PARALLEL FACTOR ANALYSIS OF GAIT WAVEFORM DATA: A MULTIMODE EXTENSION OF PRINCIPAL COMPONENT ANALYSIS

BY

NATHANIEL E. HELWIG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Master's Committee:

Assistant Professor Sungjin Hong
Professor Lawrence J. Hubert

Abstract

Gait data are typically collected in multivariate form, so some multivariate analysis is often used to understand interrelationships between observed data. Principal Component Analysis (PCA), a data reduction technique for correlated multivariate data, has been widely applied by gait analysts to investigate patterns of association in gait waveform data (e.g., interrelationships between joint angle waveforms from different subjects and/or joints). Despite its widespread use in gait analysis, PCA is for two-mode data, whereas gait data are often collected in higher-mode form. In this paper, we present the benefits of analyzing gait data via Parallel Factor Analysis (Parafac), which is a component analysis model designed for three- or higher-mode data. Using three-mode joint angle waveform data ($subjects \times time \times joints$), we demonstrate Parafac's ability to (a) determine interpretable components revealing the primary interrelationships between lower-limb joints in healthy gait and (b) identify interpretable components revealing the fundamental differences between normal and perturbed subjects' gait patterns across multiple joints. Our results offer evidence of the complex interconnections that exist between lower-limb joints and limb segments in both normal and abnormal gaits, confirming the need for the simultaneous analysis of multi-joint gait waveform data (especially when studying perturbed gait patterns).

To my loving and supportive family.

Acknowledgments

Table of Contents

List of Abbreviations

ALS   Alternating Least Squares.

GC   Gait Cycle.

KB   Knee-Braced.

NB   Non-Braced.

Parafac   Parallel Factor Analysis.

PCA   Principal Component Analysis.

PC   Principal Component.

List of Symbols

⊙   Kahtri-Rao Product.

∘   Tensor Product.

Chapter 1

Introduction

[1] Gait data are almost always collected in multivariate form, so some multivariate analysis is often employed to quantify and understand patterns of association between observed variables. Principal Component Analysis (PCA; Hotelling, 1933; Pearson, 1901) is a multivariate technique that linearly transforms a set of correlated variables into a new set of orthogonal variables known as principal components (PCs), permitting dimensionality reduction and potentially improved interpretation of data. PCA was first applied to gait data by Shiavi and Griffin (1981) to examine muscle activation patterns in healthy gait and has since been applied to accomplish several objectives in gait analysis.

One typical application of PCA to gait data is to examine interrelationships between joints, muscles, and/or body segments in healthy gait (e.g., MacLellan & McFadyen, 2010; Palta, 1985; Sadeghi et al., 1997; Wootten et al., 1990). In these applications, PCA is often applied to gait waveform data in one of two manners: (a) conduct one PCA for each subject using a *joints/muscles/segments × time* matrix (e.g., Bennett et al., 2010; Herr & Popovic, 2008; Mah et al., 1994; Raptopoulos et al., 2006); or (b) conduct one PCA for each joint/muscle/segment using a *subjects × time* matrix (e.g., Ivanenko et al., 2004; Laassel et al., 1992; Loslever et al., 1994; Sadeghi et al., 2001). In addition, some researchers (Troje, 2002) have used a two-stage PCA approach to model both within- and between-subject gait pattern variation. In all cases, the PCs from the separate analyses are examined to study mechanisms underlying healthy gait.

Another typical application of PCA in gait analysis is to investigate and summarize group differences in gait waveform data (e.g., Deluzio et al., 1997; Muniz & Nadal, 2009; Reid et al., 2010; Yamamoto et al., 1983). In these applications, one PCA is often conducted for each joint/muscle/segment using a *subjects × time* matrix, where the *subjects* mode is a collection of subjects from different groups (e.g., Astephen et al., 2008; Chester & Wrigley, 2008; Cochran et al., 1984; Donoghue et al., 2008; Lamoth et al., 2006; Lauer et al., 2005; Lee et al., 2009; McKean et al., 2007). The PCs from these separate analyses are then examined

---

[1]This paper is currently in press at *Human Movement Science* (doi:10.1016/j.humov.2011.06.011).

to understand gait patterns of different subject groups, and PC scores (see Section 2.1.1) are often used to quantify between-group differences.

Despite its widespread use in gait analysis, PCA is for the analysis of two-mode data (e.g., *subjects × time*), whereas gait data are typically collected in three- or higher-mode form (e.g., *subjects × time × joints*). Therefore, gait analysts applying PCA often (a) perform several different PCAs (as previously described) or (b) reorganize three-mode data into some two-mode form to apply one PCA (e.g., Hubley-Kozey et al., 2008; Sadeghi, 2003; Shemmell et al., 2007). The first approach produces many solutions that are relevant only to particular joints/muscles/segments (or subjects), lacking a solution that describes the data as a whole. The second approach yields a solution that only reveals a part of the data structure, lacking a full description of the data variation in all modes.

To overcome these limitations of PCA for gait analysis, Davis et al. (2002, 2003a,b, 2004) used Tucker's (1966) three-mode factor analysis model (known as the Tucker3 model) to analyze variation in three-mode gait data. Davis and colleagues applied the Tucker3 model to gait data of the form *body pose × time × effort/style*, and showed that it is possible to recognize different gait efforts (e.g., carrying a light vs. heavy load) and styles (e.g., male vs. female) via the *effort/style* weights obtained from the model. Davis et al.'s Tucker3 approach for analyzing gait waveforms offers an improvement over traditional two-mode PCA approaches, given that this Tucker3 method directly models the three-mode variation of the data. However, like the two-mode PCA model, the Tucker3 model suffers from rotational indeterminacy (Kiers, 1998; Tucker, 1966), and, even after an optimal rotation, Tucker3 solutions can be difficult to interpret due to the model's complex structure (see Harshman & Lundy, 1984a, for a discussion).

In this paper, we present an alternative approach, Parallel Factor Analysis (Parafac; Harshman, 1970), to the visualization and interpretation of gait waveform data. Parafac is a multimode extension of PCA that is designed for the analysis of three- or higher-mode data. We illustrate how Parafac can be used to analyze three-mode gait data (*subjects × time × joints*) via one parsimonious model. This approach makes it possible to (a) uniquely identify components describing individual and/or group differences in motion patterns, (b) determine the time points when the components are influential (including the directions of their effects), and (c) examine symmetry and interrelationships between joints along each component. Using experimental data with and without a simulated injury, we demonstrate Parafac's ability to describe meaningful variation within normal gait, as well as Parafac's power to identify differences between normal

and perturbed gait patterns. Our analyses reveal the dynamic interconnections that exist between lower-limb joints and segments in both healthy and perturbed gaits, confirming the need for the simultaneous analysis of multi-joint gait waveform data.

Chapter 2

Methods

2.1 Component Models

2.1.1 Principal Component Analysis

A data matrix $\mathbf{X} \equiv \{x_{ij}\}_{I \times J}$ (e.g., I subjects $\times$ J time points) can be approximated by PCA as

$$\mathbf{X} = \mathbf{AB}' + \mathbf{E}, \tag{2.1}$$

where $\mathbf{A} \equiv \{a_{ir}\}_{I \times R}$ contains mutually orthogonal columns with successively maximum sums-of-squares, and $\mathbf{B} \equiv \{b_{jr}\}_{J \times R}$ is columnwise orthonormal. Columns of $\mathbf{A}$ and $\mathbf{B}$ (sometimes called PC scores and loadings, respectively) indicate the relative weights of the rows and columns of $\mathbf{X}$ for the corresponding R PCs (where $R \leq \min(I, J)$), and $\mathbf{E} \equiv \{e_{ij}\}_{I \times J}$ is the information in $\mathbf{X}$ that cannot be explained by the R PCs. If columns of $\mathbf{X}$ are mean-centered,[1] then $I^{-1}\mathbf{A}'\mathbf{A}$ becomes a diagonal matrix with entries being the R largest eigenvalues (in descending order) of the covariance matrix of $\mathbf{X}$ ($\equiv I^{-1}\mathbf{X}'\mathbf{X}$), and columns of $\mathbf{B}$ are the corresponding eigenvectors. Thus, the PCA representation $\mathbf{AB}'$ gives the least-squares (i.e., minimizes $\sum_i \sum_j e_{ij}^2$) R-dimensional approximation of $\mathbf{X}$ (Hotelling, 1933; Pearson, 1901). The PCA model can be alternatively written for an arbitrary entry x_{ij} as

$$x_{ij} = \sum_{r=1}^{R} a_{ir} b_{jr} + e_{ij}, \tag{2.2}$$

where a_{ir} and b_{jr} indicate weights of the r-th PC for subject i and time point j, respectively, and e_{ij} denotes the entry's unexplained information.

---

[1]The bilinear form of the PCA model requires ratio scale data, so tradition (from the social sciences) dictates that variables are mean-centered before conducting PCA. However, gait data are typically ratio scale by nature, so it is not necessary to mean-center such data before applying PCA (or Parafac). In fact, mean-centering data that are already ratio scale may alter the data's bi-/trilinear structure, making the data no longer appropriate for the PCA/Parafac model. Thus, in this paper, we apply Parafac to un-centered gait data. See Harshman and Lundy (1984b) and Smilde et al. (2004) for a thorough discussion of how different forms of pre-processing (such as centering) influence PCA and Parafac solutions.

It is well-known that, without the conditions of orthogonality and variance maximization, the PCA model is not unique given that $\mathbf{A}$ and $\mathbf{B}$ can be rotated by any nonsingular orthogonal matrix $\mathbf{T}$ without loss in fit (i.e., $\mathbf{X} - \mathbf{E} = \mathbf{A}\mathbf{B}' = \mathbf{A}\mathbf{T}\mathbf{T}'\mathbf{B}' = \widetilde{\mathbf{A}}\widetilde{\mathbf{B}}'$), which is referred to as the model's *rotational indeterminacy*. This means that, given an R-dimensional PCA solution, there are infinite different orientations of the R dimensions that equally account for the data, and many of these orientations will likely lead to substantively different interpretations. Consequently, some gait analysts (e.g., Davis & Vaughan, 1993; Mandeville et al., 2009; Merkle et al., 1998; Sadeghi et al., 2000) have applied Kaiser's (1958) varimax rotation for a simpler interpretation.

