rotated such that it is normal to the Utopian plane.

On the other hand, although the enlarged Utopian plane can result in the complete Pareto frontier, it increases the probability to explore useless regions where dominated points exist. To make the computation process more efficient, unnecessary Utopia plane regions that must be projected to the dominated boundary regions need to be eliminated. As a result, the optimal shape of the enlarged Utopian plane is commonly a polygon, which is called the Utopian polygon hereafter. The technique to efficiently expand the Utopian plane has been well established by Messac and Mattson (2004).

The size of the enlarged Utopian polygon actually depends on how we define the ranges of the boundary of the feasible objective space. To adjust the size of the Utopian polygon, we can select the coordinates of either pseudo Nadir point ($\boldsymbol{y}^P$) or Nadir point ($\boldsymbol{y}^N$) to define the upper limits of the boundary region. Here, we further introduce a concept of the mixed Nadir point ($\boldsymbol{y}^{P/N}$) whose coordinates can be arbitrarily chosen from those of either pseudo Nadir point ($\boldsymbol{y}^P$) or Nadir point ($\boldsymbol{y}^N$). Using the mixed Nadir point ($\boldsymbol{y}^{P/N}$), it allows us to more flexibly control the size of the extended Utopian polygon. In Figure 6.11, we list three Utopian polygons which were adopted in this study. Their sizes were determined by the pseudo Nadir point ($\boldsymbol{y}^P$), the Nadir point ($\boldsymbol{y}^N$) and a mixed Nadir point ($\boldsymbol{y}^{P/N} = [y_1^N \quad y_2^P \quad y_3^N]^T$), respectively.

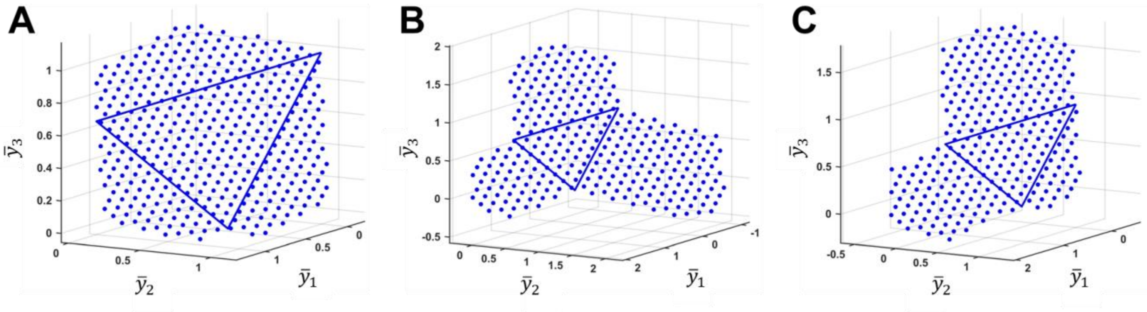


Figure 6.11: Different Utopian polygons in the 3-D objective space. They are controlled by the pseudo Nadir point ($\boldsymbol{y}^P$) (a), the Nadir point ($\boldsymbol{y}^N$) (b), and a mixed Nadir point ($\boldsymbol{y}^{P/N} = [y_1^N \quad y_2^P \quad y_3^N]^T$) (c), respectively.

Since the points (Figure 6.11) on the Utopian polygon have been pre-defined using a discretization technique (Appendix C.2.3), we can estimate the positions of the boundary points resulting from the normal projection of the points on the Utopian polygon. Therefore, a triangular mesh can be maintained to visualize the boundary of the feasible objective space during the NBI procedure, as shown in Figure 6.12. More details about the development of the MOO algorithm to generate an even representation of the complete Pareto frontier can be found in Appendix C.2.

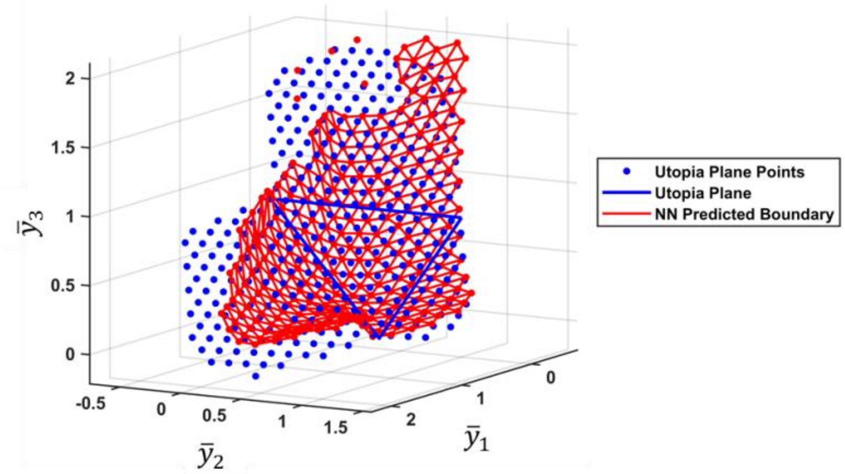


Figure 6.12: Visualization of the boundary of the feasible objective space using a triangle mesh. Using the NBI method, the points (blue) on the Utopian polygon are projected to the boundary (red). It is noted that the three anchor points determine the Utopian plane (blue triangle), which has been enlarged to a Utopian polygon defined by a mixed Nadir point.

6.3. Learning Procedure for Pareto-optimal TDA Designs

A three-layer feed-forward artificial neural network (NN) with 10 inputs and 3 outputs was created in MATLAB R2014a using the default settings of the NN toolbox (the hyperbolic tangent activation function was assigned to the hidden layer, and no activations existed at the input and output layers). An initial dataset consisting of 764 feature ($\boldsymbol{X}_{10\times764}$) – label ($\boldsymbol{T}_{3\times764}$) examples was generated by simulating the responses of the TDA-treated segmental in FE analyses, in which TDA design variables ($\boldsymbol{x}$) were randomly chosen within the ranges (Table 6.1). Subsequently, the boundary of the feasible objective space was learned 50 times, and the number of examples was boosted to 6292 in the end. The prediction accuracy of each NN output during the boundary learning was monitored using regression analysis (MATLAB R2014a). In each boundary learning, the following procedures were carried out.

6.3.1. Adaptive Neural Network Training

In NN training, the example dataset was randomly divided into a training set (70% examples), a validation set (15% examples) and a test set (15% examples), using the default setting of the NN toolbox in MATLAB R2014a. To achieve the best tradeoff of bias and variance, cross-validation has been automatically performed during the mini-batch NN training process in the NN toolbox. We further tuned the number of neurons at the hidden layer, r_h, such that the mean squared error of the test set was minimized. As the number of examples increased from 764 to 6292, the tuned number of hidden neurons increased from 15 to 75.

6.3.2. Generation of Feasible Objective Space Boundary Points

Using the NBI-based complete Pareto frontier MOO algorithm (Equations 6.9, 6.10, and 6.7), the boundary points (including dominated and non-dominated points) of the feasible objective space were generated. In each MOO, 50 ~ 300 boundary points were solved, depending on the size and the grid density of the Utopian plane (Figure 6.11). In the last 10 learnings, the mixed Nadir point Utopian polygon (Figure 6.11c) was adopted, because the boundary region of $y_2 \geq y_2^P$ was always dominated. It is important to realize the non-convexity of the cost functions, $\boldsymbol{y}(\boldsymbol{x})$. Therefore, to solve each boundary point, optimization in the NBI method (Equations 6.9, 6.10, and 6.7) was performed 20 times by random initialization of design variables; only the solution corresponding to the maximum value of the projection distance (ρ) was considered as the boundary point.

6.3.3. Validation of NN predictions and NN Sample Expansion

After MOO, the resulting TDA designs ($\boldsymbol{X}^{NBI}$) regarding all boundary points ($\boldsymbol{Y}^{NBI}$) were input into the TDA-treated segment FE model, to simulate the ROM, FJF, and PCP metrics using FE analysis, and thus generated additional labels ($\boldsymbol{T}^{NBI}$). The new feature ($\boldsymbol{X}^{NBI}$) – label ($\boldsymbol{T}^{NBI}$) dataset was fed back to the NN in the next boundary learning. It should be highlighted that we trained the NN only using the dataset of boundary points (a subset of the feasible objective space), after providing an initial sample of 764 examples by simulating the responses of random TDA designs.

6.3. Results

6.3.1. Neural Network Prediction Accuracy for Test Set

The NN prediction accuracy for the test set was evaluated using regression analysis (MATLAB R2014a). It should be noted that the test set was not used for NN training, so it is the best criterion to evaluate the NN prediction capability. The linear association between the NN predictions and the labels in the test set is measured by the correlation coefficient. For the initial dataset which was generated by simulating the responses of random TDA designs, the results of linear regression analyses for each TDA performance metric were presented in Figure 6.13. The correlation coefficients of the ROM, FJF, and PCP metrics were 0.956, 0.967, and 0.837, respectively.

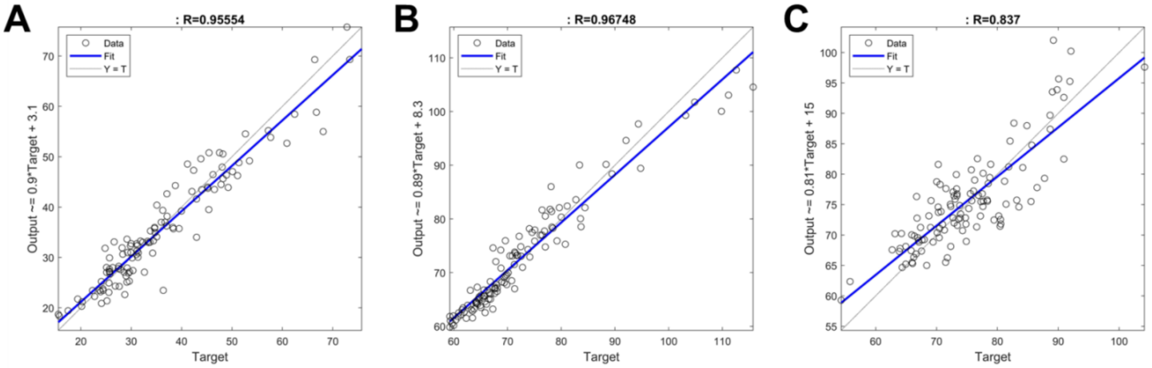


Figure 6.13: The regression analyses between the predictions and labels in the test set generated by simulating the responses of random TDA designs. (a) The ROM metric; (b) The FJF metric; (c) The PCP metric.

After machine learning of the feasible objective space boundary 50 times, the number of examples in the dataset (including the training set, validation set, and test set) was increased from

764 to 6292. During each NN training, the boundary point designs generated from the previous learning were fed to the NN, after they were validated by FE analaysis. As shown in Figure 6.14, the prediction accuracy of each performance metric in the test set was significantly improved at the last boundary learning, especially for the PCP metric. The correlation coefficients of the ROM, FJF, and PCP metrics have reached 0.998, 0.996, and 0.991, respectively.

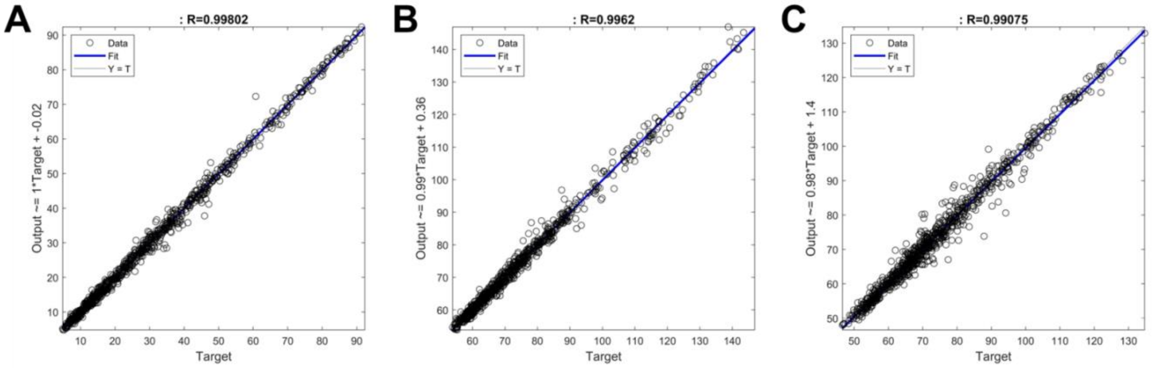


Figure 6.14: The regression analyses between the predictions and labels in the test set after 50 times of boundary learning. (a) The ROM metric; (b) The FJF metric; (c) The PCP metric.

6.3.2. NN predicted and FE predicted 3-D Pareto Frontiers

The complete Pareto frontier MOO algorithm was implemented to generate the boundary of the 3-D feasible objective space, in which each point was predicted using the NN. As shown in Figure 6.15, the dominated boundary regions (green) were filtered out, and the remaining boundary regions (red) represented the non-dominated boundary (Pareto frontier). It can be observed that the Utopian triangle consisting of three anchor points cannot cover the complete Pareto frontier in the 3-D feasible objective space. Especially, the bottom left Pareto frontier shown in Figure 6.15a is far beyond the connecting line between $\boldsymbol{y}^{2*} = [54.05\ 53.65\ 89.92]^{T}$ and $\boldsymbol{y}^{3*} = [38.90\ 79.15\ 47.22]^{T}$.

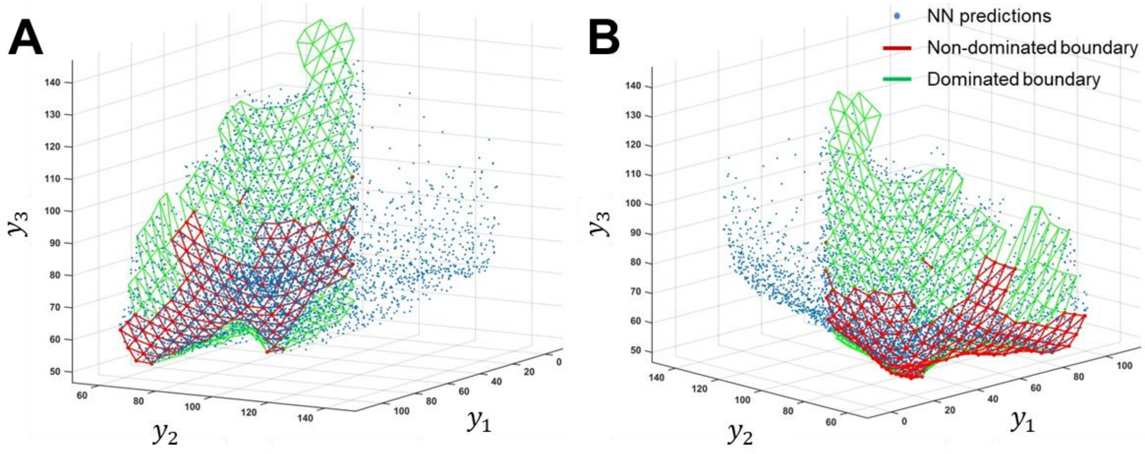


Figure 6.15: The NN predicted 3-D Pareto frontier in the 3-D feasible objective space. (a) The view from the Nadir point; (b) The view from the Utopian point.

In comparison, the FE predicted Pareto frontier in the 3-D feasible objective space was also obtained by filtering out dominated points, as presented in Figure 6.16. It should be understood that the NN predicted Pareto frontier (Figure 6.15) is a fit for the discrete FE predicted Pareto frontier (Figure 6.16). It can be found that there is a mismatch between the NN predicted and FE predicted Pareto frontiers, especially located at the border between the non-dominated boundary and the dominated boundary, where the curvature of the boundary surface changes drastically.

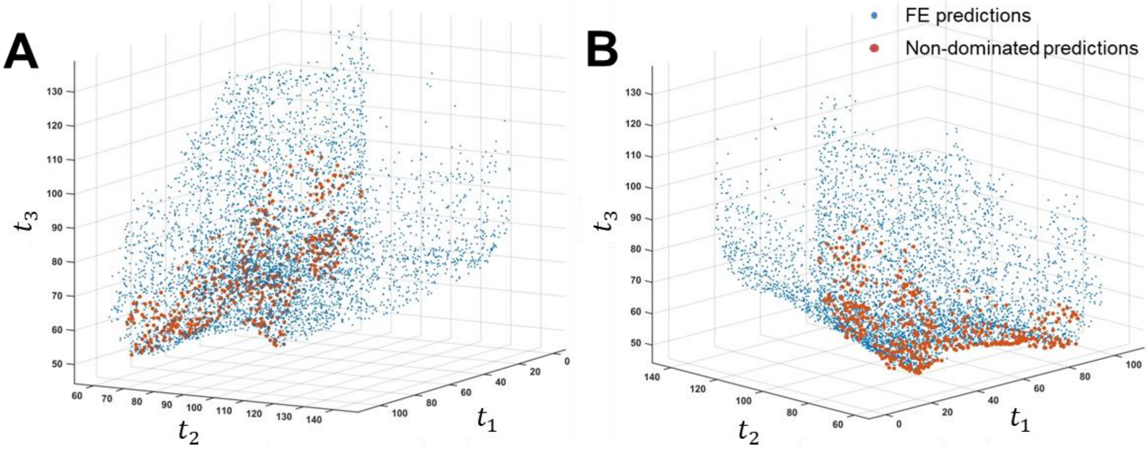


Figure 6.16: The FE predicted 3-D Pareto frontier in the 3-D feasible objective space. (a) The view from the Nadir point; (b) The view from the Utopian point.

The projection of the NN or FE predicted 3-D feasible objective space onto the coordinate plane of the ROM (y_1 or t_1) and FJF (y_2 or t_2) metrics is shown in Figure 6.17. The frontier of the projection represents a 2-D Pareto curve with respect to the ROM and FJF metrics. The NN predicted 2-D Pareto frontier (Figure 6.17a) can well fit that predicted by FE analysis (Figure 6.17b). It clearly shows that the NN predicted boundary region of $y_2 \geq y_2^P = 87.80$ is always dominated, consistent with that of $t_2 \geq t_2^P = 86.77$.

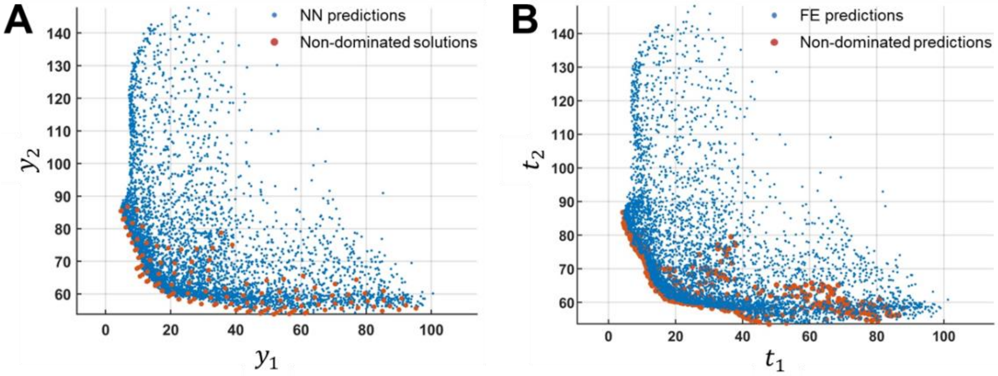


Figure 6.17: Comparison of the projections of the 3-D NN predicted and FE predicted feasible objective spaces on the coordinate plane of the ROM (y_1 or t_1) and FJF (y_2 or t_2) metrics. (a) The projection of all NN predictions on the y_1–y_2 coordinate plane; (b) The projection of all FE predictions on the t_1–t_2 coordinate plane.

6.3.3. Pareto-optimal TDA Design Variables

The single-objective optimum TDA designs including ROM-optimal, FJF-optimal, and PCP-optimal TDA designs correspond to the three anchor points ($\boldsymbol{y}^{1*}$, $\boldsymbol{y}^{2*}$, and $\boldsymbol{y}^{3*}$) in the feasible objective space. Furthermore, the best-tradeoff TDA design is defined as a Pareto-optimal TDA design, corresponding to a point in the feasible objective space in the minimum Euclidean distance away from the Utopian point ($\boldsymbol{y}^U$). These Pareto-optimal TDA designs are presented in Figure 6.18 and their design variables are listed in Table 6.2. Especially, the rim congruence of the PCP-optimal design at the anterior, lateral and posterior on the inferior MoP articulation is highest ($g_{ia} = 0.01$ mm, $g_{il} = 0.11$ mm, and $g_{ip} = 0.29$ mm), indicating that the mobile-bearing TDA design has been almost reduced to a convex-down fixed-bearing TDA design. The best-tradeoff TDA design is similar to the PCP-optimal design, but features larger gaps ($g_{ia} = 0.42$ mm, $g_{il} = 0.24$ mm, and $g_{ip} = 0.47$ mm) on the inferior rim articulation and the approximately complete rim congruence ($g_{sp} = 0.08$ mm) on the posterior of the superior rim articulation.

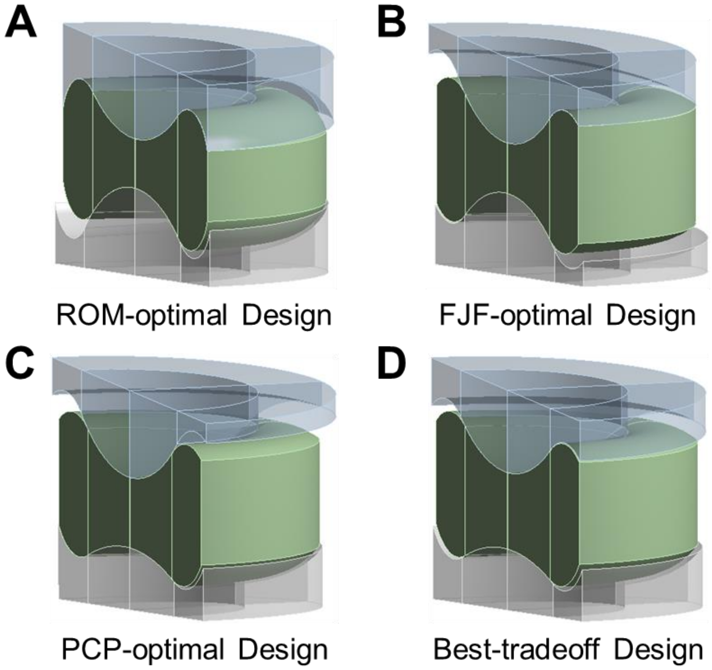


Figure 6.18: The ROM-optimal (a), FJF-optimal (b), PCP-optimal (c) and best-tradeoff (d) TDA designs.

Table 6.2: The design variables of the ROM-optimal, FJF-optimal, PCP-optimal and best-tradeoff TDA designs. The unit is mm.

	r_s	r_i	d_s	d_i	g_{sa}	g_{ia}	g_{sl}	g_{il}	g_{sp}	g_{ip}
ROM-optimal	11.39	11.00	4.10	4.05	0.01	1.20	0.48	0.16	0.01	0.42
FJF-Optimal	11.00	11.79	3.10	3.10	1.20	0.12	0.17	0.58	0.01	0.80
PCP-optimal	11.00	19.18	3.10	3.10	0.95	0.01	1.00	0.11	0.80	0.29
Best-tradeoff	11.71	20.00	3.10	3.10	0.69	0.42	0.61	0.24	0.08	0.47

6.3.4. Pareto-optimal TDA Performance Metrics

The ROM, FJF and PCP metrics of each Pareto-optimal TDA design predicted by the NN and FE analysis, are listed in Table 6.3. Generally, the NN predicted metrics of these Pareto-optimal TDA designs closely match the corresponding FE predicted metrics, except for the PCP metric ($y_3 = 99.76$ vs. $t_3 = 87.19$) of the ROM-optimal TDA design, and the ROM metric ($y_1 = 54.05$ vs. $t_1 = 47.89$) and the PCP metric ($y_3 = 80.92$ vs. $t_3 = 70.05$) of the FJF-optimal TDA design. For the best-tradeoff TDA design, the NN predicted metrics are in excellent agreement with the FE predicted metrics.

In Table 6.3, the diagonal metrics (in bold) of the ROM-optimal, FJF-optimal and PCP-optimal TDA designs represent the optimum ROM, FJF and PCP metrics using single-objective optimization. These single-objective optimal TDA designs cause a substantial sacrifice of other metrics. For example, the maximum FJF metric ($y_2 = 87.80$ vs. $t_2 = 86.77$) and PCP metric ($y_3 = 99.76$ vs. $t_3 = 87.19$) are achieved simultaneously for the ROM-optimal design, and the FJF-optimal design causes the maximum ROM metric ($y_3 = 54.05$ vs. $t_3 = 47.89$). In contrast, the metrics of the best-tradeoff TDA are within the metric ranges of these single-objective optimal TDA designs.

Table 6.3: The NN predicted and FE predicted ROM (y_1 or t_1), FJF (y_2 or t_2), and PCP (y_3 or t_3) metrics of the ROM-optimal, FJF-optimal, PCP-optimal, and best-tradeoff TDA designs.

	ROM Metric		FJF Metric		PCP Metric	
	y_1	t_1	y_2	t_2	y_3	t_3
ROM-optimal Design	**4.46**	**4.33**	87.80	86.77	99.76	87.19
FJF-optimal Design	54.05	47.89	**53.65**	**53.53**	80.92	70.05
PCP-optimal Design	38.90	37.72	79.15	77.55	**47.22**	**46.28**
Best-tradeoff Design	16.09	14.96	66.94	65.79	56.88	57.57

The competing relationship among these performance metrics is further revealed in the FE model-predicted responses of lumbar segments treated by the four Pareto-optimal TDA designs, including the simulated ROMs (Figure 6.19), FJFs (Figure 6.20), and PCPs (Figure 6.21). Although neither the closest ROMs and FJFs to those of the intact segment nor the minimum PCPs did the best-tradeoff TDA design lead to in different loading scenarios, no excessive sacrifices of these performance metrics occurred, as compared to other single-objective optimum TDA designs.

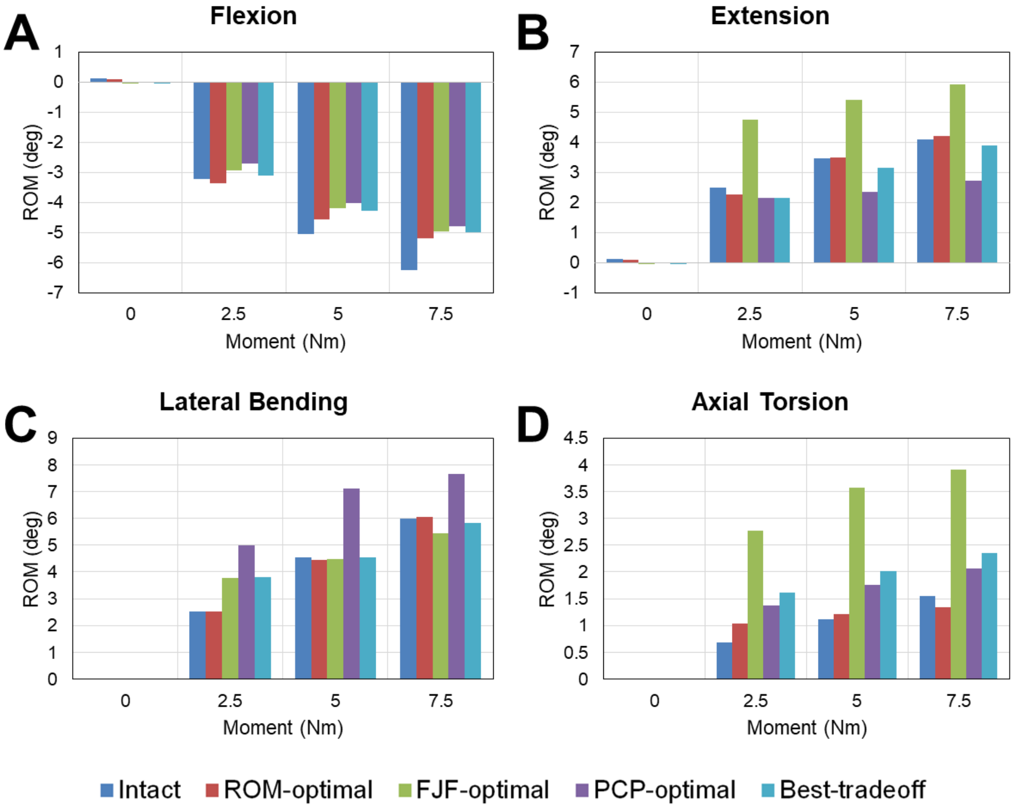


Figure 6.19: The FE model-predicted ROMs of the intact lumbar segment, as well as those treated by the ROM-optimal, FJF-optimal, PCP-optimal and best-tradeoff TDA designs in different loading scenarios. “0 Nm” indicates the ROMs after preloading. (a) Flexion; (b) Extension; (c) Lateral bending; (d) Axial torsion.