2.1.2 Parallel Factor Analysis

The Parafac model can be considered a straightforward extension of the PCA model in Equation (2.1). Suppose we have a three-mode data array $\underline{\mathbf{X}} \equiv \{x_{ijk}\}_{I \times J \times K}$ (e.g., the $I \times J$ matrix $\mathbf{X}$ measured for K joints in the body). Parafac approximates such data array as
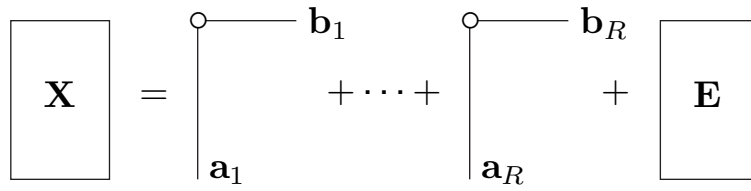
$$\mathbf{X}_k = \mathbf{A}\mathbf{C}_k\mathbf{B}' + \mathbf{E}_k \qquad \text{for} \quad k = 1, \ldots, K, \tag{2.3}$$

where $\mathbf{X}_k \equiv \{x_{ij(k)}\}_{I \times J}$ indicates the k-th frontal slice of $\underline{\mathbf{X}}$, $\mathbf{C}_k$ is a diagonal matrix containing the k-th row of $\mathbf{C} \equiv \{c_{kr}\}_{K \times R}$ (the weights of the R components for the K joints), and $\mathbf{E}_k \equiv \{e_{ij(k)}\}_{I \times J}$ is information in $\mathbf{X}_k$ that cannot be explained by the R components. Like Equation (2.2), the Parafac model can be written for an arbitrary entry x_{ijk} as

$$x_{ijk} = \sum_{r=1}^{R} a_{ir}b_{jr}c_{kr} + e_{ijk}, \tag{2.4}$$

where a_{ir}, b_{jr}, and c_{kr} indicate weights of the r-th component for subject i, time point j, and joint k, respectively, and e_{ijk} denotes the entry's unexplained information. Connections between PCA and Parafac can be further visualized via tensor product notation: PCA approximates $\mathbf{X}$ as a sum of R rank-1 second-order tensors (i.e., outer-products), and Parafac approximates $\underline{\mathbf{X}}$ as a sum of R rank-1 third-order tensors (Fig. 2.1). See Kiers (2000) and Kolda and Bader (2009) for other notational conventions for Parafac.

Although Parafac is a straightforward extension of PCA, the multilinear structure of the Parafac model (e.g., the trilinear structure in Equations (2.3) and (2.4)) postulates stronger conditions of the data, which result in important theoretical and practical differences between Parafac and PCA. For example, given the $I \times J \times K$ data array $\underline{\mathbf{X}}$ described above, one possibility to explaining the data's three-mode variation via

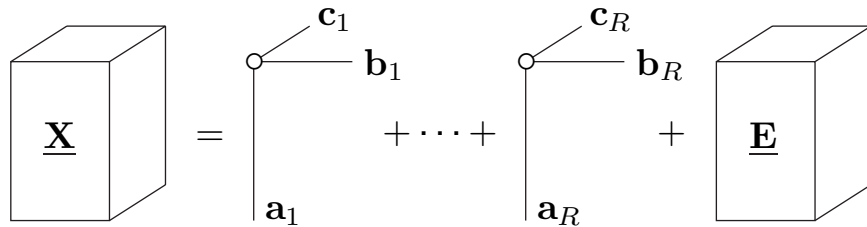## a) PCA: $\mathbf{X} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r + \mathbf{E}$

b) Parafac: $\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r + \underline{\mathbf{E}}$

Figure 2.1: The PCA (a) and Parafac (b) models using tensor product notation. The vectors $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$ denote the r-th columns $(r = 1, \ldots, R)$ of the weight matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively, and the symbol $\circ$ denotes the tensor product of the vectors.

the two-mode PCA model is to fit K different PCA models (one PCA model for each joint's $I \times J$ matrix).[2] This approach produces K different sets of PCA weights $(\mathbf{A}_k, \mathbf{B}_k$ for $k = 1, \ldots, K)$, and there is no clear method for comparing these weight matrices across the K different joints. Also, as discussed in Section 2.1.1, each of these K sets of PCA weights suffers from rotational indeterminacy, and (aside from trial-and-error) there is no way to determine which orientation(s) will lead to the most substantively meaningful solutions.

In contrast, the trilinear structure of the three-mode Parafac model assumes that the K different joints share the same subjects and time weights ($\mathbf{A}$ and $\mathbf{B}$, respectively), which are weighted by each joint's diagonal weight matrix ($\mathbf{C}_k$ for $k = 1, \ldots, K$) to give the least-squares approximation of each joint's data matrix ($\mathbf{X}_k$ for $k = 1, \ldots, K$). This idea of trilinear variation originates from Cattell's (1944) *principle of parallel proportional profiles*, which was an attempt to resolve the rotational indeterminacy problem inherent in two-mode component/factor analysis models. Cattell's principle states that "real" components should have an invariant pattern/structure across different conditions and should vary proportionally depending on the relative importance of each component in explaining the data variation in each condition. Cat-

---

[2]We assume that matricizing the three-mode array (e.g., into an $IK \times J$ matrix) obscures the three-mode variation underlying the data, so we do not discuss this possibility.

tell then suggested that conditions with salient differences should be selected so that the true component pattern/structure could be identified and the proportional variation of the components could be studied. Expanding upon this idea, Harshman (1970) formalized Cattell's principle into a functional three-mode component analysis model, accordingly named *Parallel Factor Analysis* (Parafac; see Equation (2.3)).

The most important consequence of Parafac's multilinear structure is that the Parafac model is unique (up to rescaling/counter-scaling and reordering of components), which is referred to as the model's *intrinsic axis property* (Harshman & Lundy, 1984a). For the three-mode Parafac model, this uniqueness property can be understood through Cattell's principle of parallel proportional profiles. The trilinear structure of the three-mode Parafac model requires that (a) the subjects ($\mathbf{A}$) and time ($\mathbf{B}$) weights remain invariant across the K joints and (b) the importance of components varies proportionally across the K joints via multiplication by each joint's diagonal weight matrix ($\mathbf{C}_k$ for $k = 1, \ldots, K$). Thus, in the three-mode Parafac model, the pattern/structure of the $\mathbf{A}$ and $\mathbf{B}$ weights represents the unique orientation common to the K joints, such that the $\mathbf{C}$ weights enter the model multiplicatively to produce the least-squares approximation of $\underline{\mathbf{X}}$. See Harshman (1970, 1972) for the uniqueness proof specific to the Parafac model, and see Harshman and Lundy (1996) for a generalized uniqueness proof for a family of models that includes Parafac. Finally, Kruskal (1977) offers a sufficient condition for uniqueness of the three-mode Parafac model, which is typically met by most data arrays that are not unusually small (e.g., larger than $2 \times 2 \times 2$).

In practice, Parafac's intrinsic axis property implies that, given an R-dimensional Parafac solution, there is only one orientation (i.e., rotation) of the R dimensions that best accounts for the data. Thus, unlike PCA and Tucker3 solutions, the interpretation of a Parafac solution does not depend on the post-hoc user-specified rotation of the components. Also, because of Parafac's intrinsic axis property, Parafac components are not required to be orthogonal to one another. However, constraining one or more of the Parafac weight matrices to be columnwise orthogonal (*orthogonality constraint*) produces an orthogonal solution where the variance-accounted-for (VAF) by each component is not shared with any other component. In some cases, such constraints can be useful for obtaining more interpretable solutions and/or for preventing degenerate solutions. See Harshman and Lundy (1984a,b) for a detailed overview of both theoretical and practical aspects of the Parafac model.

2.1.3 Fitting the PCA and Parafac Models

Just as the PCA solution is sought to provide the least-squares R-dimensional approximation of $\mathbf{X}$ (i.e., the solution minimizing $\sum_i \sum_j e_{ij}^2$ from Equation (2.2)), the Parafac solution is sought to provide the least-squares R-dimensional approximation of $\underline{\mathbf{X}}$ (i.e., the solution minimizing $\sum_i \sum_j \sum_k e_{ijk}^2$ from Equation (2.4)). The least-squares PCA solution is easily obtainable by the singular value decomposition, which is available in many computational environments (e.g., MATLAB or R). However, unlike PCA, there is no closed-form solution to the Parafac decomposition. Instead, the minimization is typically achieved via the alternating least-squares (ALS) algorithm (Wold, 1966), which uses ordinary least-squares regression to estimate each mode's weight matrix while holding the other modes' weight matrices constant (with the process iterating until the relative change in the model's fit is less than some predetermined tolerance).

The ALS Parafac algorithm for a three-mode array is given in Appendix A. This algorithm is known to converge and is known to be accurate under various simulated conditions. However, the algorithm's convergence can be slow for difficult problems (e.g., problems with dominant noise variance, systematic interfering variance, etc.), so the algorithm is set to stop if convergence has not been met after a sufficiently large number of iterations (e.g., 5000 iterations). To improve efficiency, various numerical techniques have been tested, such as ridge regression (Rayens & Mitchell, 1997), line search (Bro, 1998; Tomasi, 2006; Rajih et al., 2008), and gradient-based methods (Franc, 1992; Paatero, 1999; Tomasi, 2006). Although performance of an algorithm depends on the data and the fit model, the ALS algorithm in Appendix A seems to provide the most accurate Parafac solution at the cost of slow convergence (see Faber et al., 2003; Tomasi & Bro, 2006). Andersson and Bro (2000) provide an excellent open-source MATLAB toolbox with techniques for fitting various multimode component models, including the ALS algorithm for fitting the Parafac model.

2.2 Data

To demonstrate the utility of Parafac for the analysis of gait data, we utilized kinematic data from Shorter et al. (2008). Ten healthy male subjects (21 ± 2 yrs) walked on a treadmill under two conditions: (a) normal, non-braced (NB) walking; and (b) knee-braced (KB) walking such that right knee motion was restricted to full extension by a brace. The KB condition was designed to simulate the reduced flexion that would result from a perturbation such as a right knee sprain. By applying Parafac to these data, we hoped to reveal the complex interrelationships that exist between lower-limb joints during normal gait, and also understand how the reduced flexion caused by the right knee bracing altered these interrelationships. See Shorter et al. for

details of the experimental setup and data collection procedure.