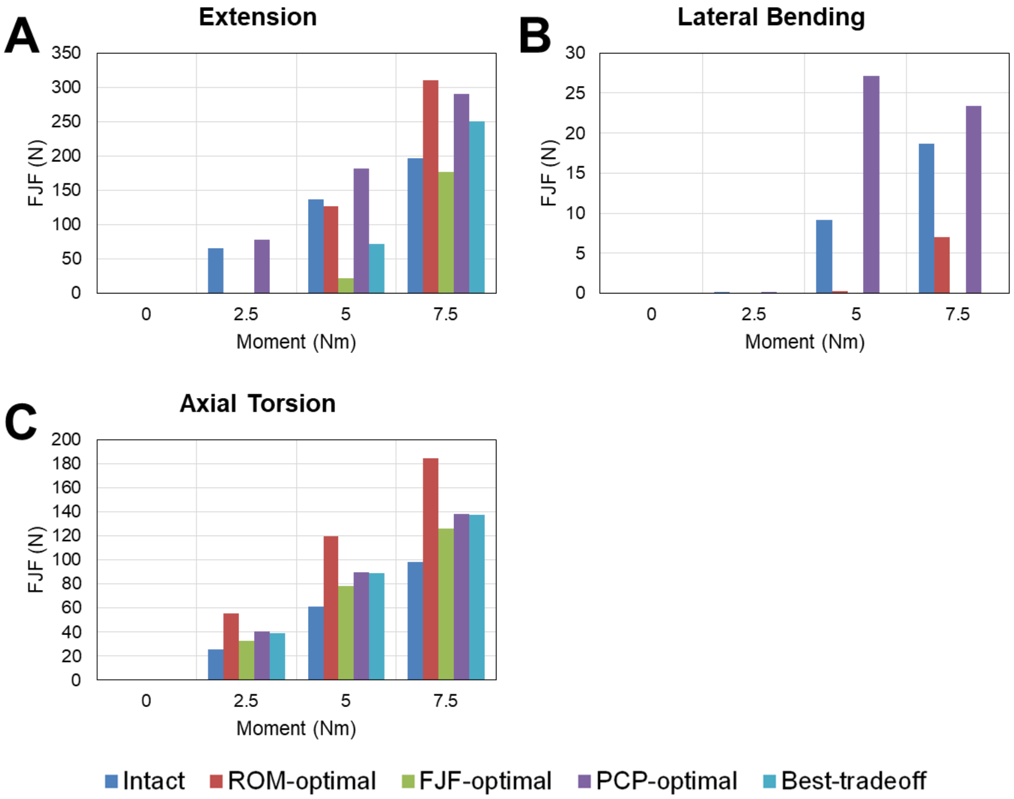


Figure 6.20: The FE model-predicted total FJFs of the left and right facet joints in the intact lumbar segment, as well as those treated by the ROM-optimal, FJF-optimal, PCP-optimal and best-tradeoff TDA designs in different loading scenarios. It is noted that all facet joints in the intact and TDA-treated segments remain separated in flexion. “0 Nm” indicates the FJFs after preloading. (a) Extension; (b) Lateral bending; (c) Axial torsion.

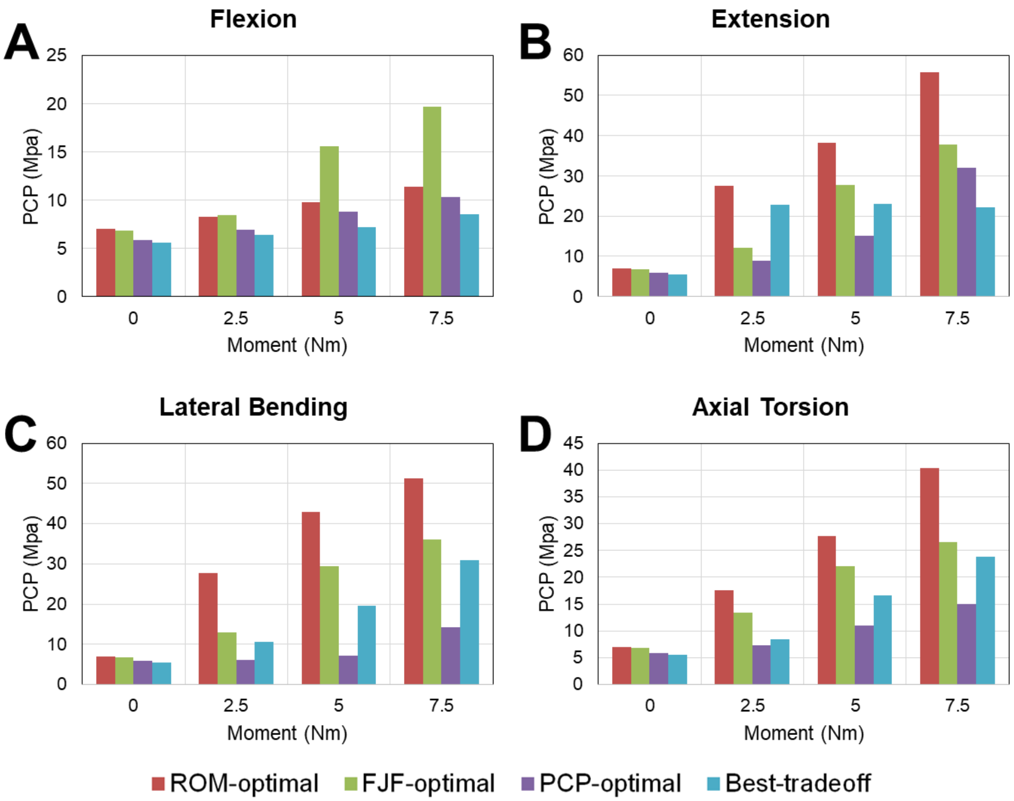


Figure 6.21: The FE model-predicted maximum PCPs on the MoP articulations of the ROM-optimal, FJF-optimal, PCP-optimal and best-tradeoff TDA designs in different loading scenarios. "0 Nm" indicates the PCPs after preloading. (a) Flexion; (b) Extension; (c) Lateral bending; (d) Axial torsion.

6.3.5. Slicing of 3-D Pareto Frontier

The slicing algorithm developed in Appendix C.2.4 was implemented in the dimensionality reduction of the 3-D objective space, to better understand the 3-D Pareto frontier. As shown in Figure 6.22, the NN predicted 3-D Pareto frontier is sliced using three planes of y_1 = 15, y_1 = 20, and y_1 = 40, respectively, such that their projections onto the y_2-y_3 coordinate plane (Figure 6.22c) are equally spaced. It indicates that along the y_1 coordinate, a larger sacrifice in the ROM metric (y_1) is required, in order to improve the others.

In each slice (Figure 6.22c), although some non-dominated points are not obtained in the region corresponding to the 3-D Pareto frontier, it reveals that a near 1:1 tradeoff between the FJF (y_2) and PCP (y_3) metric exists at the bottom ($y_3 < 65$) of each slice. At the upper portion ($y_3 >$

65) of each slice, however, this tradeoff is not balanced; the PCP metric (y_3) is substantially sacrificed to achieve a small improvement in the FJF metric (y_2).

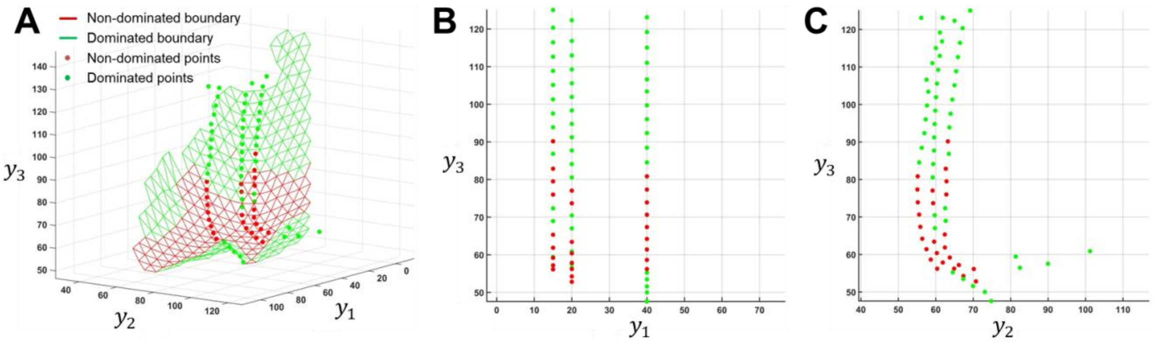


Figure 6.22: Slicing of the NN predicted 3-D Pareto frontier using three planes of y_1 = 15, y_1 = 20, and y_1 = 40. (a) The oblique view from the Nadir point; (b) The projection onto the y_1-y_3 coordinate plane, where all points located at the three slicing planes, respectively; (c) The projection onto the y_2-y_3 coordinate plane, where the slices from left to right are y_1 = 40, y_1 = 20, and y_1 = 15, respectively.

Furthermore, linear regression analysis of the PCP metric (y_3) on the Pareto frontier sliced by the plane of y_1 = 15 with respect to each design variable was performed, to investigate the correlation between independent Pareto-optimal TDA design variables and the dependent Pareto-optimal PCP metric (y_3) when the ROM metric was fixed at y_1 = 15. As presented in Figure 6.22c, nine non-dominated points on the slice of y_1 = 15 were generated using the slicing algorithm. According to the results of regression analysis (Table 6.4 and Figure 6.23), the inferior dome radius of curvature, r_i (R = -0.97), the anterior rim gap on the superior MoP articulation, g_{sa} (R = 0.94), the anterior rim gap on the inferior MoP articulation, g_{ia} (R = 0.80), and the lateral rim gap on the inferior MoP articulation, g_{il} (R = -0.95) are linearly associated with the PCP metric (y_3) in the Pareto frontier slice at y_1 = 15.

Table 6.4: The linear association between the design objectives (metrics) at the slice of y_1 = 15 and design variables. The correlation coefficient is denoted by R.

	y_1	y_2	y_3	r_s	r_i	d_s	d_i	g_{sa}	g_{ia}	g_{sl}	g_{il}	g_{sp}	g_{ip}
Design Objectives / Variables	15.00	63.24	90.16	11.00	11.00	3.10	3.10	1.20	0.84	0.40	0.02	0.01	0.58
	15.00	62.67	82.86	11.00	11.00	3.10	3.10	1.20	0.73	0.38	0.05	0.01	0.57
	15.00	62.87	79.51	15.82	11.82	3.42	3.10	1.20	0.86	1.00	0.05	0.01	0.42
	15.00	62.82	76.00	15.68	12.10	3.33	3.10	1.20	0.82	1.00	0.07	0.01	0.42
	15.00	62.42	65.34	11.00	18.32	3.10	3.10	0.93	0.66	0.38	0.18	0.01	0.73
	15.00	62.44	61.87	11.00	18.15	3.10	3.10	0.81	0.52	0.36	0.25	0.01	0.67
	15.00	63.74	59.19	11.00	19.01	3.10	3.10	0.63	0.55	0.41	0.33	0.01	0.60
	15.00	66.17	57.19	11.17	20.00	3.10	3.10	0.62	0.59	0.45	0.31	0.11	0.57
	15.00	70.16	56.13	11.00	20.00	3.10	3.10	0.59	0.69	0.44	0.26	0.36	0.53
R	0.00	-0.55	1.00	0.35	-0.97	0.37	0.00	0.94	0.80	0.32	-0.95	-0.51	-0.32

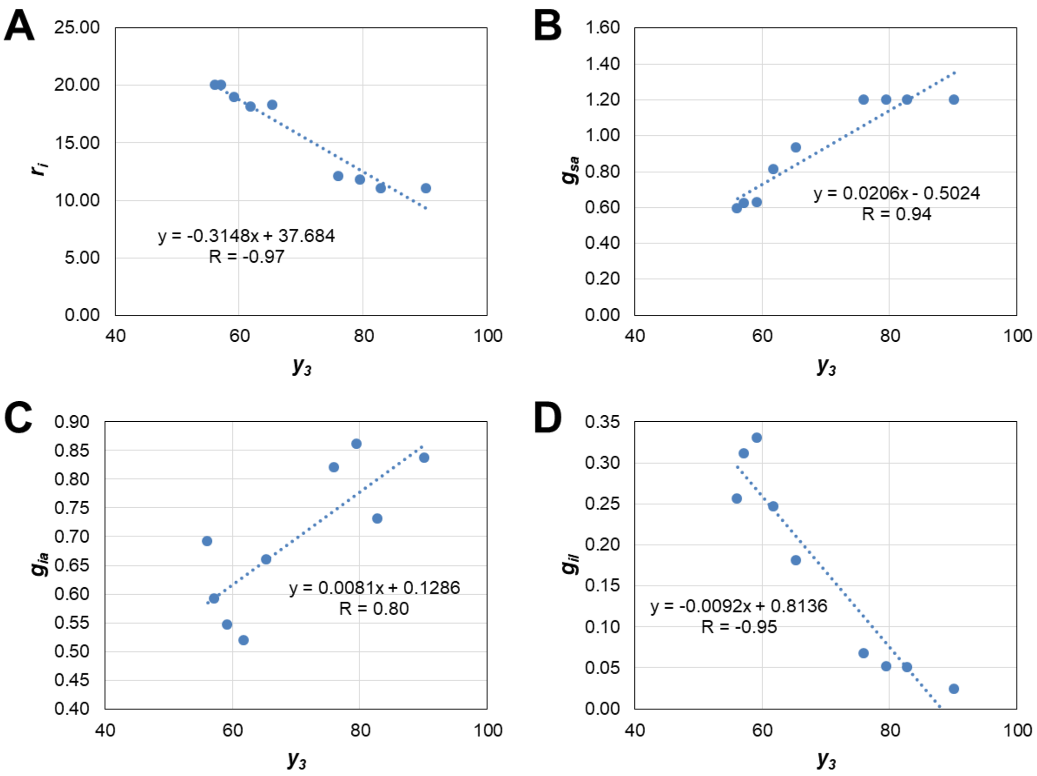


Figure 6.23: The linear regression of design variables (r_i, g_{sa}, g_{ia} and g_{il}) with respect to the design objectives (metrics) at the slice of y_1 = 15.

6.4. Discussion and Conclusion

In this study, a neural network based multiobjective optimization algorithm was developed to refine a new mobile-bearing TDA design featuring a biconcave polyethylene core, which strengths the polyethylene rim where a high risk of polyethylene fracture/wear exists due to the impingent on the MoP articulations. After multiobjective optimization, a TDA design (the best-tradeoff TDA design) was achieved without excessively compromising any one of ROM, FJF, and PCP performance metrics. It indicates that the best-tradeoff TDA design may result in a lower incidence of common complications as reported in the clinic results [71, 120–122] and retrieval studies [6, 10, 76, 77] following the surgical treatments using commercial available lumbar TDAs. Furthermore, we demonstrated that the responses of the TDA-treated lumbar segment were very sensitive to TDA designs. TDA can be optimally designed to match the native intervertebral disc biomechanics, but an arbitrary TDA design may cause abnormal postoperative spinal responses and a higher risk of implant failures. Therefore, it should be realized that design optimization of TDA is essential to compete with the motion-sacrificing surgical technique (*e.g.*, spinal fusion) which is the gold standard treatment for spinal disorders nowadays. Several aspects regarding the work are discussed as follows.

6.4.1. Fitting Accuracy of 3-D Pareto Frontier

After providing an initial sample of 764 examples, the NN predicted TDA designs ($\boldsymbol{X}^{NBI}$) corresponding to the boundary points ($\boldsymbol{Y}^{NBI}$) were simulated by FE analysis in each learning, and the resulting dataset ($\boldsymbol{X}^{NBI}$ and $\boldsymbol{T}^{NBI}$) were fed back to the NN in the next learning. Although only the validated data of boundary points were used to train the NN during the learning procedure, the NN prediction accuracy for all performance metrics was significantly increased compared to the initial prediction accuracy. It demonstrates that the NN prediction accuracy can be improved more strategically by only providing the validated boundary points of the feasible objective space, instead of attempting to cover the entire feasible objective space (*e.g.*, using random initialization of design variables), which requires more computational resources and time.

As the dataset validated by FE analysis became sufficiently large, it revealed that the region of $t_2 \geq t_2^P$ in the feasible objective space was always dominated. Hence, it is not necessary to further improve the prediction accuracy for the region of $y_2 \geq y_2^P$. On the other hand, a Utopian polygon (Figure 6.11a) defined by a pseudo Nadir point ($\boldsymbol{y}^P$) as recommended by Messac and Mattson (2004) was not large enough to cover the complete Pareto frontier (Figure 6.15 or 6.16) in the TDA multiobjective optimization problem. To satisfy both demands, we introduced a Utopian polygon defined by a mixed Nadir point ($\boldsymbol{y}^{P/N} = [y_1^N \quad y_2^P \quad y_3^N]^T$, Figure 6.11c) to generate the boundary points of the feasible 3-D objective space.

It is worth noting that we trained the NN using both the non-dominated (Pareto) and the dominated boundary points. These dominated boundary points were helpful to improve the prediction accuracy for a local region on the boundary of the feasible objective space, and facilitated the judgement of the non-dominated boundary points, since the Pareto filtering algorithm (Appendix C.2.6) was conducted by comparing the coordinates of a concerned point with those of all boundary points.

It can be found that there is a mismatch between the shapes of the Pareto frontiers predicted by the NN (Figure 6.15) and FE analysis (Figure 6.16), primarily located at the border between the non-dominated boundary and the dominated boundary, where the curvature of the boundary surface tends to be infinite. For the same reason, distinct discrepancies may occur in the NN predicted performance metrics of single-objective optimum TDA designs corresponding to three anchor points from those predicted by FE analysis (Table 6.3). However, the NN predicted metrics are in excellent agreement with the FE predicted metrics for the best-tradeoff TDA design. To better fit the FE predicted Pareto frontier, a denser label dataset at these steep local regions is required to enhance the NN prediction accuracy, such that Pareto filtering can correctly distinguish the non-dominated and dominated boundary points.

6.4.2. Pareto-optimal TDA Designs and Performance Metrics

In Chapter 5.2, we have demonstrated that the convex-down fixed bearing TDA design led to a better motion pattern and lower polyethylene stress due to the MoP articulation, compared with other commercially available TDA designs. By optimizing the PCP metric, the resulting PCP-optimal TDA design (Figure 6.18c and Table 6.2) has the highest congruence at the inferior rim articulation, similar to the convex-down fixed bearing TDA design. Therefore, it indirectly validates the robustness of the proposed NN based multiobjective optimization method.

Using single-objective optimization, the ROM-optimal, FJF-optimal, and PCP-optimal TDA designs can achieve the optimal performance for a single objective function, but other objectives are sacrificed (Table 6.3 and Figures 6.19, 6.20, and 6.21). In contrast, no performance metrics are excessively compromised for the best-tradeoff TDA design. Therefore, in order to get more confidence for choosing the candidate implant design, more performance metrics need to be considered in MOO. Our NN based MOO algorithm can be further expanded to incorporate more performance metrics, and generate the Pareto frontier in the high-dimensional ($m > 3$) objective space. Although the best-tradeoff TDA design was focused in the study, it is important to acknowledge that we do not just pull the best tradeoff out of MOO. The MOO algorithm also enables us to quantify how much we need to sacrifice in one area to improve in another, so preference can be applied more intelligently.

The development of an analytic expression for the relationship between Pareto-optimal design variables and performance metrics is attractive but challenging. As shown in Figure 6.18c and Table 6.2, different optimum points on the Pareto frontier correspond to different TDA design variables, indicating that 3 performance metrics are highly associated with 10 design variables. However, the PCP metric is linearly associated with only 4 design variables (Table 6.4 and Figure 6.23), when a slice of the Pareto frontier at $y_1 = 15$ was investigated. It implies that the function (Figure 6.6 or 6.8) mapping the Pareto-optimal design space to the Pareto-optimal objective space is highly nonlinear. To fit this complex map function, the number of the hidden neurons in the NN was tuned to 75, when the NN sample size increased to 6292. For a point on the Pareto frontier between known optimum points obtained using the NBI method, the corresponding Pareto-optimal design variables cannot be simply determined using linear interpolation. The correct approach is to densify the grid on the Utopian plane, such that denser Pareto boundary points and their corresponding design variables can be yielded using the NBI method.

6.4.3. Limitations

In multiobjective optimization of the new biconcave mobile-bearing TDA design, some assumptions and simplifications have been made and they require further validations. First, only a TDA-treated segment FE model was adopted to simulate the ROM, FJF, and PCP metrics, so the responses of adjacent segments were not be incorporated in MOO. In the previous work (Chapter 5.1) to investigate the behaviors of intact and treated multi-segment lumbar spines, we found that no differences in the simulated ROMs, FJFs, and ligament forces occurred at the adjacent levels using the load-control protocol. Therefore, we reduced the multi-segment spine model to a single segment model. It is well known that the *in-vivo* loading conditions are complicated and a hybrid loading protocol [73] may be more suitable for considering the muscular co-activation. However, we assumed that no abnormal responses at the adjacent levels existed, if TDA biomechanics were able to accurately reproduce the native intervertebral disc biomechanics. To match native disc biomechanics, the ROM and FJF metrics at the treated level have been introduced to multiobjective optimization.

Furthermore, the vertebrae and TDA metallic endplates were modeled as rigid bodies to save computational cost. This simplification has been made in previously reported lumbar spine FE models [25, 123]. Comparing the simulation results in Chapters 5.1.2 and 5.2.2, we indeed found that the formulation of rigid body caused a decrease in the segmental ROM and an increase in the polyethylene contact pressures and facet joint forces. However, it was impossible to change the general trend when the responses of the TDA-treated segment model were optimized to those of the intact segment model.

In this study, a representative lumbar segment FE model was created according to well-documented anatomic dimensions [14, 15], and the kinematical responses of the intact segment FE model (using material properties of deformable vertebrae) were calibrated to *in-vitro* experimental measurements in various loading scenarios [28] using a step-wise addition procedure of functional spinal structure [16, 33, 34]. However, we did not consider the inter-individual variations in the spine anatomy, inter-segment alignment, and material properties of spinal tissues. However, we have performed a sensitivity analysis of segment responses (ROM, FJF, and PCP) to varied ligament properties, which was documented in Appendix D.3. It was demonstrated that variations in ligament properties caused slight changes in segment responses, indicating that discs and TDAs are the main load-carrying components. In addition, it is essential to further introducing statistical modeling [92, 124–126] to incorporate these inter-individual variations into the computational FE models. The coefficients of their principal components of these probability distributions can be added to the NN inputs to fit the responses of the instrumented lumbar spine to these inter-individual variations. The future work would create a more complex neural network based surrogate model [127], and perform multiobjective optimization of patient-specific TDA designs.

Chapter 7: Summary and Conclusion

7.1. Accomplishments

The overarching goal of this work was to perform multiobjective design optimization of a new biconcave mobile-bearing total disc arthroplasty considering spinal kinematics, facet joint loading, and metal-on-polyethylene contact mechanics in an instrumented lumbar segment. A set of necessary sub-objectives were outlined in Chapter 1.2, each of which has been successfully completed in order to meet the proposed overarching goal.

7.1.1. Sub-objective 1: Investigation of Native Disc Biomechanics Using Finite Element Analysis

The material anisotropy and heterogeneity across the AF of the intervertebral disc were modeled based on a new constitutive model of collagen fibers, which enables the fitting and interpolation of experimentally measured stress-stretch curves of single annular lamellae. The pull-back computation and disc prestrain simulation were used to investigate the effect of nucleus swelling on the disc geometry, fiber orientation and mechanics. In the previously reported simulation, only the unloaded AF configuration due to the loss of the IDP was derived, without including the NP. The prestrained disc was modeled by applying an initial strain distribution (equivalent to that when an IDP was directly applied on the internal surface of the AF) to the AF. By introducing thermoelastic analysis, we demonstrated that thermal expansion can provide an accurate representation of the disc prestrain under nucleus swelling. This new technique allows us to derive the unloaded configuration of an entire disc when the IDP is lost. Furthermore, we simulated the responses of the normal (prestrained) disc and the degenerated (unloaded) disc in various physiologically relevant loading scenarios. It was concluded that the disc stiffness is regulated by the IDP and fiber orientation; a decrease in the IDP and fiber angles causes a distinct deterioration in the resistance to intervertebral rotations.

7.1.2. Sub-objective 2: Finite Element Modeling of Multi-segment Lumbar Spine

A representative intact multi-segment lumbar spine FE model was parametrically created according to well-documented anatomic dimensions. The intervertebral discs were modeled considering the material anisotropy and heterogeneity across the AF. The material properties of discs and ligaments were calibrated using a step-wise addition procedure of functional spinal structures, such that the simulated segmental kinematical responses match the *in-vitro* experimental measurements. The physiological co-activation of trunk muscles was considered by introducing a compressive follower preload. The path of the preload was optimized, such that no intervertebral rotation in the sagittal plane occurred at each level of the intact spine model. The resulting segmental ROMs, facet joint forces and ligament forces at each level of the intact lumbar spine model are in good agreement with those of well-established lumbar spine FE models in the literature.

7.1.3. Sub-objective 3: Performance Assessment of Commercially Available TDA Designs

Using the multi-segment lumbar spine FE model, it was observed that the implantation of a TDA did not influence the mechanical responses of adjacent levels regardless of different parameters adopted in a parametric TDA design, indicating that the same load can be transferred to each level throughout the treated lumbar spine. Therefore, TDA using the anterior surgical procedure causes a poorer resistance to extension (larger ROM) and changes the load-sharing of spinal tissues only at the treated segment. Due to the slight change in the responses of adjacent levels, the multi-segement lumbar spine FE model was simplified to a single segment (L3-L4) FE model, in which different TDA designs was futher evaluated in terms of segmental kinematics and MoP contact mechanics.

The convex-up fixed-bearing, convex-down fixed-bearing, and biconvex mobile-bearing commercially available TDA designs were inserted into the L3-L4 segment model to simulate the resulting responses. They represent three TDA designs whose initial instantaneous centers of rotation (ICRs) are located blow the inferior metallic EP, above the superior metallic EP, and at the center of the PE core, respectively. Compared to the fixed ICRs in the fixed-bearing TDAs, it was assumed that the unconstrained configuration in the biconvex mobile-bearing TDA leads to a physiological moving ICR which is less sensitive to implant placement, and thus lower facet joint forces [9, 110]. However, our simulation results showed that all designed TDA motion patterns were changed by lift-off, impingement, and/or unsymmetric motion, due to the interaction with the facet joint articulation.

For the convex-up TDA design, the FJ tends to be separated by a coupled posterior translation of the superior vertebra in extension and it goes more evident when the functional tissues are dissected in the anterior surgery, so it results in a larger extension rotation and a higher incidence of impingement. The exact opposite occurs for the convex-down TDA design; a larger resistance to extension can be provided by the FJs, as they close. Combining both convex-up and convex-down TDA motions can explain the unsymmetrical motions on the superior and inferior MoP articulations in the biconvex mobile-bearing TDA design. This unsymmetric motion in the biconvex TDA design causes a distinct downward bending of the PE rim, where the median rim thickness is only 2.9 mm [76] and the yield strength is 13 MPa [115]. Hence, the PE rim is potentially plastically deformed and cracked during cyclic bending in life activities. Consistent with previously reported retrieval studies [10, 76, 77], the PE rim fracture appears to be a more noticeable problem than PE wear.

7.1.4. Sub-objective 4: Proposal of a New Biconcave Mobile-bearing TDA Design Concept

Impingement, observed in both mobile- and fixed-bearing TDAs, is defined as the unintended contact between two nonbearing surfaces. Through an in-depth analysis of the segmental kinematics and MoP contact mechanics after the treatment using commercially available TDA designs, we demonstrated that the unsymmetric motion and impingement in the mobile-bearing TDA with a biconvex core are the contributing factors of the PE fracture. On the other hand, TDA using the anterior surgical procedure introduces a low-stiffness artificial joint, causing excessive segmental rotations [8, 9]. Therefore, the impingement of the metallic EPs against the rims of the PE core is inevitable. We hypothesized that impingement as a contact constraint could prevent excessive intervertebral rotations and facet joint overloading. To strengthen the PE rims, a new mobile-bearing TDA concept was proposed featuring a biconcave PE mobile core, with a smooth transition from the dome to the rim. Simulation results indicate that the proposed TDA design with a biconcave core can improve the MoP articulation and relieve the stress concentration in the PE rim. Therefore, it is expected to considerably prolong implant life *in vivo*, compared to commercially available mobile-bearing TDA designs.

7.1.5. Sub-objective 5: Development of Neural Network Based Multiobjective Optimization Algorithm

Generally, evolutionary optimization algorithms based on gradient descent require the gradient of the objective function and constraint function with respect to each design variable. In the traditional design optimization method, finite element analysis performed in each optimization iteration can only provide a black-box objective or constraint function without an analytical expression. Although the numerical gradient computed using finite difference enables the execution of an optimization program, it results in an expensive computational cost and a slow process to seek the optimal solution in design optimization. To address the problem, artificial neural network was implemented to derive analytical expressions of the objective and constraint functions and their gradients with respect to design variables. By combining artificial neural network and multiobjective optimization, we realized design optimization of a TDA simultaneously considering three performance metrics. To the best of our knowledge, there has not been a work that performs tri-objective design optimization of an implant in the literature.

7.1.6. Sub-objective 6: Multiobjective Optimization of Total Disc Arthroplasty

The biconcave mobile-bearing TDA design was further optimized using the proposed NN based MOO algorithm, considering the segmental kinematics, facet joint loading, and MoP contact mechanics simultaneously. To that end, a parametric model of the biconcave mobile-bearing TDA with ten design variables was developed in ANSYS 15.0 Workbench DesignModeler. The parametric TDA model can be reduced to two design extremities, a convex-down fixed-bearing TDA or a convex-up fixed-bearing TDA, by assigning a complete congruence on either the superior or inferior rim contact interfaces. As such, the fixed-bearing TDA configurations were also included in the design space. The parametric TDA model was inserted into a lumbar segment model, to simulate various performance metrics.

Using the NN based MOO algorithm, the complete Pareto frontier in the 3-D feasible objective space was fit and visualized by a triangular mesh. The ROM-optimal, FJF-optimal, and PCP-optimal TDA designs represent the optimal performance for a single objective function, but sacrifices in other objectives are observed. The best-tradeoff TDA design was defined as a design corresponding to an optimum point on the Pareto frontier in the minimum Euclidean distance away from the Utopian point. According to the simulated responses of the lumbar segment treated by the best-tradeoff TDA design, no performance metrics are excessively compromised, indicating that native disc biomechanics and the best MoP contact mechanics are achieved.