Joint flexion/extension angles were anatomically defined according to the triaxial system discussed in Chao (1980), such that positive angles represented flexion (hip and knee) or dorsiflexion (ankle) and negative angles represented extension (hip and knee) or plantar flexion (ankle; see Fig. 2.2). Bilateral hip, knee, and ankle flexion/extension angle waveforms were calculated from the motion-capture data using the approach
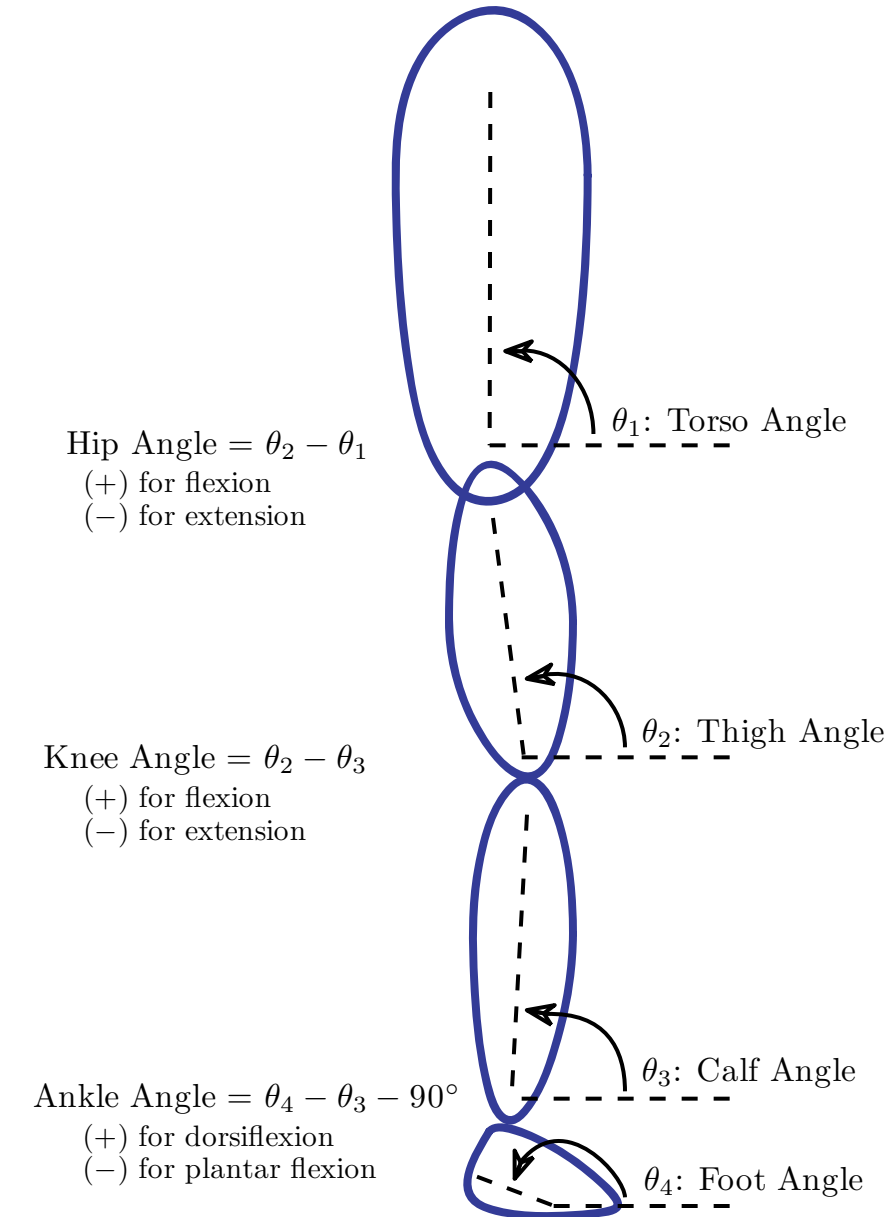

Figure 2.2: Anatomical segment and joint angle definitions. The angles are defined such that a positive joint angle represents flexion (hip and knee) or dorsiflexion (ankle) and a negative joint angle represents extension (hip and knee) or plantar flexion (ankle).

in Vaughan et al. (1999). Ten consecutive gait cycles (GCs) per condition were used per subject. Before applying Parafac, data were (a) segmented into GCs from heel strike to heel strike separately for each limb, (b) linearly length normalized to 101 time points representing 0–100% GC (see Helwig et al., 2011), and (c) averaged over the ten GC replications within each subject-condition combination to determine each subject's typical behavior in each condition (Fig. 2.3).
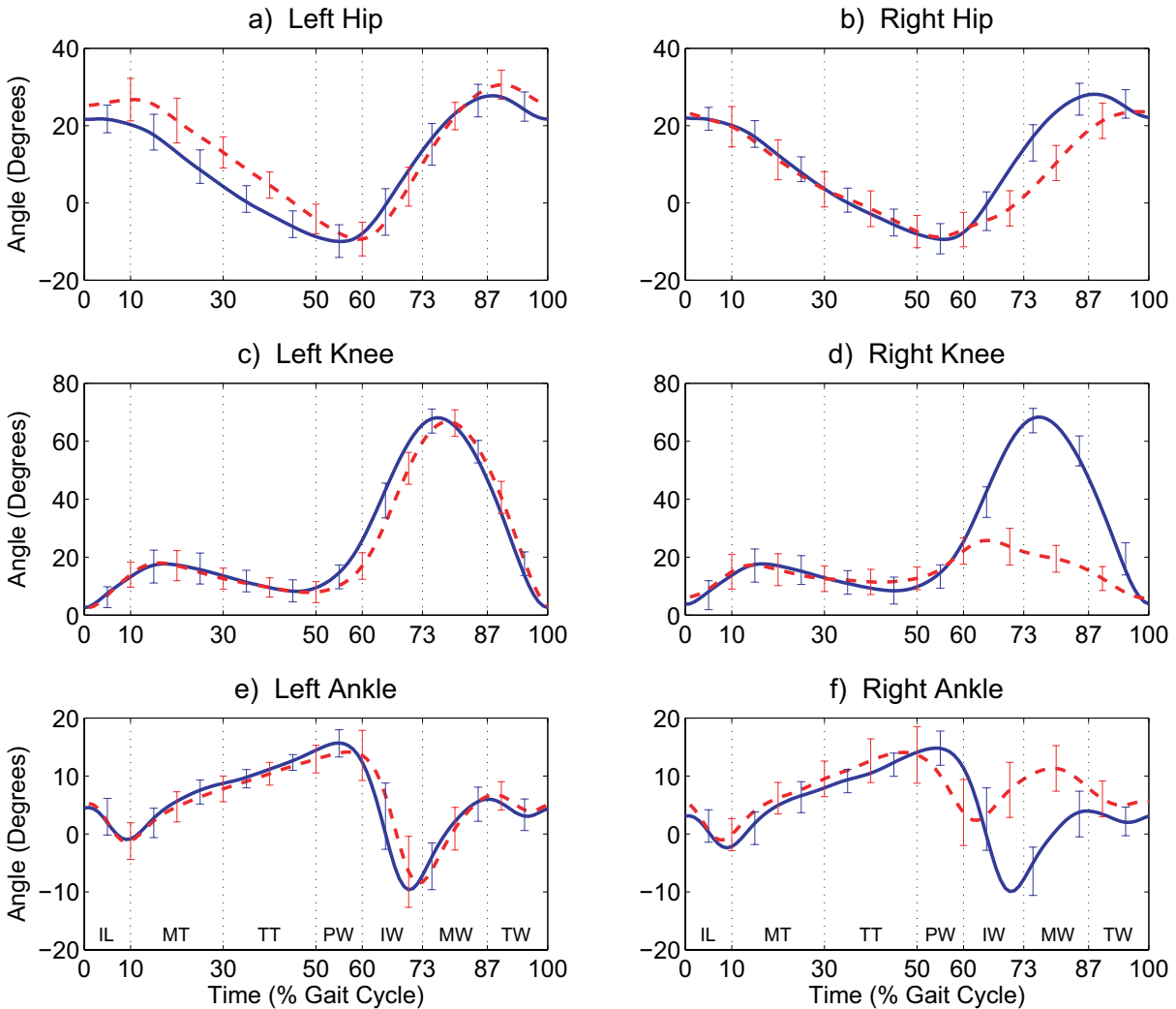
Figure 2.3: Average (with standard deviation bars) hip, knee, and ankle angle waveforms in the non-braced (blue solid lines) and knee-braced (red dashed lines) conditions. Averages (and standard deviations) are over 100 gait cycles (10 replications from 10 subjects). In each subplot, the gait cycle is divided into biomechanically-relevant subphases that are labeled in subplots (e) and (f): initial contact and loading response (IL), mid-stance (MT), terminal stance (TT), pre-swing (PW), initial swing (IW), mid-swing (MW), and terminal swing (TW).

2.3 Parallel Factor Analysis of Joint Angle Data

To assess interrelationships between joints in healthy gait, NB data were collected to form a $10 \times 101 \times 6$ array of joint angle waveforms where Mode A represented the 10 subjects, Mode B represented the 101 time points, and Mode C represented the 6 joints (left and right hip, knee, and ankle). Parafac was then used to decompose this array into weight matrices of subjects ($\mathbf{A}$), time ($\mathbf{B}$), and joints ($\mathbf{C}$) to better understand joint angle patterns in healthy gait. Also, to assess differences in joint angle patterns between the NB and KB conditions, data were collected to form a $20 \times 101 \times 6$ array of joint angle waveforms where Mode A contained a collection of the 10 NB and 10 KB subjects, and Modes B and C were the same as before. Parafac was then used to decompose this array into weight matrices of subjects ($\mathbf{A}$), time ($\mathbf{B}$), and joints ($\mathbf{C}$) to identify components that accounted for group (NB vs. KB) differences.

Like in PCA, in Parafac the number of components to extract is user-provided, so different numbers of components were tried to determine the optimal and most interpretable solution. Unlike PCA, the Parafac solution must be iteratively estimated, typically by the ALS algorithm (see Section 2.1.3). ALS Parafac solutions only guarantee local optimality, so we used 100 randomly initialized starts to determine the best solution. Also, an orthogonality constraint was imposed on $\mathbf{C}$ to prevent degenerate solutions. The orthogonality constraint was implemented by replacing Step 3 of the ALS algorithm in Appendix A with

$$3. \ \ \mathbf{C} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1/2} \ \ \text{where} \ \ \mathbf{Z} = \mathbf{X}_c(\mathbf{B} \odot \mathbf{A})$$

(Harshman & Lundy, 1994). We chose to appy the constraint on $\mathbf{C}$ (instead of $\mathbf{A}$ or $\mathbf{B}$) to ensure that each component provided non-redundant information concerning interrelationships between the joints, while allowing the components to be linearly related with respect to the subjects and time points.

Chapter 3

Results

3.1 Parafac Results for Non-Braced Data

The Parafac results for the NB data produced an interpretable three-component solution with a VAF of 98% (i.e., $(s - s_1)/s = .98$ at convergence of the ALS algorithm, where $s = \text{trace}(\mathbf{X}_a \mathbf{X}'_a)$ is the total sum-of-squares of the data, and $\mathbf{X}_a$ and s_1 are defined in Appendix A). To facilitate interpretation of the resulting weight matrices, the Parafac solution was scaled such that the subjects (Mode A) weights absorbed the scale of the data and the time (Mode B) weights absorbed the sign of the data (see Appendix B for details). We chose to absorb the data's scale into the subjects weights to facilitate the discrimination of locomotive patterns via the subjects weights (as discussed in Section 4.3). We absorbed the sign of the data into the time weights because the time weights have a meaningful bipolar interpretation (i.e., positive=flexion and negative=extension). For the remainder of the manuscript, the double-dot notation used in Appendix B will be dropped, and the scaled Parafac solution $\ddot{\mathbf{A}}$, $\ddot{\mathbf{B}}$, $\ddot{\mathbf{C}}$ will simply be referred to using the notation $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$.