7.2. Limitations to Address in Future Work

7.2.1. Continuum Modeling of Soft Tissues

In the research, performance measures were proposed to evaluate the risks of postsurgical complications associated with abnormal kinematics, facet joint overloading, and implant failures in the instrumented lumbar segment. Therefore, stresses in intervertebral discs and spinal ligaments were not a focus. The FE representations of these soft tissues were simplified using discrete FE modeling techniques which facilitate the computational efficiency. In the lumbar segment FE model, the discs and ligaments were modeled using smeared reinforcing elements and nonlinear spring elements, respectively. However, stresses in the discs and ligaments are the important indicator for spinal tissue overloading and adjacent disc degeneration. To accurately predict stresses in soft tissues, continuum modeling techniques should be implemented in future work.

7.2.2. Physiological Loading Conditions

In the parametric analysis of a socket-in-ball fixed-bearing TDA design, changes in the simulated responses of the adjacent levels in the multi-segment lumbar spine model were not observed, as different TDA design parameters were used. It is concerned that the loading protocol utilizing a follower preload followed by pure moments in different loading scenarios is too simple, although it has been widely used to simulate the spine behavior in previous reported *in-vitro* experiments and computational models. The adjacent-level responses in the TDA-treated lumbar spine FE model should be further validated using more complicated loading conditions such as the hybrid loading protocol [73] and musculoskeletal modeling to consider the physiological co-activation of trunk muscles.

7.2.3. Inter-individual Variations

In this study, a representative lumbar segment FE model was created according to well-documented anatomic dimensions [14, 15], and the kinematical responses of the intact segment FE model (using material properties of deformable vertebrae) were calibrated to *in-vitro* experimental measurements in various loading scenarios [28] using a step-wise addition procedure of functional spinal structure [16, 33, 34]. However, we did not consider the inter-individual variations in the spine anatomy, inter-segment alignment, and material properties of spinal tissues. To address the problem, statistical modeling [92, 124–126] needs to be performed to investigate these inter-individual variations. The coefficients of their principal components of these probability distributions can be added to the NN inputs to fit the responses of the instrumented lumbar spine to these inter-individual variations. The future work would create a more complex neural network based surrogate model [127], and perform multiobjective optimization of patient-specific TDA designs.

7.2.4. Polyethylene Wear

Theoretically, the instantaneous center of rotation of the biconvex mobile-bearing TDA design is mobile, resulting in the better kinematics and lower facet joint loads than fixed-bearing TDA designs. However, unsymmetric motion and impingement were observed in the biconvex mobile-bearing TDA design, when it was simulated using a computational model incorporating functional spinal structures. Due to the abnormal motion and MoP articulation, the PE rim is vulnerable to fracture, which was considered as a more serious problem to be addressed compared to PE wear. Therefore, a new biconcave mobile-bearing TDA design was proposed to strengthen the PE rim, such that it can serve as an intended bearing.

In multiobjective design optimization, the performance metric regarding the PE contact pressure was introduced, and it was assumed that the lower PE contact pressure can reduce the risks of both the PE fracture and wear. However, it is well known that PE wear is sensitive to the patterns of sliding distance and cross-shear ratio, and the resulting wear debris may induce osteolysis which is the major cause of aseptic loosening and implant failure in MoP arthroplasty [12, 128]. In future, a more comprehensive literature review will be conducted to identify the risk of PE wear in TDA, and develop a computational framework to predict the TDA performance related to PE wear.

7.3. Conclusion

In conclusion, a new mobile-bearing TDA featuring a biconcave core was optimally designed in terms of spinal kinematics, facet joint loading, and MoP contact mechanics using multiobjective design optimization. These results could not have been obtained without the numerical frameworks for simulating these performance metrics related to the complications after the TDA treatment, as well as the development of the neural network based multiobjective optimization algorithm. By searching the optimum point on the 3-D Pareto frontier in the minimum Euclidean distance from the Utopian point, the best-tradeoff TDA design was determined. Since no distinct sacrifices of each performance measure occur in the best-tradeoff TDA design, it is expected to match native disc biomechanics and improve the MoP contact mechanics. Although the impressive results have been achieved using multiobjective design optimization, TDA surgical outcomes are influenced by many more factors, such as implant designs and alignment, soft tissue functional deficits, maladies in the vertebral anatomy, surgical techniques, rehabilitation, postoperative activity level.

The future vision is multiobjective design optimization of patient-specific TDAs, in the combination with imaging techniques, *in-vivo* clinical research, material characterization, statistical analysis, and rapid prototyping technology. The proposed neural network based multiobjective optimization algorithm can be further expanded to generate the patient-specific Pareto frontier by incorporating inter-individual variations in spine anatomy, intersegment alignment, and material properties of spinal tissues. The resulting Pareto frontier allows the health care provider to determine the best tradeoff TDA design for individual patients, which would be manufactured using rapid prototyping. As these techniques are put in practice, motion-preserving techniques will be able to reproduce native disc biomechanics, and ultimately take the place of motion-sacrificing techniques in the treatment for spinal disorders.

Appendix A: Mesh Morphing Technique

In previously reported computer vision works [129, 130], implicit surface equations have been employed to model complicated anatomic geometry. They have deomonstred that both the vertebral body (VB) and intervertebral disc (IVD) can be modeled using a series of transformations governed by implicit surface equations. Using these implicit equations, visualization can be trivially achieved in programming software. In the study, the implicit surface modeling method were incorporated into a new mesh morphing technique to create lumbar VB and IVD parametric FE models in ANSYS 15.0 MAPDL. The representative transformations in mesh morphing of VB and IVD models were presented in Figures A.1 and A.2, respectively. Initially, a truncated elliptical cone was created as the basic shape of the VB/IVD model. After it was meshed, the nodal coordinates were sequentially applied wall cardioid shape transformation, wall concave/bulge transformation, endplate concave/bulge transformation, and endplate sagittal inclination transformation for either VB (Figures A.1) or IVD (Figures A.2).

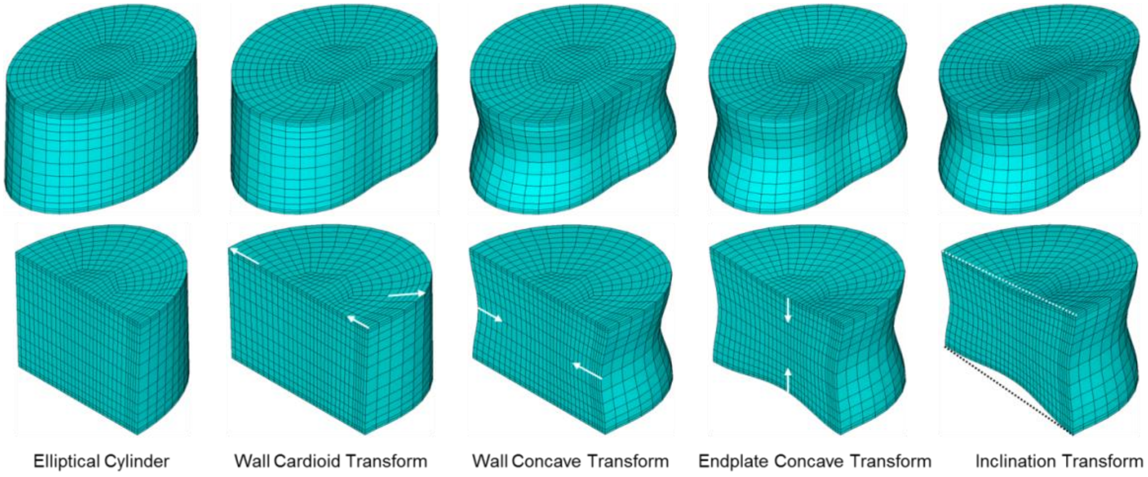


Figure A.1: The shape transformations used to model a vertebral body.

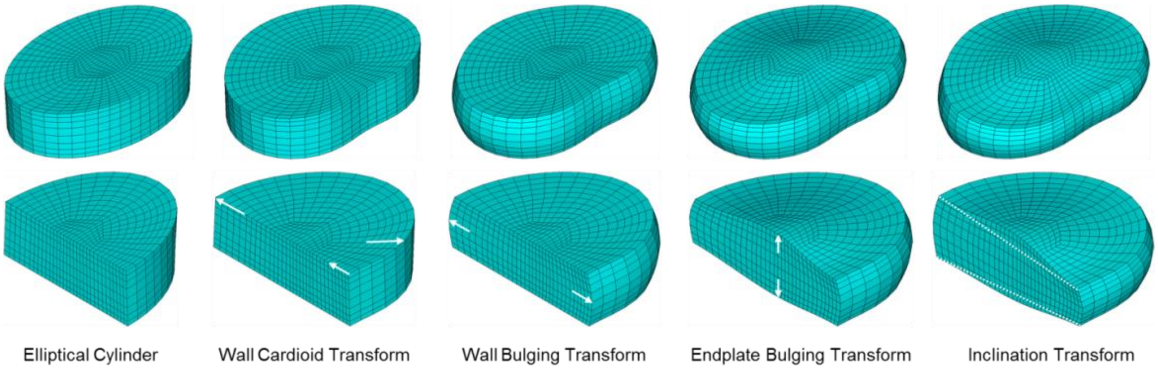


Figure A.2: The shape transformations used to model an intervertebral disc.

Previously, we have successfully implemented the mesh morphing technique to model an L4-L5 lumbar motion segment using well-documented anatomic dimension (Table A.1), as presented in Figures A.3. To ensure the endplate congruence between the VB and the IVD, the reported transformation formulations [129, 130] were modified. These modified transformation formulations used in this study for modeling both VB (Figures A.1) and IVD (Figures A.2) were described in the following.

Table A.1: Well-documented anatomic dimensions of L4-L5 lumbar motion segment [14, 32, 33]. Note that the unit is mm, except for the disc inclination angle and the volume ratio of the nucleus pulposus to the whole L4-L5 disc.

Dimensions	L4	L5	References
D_{sup}	35.5	34.7	Panjabi et al (1992)
W_{sup}	46.6	47.3	
D_{inf}	33.9	33.2	
W_{inf}	49.5	49.4	
H	24.1	22.9	
h	9.6		Weisse et al (2012)
α	7.6 °		
t_a, t_l, t_p	9.1, 8.9, 5.2		Schmidt et al (2006)
Vol. ratio of NP	44%		

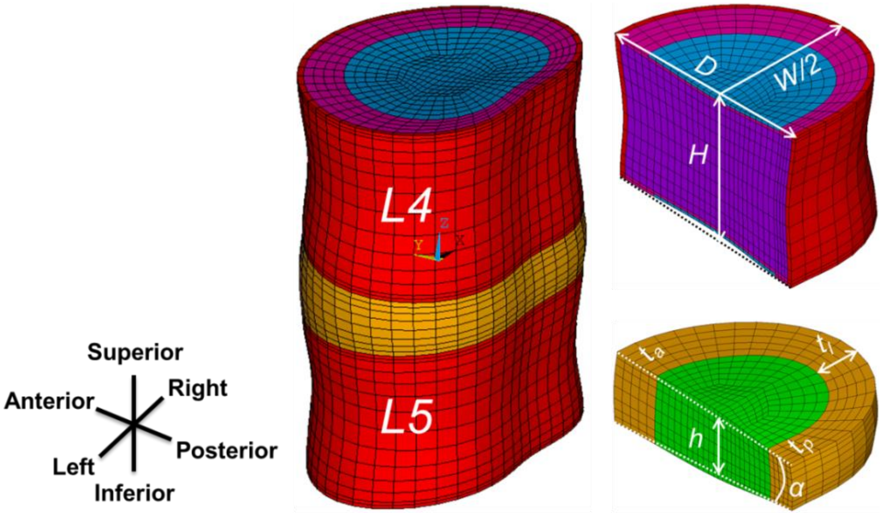


Figure A.3: Parametric FE model of a L4-L5 lumbar motion segment using reported anatomic dimensions, as listed in Table A.1.

A.1. Initialization of VB/IVD FE model: Alignment and Basic Shape

The alignment of the VB/IVD FE model can be conveniently performed by defining a local coordinate system (CS) in ANSYS, equivalent to the operations of rigid body translation and rotation. The basic shape of the VB/IVD was defined as a truncated elliptical cone. After the

VB/IVD basic shape model was meshed, the nodal coordinates are governed by an inequality equation, $F_{BS}(x,y,z) \leq 1$, in which $F_{BS}(x,y,z)$ is an implicit function of the truncated elliptical cone,

$$F_{BS}(x,y,z) = (\frac{x^2+y^2}{r^2})^{10} + (\frac{z}{h})^{20} \tag{A.1}$$

where x, y, and z are the nodal coordinates of the FE mesh in the local Cartesian CS. h is the half height of the elliptical cone. $r(z,\theta)$ is the radius of the directrix curve (the cross-section ellipse),

$$r(z,\theta) = r_{sup}(z,\theta) + r_{inf}(z,\theta)$$

$$= \frac{1}{2}\left(1+\frac{z}{h}\right)\frac{a_{sup}b_{sup}}{\sqrt{\left(a_{sup}\sin\theta\right)^2+\left(b_{sup}\cos\theta\right)^2}} + \frac{1}{2}\left(1-\frac{z}{h}\right)\frac{a_{inf}b_{inf}}{\sqrt{\left(a_{inf}\sin\theta\right)^2+\left(b_{inf}\cos\theta\right)^2}} \tag{A.2}$$

where $\theta = \arctan(y/x)$ is the polar angle measured from the semi-major axis of the cross-section ellipse, and the angle unit is [rad]. a_{sup} and b_{sup} are the major and minor radii of the superior surface of the elliptical cone, and a_{inf} and b_{inf} are the major and minor radii of the inferior surface.

A.2. Wall Kidney Shape Transformation

The cross-section of the elliptical cone in the basic shape model was transformed to a kidney shape, by superposing four Gaussian functions on the radial nodal coordinates of the basic shape model, with the magnitudes (m_i), standard deviation angles (σ_i), and polar angle coordinates (θ_i) at the anterior (i = a), posterior (i = p), left (i = l) and right (i = r), respectively.

$$r_G = r \cdot \left[1 + \sum_{i\in\{\mathrm{a,p,l,r}\}} m_i e^{-\frac{f^2(\theta-\theta_i)}{2\sigma_i^2}}\right] \tag{A.3}$$

where $f(\alpha)$ is a correction function of the polar angle coordinates in terms of the argument, $\alpha = \theta - \theta_i$,

$$f(\alpha) = \begin{cases} \alpha + 2\pi, & \alpha < -\pi \\ \alpha & -\pi \le \alpha \le \pi \\ \alpha - 2\pi, & \alpha \ge \pi \end{cases} \tag{A.4}$$

For convenience, the kidney-shaped wall transformation can be performed by defining a local cylindrical CS (r, θ, z) with the same origin as the previous local Cartesian CS.

A.3. Wall Concavity/Bulge Transformation

The wall concavity/bulge can be simply modeled by a cosine function with the amplitude of c_w (positive value for the VB and negative value for the IVD), and half of the function period is equal to the height ($2h$) of the VB/IVD basic shape model.

$$r_W = r \cdot \left[1 - c_w \cos(\frac{\pi z}{2h})\right] \tag{A.5}$$

A.4. Endplate Concavity/Bulge Transformation

The superior and inferior endplate concavity/bulge of the VB/IVD was modeled by introducing a 2D cosine function with the period equal to the diameter ($2r_W$) of the directrix curve and the amplitude of c_{es} at the superior endplate or c_{ei} at the inferior endplate (positive value for the VB and negative value for the IVD). To ensure the interface congruence of the adjacent VB and IVD, the amplitudes was corrected by dividing the half height (h) of the basic shape model. Furthermore, the peaks or valleys of the cosine functions at the superior and inferior surfaces may be shifted by (x'_{sup}, y'_{sup}) and (x'_{inf}, y'_{inf}), respectively.

$$z_E = \begin{cases} z \cdot \left[1 + \dfrac{c_{es}}{h} \cos\left(\dfrac{\pi\sqrt{\left(x - x'_{sup}\right)^2 + \left(y - y'_{sup}\right)^2}}{r_W}\right)\right], & z \ge 0 \\ z \cdot \left[1 + \dfrac{c_{ei}}{h} \cos\left(\dfrac{\pi\sqrt{\left(x - x'_{inf}\right)^2 + \left(y - y'_{inf}\right)^2}}{r_W}\right)\right], & z < 0 \end{cases} \tag{A.6}$$

A.5. Endplate Sagittal Inclination Transformation

The endplate inclination in the sagittal plane was modeled by bending deformation with the magnitudes of s_{es} and s_{ei} (positive value for the VB and negative value for the IVD), and the directions defined by the polar angles of ψ_{es} and ψ_{ei} (positive value for the superior surface and negative value for inferior surface in both VB and IVD models).

$$\begin{bmatrix} x_I \\ y_I \\ z_I \end{bmatrix} = \begin{cases} \begin{bmatrix} x + (B - b) \cdot \cos \psi_{es} \\ y + (B - b) \cdot \sin \psi_{es} \\ z - (|s_{es}| - b) \cdot \sin(\dfrac{\pi s_{es} z}{180 h_E}) \end{bmatrix}, & z \geq 0 \\ \begin{bmatrix} x + (B - b) \cdot \cos \psi_{ei} \\ y + (B - b) \cdot \sin \psi_{ei} \\ z - (|s_{ei}| - b) \cdot \sin(\dfrac{\pi s_{ei} z}{180 h_E}) \end{bmatrix}, & z < 0 \end{cases} \tag{A.7}$$

$$b = \begin{cases} \sqrt{x^2 + y^2} \cdot \cos(\psi_{es} - \theta), & z \geq 0 \\ \sqrt{x^2 + y^2} \cdot \cos(\psi_{ei} - \theta), & z < 0 \end{cases} \tag{A.8}$$

$$B = \begin{cases} |s_{es}| - (|s_{es}| - b) \cdot \cos(\dfrac{\pi s_{es} z}{180 h_E}), & z \geq 0 \\ |s_{ei}| - (|s_{ei}| - b) \cdot \cos(\dfrac{\pi s_{ei} z}{180 h_E}), & z < 0 \end{cases} \tag{A.9}$$

where $h_E = z_E|_{z=h}$ according to Equation A.6 is introduced to achieve the endplate congruence between adjacent VB and IVD.

Appendix B: Constitutive Modeling of Intervertebral Disc Annular Lamellae

The intervertebral disc (IVD) is a sophisticated soft tissue structure that consists of a gel-like center, the nucleus pulposus (NP), and an outer fibrous ring, the annulus fibrosus (AF). The AF ground substance (GS) is reinforced by collagen fiber lamellae [19–22], with a crossing pattern of fiber bundles due to the alternating fiber orientations in adjacent lamellae. The fiber angles vary radially and circumferentially across the AF [19, 20]. The relative content of type I and II collagens gradually changes across the AF, causing a non-uniform distribution of lamellar tensile properties (Brickley-Parsons and Glimcher. 1984). Therefore, the different collagen content distribution and fiber orientation within the AF result in the material anisotropy and heterogeneity of the AF composite.

To develop a computational model of the IVD, the constitutive relationship of the annular material needs to be provided. However, commercially available FE software lacks a material model to describe the nonlinear tensile properties of the collagen fibers. Therefore, the author proposed a continuum material model which introduces new formulations to model the nonlinear mechanical behavior of the collagen fibers. Furthermore, the constitutive relationship of the collagen fibers intimately interacts with the algorithms of fitting and interpolation of the fiber stress-stretch curves described in the subsequent Chapter, in order to model the material heterogeneity.

B.1. Basic Concepts in Large Deformation Solid Mechanics

B.1.1. Kinematics and Configurations

Typically, a continuum body (Figure B.1) at the initial time ($t = 0$) is undeformed, thus the configuration was chosen as the initial or reference configuration (Ω_0), equivalently represented by the coordinates ($\boldsymbol{X}$) of the continuum. After it is deformed in a motion (χ), the new configuration at time (t) was defined as the current or spatial configuration (Ω), equivalently represented by the new coordinates ($\boldsymbol{x}$). The motion χ is assumed to be uniquely invertible, and the inverse motion is denoted by χ^{-1}.

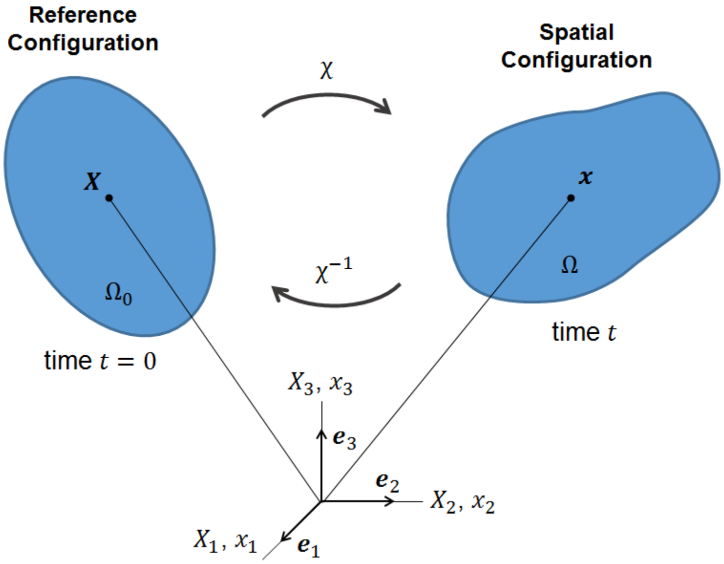


Figure B.1: Illustration of the configurations and motion of a continnum body [93], modified with permission from Springer Nature.

If all field quanties are descibed in the reference configuration ($\boldsymbol{X}$), then the description is referred to as the Lagrangian or material or referential description. The Eulerian or spatial description is used to describe these fields in the spatial configuration ($\boldsymbol{x}$). In solid mechanics, constitutive relations are preferentially derived using the referential description, so all quantities need to be applied a push-forward operation (χ_*) to obtain the spatial description. The opposite operation that maps the quantities in the spatial configuration back to the reference configuration is called a pull-back operation (χ_*^{-1}).

As a primary measure of deformations, the deformation gradient ($\boldsymbol{F}$) is introduced to characterize the behavior of motion in the neighborhood of a point, from the reference configuration to the spatial configuration,

$$d\boldsymbol{x} = \boldsymbol{F}d\boldsymbol{X} \tag{B.1}$$

Provided a fiber located at $\boldsymbol{X}$ in the reference configuration has the inital length dl_0 ($dl_0 \ll |\boldsymbol{X}|$) and the inital direction $\boldsymbol{a}_0$, the fiber can be represented by a vector $d\boldsymbol{X} = dl_0\boldsymbol{a}_0$. After a deformation, the fiber moves to a new location $\boldsymbol{x}$ with a new length dl ($dl \ll |\boldsymbol{x}|$) and a new direction $\boldsymbol{a}$, so it is represented by $d\boldsymbol{x} = dl\boldsymbol{a}$. Subsituting $d\boldsymbol{X} = dl_0\boldsymbol{a}_0$ and $d\boldsymbol{x} = dl\boldsymbol{a}$ into Equation B.1,

$$dl\boldsymbol{a} = \boldsymbol{F}dl_0\boldsymbol{a}_0 \tag{B.2}$$

The strecth ratio $\lambda = \frac{dl}{dl_0}$ is introduced to descibe the change in the fiber length. Therefore, an important relation was obtained,

$$\lambda\boldsymbol{a} = \boldsymbol{F}\boldsymbol{a}_0 \tag{B.3}$$

If the fiber direction ($\boldsymbol{a}$) in the spatial configuration is known, Equation B.3 can be reformulated to calculate the fiber direction ($\boldsymbol{a}_0$) in the reference configuration,

$$\boldsymbol{a}_0 = \frac{\boldsymbol{F}^{-1}\boldsymbol{a}}{|\boldsymbol{F}^{-1}\boldsymbol{a}|} \tag{B.4}$$

Equation B.4 can be used to pull back the fiber orientation from the spatial configuration to the reference configuration. According to the definition in Equation B.1, the deformation gradient ($\boldsymbol{F}$) can be written as,

$$\boldsymbol{F} = \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{X}} = \boldsymbol{I} + \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{X}} = \boldsymbol{I} + \mathrm{Grad}\,\boldsymbol{u} \tag{B.5}$$

where $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{X}$ is the displacement from the reference configuration ($\boldsymbol{X}$) to the spatial configuration ($\boldsymbol{x}$). $\boldsymbol{I} = \delta_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j$ is the identity tensor with the components expressed by the Kronecker delta symbol (δ_{ij}).

B.1.2. Numeric Deformation Gradient

In most commercial FE solvers, the deformation gradient ($\boldsymbol{F}$) is not an output, but it can be numerically computed for each element using a procedure reported by Wriggers (2016). For the 8-node hexahedral element, the element displacement field ($\boldsymbol{u}_e$) and the element coordinates ($\boldsymbol{X}_e$) in the reference configuration are approximated using tri-linear shape functions based on the isoparametric concept,

$$\boldsymbol{u}_e = \sum_{i=1}^{8} N_i(\xi,\eta,\zeta)\boldsymbol{u}_i = [N_1 \quad N_2 \quad \cdots \quad N_8]_{1\times 8} \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ \vdots & \vdots & \vdots \\ u_8 & v_8 & w_8 \end{bmatrix}_{8\times 3} \tag{B.6}$$

$$\boldsymbol{X}_e = \sum_{i=1}^{8} N_i(\xi,\eta,\zeta)\boldsymbol{X}_i = [N_1 \quad N_2 \quad \cdots \quad N_8]_{1\times 8} \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ \vdots & \vdots & \vdots \\ X_8 & Y_8 & Z_8 \end{bmatrix}_{8\times 3} \tag{B.7}$$

where $\boldsymbol{u}_i$ and $\boldsymbol{X}_i$ are the nodal displacements and coordinates of the 8-node element. $N_i(\xi,\eta,\zeta)$ are the shape functions given by,

$$N_i(\xi,\eta,\zeta) = \frac{1}{8}(1+\xi_i\xi)(1+\eta_i\eta)(1+\zeta_i\zeta), \quad i = 1,2,\cdots,8 \tag{B.8}$$

where ξ_i, η_i and ζ_i are the nodal coordinates of the element in the isoparametric formulation [132]. For the 8-node linear hexahedral elements without extra shape functions (ANSYS MAPDL 15.0 Theory Reference), ξ_i, η_i and ζ_i are shown in Table B.1.

Table B.1: The nodal coordinates of ξ_i, η_i and ζ_i of 8-node linear hexahedral elements.

Node	ξ_i	η_i	ζ_i
1	-1	-1	-1
2	+1	-1	-1
3	+1	+1	-1
4	-1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	+1	+1	+1
8	-1	+1	+1

Using the chain rule of differentiation, the deformation gradient (as show in Equation B.5) of each element can be expressed in terms of the nodal displacements and coordinates,

$$\boldsymbol{F}_e = \boldsymbol{I} + \text{Grad}\, \boldsymbol{u}_e = \boldsymbol{I} + \frac{\partial \boldsymbol{u}_e}{\partial \boldsymbol{X}_e} = \boldsymbol{I} + \frac{\partial \boldsymbol{u}_e}{\partial \boldsymbol{\Xi}} \frac{\partial \boldsymbol{\Xi}}{\partial \boldsymbol{X}_e} = \boldsymbol{I} + \frac{\partial \boldsymbol{u}_e}{\partial \boldsymbol{\Xi}} \boldsymbol{J}_e^{-1} \tag{B.9}$$

where $\boldsymbol{J}_e = \frac{\partial \boldsymbol{X}_e}{\partial \boldsymbol{\Xi}}$ is the Jacobian matrix of transformation from the non-dimensional coordinates $\boldsymbol{\Xi}$ to the actual element coordinates $\boldsymbol{X}_e$. Here, $\frac{\partial \boldsymbol{u}_e}{\partial \boldsymbol{\Xi}}$ and $\boldsymbol{J}_e = \frac{\partial \boldsymbol{X}_e}{\partial \boldsymbol{\Xi}}$ can be derived by taking derivatives of Equations B.6 and B.7,

$$\frac{\partial \boldsymbol{u}_e}{\partial \boldsymbol{\Xi}} = \begin{bmatrix} \frac{\partial N_1}{\partial \xi} & \frac{\partial N_2}{\partial \xi} & \cdots & \frac{\partial N_8}{\partial \xi} \\ \frac{\partial N_1}{\partial \eta} & \frac{\partial N_2}{\partial \eta} & \cdots & \frac{\partial N_8}{\partial \eta} \\ \frac{\partial N_1}{\partial \zeta} & \frac{\partial N_2}{\partial \zeta} & \cdots & \frac{\partial N_8}{\partial \zeta} \end{bmatrix}_{3\times 8} \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ \vdots & \vdots & \vdots \\ u_8 & v_8 & w_8 \end{bmatrix}_{8\times 3} \tag{B.10}$$

$$\boldsymbol{J}_e = \frac{\partial \boldsymbol{X}_e}{\partial \boldsymbol{\Xi}} = \begin{bmatrix} \frac{\partial N_1}{\partial \xi} & \frac{\partial N_2}{\partial \xi} & \cdots & \frac{\partial N_8}{\partial \xi} \\ \frac{\partial N_1}{\partial \eta} & \frac{\partial N_2}{\partial \eta} & \cdots & \frac{\partial N_8}{\partial \eta} \\ \frac{\partial N_1}{\partial \zeta} & \frac{\partial N_2}{\partial \zeta} & \cdots & \frac{\partial N_8}{\partial \zeta} \end{bmatrix}_{3\times 8} \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ \vdots & \vdots & \vdots \\ X_8 & Y_8 & Z_8 \end{bmatrix}_{8\times 3} \tag{B.11}$$

where the derivatives of the shape functions (as shown in Equation B.8) with respect to the non-dimensional coordinates $\boldsymbol{\Xi} = \{\xi, \eta, \zeta\}$ can be trivially obtained.

B.1.3. Stress Tensor and Elasticity Tensor

In this study, a continuum approach [93] will be implemented to develop a transversely isotropic hyperelastic constitutive material model for the annulus fiber lamellae, and investigate the nonlinear mechanics of the IVD in large deformation. Based on the assumption that the macroscopic nature of materials can be described as continua, the material responses can be formulated by proposing an appropriate strain energy function (Ψ). Hence, the resulting constitutive material model is also called a phenomenological model.

According to the second law of thermodynamics, the internal dissipation (D_{int}) is reduced to zero for perfectly elastic material,

$$D_{int} = w_{int} - \dot{\Psi} = 0 \tag{B.12}$$

where $w_{int} = \frac{1}{2}\boldsymbol{S}:\dot{\boldsymbol{C}}$ is the stress power (the rate of internal mechanical work per unit reference volume), defined by a work conjugate pair consisting of the second Piola-Kirchhoff stress tensor ($\boldsymbol{S}$) and the material time derivative of the right Cauchy-Green tensor ($\boldsymbol{C}$). $\dot{\Psi} = \dot{\Psi}(\boldsymbol{C})$ is the material time derivative of the strain energy function (Ψ). Using the chain rule of differentiation, $\dot{\Psi} = \frac{\partial \Psi}{\partial \boldsymbol{C}}:\dot{\boldsymbol{C}}$, Equation B.12 can be expressed as,

$$\left(\frac{1}{2}\boldsymbol{S} - \frac{\partial \Psi(\boldsymbol{C})}{\partial \boldsymbol{C}}\right):\dot{\boldsymbol{C}} = 0 \tag{B.13}$$

Equation B.13 must hold at every point of the continuum body and for all times during the process. Therefore, the second Piola-Kirchhoff stress tensor ($\boldsymbol{S}$) is deduced in terms of the right Cauchy-Green tensor ($\boldsymbol{C}$),

$$\boldsymbol{S} = 2\frac{\partial \Psi(\boldsymbol{C})}{\partial \boldsymbol{C}} \tag{B.14}$$

Furthermore, the 4th-order elasticity tensor ($\mathbb{C}$) that characterizes the gradient of $\boldsymbol{S}$ in terms of $\boldsymbol{C}$ was introduced in the concept of linearization, measuring the change in stress which results from a change in strain,

$$\mathbb{C} = 2\frac{\partial \boldsymbol{S}(\boldsymbol{C})}{\partial \boldsymbol{C}} = 4\frac{\partial^2 \Psi(\boldsymbol{C})}{\partial \boldsymbol{C}\partial \boldsymbol{C}} \tag{B.15}$$

It is noted that Equations B.14 and B.15 are derived based on the reference configuration. They are mapped to the spatial configuration using the Piola transformation, which is defined as a push-forward operation (χ_*) times a factor of J^{-1}. Here, $J = \det(\boldsymbol{F})$ measuring the volume ratio is the determinant of the deformation gradient ($\boldsymbol{F}$). Therefore, the Cauchy stress ($\boldsymbol{\sigma}$) and the elasticity tensor ($\mathbb{c}$) in the spatial configuration are expressed as,

$$\boldsymbol{\sigma} = J^{-1}\chi_*(\boldsymbol{S}) = J^{-1}\boldsymbol{FSF}^T \tag{B.16}$$

$$\mathbb{c} = J^{-1}\chi_*(\mathbb{C}), \quad c_{abcd} = J^{-1}F_{aA}F_{bB}F_{cC}F_{dD}C_{ABCD} \tag{B.17}$$

For material constitute model development, the frame indifference (objectivity) must be satisfied by introducing objective stress rates. In fact, the objectivity of the material response using above formulations have been satisfied, since the elasticity tensor $\mathbb{c}$ in Equation B.16 is with respect to the objective Oldroyd stress rate of the Kirchhoff stress tensor $\boldsymbol{\tau} = J\boldsymbol{\sigma}$,

$$\mathrm{Oldr}(\boldsymbol{\tau}) = J\mathbb{c}: \boldsymbol{d} \tag{B.18}$$

where $\boldsymbol{d}$ is the rate of deformation tensor. However, in ANSYS (see Mechanical APDL Technology Demonstration Guide Chapter 40), the user material is formulated in a co-rotational frame, thus the objective Jaumann rate of the Cauchy stress is adopted. Using the Jaumann rate ($\boldsymbol{\sigma}^{\nabla}$), the constitutive response in the co-rotated frame is given by,

$$\boldsymbol{\sigma}^{\nabla} = \mathbb{c}^J: \boldsymbol{d} \tag{B.19}$$

where $\mathbb{c}^J$ is the Jaumann tangent stiffness tensor, which is associated with the Cauchy stress ($\boldsymbol{\sigma}$) in Equation B.16 and the elasticity tensor ($\mathbb{c}$) in Equation B.17. Using the index notation, the Jaumann tangent stiffness tensor (c^J_{ijkl}) can be expressed as [133],

$$c^J_{ijkl} = c_{ijkl} + \frac{1}{2}(\delta_{ik}\sigma_{jl} + \delta_{il}\sigma_{jk} + \delta_{jk}\sigma_{il} + \delta_{jl}\sigma_{ik}) \tag{B.20}$$

where σ_{ij} is the Kronecker delta. Derivation of the Cauchy stress tensor ($\boldsymbol{\sigma}$) in Equation B.16 and the tangent stiffness tensor ($\mathbb{c}^J$) with respect to the Jaumann rate in Equation B.20 is a crucial task for development of a user material using ANSYS USERMAT FORTRAN subroutine. For hyperelastic materials, the total 36 components in c^J_{ijkl} have only 21 independent components according to the major and minor symmetries [134].

B.2. Development of User Material Subroutine for Annular Lamellae

A user material subroutine (ANSYS USERMAT subroutine) was developed to simulate the fiber reinforced annular lamellae. The main task of user material development is to derive the Cauchy stress tensor ($\boldsymbol{\sigma}$, Equation B.16) and the tangent stiffness tensor (c^{J}_{ijkl}, Equation B.20) with respect to the Jaumann rate, both of which are required in Newton-Raphson nonlinear numerical method. The strain energy function of the AF composite material is:

$$\Psi^{AF}(J, \bar{I}_1, \lambda_1, \lambda_2) = \Psi^{AF}_{vol}(J) + \Psi^{AF}_{iso}(\bar{I}_1) + \Psi^{AF}_{fib}(\lambda_{\boldsymbol{a}}) + \Psi^{AF}_{fib}(\lambda_{\boldsymbol{g}}) \tag{B.21}$$

where,

$$\Psi^{AF}_{vol}(J) = \frac{\kappa}{2}(J-1)^2 \tag{B.22}$$

$$\Psi^{AF}_{iso}(\bar{I}_1) = c_1(\bar{I}_1 - 3) + c_2(\bar{I}_1 - 3)^2 \tag{B.23}$$

$$\Psi^{AF}_{fib}(\lambda) = \begin{cases} 0, & \lambda < 1 \\ c_3\left\{\dfrac{1}{c_4}\left[e^{c_4(\lambda-1)} - 1\right] - (\lambda - 1)\right\}, & 1 \le \lambda < c_5 \\ \dfrac{E^*}{2}(\lambda - c_5)^2 + T^*(\lambda - c_5) + \Psi^*, & \lambda \ge c_5 \end{cases} \tag{B.24}$$

The calculus calculation results of stress tensors and elasticity tensors of the AF composite were presented as follows.

B.2.1. Volumetric Part, $\Psi^{AF}_{vol}(J)$

According to Equation B.22, the derived stresses and elasticity tensors of the volumetric part were listed below,

$$\boldsymbol{S}^{AF}_{vol} = 2\frac{\partial \Psi^{AF}_{vol}(J)}{\partial \boldsymbol{C}} = \kappa J(J-1)\boldsymbol{C}^{-1} \tag{B.25}$$

$$\boldsymbol{\sigma}^{AF}_{vol} = J^{-1}\chi_*\left(\boldsymbol{S}^{AF}_{vol}\right) = \kappa(J-1)\boldsymbol{I} \tag{B.26}$$

$$\mathbb{C}^{AF}_{vol} = 2\frac{\partial \boldsymbol{S}^{AF}_{vol}}{\partial \boldsymbol{C}} = \kappa J[(2J-1)\boldsymbol{C}^{-1} \otimes \boldsymbol{C}^{-1} - 2(J-1)\boldsymbol{C}^{-1} \odot \boldsymbol{C}^{-1}] \tag{B.27}$$

$$\mathbb{c}_{vol}^{AF} = J^{-1}\chi_*\left(\mathbb{C}_{vol}^{AF}\right) = \kappa J[(2J-1)\boldsymbol{I}\otimes\boldsymbol{I} - 2(J-1)\mathbb{I}] \tag{B.28}$$

where $\boldsymbol{I} = \delta_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j$ is the unit tensor with the components expressed by the Kronecker delta symbol (δ_{ij}). The operation $\boldsymbol{C}^{-1} \odot \boldsymbol{C}^{-1} = \frac{1}{2}(C_{IK}^{-1}C_{JL}^{-1} + C_{IL}^{-1}C_{JK}^{-1})$ results into a 4th-order tensor, which is pushed forward to obtain a 4th-order unit tensor $\mathbb{I} = \chi_*(\boldsymbol{C}^{-1} \odot \boldsymbol{C}^{-1}) = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$.

B.2.2. Isochoric Part, $\Psi_{iso}^{AF}(\bar{I}_1)$

According to Equation B.23, the derived stresses and elasticity tensors of the isochoric part were listed below,

$$\boldsymbol{S}_{iso}^{AF} = 2\frac{\partial\Psi_{iso}^{AF}(\bar{I}_1)}{\partial\boldsymbol{C}} = 2J^{-2/3}[c_1 + 2c_2(\bar{I}_1 - 3)](\boldsymbol{I} - \frac{I_1}{3}\boldsymbol{C}^{-1}) \tag{B.29}$$

$$\boldsymbol{\sigma}_{iso}^{AF} = J^{-1}\chi_*\left(\boldsymbol{S}_{iso}^{AF}\right) = 2J^{-1}[c_1 + 2c_2(\bar{I}_1 - 3)](\bar{\boldsymbol{b}} - \frac{\bar{I}_1}{3}\boldsymbol{I}) \tag{B.30}$$

$$\begin{aligned}\mathbb{C}_{iso}^{AF} = 2\frac{\partial\boldsymbol{S}_{iso}^{AF}}{\partial\boldsymbol{C}} &= 8c_2J^{-4/3}\left(\boldsymbol{I}\otimes\boldsymbol{I} - \frac{I_1}{3}\boldsymbol{I}\otimes\boldsymbol{C}^{-1} - \frac{I_1}{3}\boldsymbol{C}^{-1}\otimes\boldsymbol{I} + \frac{I_1^2}{9}\boldsymbol{C}^{-1}\otimes\boldsymbol{C}^{-1}\right) \\ &\quad + \frac{4}{3}J^{-2/3}[c_1 + 2c_2(\bar{I}_1 - 3)](-\boldsymbol{I}\otimes\boldsymbol{C}^{-1} - \boldsymbol{C}^{-1}\otimes\boldsymbol{I} + I_1\boldsymbol{C}^{-1}\odot\boldsymbol{C}^{-1} \\ &\quad + \frac{I_1}{3}\boldsymbol{C}^{-1}\otimes\boldsymbol{C}^{-1})\end{aligned} \tag{B.31}$$

$$\begin{aligned}\mathbb{c}_{iso}^{AF} &= J^{-1}\chi_*\left(\mathbb{C}_{iso}^{AF}\right) \\ &= 8c_2J^{-1}\left(\bar{\boldsymbol{b}}\otimes\bar{\boldsymbol{b}} - \frac{\bar{I}_1}{3}\bar{\boldsymbol{b}}\otimes\boldsymbol{I} - \frac{\bar{I}_1}{3}\boldsymbol{I}\otimes\bar{\boldsymbol{b}} + \frac{\bar{I}_1^2}{9}\boldsymbol{I}\otimes\boldsymbol{I}\right) + \frac{4}{3}J^{-2/3}[c_1 \\ &\quad + 2c_2(\bar{I}_1 - 3)](-\bar{\boldsymbol{b}}\otimes\boldsymbol{I} - \boldsymbol{I}\otimes\bar{\boldsymbol{b}} + I_1\mathbb{I} + \frac{I_1}{3}\boldsymbol{I}\otimes\boldsymbol{I})\end{aligned} \tag{B.32}$$

where $I_1 = \mathrm{tr}(\boldsymbol{C})$ is the first invariant of the right Cauchy-Green tensor. The modified left Cauchy-Green tensor is defined as $\bar{\boldsymbol{b}} = \bar{\boldsymbol{F}}\,\bar{\boldsymbol{F}}^T$, in which $\bar{\boldsymbol{F}} = J^{-1/3}\,\boldsymbol{F}$ such that $\det(\bar{\boldsymbol{F}}) = 1$.

B.2.3. Reinforcing Fiber Part, $\Psi_{fib}^{AF}(\lambda_{\boldsymbol{a}})$ and $\Psi_{fib}^{AF}(\lambda_{\boldsymbol{g}})$

According to Equation B.24, the derived stresses and elasticity tensors of the reinforcing fiber part $\Psi_{fib}^{AF}(\lambda_{\boldsymbol{a}})$ with the fiber direction $\boldsymbol{a}$ were listed below,

$$\boldsymbol{S}_{fib}^{AF}(\lambda_{\boldsymbol{a}}) = 2\frac{\partial\Psi_{fib}^{AF}(\lambda_{\boldsymbol{a}})}{\partial\boldsymbol{C}} = \frac{1}{\lambda_{\boldsymbol{a}}}\frac{\partial\Psi_{fib}^{AF}(\lambda)}{\partial\lambda}\bigg|_{\lambda=\lambda_{\boldsymbol{a}}}\boldsymbol{a}_0\otimes\boldsymbol{a}_0 \tag{B.33}$$

$$\boldsymbol{\sigma}_{fib}^{AF}(\lambda_{\boldsymbol{a}}) = J^{-1}\chi_*\left(\boldsymbol{S}_{fib}^{AF}(\lambda_{\boldsymbol{a}})\right) = \frac{\lambda_{\boldsymbol{a}}}{J}\frac{\partial\Psi_{fib}^{AF}(\lambda)}{\partial\lambda}\bigg|_{\lambda=\lambda_{\boldsymbol{a}}}\boldsymbol{a}\otimes\boldsymbol{a} \tag{B.34}$$

$$\begin{aligned}\mathbb{C}_{fib}^{AF}(\lambda_{\boldsymbol{a}}) &= 2\frac{\partial\boldsymbol{S}_{fib}^{AF}(\lambda_{\boldsymbol{a}})}{\partial\boldsymbol{C}}\\ &= \frac{1}{\lambda_{\boldsymbol{a}}^2}\left(\frac{\partial^2\Psi_{fib}^{AF}(\lambda)}{\partial\lambda^2}\bigg|_{\lambda=\lambda_{\boldsymbol{a}}} - \frac{1}{\lambda_{\boldsymbol{a}}}\frac{\partial\Psi_{fib}^{AF}(\lambda)}{\partial\lambda}\bigg|_{\lambda=\lambda_{\boldsymbol{a}}}\right)\boldsymbol{a}_0\otimes\boldsymbol{a}_0\otimes\boldsymbol{a}_0\otimes\boldsymbol{a}_0\end{aligned} \tag{B.35}$$

$$\begin{aligned}\mathbb{c}_{fib}^{AF}(\lambda_{\boldsymbol{a}}) &= J^{-1}\chi_*\left(\mathbb{C}_{fib}^{AF}(\lambda_{\boldsymbol{a}})\right)\\ &= \frac{\lambda_{\boldsymbol{a}}^2}{J}\left(\frac{\partial^2\Psi_{fib}^{AF}(\lambda)}{\partial\lambda^2}\bigg|_{\lambda=\lambda_{\boldsymbol{a}}} - \frac{1}{\lambda_{\boldsymbol{a}}}\frac{\partial\Psi_{fib}^{AF}(\lambda)}{\partial\lambda}\bigg|_{\lambda=\lambda_{\boldsymbol{a}}}\right)\boldsymbol{a}\otimes\boldsymbol{a}\otimes\boldsymbol{a}\otimes\boldsymbol{a}\end{aligned} \tag{B.36}$$

It is worth noting that both projection tensors $\boldsymbol{a}_0\otimes\boldsymbol{a}_0$ and $\boldsymbol{a}\otimes\boldsymbol{a}$ in Equations B.33 and B.34 indicate that the stress states of the fibers are uniaxial tension in both the reference and spatial configurations.

Similarly, the stress tensors (Equations B.33 and B.34) and elasticity tensors (Equations B.35 and B.36) of another reinforcing fiber part $\Psi_{fib}^{AF}(\lambda_{\boldsymbol{g}})$ in the fiber direction $\boldsymbol{g}$ can be written by replacing $\boldsymbol{a}$ by $\boldsymbol{g}$ and $\lambda_{\boldsymbol{a}}$ by $\lambda_{\boldsymbol{g}}$, as presented below:

$$\boldsymbol{S}_{fib}^{AF}(\lambda_{\boldsymbol{g}}) = 2\frac{\partial\Psi_{fib}^{AF}(\lambda_{\boldsymbol{g}})}{\partial\boldsymbol{C}} = \frac{1}{\lambda_{\boldsymbol{g}}}\frac{\partial\Psi_{fib}^{AF}(\lambda)}{\partial\lambda}\bigg|_{\lambda=\lambda_{\boldsymbol{g}}}\boldsymbol{g}_0\otimes\boldsymbol{g}_0 \tag{B.37}$$

$$\boldsymbol{\sigma}_{fib}^{AF}(\lambda_{\boldsymbol{g}}) = J^{-1}\chi_*\left(\boldsymbol{S}_{fib}^{AF}(\lambda_{\boldsymbol{g}})\right) = \frac{\lambda_{\boldsymbol{g}}}{J}\frac{\partial\Psi_{fib}^{AF}(\lambda)}{\partial\lambda}\bigg|_{\lambda=\lambda_{\boldsymbol{g}}}\boldsymbol{g}\otimes\boldsymbol{g} \tag{B.38}$$

$$\begin{aligned}\mathbb{C}^{AF}_{fib}(\lambda_{\boldsymbol{g}}) &= 2\frac{\partial \boldsymbol{S}^{AF}_{fib}(\lambda_{\boldsymbol{g}})}{\partial \boldsymbol{C}} \\ &= \frac{1}{\lambda_{\boldsymbol{g}}^2}\left(\left.\frac{\partial^2 \Psi^{AF}_{fib}(\lambda)}{\partial \lambda^2}\right|_{\lambda=\lambda_g} - \frac{1}{\lambda_{\boldsymbol{g}}}\left.\frac{\partial \Psi^{AF}_{fib}(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_g}\right)\boldsymbol{g}_0 \otimes \boldsymbol{g}_0 \otimes \boldsymbol{g}_0 \otimes \boldsymbol{g}_0\end{aligned} \tag{B.39}$$

$$\begin{aligned}\mathbb{c}^{AF}_{fib}(\lambda_{\boldsymbol{g}}) &= J^{-1}\chi_*\left(\mathbb{C}^{AF}_{fib}(\lambda_{\boldsymbol{g}})\right) \\ &= \frac{\lambda_{\boldsymbol{g}}^2}{J}\left(\left.\frac{\partial^2 \Psi^{AF}_{fib}(\lambda)}{\partial \lambda^2}\right|_{\lambda=\lambda_g} - \frac{1}{\lambda_{\boldsymbol{g}}}\left.\frac{\partial \Psi^{AF}_{fib}(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_g}\right)\boldsymbol{g} \otimes \boldsymbol{g} \otimes \boldsymbol{g} \otimes \boldsymbol{g}\end{aligned} \tag{B.40}$$

B.2.4. Resultant Stress Tensor and Elasticity Tensor of Annular Lamellae

In summary, the Cauchy stress $\boldsymbol{\sigma}^{AF}$ of the AF composite material is the superposition of each Cauchy stress part (Equations B.26, B.30, B.34 and B.38):

$$\boldsymbol{\sigma}^{AF} = \boldsymbol{\sigma}^{AF}_{vol} + \boldsymbol{\sigma}^{AF}_{iso} + \boldsymbol{\sigma}^{AF}_{fib}(\lambda_{\boldsymbol{a}}) + \boldsymbol{\sigma}^{AF}_{fib}(\lambda_{\boldsymbol{g}}) \tag{B.41}$$

The elasticity tensor $\mathbb{c}^{AF}$ with respect to Oldroyd stress rate of the AF composite material is the superposition of each elasticity tensor (Equations B.28, B.32, B.36 and B.40):

$$\mathbb{c}^{AF} = \mathbb{c}^{AF}_{vol} + \mathbb{c}^{AF}_{iso} + \mathbb{c}^{AF}_{fib}(\lambda_{\boldsymbol{a}}) + \mathbb{c}^{AF}_{fib}(\lambda_{\boldsymbol{g}}) \tag{B.42}$$

Substituting the Cauchy stress tensor (Equation B.41) and the Oldroyd elasticity tensor (Equation B.42) into Equation B.20, the Jaumann tangent stiffness tensor of the AF composite material is obtained [133, 134]:

$${c^{J}_{ijkl}}^{AF} = c^{AF}_{ijkl} + \frac{1}{2}(\delta_{ik}\sigma^{AF}_{jl} + \delta_{il}\sigma^{AF}_{jk} + \delta_{jk}\sigma^{AF}_{il} + \delta_{jl}\sigma^{AF}_{ik}) \tag{B.43}$$

B.3. Element Patch Test of Continuum and Discrete Material Models

To validate the two-family fiber reinforced material subroutine, smear reinforcing elements (REINF265) in ANSYS were employed to model the equivalent composite with crossing patterned fiber layers (Figure B.2), using a reported modeling technique [35, 97]. The matrix element (SOLID185) was assigned the modified Yeoh material properties (Equations B.22 and B.23) using the material model provided by ANSYS. Since REINF265 is a 1-D material element (ANSYS Mechanical APDL 15.0 Element Reference), a 1-D material subroutine using the formulation of the reinforcing fiber part (Equations B.24) was developed to model the single lamellar material behavior.

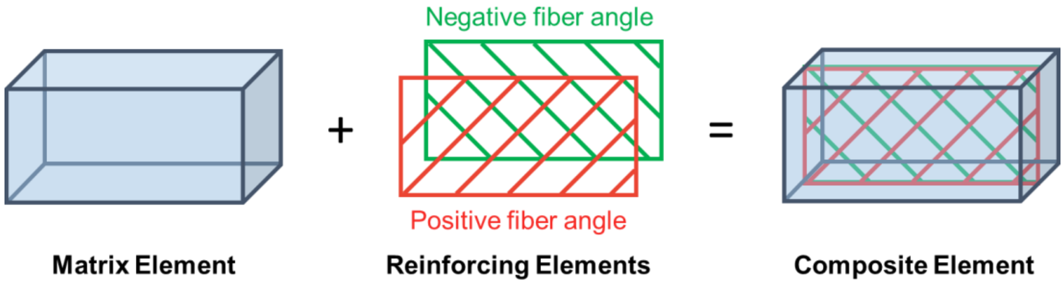


Figure B.2: Modeling of fiber reinforced composite using reinforcing elements (REINF265) [35], modified with permission from Springer Nature.

In a three-element patch test (Figure B.3), material properties (κ = 73.521 MPa, c_1 = 38.793 kPa, c_2 = 55.049 MPa, c_3 = 0.015, c_4 = 206.661, c_5 = 1.018) of the outermost annular lamellae at the disc anterior fitted from experimental data [19, 38] were assigned to both the continuum material model (two-family fiber reinforced material subroutine) and the discrete material model (solid matrix elements + reinforcing elements).

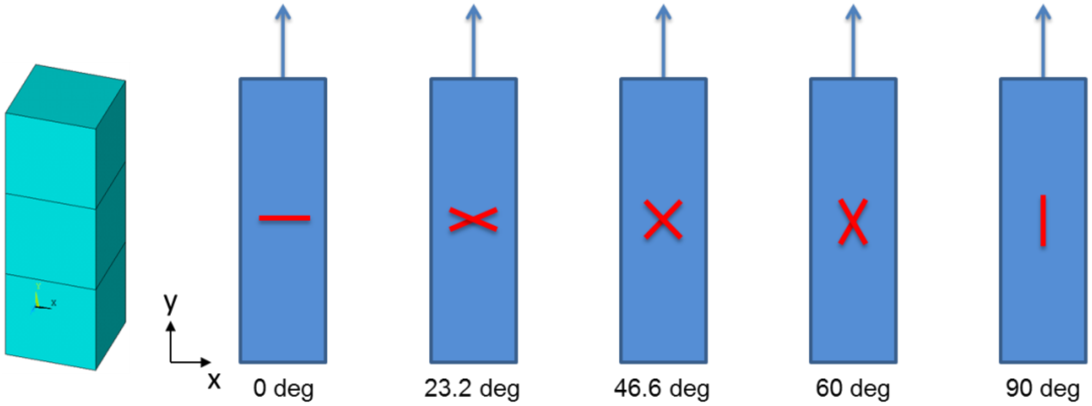


Figure B.3: Modeling of fiber reinforced composite using reinforcing elements (REINF265).

The fiber angles with respect to the X-Z plane were chosen as 0°, 23.2°, 46.6°, 60° and 90°, respectively. Especially, 23.2° and 46.6° correspond to the fiber angles within the anterior and posterior annular lamellae of lumbar discs, respectively, according to previously reported *in-vitro* measurements [19]. During loading and unloading, a displacement in the Y direction was applied on the top of the patch model, and the normal displacements at the planes of X = 0, Y = 0 and Z = 0 were constrained.

The element patch test (Figure B.4) demonstrates that the force-displacement behaviors using the continuum material model exactly match those using the well-established ANSYS material models and elements. For the fiber angle of 0° and 23.2°, the continuum material model has a better convergence capability, since fewer iterations were performed. It is worth noting that fibers in 23.2° (the fiber angle measured at the anterior lamellae in the *in-vitro* experiment) are not capable of reinforcing the matrix using both material models, as compared with fibers in 0°.

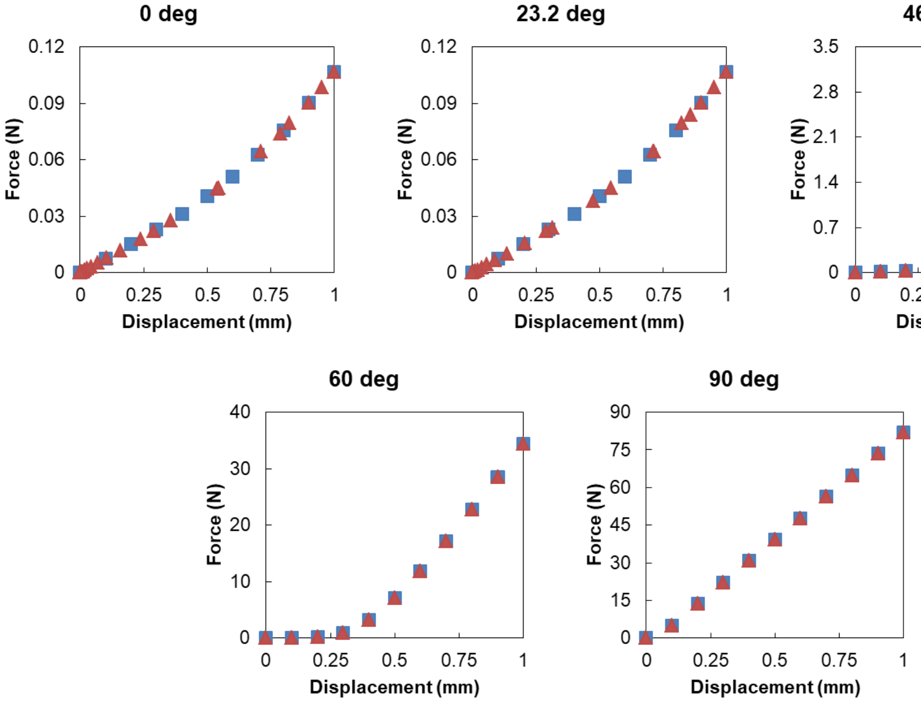


Figure B.4: Comparison of the material responses in loading and unloading using different material modeling techniques.

Appendix C: Artificial Neural Network Based Multiobjective Optimization Algorithm

In a traditional design optimization algorithm, finite element (FE) analysis is performed in each optimization iteration in order to provide a black-box cost function. Generally, evolutionary optimization algorithms (*e.g.*, gradient descent algorithm or BFGS algorithm) require the gradient of the cost function with respect to each design variable. Although it can be approximated using the finite difference method in an optimization algorithm, the implementation of the numerical gradient typically causes a computational cost multi-fold higher than that using the analytical gradient.

Moreover, it is difficult to ensure the convergence of FE analysis in each optimization iteration. Therefore, additional measures, such as assigning a large value to the cost function when FE analysis fails, need to be taken to prevent the termination of the optimization program. However, it may cause an inaccurate numerical gradient, which is a concern for the optimal solution.

At last but not least, multiple local optimums may exist in a non-convex cost function. To estimate the global optimum, it is essential to run the optimization program many times by random initialization of design variables. When the computational cost of FE analysis in each optimization iteration is expensive, it would be extremely time-consuming to run design optimization many times. In our previous work, we employed a workstation with 2 processors and 384 GB RAM. A single-objective optimization spent approximately 1 week, in which each iteration (FE analysis) took 30 min. To fit a 2-D Pareto curve, 10 optimizations were performed, and spent about 10 weeks. Thus the traditional design optimization method is not suitable for tri-objective optimization, where 50-300 optimizations need to perform depending on the grid density of the 3-D Pareto frontier.