After scaling, the resulting Parafac solution was interpreted in the following manner. The $\mathbf{A}$ weights illustrated inter-individual variation along each component, as well as the relative importance of each component; however, for the NB data, the $\mathbf{A}$ weights were not informative for interpreting the obtained Parafac components (Fig. 3.1). In contrast, the $\mathbf{B}$ and $\mathbf{C}$ weights were useful for characterizing and understanding the components. The $\mathbf{B}$ weights provided the pattern of each component's angular waveform over the course of the GC (Fig. 3.2), while the $\mathbf{C}$ weights indicated which components best accounted for each joint's behavior (Table 3.1). The first component (53% VAF) weighted highly for the hips and knees (Table 3.1, Comp1) and resembled the pattern of a typical thigh angle waveform (Fig. 3.2a). The second component (31% VAF) was characterized by large weights for the ankles and knees (Table 3.1, Comp2) and resembled the pattern of an inverted calf angle waveform (Fig. 3.2b). Lastly, the third component (14% VAF) produced moderate-to-large weights for all three joints (Table 3.1, Comp3) and resembled the pattern of a typical foot angle waveform (Fig. 3.2c).
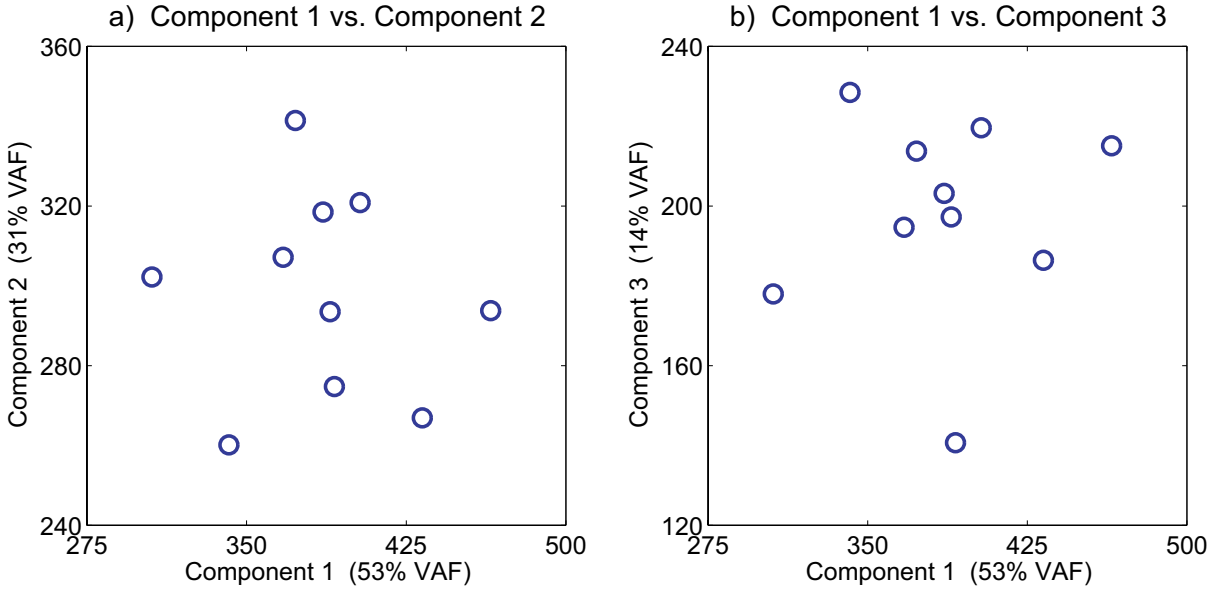
Figure 3.1: Subjects (**A**) weights from the Parafac fitting of the non-braced data. These weights illustrate inter-individual variation along each of the three components. Larger absolute values indicate that a component is more influential for explaining a given subject's behavior. Also, these weights carry the scale of the data, such that the sum (over subjects) of the squared values indicates the relative influence of each component.

Table 3.1: Joints (**C**) weights from the Parafac fitting of the non-braced data. The rows give the weights of the left and right (L, R) hip, knee, and ankle (H, K, A) on the first, second, and third components (Comp1, Comp2, Comp3). Larger absolute values indicate that a component is more influential for explaining a given joint's behavior.

Joint	Comp1	Comp2	Comp3
LH	0.60	−0.20	0.32
RH	0.60	−0.19	0.32
LK	0.35	0.48	−0.39
RK	0.37	0.47	−0.37
LA	−0.09	0.50	0.52
RA	−0.12	0.48	0.48
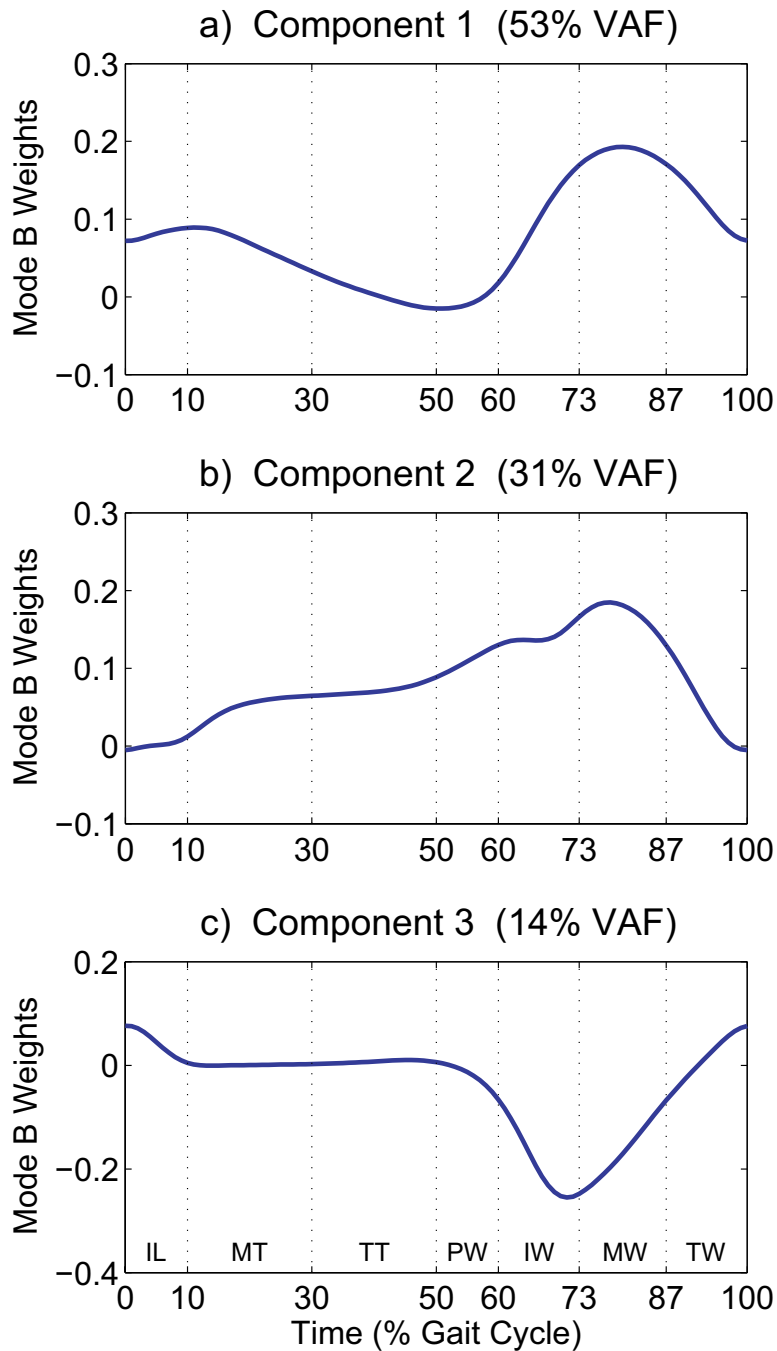
Figure 3.2: Time (**B**) weights from the Parafac fitting of the non-braced data. These weights illustrate the pattern of each component's angular waveform over the course of the gait cycle. Larger absolute values indicate that a component is more influential for explaining variation at a particular time point. For our purposes, a time point with a weight greater than 0.1 in absolute value was (somewhat arbitrarily) considered influential relative to the other time points. See Fig. 2.3 for definitions of the gait cycle subphases labeled in subplot (c).

3.2 Parafac Results for Non-Braced and Knee-Braced Data

The Parafac results for the combined NB and KB data produced an interpretable four-component solution with a VAF of 97%. Like in the previous analysis, the Parafac solution was scaled such that the subjects weights absorbed the data's scale and the time weights absorbed the data's sign. In this analysis, the **A** (Fig. 3.3; Table 3.2), **B** (Fig. 3.4), and **C** (Table 3.3) weights were used to understand the extracted components and to assess individual and group (NB vs. KB) differences along each component.
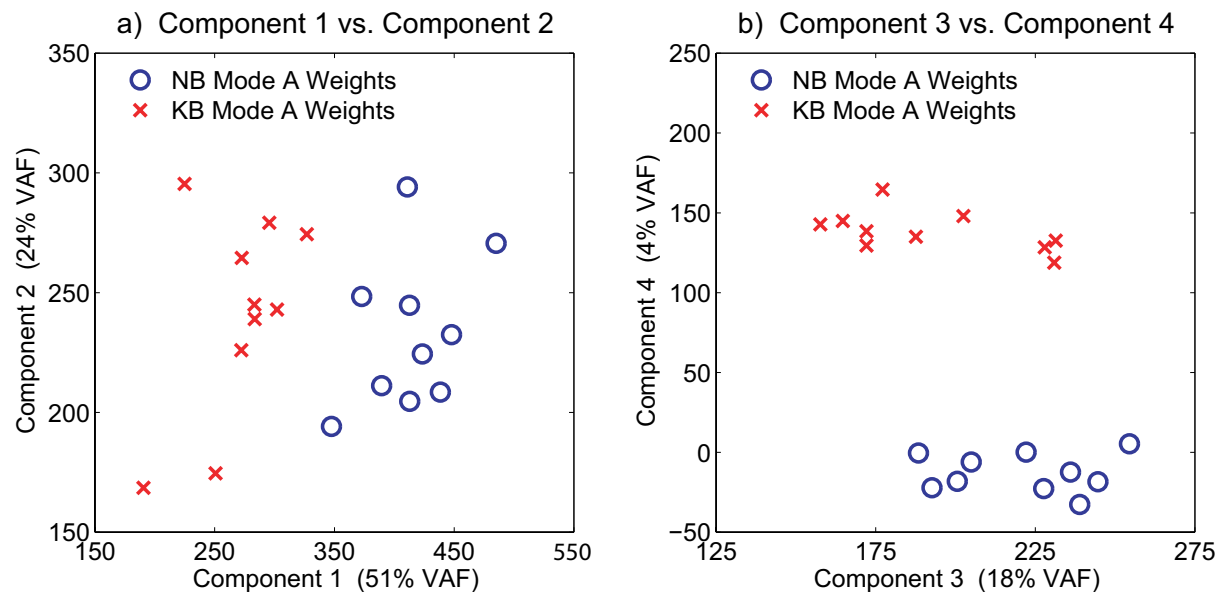
Figure 3.3: Subjects (**A**) weights from the Parafac fitting of the non-braced (NB) and knee-braced (KB) data. These weights illustrate inter-individual and between-group (NB vs. KB) variation along each of the four components. Larger absolute values indicate that a component is more influential for explaining a given subject's (or group's) behavior. Also, these weights carry the scale of the data, such that the sum (over subjects) of the squared values indicates the relative influence of each component. Note that complete separation between the NB and KB data is achieved along the first and fourth components.

Table 3.2: Means (and standard deviations) of the subjects (**A**) weights from the Parafac fitting of the non-braced (NB) and knee-braced (KB) data.

Group	Comp1	Comp2	Comp3	Comp4
NB	414.12 (38.90)	233.30 (31.64)	221.03 (23.14)	-12.82 (12.23)
KB	270.18 (39.45)	240.94 (42.18)	192.41 (28.72)	138.31 (12.68)
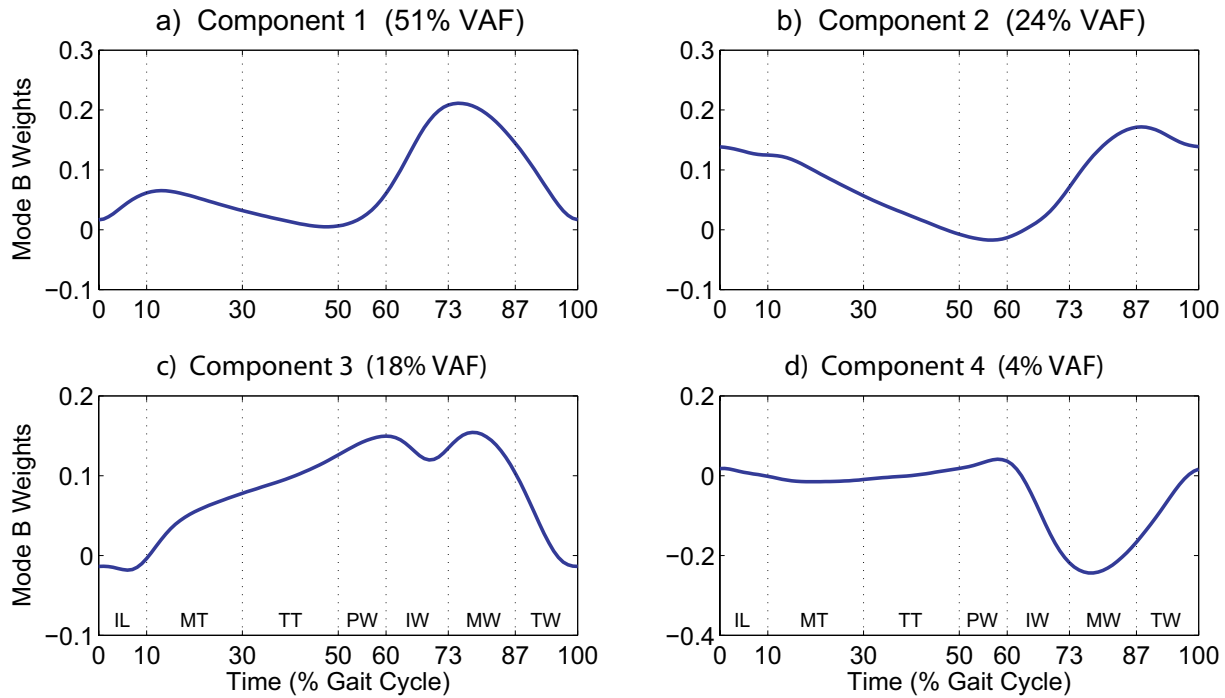
Figure 3.4: Time (**B**) weights from the Parafac fitting of the non-braced and knee-braced data. See Fig. 3.2 for a description of how these weights are interpreted, and see Fig. 2.3 for definitions of the gait cycle subphases labeled in subplots (c) and (d).

Table 3.3: Joints (**C**) weights from the Parafac fitting of the non-braced and knee-braced data. The rows give the weights of the left and right (L, R) hip, knee, and ankle (H, K, A) on the first, second, third, and fourth components (Comp1, Comp2, Comp3, Comp4). Larger absolute values indicate that a component is more influential for explaining a given joint's behavior.

Joint	Comp1	Comp2	Comp3	Comp4
LH	0.05	0.73	−0.18	−0.14
RH	0.13	0.59	−0.28	0.17
LK	0.64	0.04	0.33	−0.65
RK	0.59	0.04	0.35	0.68
LA	−0.33	0.25	0.59	0.18
RA	−0.32	0.23	0.56	−0.19

The first component (51% VAF) was characterized by large weights for the NB subjects (Fig. 3.3a; Table 3.2, Comp1) and knees (Table 3.3, Comp1) and resembled the pattern of a typical knee angle waveform (Fig. 3.4a). In contrast, the second component (24% VAF) weighted highly for the hips (Table 3.3, Comp2) and resembled the pattern of a typical hip angle waveform (Fig. 3.4b) but did not differentiate between the subject groups (Fig. 3.3a; Table 3.2, Comp2). The third component (18% VAF), which produced moderate-to-large weights for the ankles and knees (Table 3.3, Comp3) and resembled the pattern of a slightly abnormal inverted calf angle waveform (Fig. 3.4c), tended to weight more highly for the NB subjects (Fig. 3.3b; Table 3.2, Comp3). Finally, the fourth component (4% VAF) weighted highly for the KB subjects (Fig. 3.3b; Table 3.2, Comp4) and the knees in opposite directions (Table 3.3, Comp4), characterizing the differing effects of the right knee bracing on the knee angle waveforms (Fig. 3.4d).

Chapter 4

Discussion

4.1   Components in Non-Braced Walking

Almost all (about 98%) of the variation in the NB hip, knee, and ankle angle waveforms was explained by three Parafac components that seemed to represent the angular behaviors of the thigh, calf, and foot, respectively. To assess these interpretations, we attempted to reconstruct typical knee and ankle angle waveforms via linear combinations of the columns of $\mathbf{B}$ (Fig. 3.2). By definition, a knee angle waveform was formed by subtracting a calf angle waveform from a thigh angle waveform, and an ankle angle waveform was formed by subtracting a calf angle waveform (and a constant) from a foot angle waveform (see Fig. 2.2). Therefore, if the first and second Parafac components represented a thigh angle waveform and an inverted calf angle waveform (as hypothesized), then the addition of the first and second columns of $\mathbf{B}$ should produce a waveform resembling the shape of a typical knee angle waveform. Likewise, if the second and third Parafac components represented an inverted calf angle waveform and a foot angle waveform (as hypothesized), then the addition of the second and third columns of $\mathbf{B}$ should produce a waveform resembling the shape of a typical ankle angle waveform. The reconstructed waveforms formed via linear combinations of the columns of $\mathbf{B}$ closely resembled the shape patterns of the average NB waveforms (Fig. 4.1), supporting our interpretations of the components. Thus, during normal walking, variation in the system of lower-limb joint angle waveforms could be attributed to variation in (a) the thigh angular pattern (about 53% VAF), (b) the calf angular pattern (about 31% VAF), and (c) the foot angular pattern (about 14% VAF).

After identifying the components, the $\mathbf{B}$ (Fig. 3.2) and $\mathbf{C}$ (Table 3.1) weights were further examined to assess interrelationships between lower-limb joints and segments in normal gait. The first component (thigh) mainly explained the similar flexion patterns of the hip and knee from initial swing to terminal swing,[1] indicated by large $\mathbf{B}$ weights at 67–95% GC (Fig. 3.2a) and large positive $\mathbf{C}$ weights for the hips and knees (Table 3.1, Comp1). Likewise, the second component (calf) mostly described similar flexion patterns

---

[1]See Fig. 2.3 for definitions of the biomechanical subphases of the GC that are discussed here.
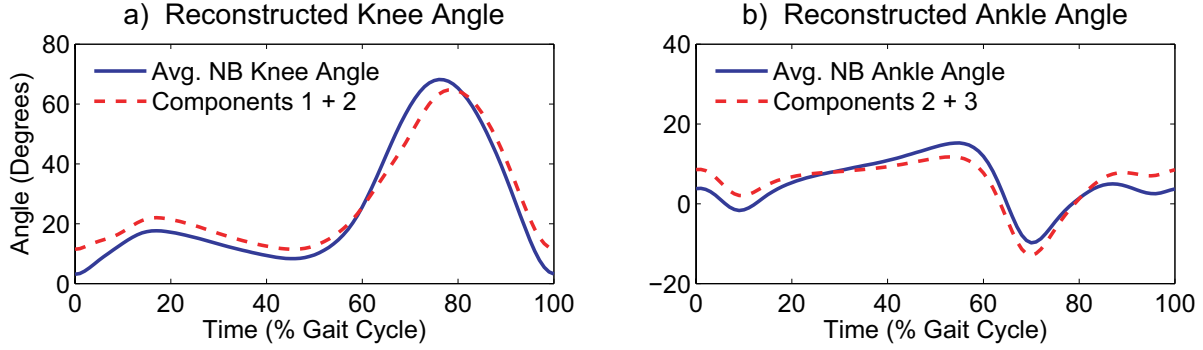
Figure 4.1: Average knee (a) and ankle (b) angles in the non-braced (NB) condition and reconstructed angles formed via linear combinations of the columns of **B** from the Parafac fitting of the NB data. Averages are over 200 gait cycles (10 replications from 2 legs of 10 subjects). Reconstructed knee (component 1 + component 2) and ankle (component 2 + component 3) angles have been scaled to have a sum-of-squares equivalent to that of the corresponding average angle (to facilitate comparison of the shape patterns).

of the knee and ankle from pre- to mid-swing, indicated by large **B** weights at 53–89% GC (Fig. 3.2b) and large positive **C** weights for the knees and ankles (Table 3.1, Comp2). Finally, the third component (foot) primarily distinguished differences between the flexion of the knee and the plantar flexion of the ankle during initial swing and mid-swing, indicated by large **B** weights at 62–85% GC (Fig. 3.2c) and the sign difference of the **C** weights between the knees and ankles (Table 3.1, Comp3).

The **C** weights (Table 3.1) were also used to examine symmetry between the left and right limbs in healthy gait. All joints displayed relatively symmetric behavior, indicated by the similar Mode C weighting patterns between corresponding left and right limb joints. The hips exhibited the most symmetry (i.e., most similar **C** weights), whereas the ankles were the least symmetrical. Also, the knees and ankles displayed a similar pattern of asymmetry: right limb joints weighted more highly on the first component (thigh), whereas left limb joints weighted more highly on the second (calf) and third (foot) components. This pattern could possibly provide an example of the interrelationships between local asymmetries that exist in normal gait to produce globally symmetric behavior in the lower-limbs (see Sadeghi, 2003).