These problems have been well addressed by developing a neural network based surrogate model to fit the response surface of FE analysis. Furthermore, we formulated the analytical gradient of a feed-forward neural network, which was incorporated into a multiobjective optimization algorithm to generate the complete Pareto frontier.

C.1. Artificial Neural Network Architecture and Related Formulations

Machine learning using an artificial neural network (NN) is inherently a multiobjective task [135]. Therefore, multiobjective optimization and multi-output neural network would be a good combination. A feed-forward three-layer neural network was employed to fit a vector-valued objective function ($\boldsymbol{y}$) of a design variable vector ($\boldsymbol{x}$) in multiobjective optimization. Theoretically, a three-layer NN can fit any continuous functions. We also hypothesized that the NN accuracy for a multivariate, multi-output regression problem would not be compromised, if there was a correlation among the NN outputs.

C.1.1. Formulation of a Single Neuron

The representation of a single neuron in a NN is shown in Figure C.1. Each neuron is allowed to receive multiple inputs but only yield one output. The output ($o_j^{(k)}$) of a neuron is a nonlinear activation of the linear combination of outputs ($o_i^{(k-1)}$) at the former layer. The output of a neuron j at layer k can be expressed as:

$$o_j^{(k)} = a(z_j^{(k)}) \tag{C.1}$$

where the superscript k in a parenthesis indicates the layer, and the subscript j the neuron at the layer k. $a(\bullet)$ is an element-wise activation function. For the purpose of introducing nonlinearity, it is often chosen to be a nonlinear function, such as the sigmoid function or hyperbolic tangent function. Particularly, the activation can be chosen as an identity function, such that $o_j^{(k)} = z_j^{(k)}$, meaning no activation.

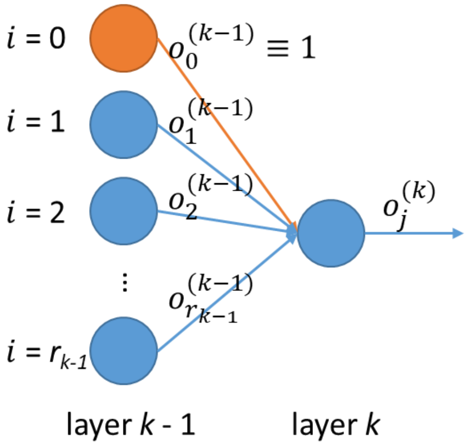


Figure C.1: Illustration of a single neuron at layer k, that processes the multiple outputs of the neurons at layer $k-1$, where the orange neuron represents the bias neuron.

The activation function is applied to the linear combination of outputs ($o_i^{(k-1)}$) at the former layer ($k-1$), which is denoted by $z_j^{(k)}$,

$$z_j^{(k)} = \sum_{i=0}^{r_{k-1}} w_{ji}^{(k)} o_i^{(k-1)} = w_{j0}^{(k)} o_0^{(k-1)} + \sum_{i=1}^{r_{k-1}} w_{ji}^{(k)} o_i^{(k-1)} = b_j^{(k)} + \sum_{i=1}^{r_{k-1}} w_{ji}^{(k)} o_i^{(k-1)} \tag{C.2}$$

where $w_{ji}^{(k)}$ is NN weights of outputs ($o_i^{(k-1)}$) at the former layer ($k-1$). By holding the output of the bias neuron always equal to 1, *i.e.*, $o_0^{(k-1)} \equiv 1$, we introduced a bias term, $b_j^{(k)}$, representing the constant intercept of the linear combination. Therefore, the bias term is always equal to the weight of the bias neuron,

$$b_j^{(k)} = w_{j0}^{(k)} \tag{C.3}$$

C.1.2. Formulation of a Three-layer Multi-output Feed-forward Neural Network

For a feed-forward NN, information always moves toward the direction of outputs, and never goes backwards. The architecture of a three-layer multi-output feed-forward NN was presented in Figure C.2. It consists of an input layer ($k = 0$) with r_0 neurons, a hidden layer ($k = 1$) with $r_1 = r_h$ neurons, and an output layer ($k = 2$) with r_2 neurons. For regression problems, nonlinear activation functions (*e.g.*, sigmoid / hyperbolic tangent functions) are adopted in neurons at the hidden layer. However, the identity activation function was used in both the input layer and the output layer, thus causing no activation.

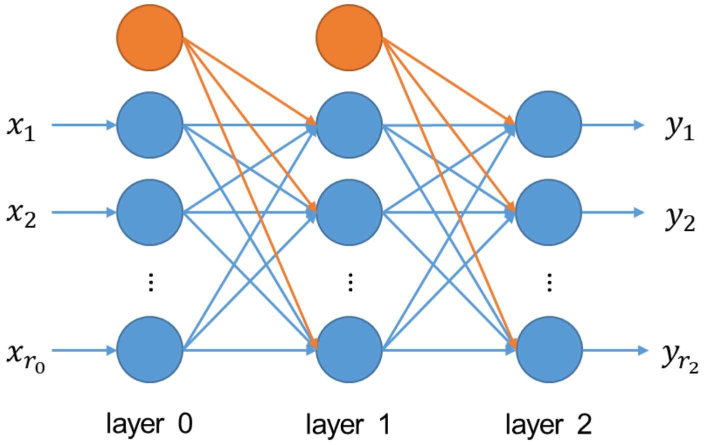


Figure C.2: The architecture of a three-layer multi-output feed-forward NN.

In this study, each neuron at the hidden layer was assigned a hyperbolic tangent activation function. The expressions of the hyperbolic tangent function and its derivative are shown below:

$$a(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{C.4}$$

$$a'(x) = \frac{\mathrm{d}\tanh(x)}{\mathrm{d}x} = 1 - a^2(x) \tag{C.5}$$

Using Equations C.1 and C.2, we can derive the outputs (predictions) of the NN (Figure C.2) as a function of the NN inputs along the inverse direction from the output layer to the input layer,

$$
\begin{aligned}
y_k = o_k^{(2)} &= \sum_{j=0}^{r_1} w_{kj}^{(2)} o_j^{(1)} = b_k^{(2)} + \sum_{j=1}^{r_1} w_{kj}^{(2)} o_j^{(1)} = b_k^{(2)} + \sum_{j=1}^{r_1} w_{kj}^{(2)} a(\sum_{i=0}^{r_0} w_{ji}^{(1)} o_i^{(0)}) \\
&= b_k^{(2)} + \sum_{j=1}^{r_1} w_{kj}^{(2)} a(b_j^{(1)} + \sum_{i=1}^{r_0} w_{ji}^{(1)} o_i^{(0)}) = b_k^{(2)} + \sum_{j=1}^{r_1} w_{kj}^{(2)} a(b_j^{(1)} + \sum_{i=1}^{r_0} w_{ji}^{(1)} x_i)
\end{aligned}
\tag{C.6}
$$

Alternatively, with the vectorized implementation, we have:

$$
\boldsymbol{y} = \boldsymbol{b}^{(2)} + \boldsymbol{w}^{(2)} a(\boldsymbol{b}^{(1)} + \boldsymbol{w}^{(1)} \boldsymbol{x}) \tag{C.7}
$$

In Equation C.6 or C.7, it is noted that both the input layer and the output layer has no activation. The element-wise hyperbolic tangent activation function (Equation C.4) was only applied to the hidden layer with r_1 neurons.

C.1.3. Formulation of the Gradients of the NN Outputs with respect to the NN Inputs

A numeric optimization program, such as the gradient descent algorithm or BFGS algorithm, can be efficiently executed, only when the analytic gradients of the cost functions with respect to the design variables are provided. Moreover, due to the existence of local optimums for a nonconvex cost function, an optimization needs to be performed many times, in order to estimate the global optimum. According to Equation C.6, it can be observed that the NN outputs with respect to the inputs are continuously differentiable, so their derivatives or gradients exist.

We further write the NN outputs (Equation C.6) with respect to the inputs using the index notation, as we have adopted in Chapter 3 to conduct the tensor calculus,

$$
y_k = b_k^{(2)} + w_{kj}^{(2)} a(b_j^{(1)} + w_{ji}^{(1)} x_i) \tag{C.8}
$$

By taking derivatives of Equation 8, the gradients of the NN outputs with respect to the NN inputs is obtained,

$$
\begin{aligned}
\frac{\partial y_k}{\partial x_i} &= \frac{\partial}{\partial x_i}\left[w_{kj}^{(2)} a(b_j^{(1)} + w_{ji}^{(1)} x_i)\right] = w_{kj}^{(2)} \frac{\partial}{\partial x_i} a(b_j^{(1)} + w_{ji}^{(1)} x_i) \\
&= w_{kj}^{(2)} \left[a'\left(z_{\underline{j}}^{(1)}\right) \frac{\partial}{\partial x_i}\left(z_{\underline{j}}^{(1)}\right)\right] = w_{kj}^{(2)} \left[a'\left(z_{\underline{j}}^{(1)}\right) w_{\underline{j}i}^{(1)}\right]
\end{aligned}
\tag{C.9}
$$

where for simplicity, we let $z_{\underline{j}}^{(1)} = b_{\underline{j}}^{(1)} + w_{ji}^{(1)} x_i$. The underline blew the subscript j indicates the element-wise multiplication, *i.e.*, the repeated index does not undergo the summation contraction. $a'(\bullet)$ is the derivative of the element-wise hyperbolic tangent activation function (Equation C.5).

Alternatively, with the matriculated implementation,

$$\operatorname{grad} \boldsymbol{y}(\boldsymbol{x}) = \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{w}^{(2)} \left[a'\left(\boldsymbol{b}^{(1)} + \boldsymbol{w}^{(1)} \boldsymbol{x}\right) \otimes \boldsymbol{e}_{r_1} \circ \boldsymbol{w}^{(1)} \right] \qquad \text{(C.10)}$$

where $\otimes$ denotes the outer product. $\circ$ denotes the component-wise product (no summation contraction). $\boldsymbol{e}_{r_1}$ is a vector of all ones with the dimension of r_1, and r_1 is the number of neurons at the hidden layer.

C.1.4. Data Scaling

As shown in Figure C.3, the derivative of the hyperbolic tangent active function vanishes, outside the range between -2 and +2. It causes a slow NN training (*i.e.*, optimization of the NN weights using the error function which is derived from the error backpropagation procedure; for the NN architecture in this study, please be referred to [136]), due to the small sensitivity of the error function to the change in NN weights. Here, we scaled both the NN inputs and outputs/predictions to [-1, +1] during NN training. After training, the normalized outputs/predictions (Equation C.6) were converted back to the true scale. Correspondingly, it should be noted that a difference in the formulation of gradients (Equation C.9) exists between using true scale and using normalized scale.

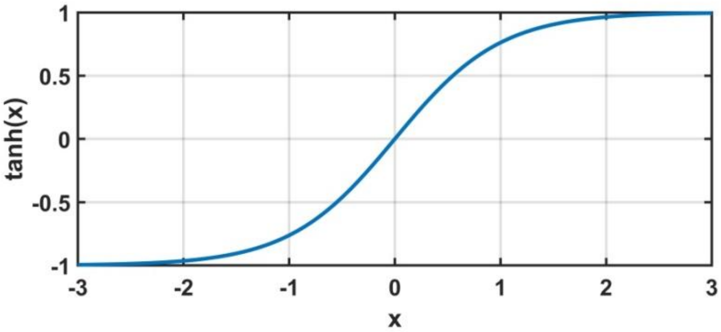


Figure C.3: The plot of the hyperbolic tangent function in Equation C.4.

The sample size (the number of input-output pairs or feature-target pairs or examples) in the training dataset is denoted by p. The dimension of each feature is n, equal to the number of NN inputs, *i.e.*, $n = r_0$. The dimension of each target/label is m, equal to the number of NN outputs, *i.e.*, $m = r_2$. Therefore, we can define a feature matrix (X) and a target matrix (T) to describe the training dataset.

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}^{(1)} & \boldsymbol{x}^{(2)} & \dots & \boldsymbol{x}^{(p)} \end{bmatrix} = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(p)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{(1)} & x_n^{(2)} & \cdots & x_n^{(p)} \end{bmatrix} \tag{C.11}$$

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{t}^{(1)} & \boldsymbol{t}^{(2)} & \dots & \boldsymbol{t}^{(p)} \end{bmatrix} = \begin{bmatrix} t_1^{(1)} & t_1^{(2)} & \cdots & t_1^{(p)} \\ t_2^{(1)} & t_2^{(2)} & \cdots & t_2^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ t_m^{(1)} & t_m^{(2)} & \cdots & t_m^{(p)} \end{bmatrix} \tag{C.12}$$

The maxima and minima of the feature are written as,

$$\boldsymbol{x}_{\max} = \max_{\text{row}} \boldsymbol{X} \quad , \quad \boldsymbol{x}_{\min} = \min_{\text{row}} \boldsymbol{X} \tag{C.13}$$

where $\max\limits_{\text{row}}$ and $\min\limits_{\text{row}}$ indicate the operation to calculate the maxima and minima of a matrix by rows, thus generating column vectors.

The maxima and minima of the target are written as,

$$\boldsymbol{t}_{\max} = \max_{\text{row}} \boldsymbol{T} \quad , \quad \boldsymbol{t}_{\min} = \min_{\text{row}} \boldsymbol{T} \tag{C.14}$$

Now, we scale both the NN inputs and outputs to [-1, +1],

$$\tilde{\boldsymbol{x}} = \frac{2(\boldsymbol{x} - \boldsymbol{x}_{min})}{\boldsymbol{x}_{max} - \boldsymbol{x}_{min}} - 1 \quad or \quad \boldsymbol{x} = \frac{\boldsymbol{x}_{max} - \boldsymbol{x}_{min}}{2}\tilde{\boldsymbol{x}} + \boldsymbol{x}_{min} + 1 \tag{C.15}$$

$$\tilde{\boldsymbol{y}} = \frac{2(\boldsymbol{y} - \boldsymbol{t}_{min})}{\boldsymbol{t}_{max} - \boldsymbol{t}_{min}} - 1 \quad or \quad \boldsymbol{y} = \frac{\boldsymbol{t}_{max} - \boldsymbol{t}_{min}}{2}\tilde{\boldsymbol{y}} + \boldsymbol{t}_{min} + 1 \tag{C.16}$$

where $\boldsymbol{x}, \boldsymbol{y}$ and $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}$ the NN inputs and outputs on the true scale and the normalized scale, respectively. It is worth noting that all operations in Equations C.15 and C.16 are element-wise vector operations.

Correspondingly, the relationship between the gradients (Equation C.9) on the true scale ($\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$) and the normalized scale ($\frac{\partial \tilde{\boldsymbol{y}}}{\partial \tilde{\boldsymbol{x}}}$) can be established,

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \frac{(\boldsymbol{t}_{max} - \boldsymbol{t}_{min}) \otimes \boldsymbol{e}_n}{\boldsymbol{e}_m \otimes (\boldsymbol{x}_{max} - \boldsymbol{x}_{min})} \frac{\partial \tilde{\boldsymbol{y}}}{\partial \tilde{\boldsymbol{x}}} \tag{C.17}$$

where the division and multiplication are the component-wise operations for the matrices with the dimension of $m \times n$, after the outer products ($\otimes$) of vectors are computed. $\boldsymbol{e}_m$ and $\boldsymbol{e}_n$ are the vectors whose elements are all ones with the dimension of m and n, respectively.

C.1.5. Numerical Gradients

To validate the derivation of analytic gradients (Equation C.17), they can be compared with the numerical gradients. Using the two-sided finite difference method, the numerical gradients of the NN outputs/predictions (y_j) with respect to the NN inputs (x_i) can be calculated,

$$\frac{\partial y_j}{\partial x_i} \approx \frac{y_j(x_1, x_2, \cdots, x_i + \varepsilon, \cdots, x_n) - y_j(x_1, x_2, \cdots, x_i - \varepsilon, \cdots, x_n)}{2\varepsilon} \tag{C.18}$$

where the NN outputs/predictions are denoted by $y_j(x_i)$, $j = 1, 2, \cdots, m$ and $i = 1, 2, \cdots, n$. ε is a very small number; here, it was set to be $\varepsilon = 10^{-4}$. If the analytical gradient is correctly formulated, the difference from the numerical gradient should be smaller than 10^{-9}.

C.2. Complete Pareto Frontier Multiobjective Optimization Algorithm

C.2.1. Statement of Generic Multiobjective Optimization Problem

The generic multiobjective optimization (MOO) problem can be stated as follows [88, 119, 137]:

$$\overline{\min_{\boldsymbol{x}}}\, \boldsymbol{y}\,(\boldsymbol{x}) \tag{C.19}$$

subject to

$$\boldsymbol{x}_{lb} \leq \boldsymbol{x} \leq \boldsymbol{x}_{ub} \tag{C.20}$$

$$\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0} \tag{C.21}$$

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \tag{C.22}$$

where $\overline{\min}$ denotes multiobjective optimization, instead of simultaneously / individually minimizing each objective. $\boldsymbol{y}$ is an m-dimensional (m-D) vector of design objectives in the objective space ($\mathbb{R}_{\boldsymbol{y}}^{m}$). $\boldsymbol{x}$ is an n-dimensional (n-D) vector of design variables in the design space ($\mathbb{R}_{\boldsymbol{x}}^{n}$), with a lower boundary ($\boldsymbol{x}_{lb}$) and upper boundary ($\boldsymbol{x}_{ub}$), respectively. $\boldsymbol{g}$ and $\boldsymbol{h}$ are inequality and equality constraint vectors in the MOO problem, respectively. The set of design variables that satisfy Equations C.20, C.21 and C.22 is the feasible design space ($\Omega_{\boldsymbol{x}}$), corresponding to the feasible objective space ($\Omega_{\boldsymbol{y}}$). In this study, continuously differentiable objective/cost functions $\boldsymbol{y}(\boldsymbol{x})$ are fitted using the neural network described in the previous section. Therefore, the design variables ($\boldsymbol{x}$) and design objectives ($\boldsymbol{y}$) exactly match the NN inputs and outputs/predictions (Equation C.6).

In this study, we combined the previously reported normal-boundary intersection (NBI) algorithm [118] and the normal constraint (NC) algorithm [88, 137] to develop a MOO algorithm, which enables an even representation of the complete Pareto frontier.

C.2.2. Basic Concepts in Multiobjective Optimization

1) Anchor Points

Using single-objective optimization, the optimal solutions ($\boldsymbol{x}^{j*}$) of the design variables in the feasible design space ($\Omega_{\boldsymbol{x}}$) corresponding to respective individual objectives (y_j) can be obtained. The corresponding points in the feasible objective space ($\Omega_{\boldsymbol{y}}$) are defined as the anchor points ($\boldsymbol{y}^{j*}$),

$$\boldsymbol{y}^{j*} = \boldsymbol{y}(\boldsymbol{x}^{j*}) = [y_1(\boldsymbol{x}^{j*}) \quad y_2(\boldsymbol{x}^{j*}) \quad \cdots \quad y_m(\boldsymbol{x}^{j*})]^T \quad , \quad j = 1, 2, \cdots, m \tag{C.23}$$

For example, in a three-dimensional (3-D, $m = 3$) objective space (Figure C.4a), there are three anchor points ($\boldsymbol{y}^{1*}$, $\boldsymbol{y}^{2*}$, and $\boldsymbol{y}^{3*}$).

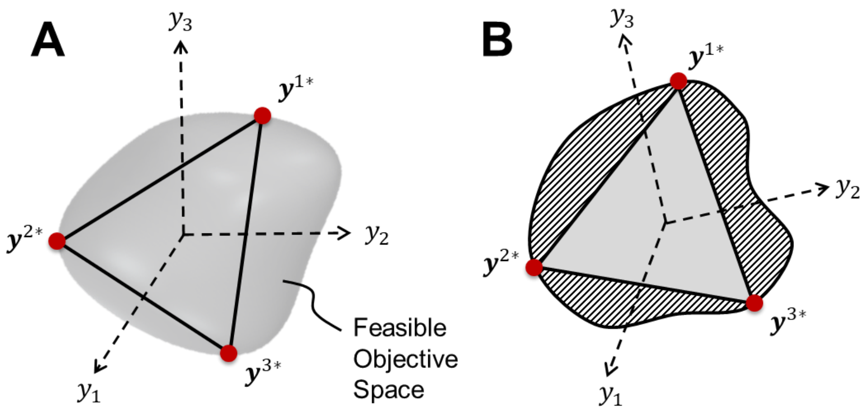


Figure C.4: Illustration of the Utopian polygon in the 3-D feasible objective space, modified from [88] with Dr. Achille Messac's permission. (a) The Utopian plane defined by the triangle section consisting of three anchor points in the 3-D objective space; (b) Unobtainable Pareto solutions (hatched regions) which are located outside the normal projection of the Utopian polygon. The view direction has been rotated such that it is normal to the Utopian plane.

2) Utopian Plane

By sequentially connecting anchor points ($\boldsymbol{y}^{j*}$), a Utopian polygon is formed on a plane which is termed as the Utopian plane. In a 3-D objective space, the Utopian plane is determined by a triangle consisting of the three anchor points (Figure C.4a).

3) Pay-off Matrix

We defined a $m \times m$ matrix whose components are the coordinates of anchor points in the objective space. This matrix is sometimes known as the pay-off matrix ($\boldsymbol{\Phi}$).

$$\boldsymbol{\Phi} = [\boldsymbol{y}^{1*} \quad \boldsymbol{y}^{2*} \quad \cdots \quad \boldsymbol{y}^{m*}] = \begin{bmatrix} y_1^{1*} & y_1^{2*} & \cdots & y_1^{m*} \\ y_2^{1*} & y_2^{2*} & \cdots & y_2^{m*} \\ \vdots & \vdots & \ddots & \vdots \\ y_m^{1*} & y_m^{2*} & \cdots & y_m^{m*} \end{bmatrix} \tag{C.24}$$

4) Utopia Point

The Utopia point ($\boldsymbol{y}^U$) is a point in the objective space when all objectives are simultaneously at their best values, corresponding to the diagonal components of the pay-off matrix,

$$\boldsymbol{y}^U = [y_1(\boldsymbol{x}^{1*}) \quad y_2(\boldsymbol{x}^{2*}) \quad \cdots \quad y_m(\boldsymbol{x}^{m*})]^T \tag{C.25}$$

5) Pseudo Nadir Point

The pseudo Nadir point ($\boldsymbol{y}^P$) is defined as a point in the objective space, each of whose coordinates is the maximum of the components in each row of the pay-off matrix.

$$\boldsymbol{y}^P = \max_{\text{row}} \boldsymbol{\Phi} \tag{C.26}$$

6) Nadir Point

Differently, the Nadir point ($\boldsymbol{y}^N$) is located where all objectives are simultaneously at their worst values in the objective space,

$$\boldsymbol{y}^N = [y_1^N \quad y_2^N \quad \cdots \quad y_m^N]^T \tag{C.27}$$

where each coordinate, y_j^N, is defined as the maximum of the jth objective which is obtained using single-objective optimization,

$$y_j^N = \max_{\boldsymbol{x}} y_j(\boldsymbol{x}) = \min_{\boldsymbol{x}} [-y_j(\boldsymbol{x})] \quad , \quad j = 1, 2, \cdots, m \tag{C.28}$$

subject to Equations C.20, C.21 and C.22.

Since we have introduced neural network, we can compute NN predictions ($\boldsymbol{Y}$) given by features ($\boldsymbol{X}$) with a sufficiently large sample size (p). By the matriculated implementation of Equation C.6 or 7.7, we can compute multiple NN predictions in parallel:

$$\boldsymbol{Y} = \boldsymbol{b}^{(2)} \otimes \boldsymbol{e}_p + \boldsymbol{w}^{(2)} a(\boldsymbol{b}^{(1)} \otimes \boldsymbol{e}_p + \boldsymbol{w}^{(1)} \boldsymbol{X}) \tag{C.29}$$

where $\boldsymbol{e}_p$ is a vector of all ones with the dimension of p. As a result, $\boldsymbol{Y}$ can be written as,

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}^{(1)} & \boldsymbol{y}^{(2)} & \cdots & \boldsymbol{y}^{(p)} \end{bmatrix} = \begin{bmatrix} y_1^{(1)} & y_1^{(2)} & \cdots & y_1^{(p)} \\ y_2^{(1)} & y_2^{(2)} & \cdots & y_2^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ y_m^{(1)} & y_m^{(2)} & \cdots & y_m^{(p)} \end{bmatrix} \tag{C.30}$$

Therefore, the Nadir point can be estimated using the maximum NN predictions for a large sample:

$$\boldsymbol{y}^N = \max_{\text{row}} \boldsymbol{Y} \tag{C.31}$$

C.2.3. Discretization of Utopian Plane Covering Complete Pareto Frontier

In a high-dimensional ($m \geq 3$) objective space, the entire set of Pareto solutions cannot be completely covered by the normal projection of the Utopian polygon consisting of anchor points [119], as illustrated by an example in the 3-D objective space in Figure C.4b. To address the problem, the Utopian polygon needs to be enlarged, such that a sufficiently large Utopian plane can cover the complete Pareto frontier (the entire set of Pareto solutions).

By implementing a parameterization method [88], a point ($\boldsymbol{y}^{utp}$) on the extended Utopian plane section can be written as,

$$\boldsymbol{y}^{utp} = \boldsymbol{\Phi\eta} \quad or \quad y_i^{utp} = \Phi_{ij}\eta_j \tag{C.32}$$

where $\boldsymbol{\Phi}$ is the $m \times m$ square pay-off matrix with the components, Φ_{ij}. $\boldsymbol{\eta}$ is a non-dimensional parameter vector with its elements, η_j. The non-dimensional parameter η_j is required to satisfy:

$$\eta_j^{lb} \leq \eta_j \leq \eta_j^{ub} \quad , \quad j = 1, 2, \cdots, m \tag{C.33}$$

and

$$\sum_{j=1}^{m} \eta_j = 1 \tag{C.34}$$

In a 3-D objective space, all points ($\boldsymbol{y}$) that satisfy Equations C.32, C.33 and C.34 form a discrete Utopian hexagon that enlarges the Utopian triangle consisting of the three anchor points, as shown in Figure C.5.

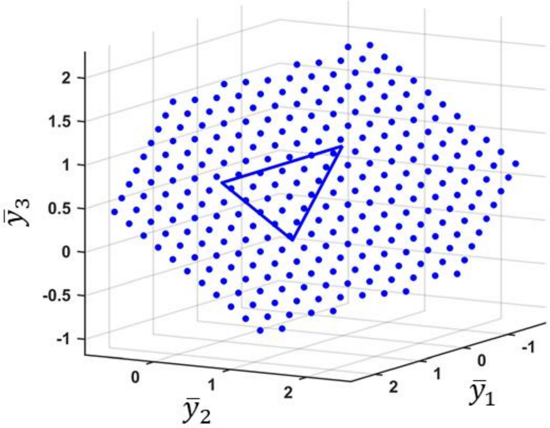


Figure C.5: The discrete Utopian hexagon that enlarged the Utopian triangle in a 3-D objective space.

To save computational cost in MOO, it is necessary to determine the lower boundary (η_j^{lb}) and upper boundary (η_j^{ub}) of the non-dimensional parameters, as well as filter out redundant points that must be projected to dominated regions along the normal of the Utopian plane, such that a discrete polygon region enlarged to the appropriate size on the Utopian plane is obtained, as shown in Figure C.6. These optimization techniques that eliminate unnecessary regions of the Utopian plane section and check of point usefulness have been proposed and well-documented in the publication of Messac and Mattson (2004). To avoid repetition, we will not reiterate them here.

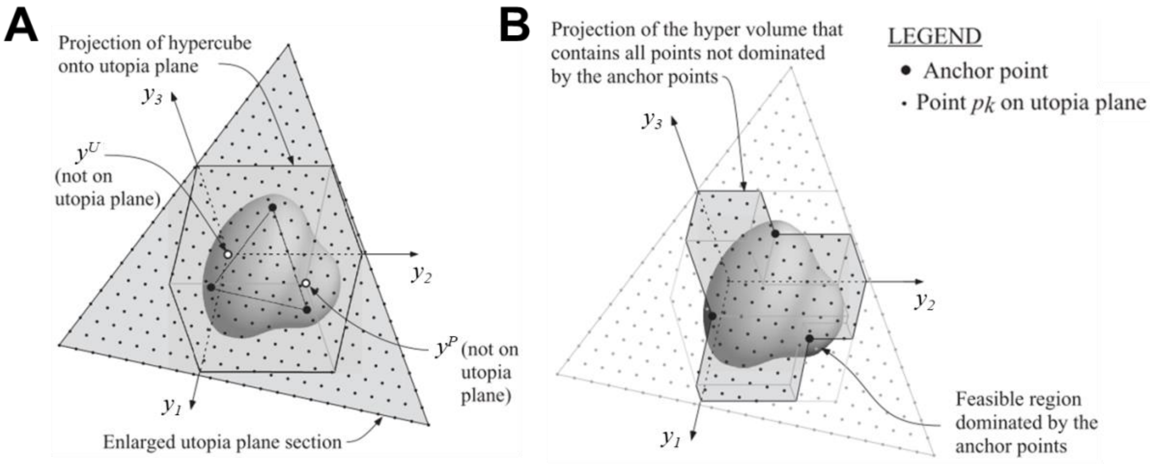


Figure C.6: Illustration of how to select an appropriate Utopian polygon to eliminate unnecessary regions, modified from [88] with Dr. Achille Messac's permission. (a) Enlargement of the Utopian triangle comprising the three anchor points; (b) Projection of a hypercube defined by y^U and y^P to determine the Utopian polygon.