It is worth noting that the Parafac fitting of the NB joint angle data was able to determine the source components (i.e., the thigh, calf, and foot angle waveforms) that defined subjects' hip, knee, and ankle angle waveforms during normal walking. The thigh component (53% VAF) was the most influential of the three source components, which was expected because the thigh angle waveform composed the definitions of the hip and knee angle waveforms (the two joint angle waveforms with the largest variances; see Figs. 2.2 & 2.3). Likewise, it was expected that the calf component (31% VAF) would be the second most influential source

component, given that the (inverted) calf angle waveform composed the definitions of the knee and ankle angle waveforms (see Fig. 2.2). Finally, the foot component (14% VAF) was the least influential of the three source components, given that the foot angle waveform composed only the definition of the ankle angle waveform (see Fig. 2.2). However, it is interesting that, although the foot angle waveform was not directly involved in the definitions of the hip or knee angle waveforms, the foot angular pattern was still rather useful for reconstructing the waveforms of the hips and knees (indicated by the moderate sized weights for the hips and knees in Table 3.1, Comp3). This offered evidence of the complex interrelationships that exist between lower-limb joints and segments during healthy gait, which confirmed the need for the simultaneous analysis of multi-joint gait waveform data.

In summary, Parafac made it possible to separately examine (in a single solution) the variation in the joint angle waveforms over the subjects, the time points, and the joints. This ability to decompose the variation simultaneously in all three modes made it possible to directly study individual differences (using the **A** weights) concerning interrelationships between lower-limb joints (using the **C** weights) throughout all time points of the GC (using the **B** weights). By simultaneously analyzing angular data from all six lower-limb joints, Parafac was able to reveal various interconnections between seemingly unconnected joints and limb segments (e.g., between the hip and foot). Thus, as demonstrated, the comprehensive results provided by Parafac were quite useful for understanding the dynamic interrelationships characterizing lower-limb angular patterns during healthy gait.

4.2   Components in Non-Braced and Knee-Braced Walking

When analyzing the combined NB and KB hip, knee, and ankle angle waveforms, most (about 97%) of the variation was explained by four Parafac components that were useful for characterizing and understanding differences between the NB and KB data. The behavior of the NB subjects was mostly described by the first three components (indicated by the fact that the NB subjects had relatively small weights on the fourth component; see Fig. 3.3b); however, unlike the previous analysis that included only the NB data, the first three components in this combined analysis seemed to represent the angular behaviors of (a) the knee (about 51% VAF), (b) the hip (about 24% VAF), and (c) the calf (about 18% VAF). Also unlike the previous analysis, a fourth component was required to account for the specific effects of the bracing (about 4% VAF).

The **A** (Fig. 3.3; Table 3.2), **B** (Fig. 3.4), and **C** (Table 3.3) weights were used to summarize and under-

stand differences between the NB and KB data. The first component (knee) mainly distinguished differences between knee flexion patterns of the NB and KB data from initial swing to terminal swing, indicated by large **A** weights for the NB group (Fig. 3.3a; Table 3.2, Comp1), large **B** weights at 63–91% GC (Fig. 3.4a), and large **C** weights for the knees (Table 3.3, Comp1). In contrast, the second component (hip) mostly explained the similar hip flexion patterns in the NB and KB data from initial contact to mid-stance and from mid-swing to terminal swing, indicated by the similar **A** weights for the NB and KB groups (Fig. 3.3a; Table 3.2, Comp2), large **B** weights at 0–19% and 76–100% GC (Fig. 3.4b), and large **C** weights for the hips (Table 3.3, Comp2). The third component (calf) primarily highlighted differences between the NB and KB knee and ankle flexion patterns from mid-stance to mid-swing, indicated by large **A** weights for the NB group (Fig. 3.3b; Table 3.2, Comp3), large **B** weights at 41–87% GC (Fig. 3.4c), and moderate-to-large **C** weights for the knees and ankles (Table 3.3, Comp3). Finally, the fourth component (bracing effect) mainly accounted for the reduced flexion of the right (but not left) knee in the KB data throughout swing, indicated by large **A** weights for the KB group (Fig. 3.3b; Table 3.2, Comp4), large **B** weights at 67–91% GC (Fig. 3.4d), and the sign difference of the **C** weights between the left and right knees (Table 3.3, Comp4).

The **C** weights (Table 3.3) were also used to examine patterns of asymmetry generated by the knee bracing. Increased asymmetry (compared to the previous analysis) was evident throughout all four components, but we focused on the fourth component (Table 3.3, Comp4) because it represented the specific bracing effects. The largest amount of asymmetry existed at the knees (as expected), whereas the hips exhibited the most symmetrical behavior. Also, all three joints displayed a similar pattern of asymmetry: right limb joints weighted more highly and in opposite directions than corresponding left limb joints. This pattern indicated that the right knee bracing had a main (limiting) effect on the right limb's behavior and a compensating effect on the behavior of the left limb.

Given that all subjects shared the same estimated **B** and **C** weights, Parafac made it possible to quantify individual and/or group differences in multi-joint gait waveform patterns simply by evaluating subjects' **A** weights. Based on the **A** weights, it was apparent that NB and KB subjects substantially differed on the first (knee), third (calf), and fourth (bracing effect) components (see Table 3.2). Joining this information with that from the **B** and **C** weights, it was possible to determine precisely how the NB and KB joint angular patterns differed. The KB subjects had reduced weights on the first component (see Fig. 3.3a), reflecting the KB subjects' reduced knee and ankle flexion (indicated by the **C** weights; see Table 3.3, Comp1) during the swing subphases of the GC (indicated by the **B** weights; see Fig. 3.4a). Likewise, the KB subjects

21

had reduced weights on the third component (see Fig. 3.3b), reflecting the KB subjects' reduced knee and ankle flexion (indicated by the **C** weights; see Table 3.3, Comp3) from the mid-stance subphase of the GC through mid-swing (indicated by the **B** weights; see Fig. 3.4c). Lastly, the KB subjects had large weights on the fourth component (see Fig. 3.3b), reflecting the specific effects of the right knee bracing beyond the overall reduced flexion patterns described by the first and third components. More specifically, the fourth component described the differing effects of the bracing on the right limb versus the left limb (indicated by the **C** weights; see Table 3.3, Comp4), such that the fourth component acted to further reduce the KB subjects' right knee and right ankle flexion, while increasing KB subjects' left knee and left ankle flexion to be comparable with those of the NB subjects (indicated by the **B** weights; see Fig. 3.4d). Thus, taken together, the first, third, and fourth components suggested that the KB subjects had substantially reduced right (but not left) knee and ankle flexion, particularly during the swing subphases of the GC.

In summary, the Parafac representation of the combined NB and KB data was able to produce substantively meaningful components that described the primary differences between the two groups' lower-limb angular patterns. Parafac made it possible to determine which components best accounted for the group differences (using the **A** weights), how these differences manifested themselves throughout the GC (using the **B** weights), and the manners in which lower-limb joints interacted to produce these differences (using the **C** weights). By simultaneously analyzing angular data from all six lower-limb joints of both the NB and KB subjects, Parafac revealed that the right knee bracing caused altered flexion/extension patterns throughout several lower-limb joints and segments. Thus, these results confirm the need for the holistic analysis of multi-joint gait waveform data when attempting to understand a perturbed gait pattern.

4.3   Parafac as a Clinical Tool

PCA is widely used in clinical settings for diagnosing and evaluating group differences in locomotive patterns. One common clinical use of PCA is the Gillette Gait Index (Schutte et al., 2000), which calculates a subject's deviation from normal gait based on the PCs of healthy control subjects. Another common clinical application involves using analysis of variance models on subjects' PC scores to determine whether components extracted from locomotive waveform data account for significant group differences (e.g., Chester & Wrigley, 2008; Hubley-Kozey et al., 2008; McKean et al., 2007). In addition, some researchers (e.g., Astephen & Deluzio, 2005; Decker et al., 2007; Deluzio & Astephen, 2007; Troje, 2002) have utilized Fisher's (1936) discriminant function analysis on subjects' PC scores to identify hierarchies of components that best

account for group differences in locomotive waveform data.

Like PCA, Parafac can be used as a basis for a diagnostic or evaluative tool in clinical settings. Suppose we have three-mode locomotive data (*subjects* × *time* × *joints*) where the first mode is an aggregation of a sufficiently large number of control subjects and perturbed subjects (e.g., subjects with lower-limb osteoarthritis). First, the three-mode data can be considered a calibration sample and analyzed by Parafac to produce a solution ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$) that is representative of the population of interest (e.g., the controls and the osteoarthritics). Next, a discriminant function analysis of the $\mathbf{A}$ weights can be used to determine the discriminant function that best differentiates the subject groups. Then, for any new subject with data measured in the same manner as the calibration sample (e.g., 101 time points × 6 joints), new Mode A weights $\widetilde{\mathbf{a}} \equiv \{\widetilde{a}_r\}_{R \times 1}$ can be estimated as

$$\widetilde{\mathbf{a}}' = \widetilde{\mathbf{x}}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}$$

where $\widetilde{\mathbf{x}}' \equiv \{\widetilde{x}_{11}, \ldots, \widetilde{x}_{J1}, \widetilde{x}_{12}, \ldots, \widetilde{x}_{J2}, \ldots, \widetilde{x}_{JK}\}_{1 \times JK}$ is a vectorized version of $\widetilde{\mathbf{X}} \equiv \{\widetilde{x}_{jk}\}_{J \times K}$ (the J time points × K joints matrix of the new subject's data), $\mathbf{Z} = \mathbf{C} \odot \mathbf{B}$, and $\mathbf{B}$ and $\mathbf{C}$ are the representative Parafac weights determined from the calibration sample. Finally, the new subject's estimated weights $\widetilde{\mathbf{a}}$ can be input into the discriminant function for a statistical classification of the subject into either the control or perturbed group. Thus, Parafac can reduce a new subject's data matrix of J time points × K joints to a vector of R component weights. Also, the error variance that is not explained by the Parafac model is likely irrelevant to the discrimination of the control subjects versus the perturbed subjects; if this is case, the discriminant function derived from the Parafac $\mathbf{A}$ weights will provide more accurate and reliable classification than one based on the original data (which is more noisy and error-prone).

Chapter 5

Conclusions

The ability to meaningfully analyze multivariate locomotive data is a necessary task in gait and motion analyses. In this study, we demonstrated the power of the Parafac model for analyzing gait data in three-mode form (*subjects × time × joints*). The Parafac representation of the NB data was able to (a) extract the source components (i.e., segment angle waveforms) that determined the system of lower-limb joint angle waveforms and (b) reveal both hypothesized and unexpected interconnections between lower-limb joints and segments during normal gait. In addition, the Parafac representation of the combined NB and KB data produced interpretable components revealing (a) the fundamental differences between the groups' gait patterns across multiple joints and (b) the need for the simultaneous analysis of multi-joint gait waveform data when studying abnormal patterns of locomotion.

As it is formulated here, Parafac is a descriptive model that explains systematic variation underlying gait data. Unlike parametric statistical models, a typical application of Parafac lacks information on the statistical reliability of the model's parameter estimates (i.e., component weights). However, if the number of observations (subjects in this case) is sufficiently large, the bootstrap (Efron & Tibshirani, 1993) has been found useful for obtaining reliability estimates of Parafac component weights (Kiers, 2004). Thus, if data are available from many subjects, the bootstrap can be used to form confidence intervals around the obtained Parafac weights in each mode. Indeed, if a sufficient amount of data are available, the bootstrap can be useful for various uncertainties encountered in applications of Parafac such as the number of components to extract (Hong et al., 2006; Timmerman & Kiers, 2000) and choosing between alternative models (Kroonenberg, 2008).