The lower boundary (η_j^{lb}) and upper boundary (η_j^{ub}) of the non-dimensional parameters actually depend on how we define the range of the feasible objective space. Therefore, the selection of either pseudo Nadir point ($\boldsymbol{y}^P$) or Nadir point ($\boldsymbol{y}^N$) will correspondingly influence the optimal size of the Utopian polygon region (Figure C.6). Here, we further introduce a concept of the mixed Nadir point ($\boldsymbol{y}^{P/N}$) whose coordinates can be arbitrarily chosen from either pseudo Nadir point ($\boldsymbol{y}^P$) or Nadir point ($\boldsymbol{y}^N$). Using the mixed Nadir point ($\boldsymbol{y}^{P/N}$), it allows us to more flexibly control the size of the Utopian polygon (Figure C.6). For example, the Utopian polygons (Figure C.7) represent three cases when the pseudo Nadir point ($\boldsymbol{y}^P$), the Nadir point ($\boldsymbol{y}^N$) or a mixed Nadir point ($\boldsymbol{y}^{P/N} = [y_1^N \quad y_2^P \quad y_3^N]^T$) is adopted.

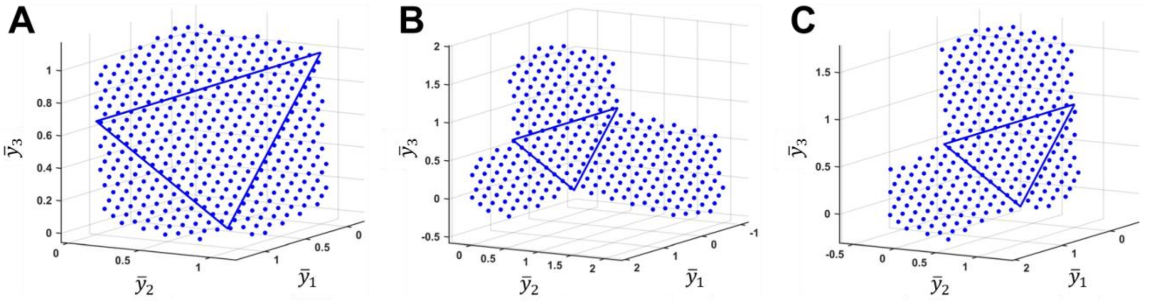


Figure C.7: Different Utopian polygons in the objective space. They are controlled by the pseudo Nadir point ($\boldsymbol{y}^P$) (a), the Nadir point ($\boldsymbol{y}^N$) (b) and a mixed Nadir point ($\boldsymbol{y}^{P/N} = [y_1^N \quad y_2^P \quad y_3^N]^T$) (c), respectively.

C.2.4. Discretization of Slicing Intersection Line on Utopian Plane

A complex 3-D curved surface can be better understood using 2-D contours. Similarly, the author proposed an algorithm to slice the 3-D Pareto frontier. As illustrated in Figure C.8, an intersection line can be obtained by slicing the Utopian plane using a plane. Here, we consider a slicing plane parallel to the y_2–y_3 coordinate plane located at y_1^{slc}. Without loss of generality, the size of the slicing plane is regulated by the coordinates of either pseudo Nadir point or Nadir point, $y_2^{P/N}$ and $y_3^{P/N}$ in this case.

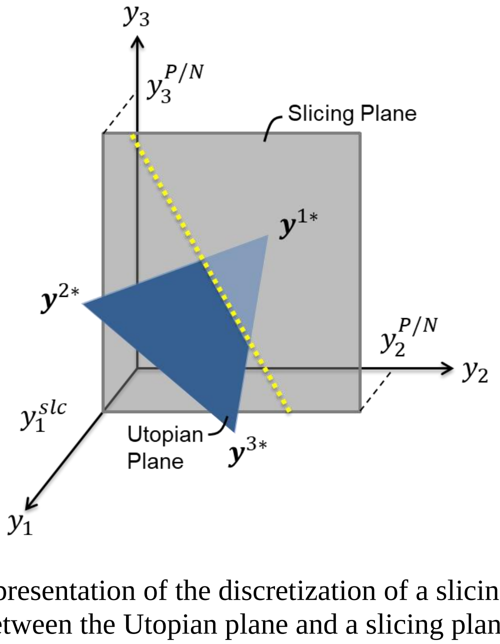


Figure C.8: The graphic representation of the discretization of a slicing intersection line (yellow) between the Utopian plane and a slicing plane.

The equation of the Utopian plane can be expressed in terms of the coordinates (y_1, y_2, y_3 when m = 3) of the objective space:

$$ay_1 + by_2 + cy_3 + d = 0 \tag{C.35}$$

where a, b, c, d are unknown parameters which need to be solved. It is noted that the Utopian plane can be completed determined by the three anchor points ($\boldsymbol{y}^{1*}$, $\boldsymbol{y}^{2*}$, $\boldsymbol{y}^{3*}$, when m = 3), with an additional notion that $[a\ b\ c]^T$ is the vector normal to the Utopian plane. The unit vector ($\boldsymbol{n}^{utp}$) normal to the Utopian plane can be obtained by,

$$\boldsymbol{n}^{utp} = [a\ b\ c]^T = \frac{\boldsymbol{v}_2 \times \boldsymbol{v}_1}{|\boldsymbol{v}_2 \times \boldsymbol{v}_1|} \tag{C.36}$$

where $\boldsymbol{v}_1 = \boldsymbol{y}^{2*} - \boldsymbol{y}^{1*}$ and $\boldsymbol{v}_2 = \boldsymbol{y}^{3*} - \boldsymbol{y}^{2*}$. By substituting $\boldsymbol{n}^{utp}$ (Equation C.36) and any one of anchor points (*e.g.*, $\boldsymbol{y}^{1*}$) into Equation C.35, the remaining parameter d is obtained,

$$d = -\boldsymbol{n}^{utp} \cdot \boldsymbol{y}^{1*} \tag{C.37}$$

The equation of the slicing intersection line can be obtained by combing the equations of the Utopian plane (Equation C.35) and the slicing plane ($y_1 = y_1^{slc}$),

$$y_2 = \alpha y_3 + \beta \tag{C.38}$$

where

$$\alpha = -\frac{c}{b} \quad , \quad \beta = -\frac{a y_1^{slc} + d}{b} \tag{C.39}$$

The slicing intersection line can be discretized by substituting evenly spaced coordinates (*e.g.*, y_2). Since the size of the slicing plane is regulated by $y_2^{P/N}$ and $y_3^{P/N}$, we have the following inequalities (Equation C.40):

$$\begin{cases} y_2^U \leq y_2 \leq y_2^{P/N} \\ y_3^U \leq y_3 \leq y_3^{P/N} \end{cases} \tag{C.40}$$

Alternatively,

$$\begin{cases} y_2^U \leq y_2 \leq y_2^{P/N} \\ \alpha y_3^U + \beta \leq y_2 \leq \alpha y_3^{P/N} + \beta \end{cases} \tag{C.41}$$

Therefore, the evenly spaced coordinates of y_2 should be chosen from the range below:

$$\max(y_2^U, \alpha y_3^U + \beta) \leq y_2 \leq \min(y_2^{P/N}, \alpha y_3^{P/N} + \beta) \tag{C.42}$$

C.2.5. Implementation of Normal-boundary Intersection Optimization Algorithm

In this study, we are interested in the boundary ($\partial \mathit{\Omega}_{\boldsymbol{y}}$) of the feasible objective space ($\mathit{\Omega}_{\boldsymbol{y}}$, Figure C.4a), which is represented by evenly spaced boundary points. The boundary points can be obtained by projecting the discretized Utopian plane to the boundary. The technique is well known as the normal-boundary intersection (NBI), which was proposed by Das and Dennis (1998). It is an efficient approach to address the generic MOO problem (Equation C.19).

The NBI can be stated as an optimization problem, in which the solution is the design variable vector ($\boldsymbol{x}$) corresponding to points on the boundary ($\partial \mathit{\Omega}_{\boldsymbol{y}}$) of the feasible objective space ($\mathit{\Omega}_{\boldsymbol{y}}$):

$$\max_{\boldsymbol{x},\rho} \rho \quad \Rightarrow \quad \min_{\boldsymbol{x},\rho} (-\rho) \tag{C.43}$$

subject to

$$\boldsymbol{y}^{utp} + \rho \hat{\boldsymbol{n}} = \boldsymbol{y}(\boldsymbol{x}) \tag{C.44}$$

and Equations C.20, C.21, and C.22.

where an auxiliary scalar variable ($\rho \in [-\infty, \infty]$) is introduced. $\boldsymbol{y}^{utp}$ represents a point on the Utopian plane, which has been introduced Equation C.32. $\hat{\boldsymbol{n}}$ is the unit projection vector in a direction towards the boundary ($\partial \mathit{\Omega}_{\boldsymbol{y}}$). $\boldsymbol{y}(\boldsymbol{x})$ is a point within the feasible design space ($\mathit{\Omega}_{\boldsymbol{y}}$), which is fitted using NN (Equation C.7). Therefore, the physical meaning of the NBI is that the point ($\boldsymbol{y}^{utp}$) on the Utopian plane is moved along the direction of $\hat{\boldsymbol{n}}$ as far as possible within the feasible objective space ($\mathit{\Omega}_{\boldsymbol{y}}$); the farthest location where $\boldsymbol{y}^{utp}$ can arrive represent a point on the boundary ($\partial \mathit{\Omega}_{\boldsymbol{y}}$).

To project the point ($\boldsymbol{y}^{utp}$) on the Utopian plane onto the boundary ($\partial\Omega_{\boldsymbol{y}}$) of the feasible objective space, the projection vector ($\hat{\boldsymbol{n}}$) can be chosen as the unit vector ($\boldsymbol{n}^{utp}$) normal to the Utopian plane,

$$\hat{\boldsymbol{n}} = \boldsymbol{n}^{utp} \tag{C.45}$$

where $\boldsymbol{n}^{utp}$ has been obtained in Equation C.36.

Similarly, a boundary slice can be obtained by projecting the discretized intersection line (Figure C.8) on the onto the boundary ($\partial\Omega_{\boldsymbol{y}}$) using the NBI method as well. However, it requires an additional condition that the projection vector ($\hat{\boldsymbol{n}}$) should be within the slicing plane. By denoting the unit normal vector of the slicing plane as $\boldsymbol{n}^{slc}$, we can define the projection vector for the specific implementation:

$$\hat{\boldsymbol{n}} = (\boldsymbol{I} - \boldsymbol{n}^{slc} \otimes \boldsymbol{n}^{slc})\boldsymbol{n}^{utp} \tag{C.46}$$

where $\boldsymbol{I}$ is the identity tensor.

C.2.6. Pareto Filtering

The non-dominated boundary known as the Pareto frontier is a subset of the boundary ($\partial\Omega_{\boldsymbol{y}}$) of the feasible objective space under the extended Utopian polygon (Figure C.6). It cannot be guaranteed that the boundary ($\partial\Omega_{\boldsymbol{y}}$) generated by the NBI method is non-dominated / Pareto-optimal. The non-dominated and dominated boundaries can be well explained considering an example of bi-objective optimization, as shown in Figure C.9. When the boundary ($\partial\Omega_{\boldsymbol{y}}$) under the Utopian plane (the connection line between the two anchor points $\boldsymbol{y}^{1*}$ and $\boldsymbol{y}^{2*}$ in the 2-D objective space) is convex, the boundary is naturally non-dominated (Figure C.9a). However, a concave region of the boundary is typically dominated (Figure C.9b). Therefore, after executing the NBI algorithm, it is essential to filter out the dominated points to represent the non-dominated boundary or Pareto frontier.

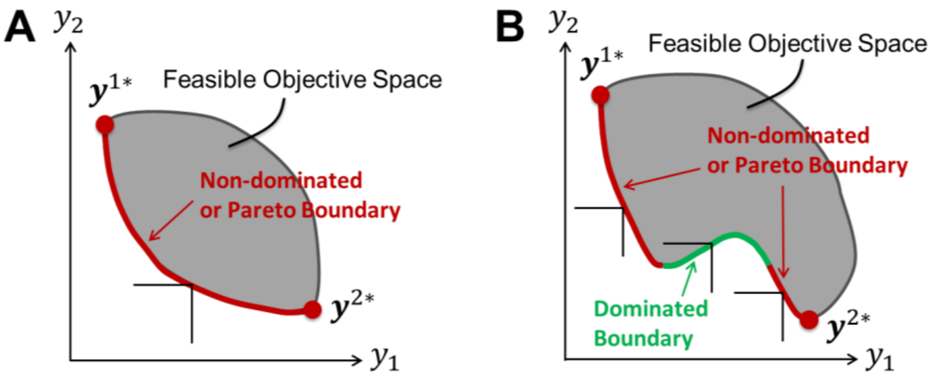


Figure C.9: Illustration of non-dominated (or Pareto) boundary and dominated boundary. (a) A convex boundary; (b) A boundary consisting of convex and concave regions.

A Pareto filtering algorithm for tri-objective optimization is widely available in the literature [88, 119]. Essentially, the dominated points defined below need to be excluded from the set of points on the boundary ($\partial\Omega_{\boldsymbol{y}}$), such that the remaining boundary points represent the Pareto frontier.

$$\{\boldsymbol{y} \in \partial\Omega_{\boldsymbol{y}} \mid y_1 > y_1' \text{ or } y_2 > y_2' \text{ or } y_3 > y_3', \text{ for all } \boldsymbol{y}' \in \partial\Omega_{\boldsymbol{y}}\} \quad \text{(C.47)}$$

More generally, these dominated points can be defined as,

$$\{\boldsymbol{y} \in \Omega_{\boldsymbol{y}} \mid y_1 > y_1' \text{ or } y_2 > y_2' \text{ or } y_3 > y_3', \text{ for all } \boldsymbol{y}' \in \Omega_{\boldsymbol{y}}\} \quad \text{(C.48)}$$

where $\Omega_{\boldsymbol{y}}$ is the feasible objective space.

C.2.7. Normalization of Design Variables and Design Objectives

Similar to the normalization of the NN inputs and outputs, the design variables and deign objectives in MOO were also normalized in order to enhance the optimization efficiency. Here, we scale the design variables to [0, 1], and the design objectives to [0, C], $C \geq 1$ due to using the pseudo Nadir point ($\boldsymbol{y}^P$) instead of the Nadir point ($\boldsymbol{y}^N$). By denoting the normalized design variables and design objectives as $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ in MOO, we have the following relationships:

$$\bar{\boldsymbol{x}} = \frac{\boldsymbol{x} - \boldsymbol{x}_l}{\boldsymbol{x}_u - \boldsymbol{x}_l} \quad or \quad \boldsymbol{x} = (\boldsymbol{x}_u - \boldsymbol{x}_l)\bar{\boldsymbol{x}} + \boldsymbol{x}_l \quad \text{(C.49)}$$

$$\overline{\boldsymbol{y}} = \frac{\boldsymbol{y} - \boldsymbol{y}^U}{\boldsymbol{y}^P - \boldsymbol{y}^U} \quad or \quad \boldsymbol{y} = (\boldsymbol{y}^P - \boldsymbol{y}^U)\overline{\boldsymbol{y}} + \boldsymbol{y}^U \tag{C.50}$$

where all operations are element-wise vector operations.

The normalization of design variables and design objectives causes a change in the conversion of gradients between using the true scale and the normalized scale,

$$\frac{\partial \overline{\boldsymbol{y}}}{\partial \overline{\boldsymbol{x}}} = \frac{\boldsymbol{e}_m \otimes (\boldsymbol{x}_u - \boldsymbol{x}_l)}{(\boldsymbol{y}^P - \boldsymbol{y}^U) \otimes \boldsymbol{e}_n} \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} \tag{C.51}$$

where the division and multiplication are the component-wise operations for the matrices with the dimension of $m \times n$, after the outer products ($\otimes$) of vectors are computed. $\boldsymbol{e}_m$ and $\boldsymbol{e}_n$ are the vectors whose elements are all ones with the dimension of m and n, respectively. It should be noted that $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$ has been obtained according to the continuous differentiability of the NN (Equation C.9). In the MOO algorithm described above, we have used normalized design variables and normalized design objectives. When implementing the MOO algorithm, all solutions should be scaled back to the true scale after finishing the algorithm execution.

Appendix D: Sensitivity Analyses of Lumbar Motion Segment Finite Element Models

D.1. Convergence Study of the Contact Element Size on Metal-on-Polyethylene Articulation

In design optimization of the biconcave-core TDA, the peak polyethylene contact pressure was chosen as a performance metric, thus a convergence study for the contact element size was performed to ensure accurate simulation results. As depicted in Figure D.1, the PE core of a biconcave-core TDA design with design variables listed in Table D.1 was meshed using 20-node brick elements and assigned J_2 plasticity material properties. Superior and inferior MoP articulations were modeled by contact with a coefficient of friction of 0.02 [73, 101, 122]. The average size of contact elements on the superior and inferior surfaces of the PE core was set to 1.2, 1.0, 0.8, and 0.6 mm. A postoperative L3-L4 lumbar segment FE model implanted by the biconcave TDA (Figure D.1) was applied a compressive follower preload of 500 N along the curve of the lumbar, followed by followed by a moment linearly ramping up to 7.5 Nm in flexion, extension, lateral bending and axial torsion. To save computational cost, lumbar vertebrae and TDA metallic endplates both were modeled as rigid bodies.

Table D.1: Design variables (mm) of the parametric biconcave-core TDA design (*Section 5.3*) used in the mesh convergence study.

r_s	r_i	d_s	d_i	g_a	g_l	g_p
12.97	14.05	3.80	3.90	1.25	0.30	0.43

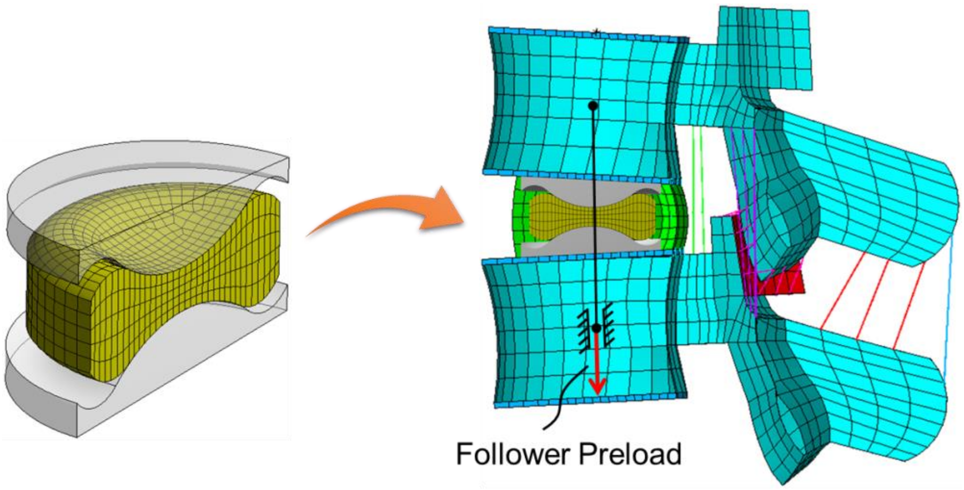


Figure D.1: Illustration of a postoperative L3-L4 lumbar motion segment FE model, into which a biconvex-core TDA was inserted. To save computational cost, lumbar vertebrae and TDA metallic endplates were modeled as rigid bodies.

As shown in Figure D.2, simulated polyethylene contact pressure (PCP) distributions (element solutions) on the superior and inferior of the biconcave PE core in different loading scenarios at 7.5 Nm became more continuous, as the size of contact elements was refined from 1.2 to 0.6 mm. No distinct difference in PCP distributions was observed between the PE cores with element sizes of 0.8 and 0.6 mm. For maximum PCPs (Figure D.3), they converged when the contact element size was decreased to 0.8 mm in flexion and lateral bending, but kept changing as the contact element size was decreased from 0.8 to 0.6 mm in extension and axial torsion. In terms of computational cost, the total time for simulation of the follower preload and four loading scenarios was exponentially increased from 31 to 106 minutes, as the contact element size was decreased from 0.8 to 0.6 mm (Table D.2). Since design optimization typically required 200 ~ 300 iterations (each iteration performed simulation of preload plus 4 loading scenarios), the element size of PE contact elements was chosen as 0.8 mm for the biconcave-core TDA to balance the computation cost and prediction accuracy.

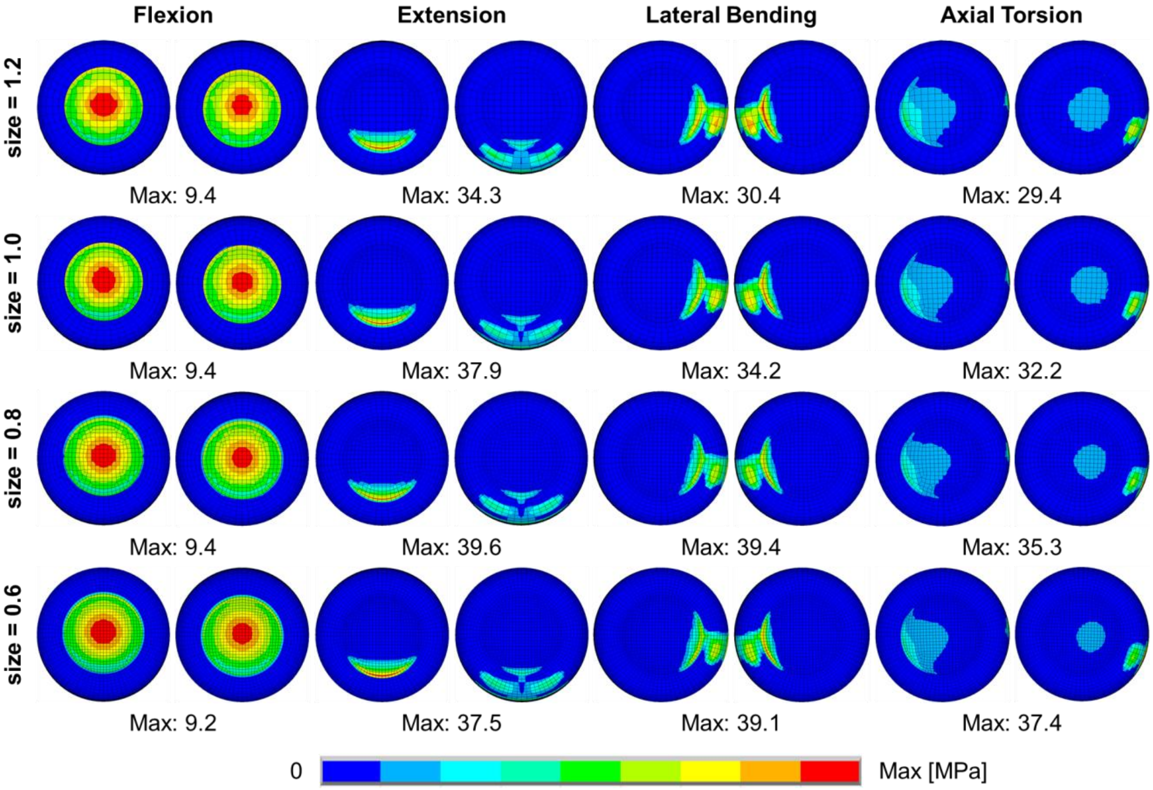


Figure D.2: Simulated contact pressure distributions (MPa) on the superior (left) and inferior (right) of the biconcave PE core in different loading scenarios at 7.5 Nm, while the size of contact elements was changed from 1.2 to 0.6 mm.

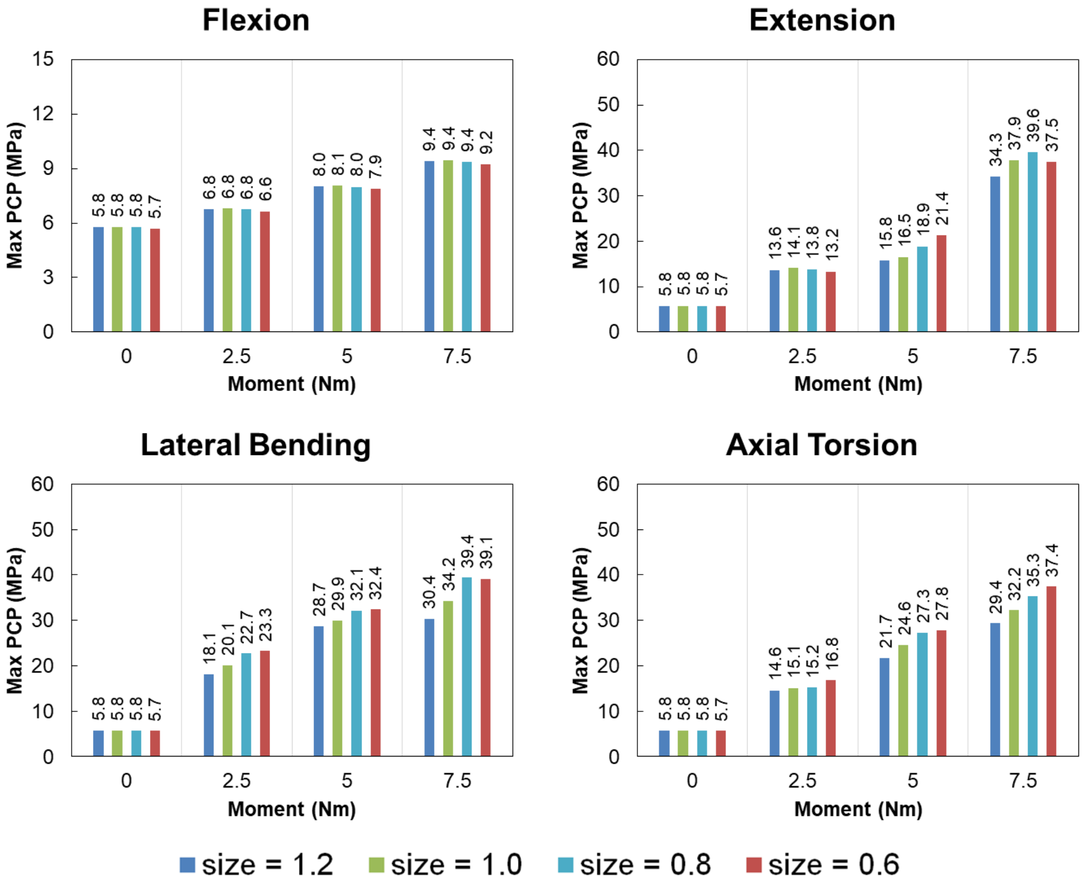


Figure D.3: Maximum polyethylene contact pressures (PCPs) in flexion, extension, lateral bending and axial torsion, while the size of contact elements was changed from 1.2 to 0.6 mm.

Table D.2: Computational cost of the lumbar segment FE model for different mesh sizes of PE contact elements.

Mesh size (mm)	Time for simulation of the follower preload (min)	Time for in-parallel simulation of 4 loading scenarios (min)	Total time (min)
1.2	2.38	8.50	10.88
1.0	3.53	17.05	20.58
0.8	4.32	26.70	31.02
0.6	8.12	98.15	106.27

D.2. Comparison of Using Deformable versus Rigid Lumbar Segment Finite Element Models

To save computational cost in TDA design optimization, lumbar vertebrae and TDA metallic endplates were simplified to rigid bodies. Here, we scrutinize the effect of the rigid body simplification on simulated mechanical responses of the L3-L4 lumbar segment, including range of motion (ROM), facet joint force (FJF) and polyethylene contact pressure (PCP), which are required to formulate proposed TDA performance metrics in design optimization. Deformable and rigid FE models of intact and treated segments were presented in Figure D.4. Without loss of generality, both deformable and rigid treated segments (Figure D.4c,d) were modeled using a commercially available biconvex-core TDA with a dome radius of curvature of 10 mm and a rim thickness of 2.5 mm.

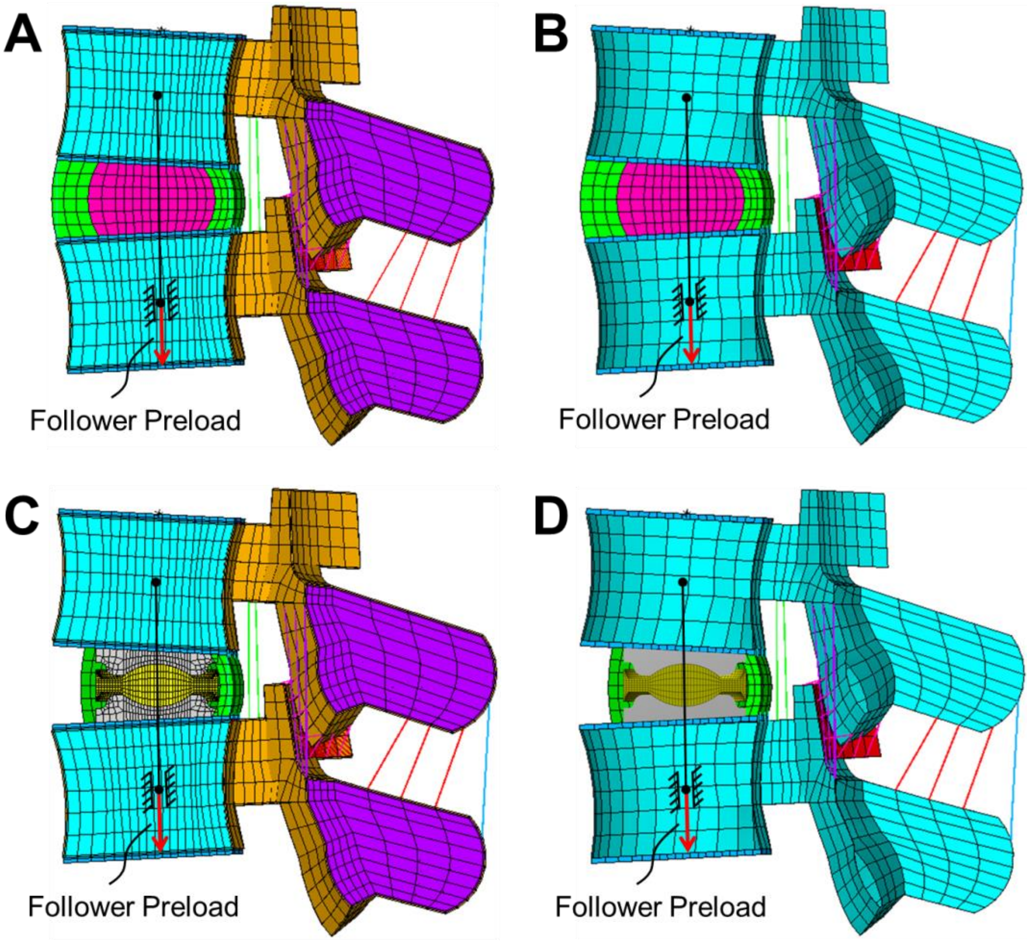


Figure D.4: Different L3-L4 lumbar segment finite element models. (a) Deformable intact segment model; (b) Rigid intact segment model; (c) Deformable treated segment model; (d) Rigid treated segment model.

In general, rigid intact and treated segment models constantly resulted in greater ROMs (Figure D.5) and FJFs (Figure D.6) in all loading scenarios, compared to deformable segment models. It was observed that PCPs in rigid segment models were always larger than those in deformable segment models during flexion and extension, as the moment applied increased, but lacked consistency in lateral bending and axial torsion (Figure D.7). Furthermore, the simulation time to run the rigid TDA-treated segment model (128 min) was distinctly shorter than that using the deformable treated segment model (325 min), as listed in Table D.3. Therefore, the vertebrae and TDA metallic endplates were modeled as rigid bodies in the treated segment model, which was used in TDA design optimization.

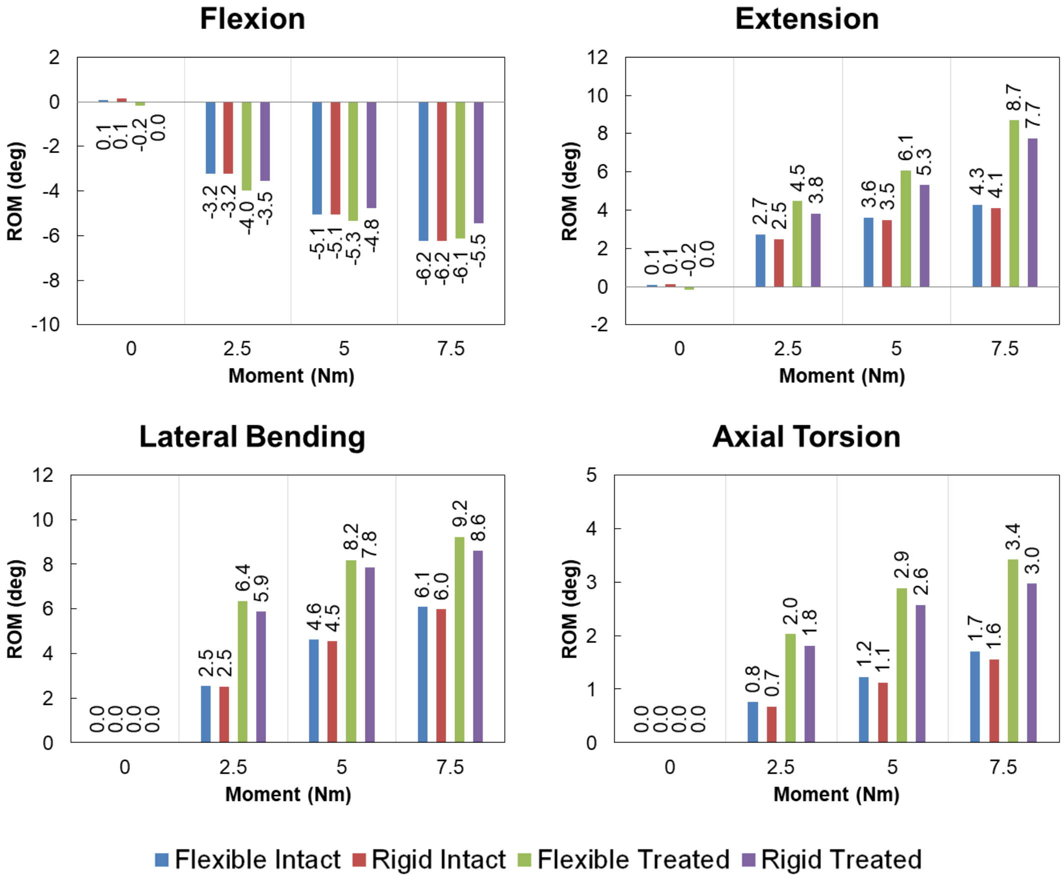


Figure D.5: Ranges of motion of segment models in different loading scenarios (0 Nm represents preloading).

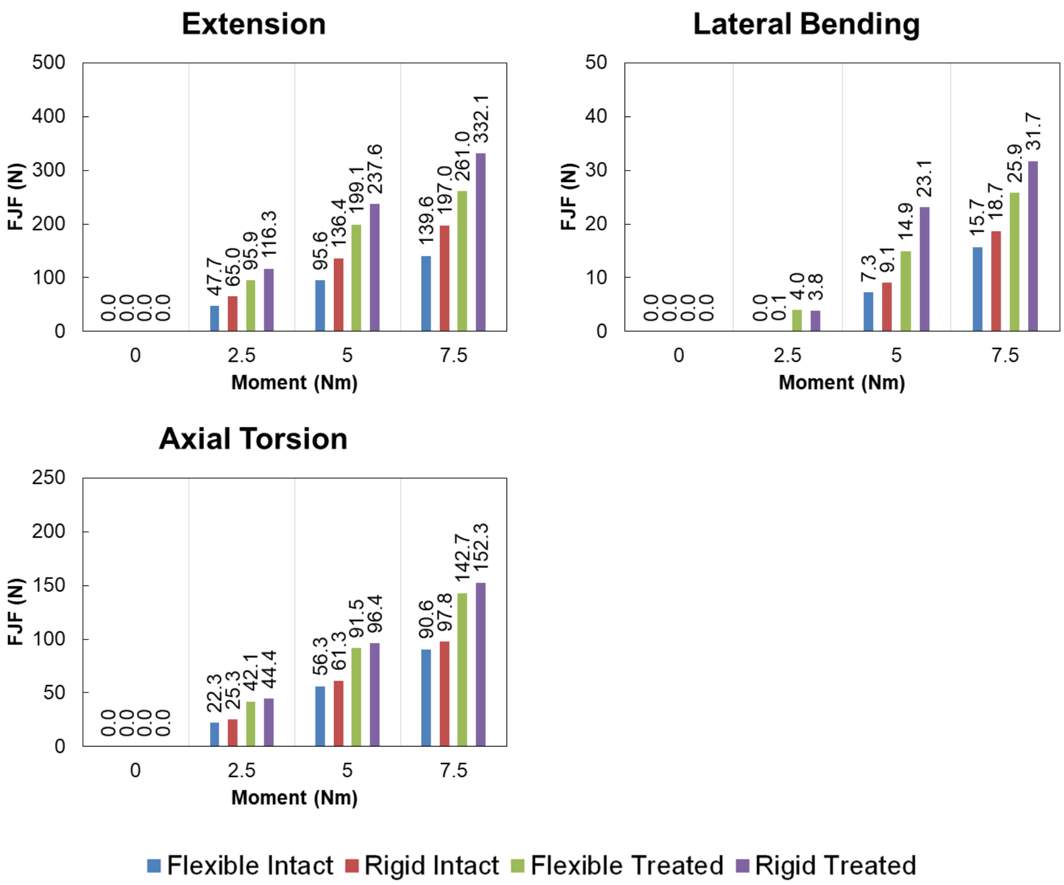


Figure D.6: Facet joint forces in segment models in different loading scenarios (0 Nm represents preloading). It is noted that no facet joint forces occur during flexion for all segment models.

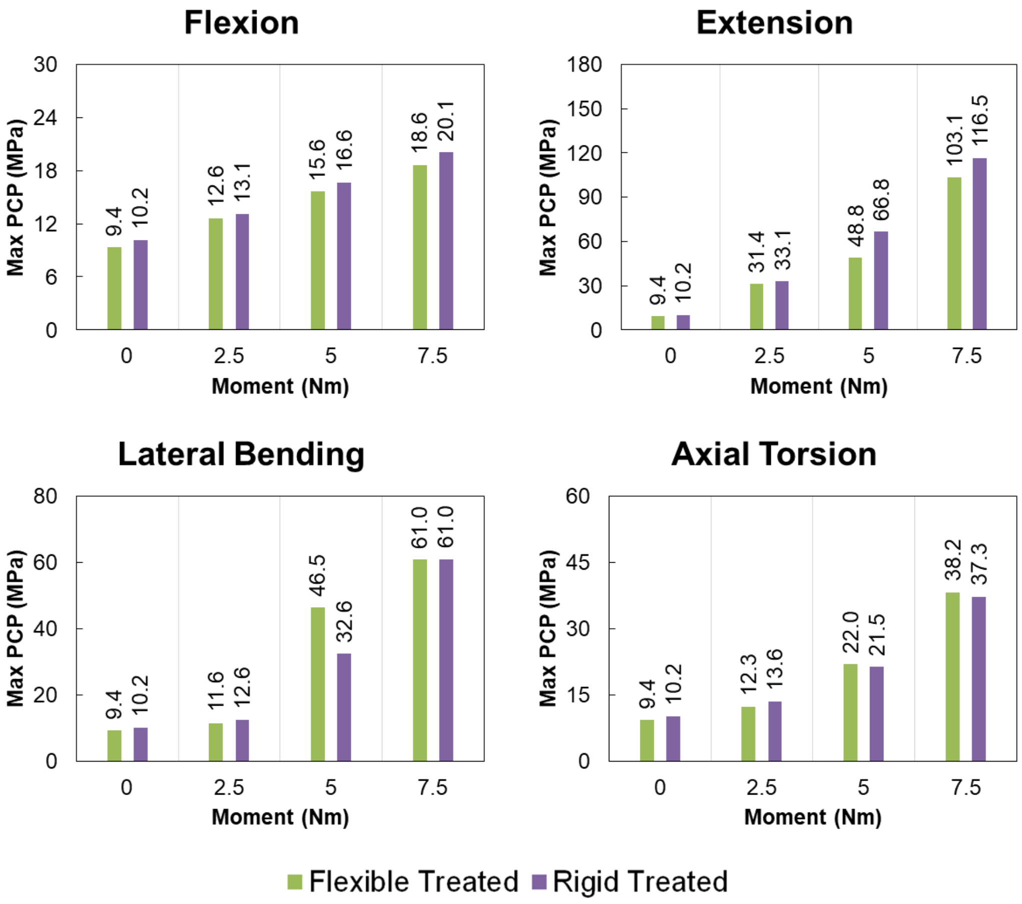


Figure D.7: Maximum PE contact pressures in TDA-treated segment models during different loading scenarios (0 Nm represents preloading).

Table D.3: Comparison of computational cost using deformable and rigid segment models.

Segment Models	Time for simulation of the follower preload (min)	Time for in-parallel simulation of 4 loading scenarios (min)	Total time (min)
Flexible Intact	22.17	65.50	87.67
Rigid Intact	0.92	3.07	3.99
Flexible Treated	26.00	298.62	324.62
Rigid Treated	6.38	122.05	128.43

D.3. Sensitivities of Lumbar Segment Responses to Changes in Tensile Properties of Spinal Ligaments

Spinal ligaments play a role in the maintenance of spinal stability. However, it is not well known whether the inter-subject material properties of spinal ligaments cause a distinct difference in the spinal responses (ROM, FJF, and PCP) after TDA. Here, previously developed rigid L3-L4 intact and TDA-treated segment FE models (Figure D.8) were employed to investigate the sensitivities of these responses to changes in tensile properties of spinal ligaments in various loading scenarios (flexion, extension, lateral bending, and axial torsion). For the treated segment model (Figure D.8b), the best-trade-off biconcave TDA design obtained using multiobjective optimization (*Chapter 6*) was implanted.

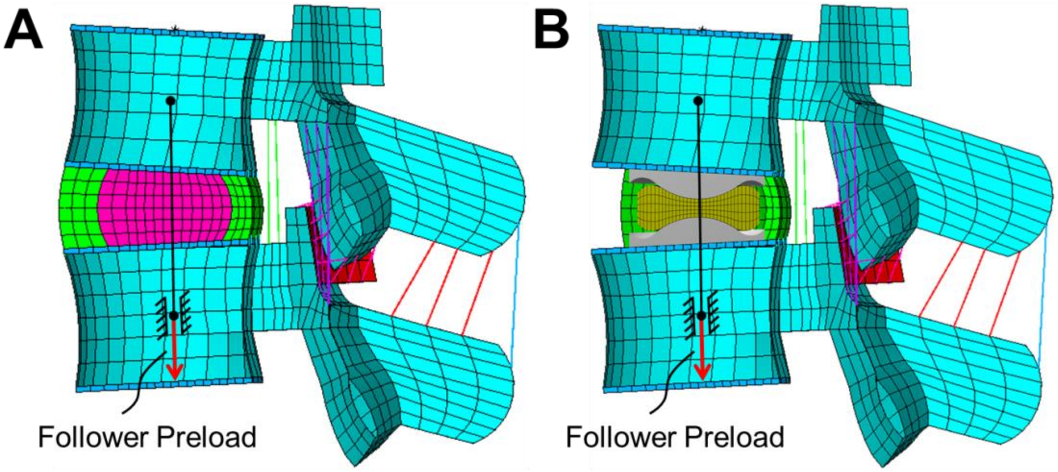


Figure D.8: The L3-L4 intact (a) and treated (b) FE segment models, in which vertebrae and TDA metallic endplates were modeled as rigid bodies.

As shown in Figure D.9a, the calibrated tensile curves of seven spinal ligaments (*Section 4.2.3*) were chosen as the baseline tensile curves. To simulate the inter-subject variation in tensile properties of spinal ligaments, the baseline tensile curve of each ligament was stiffened or softened in a 25% increment, by multiplying by a scale factor (λ = 50%, 75%, 125%, and 150%), as formulated in Equation D.1:

$$F'(D) = \lambda \cdot F(D) \tag{D.1}$$

where $F(D)$ and $F'(D)$ represent the baseline and varied force-deflection curves, respectively. The resulting tensile curves varied in a 25% increment are illustrated in Figure D.9b.

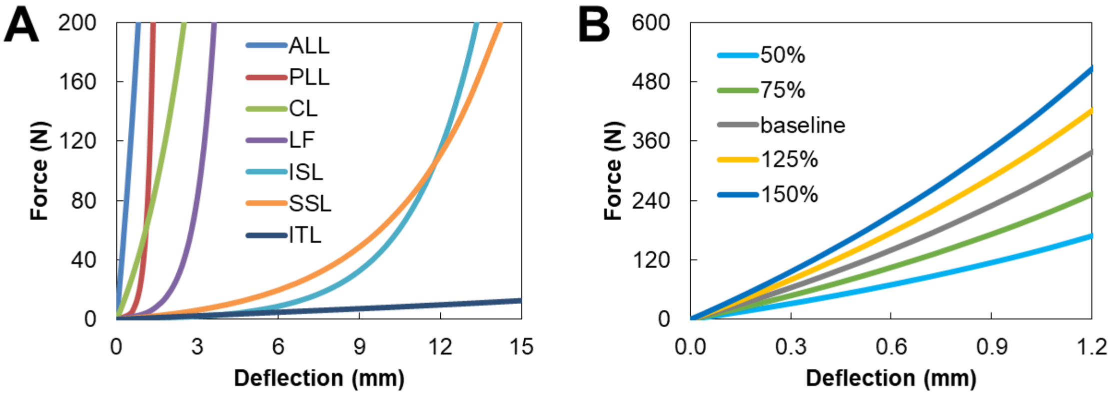


Figure D.9: The baseline and varied ligament tensile curves. (a) The calibrated tensile (force-deflection) curves of 7 spinal ligaments (*Section 4.2.3*). (b) Different ligament tensile curves yielded by elevating or declining the calibrated (baseline) tensile curve in a 25% increment. Here, only the resulting tensile curves of the anterior longitudinal ligament (ALL) are presented.

For the intact segment model, our simulation results showed that the decline of ligament tensile curves increased both ROMs (Figure D.10 and Table D.4) and FJFs (Figure D.11 and Table D.5) in all loading scenarios; the opposite changes in ROMs and FJFs occurred when ligament tensile curves elevated. Moreover, it was noticed that changes in ligament tensile properties resulted in more distinct changes in ROMs and FJFs in flexion and extension, whereas only slight changes in ROMs and FJFs occurred in lateral bending and axial torsion.

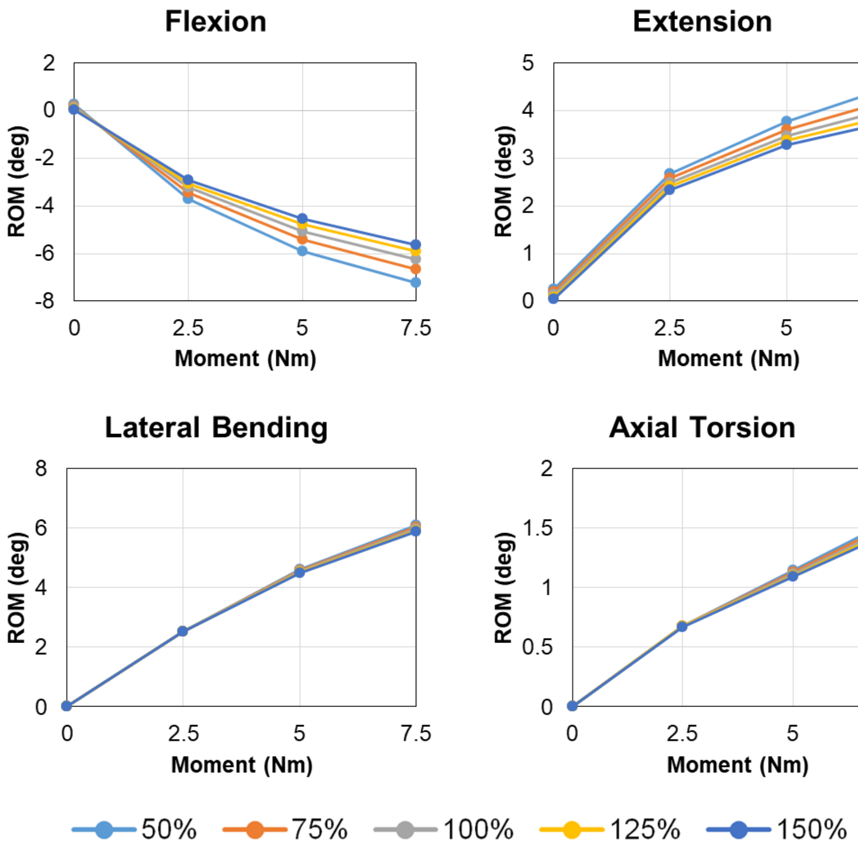


Figure D.10: The effect of changes in tensile properties of spinal ligaments on segmental ROMs simulated by the rigid intact segment model in different loading scenarios ("0 Nm" represents preloading).

Table D.4: The resulting changes in ROMs of the rigid intact segment model in response to 50% decline and elevation of ligament tensile curves in different loading scenarios, when the moments applied were ramped up to 7.5 Nm.

Loading Scenarios	50% Decline of Tensile Curves	50% Elevation of Tensile Curves
Flexion	+15.9%	-10.0%
Extension	+11.7%	-6.7%
Lateral Bending	+2.0%	-1.7%
Axial Torsion	+4.1%	-2.8%

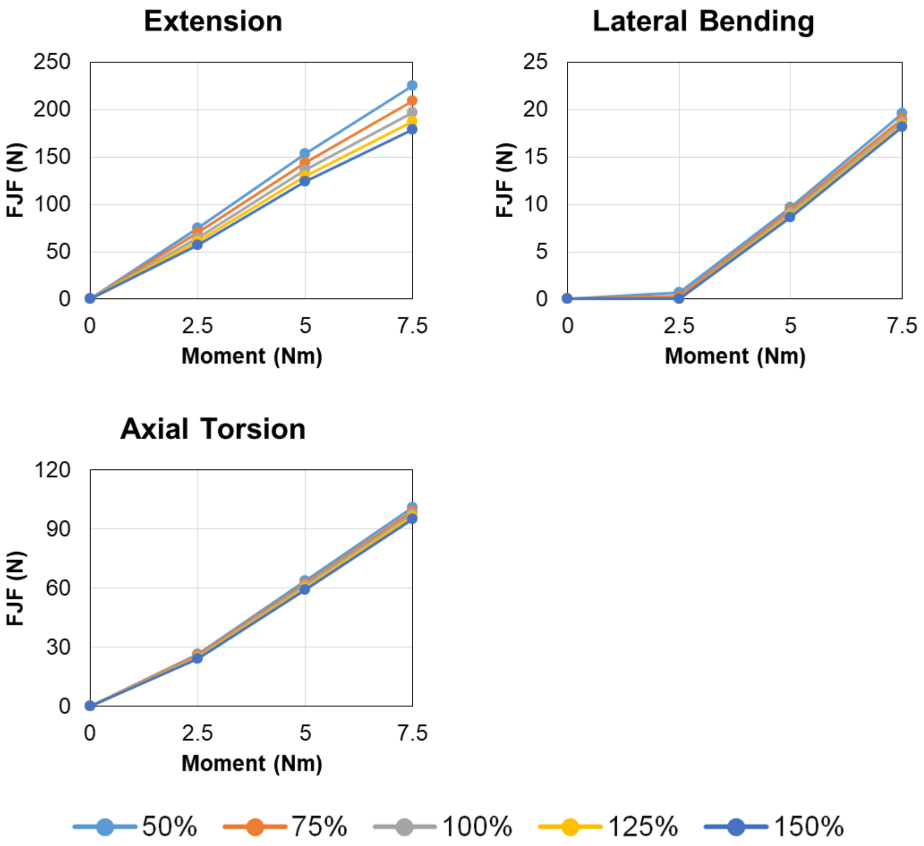


Figure D.11: The effect of changes in tensile properties of spinal ligaments on FJFs simulated by the rigid intact segment model in different loading scenarios (“0 Nm” represents preloading). It is noted that no FJF exists during flexion.

Table D.5: The resulting changes in FJFs of the rigid intact segment model in response to 50% decline and elevation of ligament tensile curves in different loading scenarios, when the moments applied were ramped up to 7.5 Nm. It is noted that no FJF exists during flexion.

Loading Scenarios	50% Decline of Tensile Curves	50% Elevation of Tensile Curves
Extension	+14.6%	-9.1%
Lateral Bending	+4.7%	-2.8%
Axial Torsion	+3.1%	-2.6%

Similar to the intact segment model, the decline and elevation of ligament tensile curves caused an increase and decrease in both ROMs (Figure D.12 and Table D.6) and FJFs (Figure D.13 and Table D.7) during all loading scenarios, respectively, after the best-trade-off biconcave TDA design was implanted. Simulated ROMs are more sensitive to changes in ligament tensile properties in flexion and axial torsion. Extension was the only loading scenario that led to a distinct difference in FJFs due to varied ligament tensile properties. For PCPs (Figure D.8 and Table D.8), only consistent changes in response to varied ligament tensile properties were observed in lateral bending and axial torsion; that is, PCPs relieved as ligaments were stiffened. In flexion and extension, the trends of PCP changes as ligament tensile properties changed were unpredictable. The main reason was that the contact region on the biconcave PE core changed in different rotation angles in each loading scenario. For example, the decline of ligament tensile curves from the baseline to 50% gradually increased the ROM from 5.0° to 5.6° in flexion; at 5.6° in flexion, the superior metallic endplate contacted the superior edge of the PE core, thus the maximum PCP soared to 19.8 MPa, compared to 8.4 MPa using the baseline ligament curves.

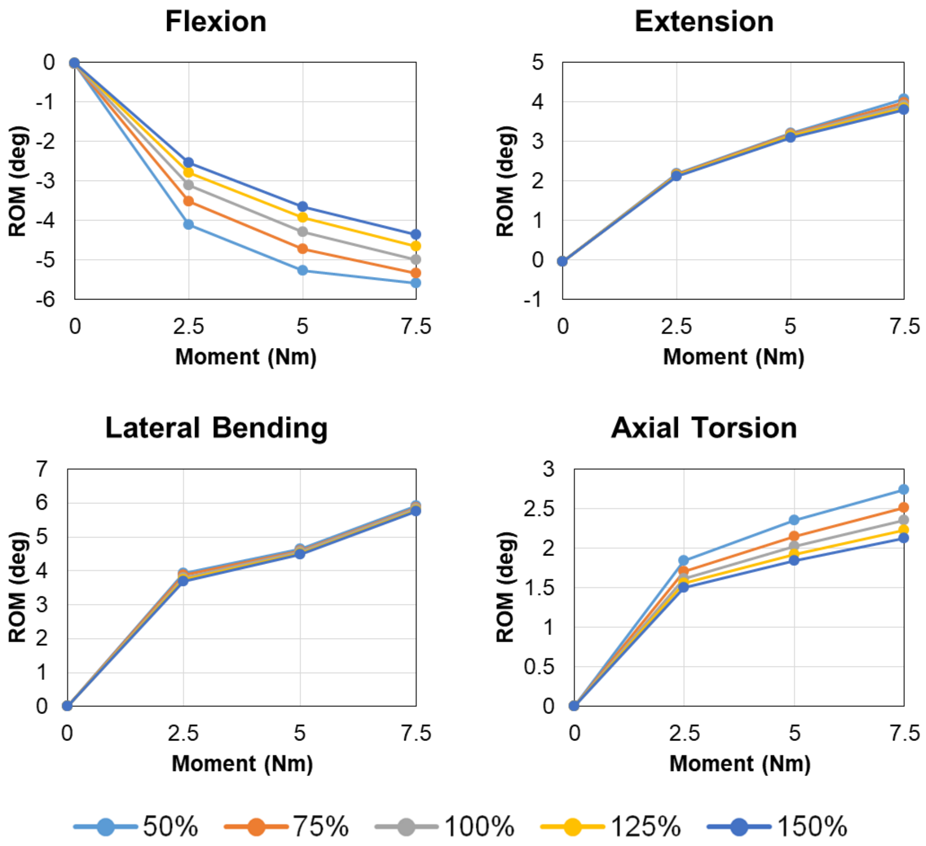


Figure D.12: The effect of changes in tensile properties of spinal ligaments on segmental ROMs simulated by the rigid treated segment model in different loading scenarios (“0 Nm” represents preloading).

Table D.6: The resulting changes in ROMs of the rigid treated segment model in response to 50% decline and elevation of ligament tensile curves in different loading scenarios, when the moments applied were ramped up to 7.5 Nm.

Loading Scenarios	50% Decline of Tensile Curves	50% Elevation of Tensile Curves
Flexion	+11.6%	-12.7%
Extension	+4.3%	-2.8%
Lateral Bending	+1.7%	-1.4%
Axial Torsion	+16.6%	-9.4%

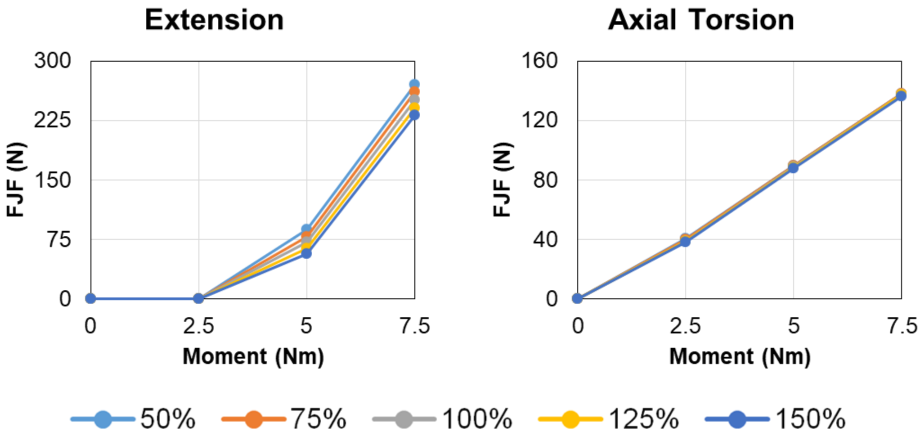


Figure D.13: The effect of changes in tensile properties of spinal ligaments on FJFs simulated by the rigid treated segment model in different loading scenarios ("0 Nm" represents preloading). It is noted that no FJF occurs during flexion and lateral bending for the best-tread-off biconcave TDA design.

Table D.7: The resulting changes in FJFs of the rigid treated segment model in response to 50% decline and elevation of ligament tensile curves in different loading scenarios, when the moments applied were ramped up to 7.5 Nm. It is noted that no FJF occurs during flexion and lateral bending for the best-tread-off biconcave TDA design.