Parafac offers a novel method for analyzing multivariate gait and motion data. This general approach can simultaneously analyze locomotive waveform data from different subjects and joints/muscles/segments via one parsimonious model, which makes it possible to (a) identify uniquely interpretable components revealing individual and/or group differences in motion patterns, (b) determine the time points during which the com-

ponents are influential, and (c) investigate interrelationships and symmetry between joints/muscles/segments along each component. Thus, Parafac solutions provide detailed functional descriptions of locomotive behavior that could be useful in clinical applications for diagnosing, evaluating, and treating deviant motion patterns.

Appendix A

ALS Parafac Algorithm

The ALS Parafac algorithm extracting R components from the array $\underline{\mathbf{X}} \equiv \{x_{ijk}\}_{I \times J \times K}$ is given below.

ALS Parafac Algorithm : $\quad \underline{\mathbf{X}} \equiv \{x_{ijk}\}_{I \times J \times K}$ and R Components

Initialization :

    a. $\mathbf{X}_a = \underline{\mathbf{X}}$ matricized to an $I \times JK$ matrix (where j runs faster than k)

    b. $\mathbf{X}_b = \underline{\mathbf{X}}$ matricized to a $J \times KI$ matrix (where k runs faster than i)

    c. $\mathbf{X}_c = \underline{\mathbf{X}}$ matricized to a $K \times IJ$ matrix (where i runs faster than j)

    d. Set $\mathbf{B}$ and $\mathbf{C}$ to random $J \times R$ and $K \times R$ matrices (respectively)

    e. $s_0 = \text{trace}(\mathbf{X}_a \mathbf{X}_a')$ and $t = 10^{-7}$ (or some tollerably small number)

Iterative Procedure :

    1. $\mathbf{A} = \mathbf{X}_a \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ where $\mathbf{Z} = \mathbf{C} \odot \mathbf{B}$

    2. $\mathbf{B} = \mathbf{X}_b \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ where $\mathbf{Z} = \mathbf{A} \odot \mathbf{C}$

    3. $\mathbf{C} = \mathbf{X}_c \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ where $\mathbf{Z} = \mathbf{B} \odot \mathbf{A}$

    4. $s_1 = \text{trace}(\mathbf{E}_a \mathbf{E}_a')$ where $\mathbf{E}_a = \mathbf{X}_a - \mathbf{A}(\mathbf{C} \odot \mathbf{B})'$

    5. If $(s_0 - s_1)/s_0 < t$, stop the iterative procedure (algorithm converged)

        Else set $s_0 = s_1$ and go to Step 1

In the above algorithm, the symbol $\odot$ denotes the Khatri-Rao product (i.e., columnwise Kronecker product) of the weight matrices. Like most iterative algorithms, this algorithm only guarantees a locally optimal solution, due to the solution's dependence on the initialized $\mathbf{B}$ and $\mathbf{C}$ matrices. Thus, to increase the chance of obtaining the globally optimal solution (minimizing $\sum_i \sum_j \sum_k e_{ijk}^2$ with respect to the entire solution space), the algorithm should be applied several times with different randomly initialized $\mathbf{B}$ and $\mathbf{C}$ matrices each time.

Appendix B

Scaling the Parafac Solution

Given the Parafac model in Equation (2.3), we can take any $R \times R$ nonsingular diagonal matrices $\mathbf{D}_b$ and $\mathbf{D}_c$ (and their corresponding inverses $\mathbf{D}_b^{-1}$ and $\mathbf{D}_c^{-1}$) and rescale the Parafac weight matrices such as

$$\mathbf{X}_k = \mathbf{A}\mathbf{C}_k\mathbf{B}' + \mathbf{E}_k \qquad \text{for} \quad k = 1, \ldots, K,$$
$$= \widetilde{\mathbf{A}}\widetilde{\mathbf{C}}_k\widetilde{\mathbf{B}}' + \mathbf{E}_k \qquad \text{for} \quad k = 1, \ldots, K,$$

where $\widetilde{\mathbf{A}} = \mathbf{A}\mathbf{D}_b\mathbf{D}_c$, $\widetilde{\mathbf{B}} = \mathbf{B}\mathbf{D}_b^{-1}$, and $\widetilde{\mathbf{C}}_k = \mathbf{C}_k\mathbf{D}_c^{-1}$ for $k = 1, \ldots, K$. As can be verified, this rescaling does not alter the fit of a Parafac solution, which is referred to as the *scaling indeterminacy* of the Parafac model. However, using the rescaling described above, all of the weights within a given column of a given weight matrix are uniformly rescaled, so the relative importance of the levels within any mode will not change as a result of this rescaling. Thus, Parafac's scaling indeterminacy only affects the metric of the components in each mode, and does influence the substantive meaning of the Parafac results. For these reasons, Parafac's scaling indeterminacy can be viewed as a convenience, given that the resulting Parafac weight matrices can be rescaled depending on the desired interpretation for a particular mode's weights. Below, we suggest a scaling procedure tailored for Parafac solutions from gait data of the form *subjects × time × joints*.

Upon convergence of the ALS algorithm in Appendix A, the Parafac solution $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ can first be scaled by taking

$$\dot{\mathbf{B}} = \mathbf{B}\mathbf{D}_b^{-1} \qquad \text{where} \quad d_{rr^*}^{(b)} = \begin{cases} \sqrt{\sum_{j=1}^J b_{jr}^2} & \text{if } r = r^* \\ 0 & \text{if } r \neq r^* \end{cases} \qquad \text{for } r, r^* = 1, \ldots, R$$

$$\dot{\mathbf{C}} = \mathbf{C}\mathbf{D}_c^{-1} \qquad \text{where} \quad d_{rr^*}^{(c)} = \begin{cases} \sqrt{\sum_{k=1}^K c_{kr}^2} & \text{if } r = r^* \\ 0 & \text{if } r \neq r^* \end{cases} \qquad \text{for } r, r^* = 1, \ldots, R$$

$$\dot{\mathbf{A}} = \mathbf{A}\mathbf{D}_b\mathbf{D}_c$$

to absorb the scale of the data into $\dot{\mathbf{A}}$. This creates a Parafac solution $\dot{\mathbf{A}}$, $\dot{\mathbf{B}}$, $\dot{\mathbf{C}}$ where the sum of the squared entries in column r of $\dot{\mathbf{A}}$ is directly proportional to the relative influence of the r-th component (for $r = 1, \ldots, R$). If the Parafac solution is constrained to be orthogonal, the quantity $\sum_{i=1}^{I} \dot{a}_{ir}^2 / s$ provides the VAF by the r-th component (for $r = 1, \ldots, R$), where $s = \text{trace}(\mathbf{X}_a \mathbf{X}_a')$ is the total data sum-of-squares and $\mathbf{X}_a$ was defined in Appendix A.

Next, the Parafac solution $\dot{\mathbf{A}}$, $\dot{\mathbf{B}}$, $\dot{\mathbf{C}}$ can be further scaled by taking

$$\ddot{\mathbf{A}} = \dot{\mathbf{A}} \mathbf{D}_a \qquad \text{where} \quad d_{rr^*}^{(a)} = \begin{cases} 1 & \text{if } r = r^* \text{ and } \sum_{i=1}^{I} \dot{a}_{ir}^3 \geq 0 \\ -1 & \text{if } r = r^* \text{ and } \sum_{i=1}^{I} \dot{a}_{ir}^3 < 0 \qquad \text{for } r, r^* = 1, \ldots, R \\ 0 & \text{if } r \neq r^* \end{cases}$$

$$\ddot{\mathbf{C}} = \dot{\mathbf{C}} \mathbf{D}_c \qquad \text{where} \quad d_{rr^*}^{(c)} = \begin{cases} 1 & \text{if } r = r^* \text{ and } \sum_{k=1}^{K} \dot{c}_{kr}^3 \geq 0 \\ -1 & \text{if } r = r^* \text{ and } \sum_{k=1}^{K} \dot{c}_{kr}^3 < 0 \qquad \text{for } r, r^* = 1, \ldots, R \\ 0 & \text{if } r \neq r^* \end{cases}$$

$$\ddot{\mathbf{B}} = \dot{\mathbf{B}} \mathbf{D}_a \mathbf{D}_c$$

to absorb the sign of the data into $\ddot{\mathbf{B}}$. This produces a solution where the sums (of the cubes) of the columns of $\ddot{\mathbf{A}}$ and $\ddot{\mathbf{C}}$ are nonnegative, while the sums (of the cubes) of the columns of $\ddot{\mathbf{B}}$ reflect the signs of the corresponding components. Cubic terms are used in the summation (as oppose to first order terms) so that a few dominant weights (i.e., those with larger absolute values) contribute more to determining the signs of the corresponding columns of $\ddot{\mathbf{B}}$ than do many other small weights.

References

Andersson, C. A., & Bro., R. (2000). The n-way toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems, 52*, 1–4.

Astephen, J. L., & Deluzio, K. J. (2005). Changes in frontal plane dynamics and the loading response phase of the gait cycle are characteristic of severe knee osteoarthritis application of a multidimensional analysis technique. *Clinical Biomechanics, 20*, 209–217.

Astephen, J. L., Deluzio, K. J., Caldwell, G. E., Dunbar, M. J., & Hubley-Kozey, C. L. (2008). Gait and neuromuscular pattern changes are associated with differences in knee osteoarthritis severity levels. *Journal of Biomechanics, 41*, 868–876.

Bennett, B. C., Russell, S. D., Sheth, P., & Abel, M. F. (2010). Angular momentum of walking at different speeds. *Human Movement Science, 29*, 114–124.

Bro, R. (1998). *Multi-way analysis in the food industry: Models, algorithms, and applications* (Unpublished doctoral dissertation). University of Amsterdam, Amsterdam, The Netherlands.

Cattell, R. B. (1944). "Parallel Proportional Profiles" and other principles for determining the choice of factors by rotation. *Psychometrika, 9*, 267–283.

Chao, E. Y. (1980). Justification of triaxial goniometer for the measurement of joint rotation. *Journal of Biomechanics, 13*, 989–1006.

Chester, V. L., & Wrigley, A. T. (2008). The identification of age-related differences in kinetic gait parameters using principal component analysis. *Clinical Biomechanics, 23*, 212–220.

Cochran, G. V. B., Wootten, M. E., & Kadaba, M. P. (1984). Representation of dynamic electromyographic data using principal component analysis. *Annals of the New York Academy of Sciences, 435*, 392–395.

Davis, B. L., & Vaughan, C. L. (1993). Phasic behavior of EMG signals during gait: Use of multivariate statistics. *Journal of Electromyography and Kinesiology, 3*, 51–60.

Davis, J. W., Gao, H., & Kannappan, V. S. (2002). A three-mode expressive feature model of action effort. In *Proceedings of the IEEE Workshop on Motion and Video Computing*, 139–144.

Davis, J. W., & Gao, H. (2003). Recognizing human action efforts: An adaptive three-mode PCA framework. In *Proceedings of the IEEE International Conference on Computer Vision, 2*, 1463–1469.