Loading Scenarios	50% Decline of Tensile Curves	50% Elevation of Tensile Curves
Extension	+8.0%	-7.6%
Axial Torsion	+0.5%	-0.8%

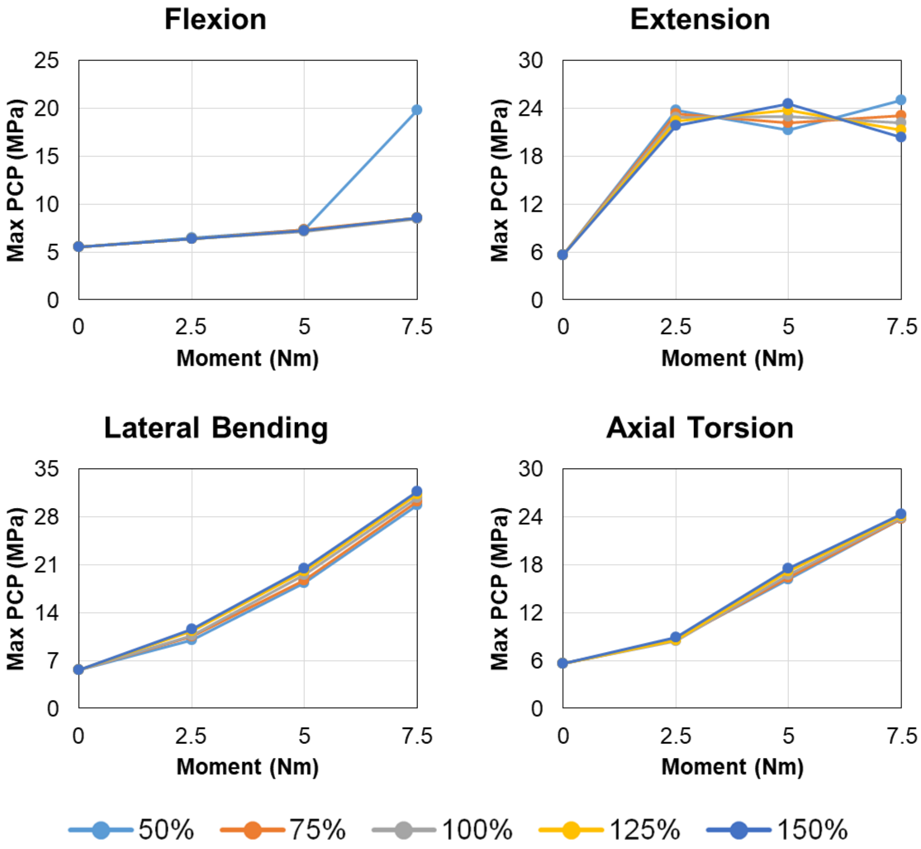


Figure D.14: The effect of changes in tensile properties of spinal ligaments on PCPs simulated by the rigid treated segment model in different loading scenarios (“0 Nm” represents preloading).

Table D.8: The resulting changes in PCPs of the rigid treated segment model in response to 50% decline and elevation of ligament tensile curves in different loading scenarios, when the moments applied were ramped up to 7.5 Nm.

Loading Scenarios	50% Decline of Tensile Curves	50% Elevation of Tensile Curves
Flexion	+135.3%	+1.8%
Extension	+12.9%	-8.0%
Lateral Bending	-3.7%	+2.6%
Axial Torsion	-0.8%	+1.6%

Bibliography

[1] Vällfors B. Acute, subacute and chronic low back pain: clinical symptoms, absenteeism and working environment. *Scand J Rehabil Med Suppl* 1985; 11: 1–98.

[2] Maetzel A, Li L. The economic burden of low back pain: a review of studies published between 1996 and 2001. *Best Pract Res Clin Rheumatol* 2002; 16: 23–30.

[3] Urban JPG, Roberts S. Degeneration of the intervertebral disc. *Arthritis Res Ther* 2003; 5: 120–130.

[4] Adams M, Roughley PJ. What is intervertebral disc degeneration, and what causes it? *Spine (Phila Pa 1976)* 2006; 31(18): 2151–2161.

[5] McNally DS, Adams MA. Internal intervertebral disc mechanics as revealed by stress profilometry. *Spine* 1992; 17: 66–73.

[6] Kurtz SM, Patwardhan A, MacDonald D, et al. What is the correlation of in vivo wear and damage patterns with in vitro TDR motion response? *Spine (Phila Pa 1976)* 2008; 33: 481–9.

[7] Rundell SA, Day JS, Isaza J, et al. Lumbar total disc replacement impingement sensitivity to disc height distraction, spinal sagittal orientation, implant position, and implant lordosis. *Spine (Phila Pa 1976)* 2012; 37: E590–E598.

[8] Marchi L, Oliveira L, Coutinho E, et al. The importance of the anterior longitudinal ligament in lumbar disc arthroplasty: 36-Month follow-up experience in extreme lateral total disc replacement. *Int J Spine Surg* 2012; 6: 18–23.

[9] Galbusera F, Bellini CM, Zweig T, et al. Design concepts in lumbar total disc arthroplasty. *Eur Spine J* 2008; 17: 1635–1650.

[10] Kurtz SM, Toth JM, Siskey R, et al. The Latest Lessons Learned from Retrieval Analyses of Ultra-High Molecular Weight Polyethylene, Metal-on-Metal, and Alternative Bearing Total Disc Replacements. *Semin Spine Surg* 2012; 24: 57–70.

[11] Riveros F, Chandra S, Finol EA, et al. A pull-back algorithm to determine the unloaded vascular geometry in anisotropic hyperelastic AAA passive mechanics. *Ann Biomed Eng* 2013; 41: 694–708.

[12] Willing R, Kim IY. Quantifying the competing relationship between durability and kinematics of total knee replacements using multiobjective design optimization and validated computational models. *J Biomech* 2012; 45: 141–147.

[13] Chao C-K, Lin J, Putra ST, et al. A neurogenetic approach to a multiobjective design optimization of spinal pedicle screws. *J Biomech Eng* 2010; 132: 091006.

[14] Panjabi MM, Goel V, Oxland T, et al. Human lumbar vertebrae: quantitative three-dimensional anatomy. *Spine* 1992; 17: 299–306.

[15] Panjabi MM, Oxland T, Takata K, et al. Articular Facets of the Human Spine Quantitative Three-Dimensional Anatomy. *Spine (Phila Pa 1976)* 1993; 18: 1298–1310.

[16] Noailly J, Wilke HJ, Planell JA, et al. How does the geometry affect the internal biomechanics of a lumbar spine bi-segment finite element model? Consequences on the validation process. *J Biomech* 2007; 40: 2414–2425.

[17] Bartel DL, Davy DT, Keaveny TM. *Orthopaedic biomechanics : mechanics and design in musculoskeletal systems*. Pearson/Prentice Hall, 2006.

[18] Iatridis J, Wedenbaum M, Setton L, et al. Is the nucleus pulposus a solid or a fluid. *Spine (Phila Pa 1976)* 1996; 21: 1174–1184.

[19] Holzapfel GA, Schulze-Bauer CAJ, Feigl G, et al. Single lamellar mechanics of the human lumbar anulus fibrosus. *Biomech Model Mechanobiol* 2005; 3: 125–140.

[20] Zhu D, Gu G, Wu W, et al. Micro-structure and mechanical properties of annulus fibrous of the L4-5 and L5-S1 intervertebral discs. *Clin Biomech* 2008; 23: 74–82.

[21] Skaggs DL, Weidenbaum M, Iatridis JC, et al. Regional variation in tensile properties and biochemical composition of the human lumbar anulus fibrosus. *Spine (Phila Pa 1976)* 1994; 19: 1310–1319.

[22] Cassidy J, Hiltner A, Baer E. Hierarchical structure of the intervertebral disc. *Connect Tissue Res* 1989; 2: 75–88.

[23] Brickley-Parsons D, Glimcher MJ. Is the chemistry of collagen in intervertebral discs an expression of Wolff's Law? A study of the human lumbar spine. *Spine* 1984; 9: 148–63.

[24] Hollingsworth NT, Wagner DR. Modeling shear behavior of the annulus fibrosus. *J Mech Behav Biomed Mater* 2011; 4: 1103–1114.

[25] Dreischarf M, Zander T, Shirazi-Adl A, et al. Comparison of eight published static finite element models of the intact lumbar spine: Predictive power of models improves when combined together. *J Biomech* 2014; 47: 1757–1766.

[26] Marini G, Huber G, Püschel K, et al. Nonlinear dynamics of the human lumbar intervertebral disc. *J Biomech* 2015; 48: 479–488.

[27] Jacobs NT, Cortes DH, Peloquin JM, et al. Validation and application of an intervertebral disc finite element model utilizing independently constructed tissue-level constitutive formulations that are nonlinear, anisotropic, and time-dependent. *J Biomech* 2014; 47: 2540–2546.

[28] Heuer F, Schmidt H, Klezl Z, et al. Stepwise reduction of functional spinal structures

increase range of motion and change lordosis angle. *J Biomech* 2007; 40: 271–280.

[29] Weiss JA. *A Constitutive Model and Finite Element Representation for Transversely Isotropic Soft Tissues*. The University of Utah, 1994.

[30] Ambard D, Cherblanc F. Mechanical behavior of annulus fibrosus: A microstructural model of fibers reorientation. *Ann Biomed Eng* 2009; 37: 2256–2265.

[31] Malandrino A, Noailly J, Lacroix D. Regional annulus fibre orientations used as a tool for the calibration of lumbar intervertebral disc finite element models. *Comput Methods Biomech Biomed Engin* 2013; 5842: 37–41.

[32] Weisse B, Aiyangar AK, Affolter C, et al. Determination of the translational and rotational stiffnesses of an L4-L5 functional spinal unit using a specimen-specific finite element model. *J Mech Behav Biomed Mater* 2012; 13: 45–61.

[33] Schmidt H, Heuer F, Simon U, et al. Application of a new calibration method for a three-dimensional finite element model of a human lumbar annulus fibrosus. *Clin Biomech* 2006; 21: 337–344.

[34] Schmidt H, Heuer F, Drumm J, et al. Application of a calibration method provides more realistic results for a finite element model of a lumbar spinal segment. *Clin Biomech* 2007; 22: 377–384.

[35] Noailly J, Planell JA, Lacroix D. On the collagen criss-cross angles in the annuli fibrosi of lumbar spine finite element models. *Biomech Model Mechanobiol* 2011; 10: 203–219.

[36] Panzer MB, Cronin DS. C4-C5 segment finite element model development, validation, and load-sharing investigation. *J Biomech* 2009; 42: 480–490.

[37] Li G, Wang S, Passias P, et al. Segmental in vivo vertebral motion during functional human lumbar spine activities. *Eur Spine J* 2009; 18: 1013–1021.

[38] Marini G, Studer H, Huber G, et al. Geometrical aspects of patient-specific modelling of the intervertebral disc: collagen fibre orientation and residual stress distribution. *Biomech Model Mechanobiol* 2016; 15: 543–560.

[39] Marini G, Ferguson SJ. Modelling the influence of heterogeneous annulus material property distribution on intervertebral disk mechanics. *Ann Biomed Eng* 2014; 42: 1760–1772.

[40] Schmidt H, Kettler A, Heuer F, et al. Intradiscal pressure, shear strain, and fiber strain in the intervertebral disc under combined loading. *Spine (Phila Pa 1976)* 2007; 32: 748–55.

[41] Newell N, Little JP, Christou A, et al. Biomechanics of the human intervertebral disc: A review of testing techniques and results. *J Mech Behav Biomed Mater* 2017; 69: 420–434.

[42] Heuer F, Schmidt H, Wilke HJ. Stepwise reduction of functional spinal structures increase vertebral translation and intradiscal pressure. *J Biomech* 2007; 40: 795–803.

[43] Yang KH, King AI. Mechanism of facet load transmission as a hypothesis for low-back

pain. *Spine (Phila Pa 1976)* 1984; 9: 557–565.

[44] Adams MA, Freeman BJC, Morrison HP, et al. Mechanical initiation of intervertebral disc degeneration. *Spine (Phila Pa 1976)* 2000; 25: 1625–1636.

[45] Adams MA, Lama P, Zehra U, et al. Why do some intervertebral discs degenerate, when others (in the same spine) do not? *Clin Anat* 2015; 28: 195–204.

[46] Chang U-K, Kim DH, Lee MC, et al. Changes in adjacent-level disc pressure and facet joint force after cervical arthroplasty compared with cervical discectomy and fusion. *J Neurosurg Spine* 2007; 7: 33–39.

[47] Pfirrmann CWA, Metzdorf A, Zanetti M, et al. Magnetic Resonance Classification of Lumbar Intervertebral Disc Degeneration. *Spine (Phila Pa 1976)* 2001; 26: 1873–1878.

[48] Rohlmann A, Zander T, Schmidt H, et al. Analysis of the influence of disc degeneration on the mechanical behaviour of a lumbar motion segment using the finite element method. *J Biomech* 2006; 39: 2484–2490.

[49] Panjabi MM. The stabilizing system of the spine: Part I. function, dysfunction, adaptation, and enhancement. *J Spinal Disord* 1992; 5: 383–389.

[50] Turner JA, Ersek M, Herron L, et al. Surgery for lumbar spinal stenosis. Attempted meta-analysis of the literature. *Spine (Phila Pa 1976)* 1992; 17: 1–8.

[51] Katz JN, Harris MB. Lumbar Spinal Stenosis. *N Engl J Med* 2008; 358: 818–825.

[52] Fujiwara A, Tamai K, Yamato M, et al. The relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine: an MRI study. *Eur Spine J* 1999; 8: 396–401.

[53] Natarajan RN, Andersson GBJ. Lumbar disc degeneration is an equally important risk factor as lumbar fusion for causing adjacent segment disc disease. *J Orthop Res* 2017; 35: 123–130.

[54] Hilibrand AS, Robbins M. Adjacent segment degeneration and adjacent segment disease: The consequences of spinal fusion? *Spine J* 2004; 4: 190–194.

[55] Denozière G, Ku DN. Biomechanical comparison between fusion of two vertebrae and implantation of an artificial intervertebral disc. *J Biomech* 2006; 39: 766–775.

[56] Xiao Z, Wang L, Gong H, et al. Biomechanical evaluation of three surgical scenarios of posterior lumbar interbody fusion by finite element analysis. *Biomed Eng Online* 2012; 11: 1–11.

[57] Whitecloud TS, Davis JM, Olive PM. Operative Treatment of the Degenerated Segment Adjacent to a Lumbar Fusion. *Spine* 1994; 19: 531–536.

[58] Mannion AF, Leivseth G, Brox J-I, et al. ISSLS Prize Winner: Long-Term Follow-up Suggests Spinal Fusion Is Associated With Increased Adjacent Segment Disc Degeneration But Without Influence on Clinical Outcome: Results of a Combined Follow-up From 4 Randomized Controlled Trials. *Spine (Phila Pa 1976)* 2014; 39: 1373–83.

[59] Lee CK. Accelerated degeneration of the segment adjacent to a lumbar fusion. *Spine* 1988; 13: 375–7.

[60] Chen C-S, Cheng C-K, Liu C-L, et al. Stress analysis of the disc adjacent to interbody fusion in lumbar spine. *Med Eng Phys* 2001; 23: 485–493.

[61] Chen SH, Zhong ZC, Chen CS, et al. Biomechanical comparison between lumbar disc arthroplasty and fusion. *Med Eng Phys* 2009; 31: 244–253.

[62] St. John KR. The use of polyurethane materials in the surgery of the spine: A review. *Spine J* 2014; 14: 3038–3047.

[63] Shih S-L, Chen C-S, Lin H-M, et al. Effect of spacer diameter of the Dynesys dynamic stabilization system on the biomechanics of the lumbar spine: a finite element analysis. *J Spinal Disord Tech* 2012; 25: E140–E149.

[64] Shih S-L, Liu C-L, Huang L-Y, et al. Effects of cord pretension and stiffness of the Dynesys system spacer on the biomechanics of spinal decompression- a finite element study. *BMC Musculoskelet Disord* 2013; 14: 191.

[65] Kiapour A, Ambati D, Hoy RW, et al. Effect of Graded Facetectomy on Biomechanics of Dynesys Dynamic Stabilization System. *Spine (Phila Pa 1976)* 2012; 37: E581–E589.

[66] Schmidt H, Galbusera F, Rohlmann A, et al. Effect of multilevel lumbar disc arthroplasty on spine kinematics and facet joint loads in flexion and extension: A finite element analysis. *Eur Spine J* 2012; 21: S663–S674.

[67] Tropiano P, Huang RC, Girardi FP, et al. Lumbar total disc replacement: Seven to eleven-year follow-up. *J Bone Jt Surg - Ser A* 2005; 87: 490–496.

[68] Veruva SY, Steinbeck MJ, Toth J, et al. Which Design and Biomaterial Factors Affect Clinical Wear Performance of Total Disc Replacements? A Systematic Review. *Clin Orthop Relat Res* 2014; 472: 3759–3769.

[69] Borkowski P, Marek P, Krzesiński G, et al. Finite element analysis of an elastomeric artificial disc in lumbar spine. *Acta Bioeng Biomech* 2012; 14: 59–66.

[70] Rischke B, Ross RS, Jollenbeck BA, et al. Preclinical and clinical experience with a viscoelastic total disc replacement. *SAS J* 2011; 5: 97–107.

[71] Serhan H, Mhatre D, Defossez H, et al. Motion-preserving technologies for degenerative lumbar spine: The past, present, and future horizons. *SAS J* 2011; 5: 75–89.

[72] Szpalski M, Gunzburg R, Mayer M. Spine arthroplasty: a historical review. *Eur Spine J* 2002; 11 Suppl 2: S65-84.

[73] Goel VK, Grauer JN, Patel TC, et al. Effects of charité artificial disc on the implanted and adjacent spinal segments mechanics using a hybrid testing protocol. *Spine (Phila Pa 1976)* 2005; 30: 2755–2764.

[74] de Kleuver M, Oner FC, Jacobs WCH. Total disc replacement for chronic low back pain:

background and a systematic review of the literature. *Eur Spine J* 2003; 12: 108–116.

[75] van Ooij A, Oner F, Verbout A. Complications of artificial disc replacement: a report of 27 patients with the SB Charité disc. *J Spinal Disord Tech* 2003; 16: 369–83.

[76] Kurtz SM, van Ooij A, Ross R, et al. Polyethylene wear and rim fracture in total disc arthroplasty. *Spine J* 2007; 7: 12–21.

[77] Kurtz SM, Peloza J, Siskey R, et al. Analysis of a retrieved polyethylene total disc replacement component. *Spine J* 2005; 5: 344–350.

[78] Wang S, Park WM, Gadikota HR, et al. A combined numerical and experimental technique for estimation of the forces and moments in the lumbar intervertebral disc. *Comput Methods Biomech Biomed Engin* 2013; 16: 1278–1286.

[79] Gardiner JC, Weiss JA. Subject-specific finite element analysis of the human medial collateral ligament during valgus knee loading. *J Orthop Res* 2003; 21: 1098–1106.

[80] Peña E, Martinez MA, Calvo B, et al. On the numerical treatment of initial strains in biological soft tissues. *Int J Numer Methods Eng* 2006; 68: 836–860.

[81] Studer HP, Riedwyl H, Amstutz CA, et al. Patient-specific finite-element simulation of the human cornea: A clinical validation study on cataract surgery. *J Biomech* 2013; 46: 751–758.

[82] Patwardhan AG, Havey RM, Meade KP, et al. A follower load increases the load-carrying capacity of the lumbar spine in compression. *Spine (Phila Pa 1976)* 1999; 24: 1003–1009.

[83] Ferguson SJ, Visser JMA, Polikeit A. The long-term mechanical integrity of non-reinforced PEEK-OPTIMA polymer for demanding spinal applications: Experimental and finite-element analysis. *Eur Spine J* 2006; 15: 149–156.

[84] Chen C-S, Huang C-H, Shih S-L. Biomechanical evaluation of a new pedicle screw-based posterior dynamic stabilization device (Awesome Rod System) - a finite element analysis. *BMC Musculoskelet Disord* 2015; 16: 81.

[85] Goel VK, Mehta A, Jangra J, et al. Anatomic facet replacement system (AFRS) restoration of lumbar segment mechanics to intact: A finite element study and in vitro cadaver investigation. *SAS J* 2007; 1: 46–54.

[86] Kim HJ, Kang KT, Chun HJ, et al. The influence of intrinsic disc degeneration of the adjacent segments on its stress distribution after one-level lumbar fusion. *Eur Spine J* 2015; 24: 827–837.

[87] Rundell SA, Auerbach JD, Balderston RA, et al. Total Disc Replacement Positioning Affects Facet Contact Forces and Vertebral Body Strains. *Spine (Phila Pa 1976)* 2008; 33: 2510–2517.

[88] Messac A, Mattson CA. Normal Constraint Method with Guarantee of Even Representation of Complete Pareto Frontier. *AIAA J* 2004; 42: 2101–2111.

[89] Allaire BT, Depaolis Kaluza MC, Bruno AG, et al. Evaluation of a new approach to compute intervertebral disc height measurements from lateral radiographic views of the spine. *Eur Spine J* 2016; 1–6.

[90] Van Der Houwen EB, Baron P, Veldhuizen AG, et al. Geometry of the intervertebral volume and vertebral endplates of the human spine. *Ann Biomed Eng* 2010; 38: 33–40.

[91] Wilke HJ, Neef P, Caimi M, et al. New in vivo measurements of pressures in the intervertebral disc in daily life. *Spine (Phila Pa 1976)* 1999; 24: 755–62.

[92] Peloquin JM, Yoder JH, Jacobs NT, et al. Human L3L4 intervertebral disc mean 3D shape, modes of variation, and their relationship to degeneration. *J Biomech* 2014; 47: 2452–2459.

[93] Holzapfel GA. *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*. Wiley, 2000.

[94] Quapp KM, Weiss JA. Material Characterization of Human Medial Collateral Ligament. *ASME J Biomed Eng* 1997; 120: 757–763.

[95] Weiss JA, Maker BN, Govindjee S. Finite element implementation of incompressible, transversely isotropic hyperelasticity. *Comput Methods Appl Mech Eng* 1996; 135: 107–128.

[96] Wilke HJ, Wenger K, Claes L. Testing criteria for spinal implants: Recommendations for the standardization of in vitro stability testing of spinal implants. *Eur Spine J* 1998; 7: 148–154.

[97] Goel VK, Monroe BT, Gilbertson LG, et al. Interlaminar shear stresses and laminae separation in a disc. Finite element analysis of the L3-L4 motion segment subjected to axial compressive loads. *Spine (Phila Pa 1976)* 1995; 20: 689–698.

[98] Lavaste F, Skalli W, Robin S, et al. Three-dimensional geometrical and mechanical modelling of the lumbar spine. *J Biomech* 1992; 25: 1153–1164.

[99] Niemeyer F, Wilke HJ, Schmidt H. Geometry strongly influences the response of numerical models of the lumbar spine-A probabilistic finite element analysis. *J Biomech* 2012; 45: 1414–1423.

[100] Little JP, Pearcy MJ, Pettet GJ. Parametric equations to represent the profile of the human intervertebral disc in the transverse plane. *Med Biol Eng Comput* 2007; 45: 939–945.

[101] Schmidt H, Midderhoff S, Adkins K, et al. The effect of different design concepts in lumbar total disc arthroplasty on the range of motion, facet joint forces and instantaneous center of rotation of a L4-5 segment. *Eur Spine J* 2009; 18: 1695–1705.

[102] Lotz JC, Fields AJ, Liebenberg EC. The role of the vertebral end plate in low back pain. *Glob Spine J* 2013; 3: 153–163.

[103] DeWit JA, Cronin DS. Cervical spine segment finite element model for traumatic injury prediction. *J Mech Behav Biomed Mater* 2012; 10: 138–150.

[104] Sharma M, Langrana NA, Rodriguez J. Role of ligaments and facets in lumbar spinal stability. *Spine* 1995; 20: 887–900.

[105] Marchand F, Ahmed AM. Investigation of the Laminate Structure of Lumbar Disc Anulus Fibrosus. *Spine (Phila Pa 1976)* 1990; 15: 402–410.

[106] Renner SM, Natarajan RN, Patwardhan AG, et al. Novel model to analyze the effect of a large compressive follower pre-load on range of motions in a lumbar spine. *J Biomech* 2007; 40: 1326–1332.

[107] Dreischarf M, Zander T, Bergmann G, et al. A non-optimized follower load path may cause considerable intervertebral rotations. *J Biomech* 2010; 43: 2625–2628.

[108] Rohlmann A, Zander T, Rao M, et al. Applying a follower load delivers realistic results for simulating standing. *J Biomech* 2009; 42: 1520–1526.

[109] Schmidt H, Shirazi-Adl A, Galbusera F, et al. Response analysis of the lumbar spine during regular daily activities-A finite element analysis. *J Biomech* 2010; 43: 1849–1856.

[110] Huang RC, Girardi FP, Cammisa FP, et al. The implications of constraint in lumbar total disc replacement. *J Spinal Disord Tech* 2003; 16: 412–7.

[111] García Vacas F, Ezquerro Juanco F, Pérez De La Blanca A, et al. The flexion-extension response of a novel lumbar intervertebral disc prosthesis: A finite element study. *Mech Mach Theory* 2014; 73: 273–281.

[112] Galbusera F, Bellini CM, Brayda-Bruno M, et al. Biomechanical studies on cervical total disc arthroplasty: A literature review. *Clin Biomech* 2008; 23: 1095–1104.

[113] Schmidt H, Heuer F, Claes L, et al. The relation between the instantaneous center of rotation and facet joint forces - A finite element analysis. *Clin Biomech* 2008; 23: 270–278.

[114] O’Leary P, Nicolakis M, Lorenz MA, et al. Response of Charité total disc replacement under physiologic loads: Prosthesis component motion patterns. *Spine J* 2005; 5: 590–599.

[115] Bergström JS, Rimnac CM, Kurtz SM. Prediction of multiaxial mechanical behavior for conventional and highly crosslinked UHMWPE using a hybrid constitutive model. *Biomaterials* 2003; 24: 1365–1380.

[116] Kim K-T, Lee S-H, Suk K-S, et al. Biomechanical changes of the lumbar segment after total disc replacement : charite®, prodisc® and maverick® using finite element model study. *J Korean Neurosurg Soc* 2010; 47: 446–53.

[117] Hyde PJ, Fisher J, Hall RM. Wear characteristics of an unconstrained lumbar total disc replacement under a range of in vitro test conditions. *J Biomed Mater Res - Part B Appl Biomater* 2017; 105: 46–52.

[118] Das I, Dennis JE. Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems. *SIAM J Optim* 1998; 8: 631–657.

[119] Kim IY, De Weck OL. Adaptive weighted sum method for multiobjective optimization: A new method for Pareto front generation. *Struct Multidiscip Optim* 2006; 31: 105–116.

[120] Shim CS, Lee S-H, Shin H-D, et al. CHARITE versus ProDisc: a comparative study of a minimum 3-year follow-up. *Spine (Phila Pa 1976)* 2007; 32: 1012–1018.

[121] Salzmann SN, Plais N, Shue J, et al. Lumbar disc replacement surgery—successes and obstacles to widespread adoption. *Curr Rev Musculoskelet Med* 2017; 10: 153–159.

[122] Moumene M, Geisler FH. Comparison of Biomechanical Function at Ideal and Varied Surgical Placement for Two Lumbar Artificial Disc Implant Designs. *Spine (Phila Pa 1976)* 2007; 32: 1840–1851.

[123] Shirazi-Adl A. Biomechanics of the lumbar spine in sagittal/lateral moments. *Spine (Phila Pa 1976)* 1994; 19: 2407–2414.

[124] Baldwin MA, Langenderfer JE, Rullkoetter PJ, et al. Development of subject-specific and statistical shape models of the knee using an efficient segmentation and mesh-morphing approach. *Comput Methods Programs Biomed* 2010; 97: 232–240.

[125] Bryan R, Surya Mohan P, Hopkins A, et al. Statistical modelling of the whole human femur incorporating geometric and material properties. *Med Eng Phys* 2010; 32: 57–65.

[126] Rao C, Fitzpatrick CK, Rullkoetter PJ, et al. A statistical finite element model of the knee accounting for shape and alignment variability. *Med Eng Phys* 2013; 35: 1450–1456.

[127] Taylor M, Perilli E, Martelli S. Development of a surrogate model based on patient weight, bone mass and geometry to predict femoral neck strains and fracture loads. *J Biomech* 2017; 55: 121–127.

[128] Willing R, Kim IY. A holistic numerical model to predict strain hardening and damage of UHMWPE under multiple total knee replacement kinematics and experimental validation. *J Biomech* 2009; 42: 2520–2527.

[129] Štern D, Likar B, Pernuš F, et al. Parametric modelling and segmentation of vertebral bodies in 3D CT and MR spine images. *Phys Med Biol* 2011; 56: 7505–7522.

[130] Korez R, Likar B, Pernuš F, et al. Parametric modeling of the intervertebral disc space in 3D: Application to CT images of the lumbar spine. *Comput Med Imaging Graph* 2014; 38: 596–605.

[131] Wriggers P. Discretization Methods for Solids Undergoing Finite Deformations. In: *Advanced Finite Element Technologies*. Springer International Publishing, pp. 17–51.

[132] Yang TY. *Finite element structural analysis*. Prentice-Hall, 1986.

[133] Wang Y, Gerling GJ. Automatic finite element implementation of hyperelastic material with a double numerical differentiation algorithm. *arXiv Comput Eng Financ Sci* 2016; 1–19.

[134] Sun W, Chaikof EL, Levenston ME. Numerical approximation of tangent moduli for finite

element implementations of nonlinear hyperelastic material models. *J Biomech Eng* 2008; 130: 061003.

[135] Jin Y, Sendhoff B. Pareto-based multiobjective machine learning: An overview and case studies. *IEEE Trans Syst Man Cybern Part C Appl Rev* 2008; 38: 397–415.

[136] McGonagle J, Shaikouski G, Williams C, et al. Backpropagation. *Brilliant.org*, https://brilliant.org/wiki/backpropagation/ (2018, accessed 16 November 2018).

[137] Messac A, Ismail-Yahaya A, Mattson CA. The normalized normal constraint method for generating the Pareto frontier. *Struct Multidiscip Optim* 2003; 25: 86–98.