Davis, J. W., & Gao, H. (2003). An expressive three-mode principal components model of human action style. *Image and Vision Computing, 21*, 1001–1016.

Davis, J. W., & Gao, H. (2004). An expressive three-mode principal components model for gender recognition. *Journal of Vision, 4*, 362–377.

Decker, L., Berge, C., Renous, S., & Penin, X. (2007). An alternative approach to normalization and evaluation for gait patterns: Procrustes analysis applied to the cyclograms of sprinters and middle-distance runners. *Journal of Biomechanics, 40*, 2078–2087.

Deluzio, K. J., & Astephen, J. L. (2007). Biomechanical features of gait waveform data associated with knee osteoarthritis: An application of principal component analysis. *Gait & Posture, 25*, 86–93.

Deluzio, K. J., Wyss, U. P., Zee, B., Costigan, P. A., & Sorbie, C. (1997). Principal component models of knee kinematics and kinetics: Normal vs. pathological gait patterns. *Human Movement Science, 16*, 201–217.

Donoghue, O. A., Harrison, A. J., Coffey, N., & Hayes, K. (2008). Functional data analysis of running kinematics in chronic Achilles tendon injuries. *Medicine & Science in Sports & Exercise, 40*, 1323–1335.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* Boca Raton, Florida: Chapman & Hall.

Faber, N. M., Bro, R., & Hopke, P. K. (2003). Recent developments in CANDECOMP/PARAFAC algorithms: A critical review. *Chemometrics and Intelligent Laboratory Systems, 65*, 119–137.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Journal of Eugenics, 7*, 179–188.

Franc, A. (1992). *Étude algebrique des multitableaux: Apports de l'algèbre tensorielle* (Unpublished doctoral dissertation). Université de Montpellier II, Montpellier, France.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics, 16*, 1–84.

Harshman, R. A. (1972). Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers in Phonetics, 22*, 111–117.

Harshman, R. A., & Lundy, M. E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder Jr., J. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 122–215). New York, New York: Praeger.

Harshman, R. A., & Lundy, M. E. (1984). Data preprocessing and the extended PARAFAC model. In H. G. Law, C. W. Snyder Jr., J. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 216–284). New York, New York: Praeger.

Harshman, R. A., & Lundy, M. E. (1994). PARAFAC: Parallel factor analysis. *Computational Statistics and Data Analysis, 18*, 39–72.

Harshman, R. A., & Lundy, M. E. (1996). Uniqueness proof for a family of models sharing features of Tucker's three-mode factor analysis and PARAFAC/CANDECOMP. *Psychometrika, 61*, 133–154.

Helwig, N. E., Hong, S., Hsiao-Wecksler, E. T., & Polk, J. (2011). Methods to temporally align gait cycle data. *Journal of Biomechanics, 44*, 561–566.

Herr, H., & Popovic, M. (2008). Angular momentum in human walking. *The Journal of Experimental Biology, 211*, 467–481.

Hong, S., Mitchell, S. K., & Harshman, R. A. (2006). Bootstrap scree tests: A Monte Carlo simulation and applications to published data. *British Journal of Mathematical and Statistical Psychology, 59*, 35–57.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417–441, 498–520.

Hubley-Kozey, C., Deluzio, K., & Dunbar, M. (2008). Muscle co-activation patterns during walking in those with severe knee osteoarthritis. *Clinical Biomechanics, 23*, 71–80.

Ivanenko, Y. P., Poppele, R. E., & Lacquaniti, F. (2004). Five basic muscle activation patterns account for muscle activity during human locomotion. *Journal of Physiology, 556*, 267–282.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation of factor analysis. *Psychometrika, 23,* 187–200.

Kiers, H. A. L. (1998). Joint orthomax rotation of the core and component matrices resulting from three-mode principal components analysis. *Journal of Classification, 15,* 245–263.

Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics, 14,* 105–122.

Kiers, H. A. L. (2004). Bootstrap confidence intervals for three-way methods. *Journal of Chemometrics, 18,* 22–36.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review, 51,* 455–500.

Kroonenberg, P. M. (2008). *Applied multiway data analysis* (chapter 8). Hoboken, New Jersey: John Wiley & Sons, Inc.

Kruskal, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications, 18,* 95–138.

Laassel, E. M., Loslever, P., & Angue, J. C. (1992). Patterns of relations between lower limb angle excursions during normal gait. *Journal of Biomedical Engineering, 14,* 313–320.

Lamoth, C. J. C., Daffertshofer, A., Meijer, O. G., & Beek, P. J. (2006). How do persons with chronic low back pain speed up and slow down? Trunk-pelvis coordination and lumbar erector spinae activity during gait. *Gait & Posture, 23,* 230–239.

Lauer, R. T., Stackhouse, C., Shewokis, P. A., Smith, B. T., Orlin, M., & McCarthy, J. J. (2005). Assessment of wavelet analysis of gait in children with typical development and cerebral palsy. *Journal of Biomechanics, 38,* 1351–1357.

Lee, M., Roan, M., & Smith, B. (2009). An application of principal component analysis for lower body kinematics between loaded and unloaded walking. *Journal of Biomechanics, 42,* 2226–2230.

Loslever, P., Laassel, E. M., & Angue, J. C. (1994). Combined statistical study of joint angles and ground reaction forces using component and multiple correspondence analysis. *IEEE Transactions on Biomedical Engineering, 41,* 1160–1167.

MacLellan, M. J., & McFadyen, B. J. (2010). Segmental control for adaptive locomotor adjustments during obstacle clearance in healthy young adults. *Experimental Brain Research, 202,* 307–318.

Mah, C. D., Hulliger, M., Lee, R. G., & O'Callaghan, I. S. (1994). Quantitative analysis of human movement synergies: Constructive pattern analysis for gait. *Journal of Motor Behavior, 26,* 83–102.

Mandeville, D. S., Osternig, L. R., Lantz, B. A., Mohler, C. G., & Chou, L. (2009). A multivariate statistical ranking of clinical and gait measures before and after total knee replacement. *Gait & Posture, 30,* 197–200.

McKean, K. A., Landry, S. C., Hubley-Kozey, C. L., Dunbar, M. J., Stanish, W. D., & Deluzio, K. J. (2007). Gender differences exist in osteoarthritic gait. *Clinical Biomechanics, 22,* 400–409.

Merkle, L. A., Layne, C. S., Bloomberg, J. J., & Zhang, J. J. (1998). Using factor analysis to identify neuromusuclar synergies during treadmill walking. *Journal of Neuroscience Methods, 82,* 207–214.

Muniz, A. M. S., & Nadal, J. (2009). Application of principal component analysis in vertical ground reaction force to discriminate normal and abnormal gait. *Gait & Posture, 29,* 31–35.

Paatero, P. (1999). The multilinear engine: A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *Journal of Computational and Graphical Statistics, 8,* 854–888.

Patla, A. E. (1985). Some characteristics of EMG patterns during locomotion: Implications for the locomotor control process. *Journal of Motor Behavior, 17*, 443–461.

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

Rajih, M., Comon, P., & Harshman, R. A. (2008). Enhanced line search: A novel method to accelerate PARAFAC. *SIAM Journal of Matrix Analysis and Applications, 30*, 1128–1147.

Raptopoulos, L. S. C., Dutra, M. S., Pinto, F. A. D. N. C., & Filho, A. C. D. P. (2006). Alternative approach to modal gait analysis through the Karhunen-Loève decomposition: An application in the sagittal plane. *Journal of Biomechanics, 39*, 2898–2906.

Rayens, W. S., & Mitchell, B. C. (1997). Two-factor degeneracies and a stabilization of PARAFAC. *Chemometrics and Intelligent Laboratory Systems, 38*, 173–181.

Reid, S. M., Graham, R. B., & Costigan, P. A. (2010). Differentiation of young and older adult stair climbing gait using principal component analysis. *Gait & Posture, 31*, 197–203.

Sadeghi, H. (2003). Local or global asymmetry in gait of people without impairments. *Gait & Posture, 17*, 197–204.

Sadeghi, H., Allard, P., & Duhaime, M. (1997). Functional gait asymmetry in able-bodied subjects. *Human Movement Science, 16*, 243–258.

Sadeghi, H., Prince, F., Sadeghi, S., & Labelle, H. (2000). Principal component analysis of the power developed in the flexion/extension muscles of the hip in able-bodied gait. *Medical Engineering & Physics, 22*, 703–710.

Sadeghi, H., Sadeghi, S., Price, F., Allard, P., Labelle, H., & Vaughan, C. L. (2001). Functional roles of ankle and hip sagittal muscle moments in able-bodied gait. *Clinical Biomechanics, 16*, 688–695.

Schutte, L. M., Narayanan, U., Stout, J. L., Selber, P., Gage, J. R., & Schwartz, M. H. (2000). An index for quantifying deviations from normal gait. *Gait & Posture, 11*, 25–31.

Shemmell, J., Johansson, J., Portra, V., Gottlieb, G. L., Thomas, J. S., & Corcos, D. M. (2007). Control of interjoint coordination during the swing phase of normal gait at different speeds. *Journal of NeuroEngineering and Rehabilitation, 4*, 10.

Shiavi, R., & Griffin, P. (1981). Representing and clustering electromyographic gait patterns with multivariate techniques. *Medical & Biological Engineering & Computing, 19*, 605–611.

Shorter, K. A., Polk, J. D., Rosengren, K. S., & Hsiao-Wecksler, E. T. (2008). A new approach to detecting asymmetries in gait. *Clinical Biomechanics, 23*, 459–467.

Smilde, A., Bro, R., & Geladi, P. (2004). *Multi-way analysis: Applications in the chemical sciences* (chapter 9). West Sussex, England: John Wiley & Sons, Ltd.

Timmerman, M. E., & Kiers, H. A. L. (2000). Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology, 53*, 1–16.

Tomasi, G. (2006). *Practical and computational aspects in chemometric data analysis* (Unpublished doctoral dissertation). Royal Veterinary and Agricultural University, Frederiksberg, Denmark.

Tomasi, G., & Bro, R. (2006). A comparison of algorithms for fitting the PARAFAC model. *Computational Statistics & Data Analysis, 50*, 1700–1734.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision, 2*, 371–387.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31*, 279–311.

Vaughan, C. L., Davis, B. L., & O'Connor, J. C. (1999). *Dynamics of human gait, 2nd edition* (chapter 3, appendix B). Cape Town, South Africa: Kiboho.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiah (Ed.). *Multivariate analysis* (pp. 391–420). New York, New York: Academic Press, Inc.

Wootten, M. E., Kadaba, M. P., & Cochran, G. V. B. (1990). Dynamic electromyography. II. Normal patterns during gait. *Journal of Orthopaedic Research, 8*, 259–265.

Yamamoto, S., Suto, Y., Kawamura, H., Hashizume, T., & Kakurai S. (1983). Quantitative gait evaluation of hip diseases using principal component analysis. *Journal of Biomechanics, 16*, 717–726.