Dynamic Simulation of Human Motion: Numerically Efficient Inclusion of Muscle Physiology by Convex Optimization.

Goele Pipeleers
Div. PMA, Dept. of Mechanical Engineering, Katholieke Universiteit Leuven, Celestijnenlaan 300B, B-3001 Heverlee, Belgium

Bram Demeulenaere
Div. PMA, Dept. of Mechanical Engineering, Katholieke Universiteit Leuven, Celestijnenlaan 300B, B-3001 Heverlee, Belgium

Ilse Jonkers
Dept. of Kinesiology, Katholieke Universiteit Leuven, Tervuursevest 101, B-3001 Heverlee, Belgium

Pieter Spaepen
Div. BMGO, Dept. of Mechanical Engineering, Katholieke Universiteit Leuven, Celestijnenlaan 300B, B-3001 Heverlee, Belgium

Georges Van der Perre
Div. BMGO, Dept. of Mechanical Engineering, Katholieke Universiteit Leuven, Celestijnenlaan 300B, B-3001 Heverlee, Belgium

Arthur Spaepen
Lab. for Ergonomics, Dept. of Kinesiology, Katholieke Universiteit Leuven, Tervuursevest 101, B-3001 Heverlee, Belgium

Jan Swevers
Div. PMA, Dept. of Mechanical Engineering, Katholieke Universiteit Leuven, Celestijnenlaan 300B, B-3001 Heverlee, Belgium

Joris De Schutter
Div. PMA, Dept. of Mechanical Engineering, Katholieke Universiteit Leuven, Celestijnenlaan 300B, B-3001 Heverlee, Belgium

Abstract. Determining the muscle forces that underlie some experimentally observed human motion, is a challenging biomechanical problem, both from an experimental and a computational point of view. No non-invasive method is currently available for experimentally measuring muscle forces. The alternative of computing them from the observed motion is complicated by the inherent overactuation of the human body: it has many more muscles than strictly needed for driving all the degrees of freedom of the skeleton. As a result, the skeleton's equations of motion do not suffice to determine the muscle forces unambiguously. Therefore, muscle force determination is often reformulated as a (large-scale) optimization problem.

Generally, the optimization approaches are classified according to the formalism, inverse or forward, adopted for solving the skeleton's equations of motion. Classical inverse approaches are fast but do not take into account the constraints imposed by muscle physiology. Classical forward approaches, on the other hand, do take the muscle physiology into account but are extremely costly from a computational point of view.

The present paper makes a double contribution. First, it proposes a novel inverse approach that results from including muscle physiology (both activation and contraction dynamics) in the inverse dynamic formalism. Second, the efficiency with which the corresponding optimization problem is solved is increased by using convex optimization techniques. That is, an approximate convex program is formulated and solved in order to provide a hot-start for the exact nonconvex program. The key element in this approximation is a (global) linearization of muscle physiology based on techniques from experimental system identification. This approach is applied to the study of muscle forces during gait. Although the results for gait are promising, experimental study of faster motions is needed to demonstrate the full power and advantages of the proposed methodology, and therefore is the subject of subsequent research.

Keywords: convex optimization, biomechanics, motion analysis, musculoskeletal modelling, dynamic simulation

Table of Contents

List of Symbols

Table I. List of Symbols

a_j	activation of muscle j, [-]
α_j	pennation angle of muscle j, [°]
f_s	sample frequency of the simulation, [Hz]
$F_{mt,j}$	musculotendon force of muscle j, [N]
j	index indicating the muscles of the musculoskeletal model; $j = \{1, \ldots, J\}$, [-]
J	total number of muscles included in the musculoskeletal model, [-]
k	index indicating the time instants of the simulation; $k = \{1, \ldots, K\}$, [-]
K	total number of time instants of the simulation, [-]
$l_{m,j}$	muscle fiber length of muscle j, [m]
$l_{mt,j}$	musculotendon length of muscle j, [m]
$l_{t,j}$	tendon length of muscle j, [m]
n	index indicating the degrees of freedom of the musculoskeletal model; $n = \{1, \ldots, N\}$, [-]
N	total number of degrees of freedom of the musculoskeletal model, [-]
T_s	sample period of the simulation, [s]
$\tau_{act,j}$	activation time constant of muscle j, [s]
$\tau_{deact,j}$	deactivation time constant of muscle j, [s]
u_j	excitation of muscle j, [-]
$v_{m,j}$	muscle fibre lengthening speed of muscle j, [m/s]
$\mathbf{c}(\mathbf{q}, \dot{\mathbf{q}})$	$\in \mathbb{R}^N$, vector of generalized coriolis and centrifugal forces, [N, N·m]
$\mathbf{F_{mt}}$	$\in \mathbb{R}^J$, vector of the musculotendon forces, [N]
$\mathbf{g}(\mathbf{q})$	$\in \mathbb{R}^N$, vector of generalized gravitational forces, [N, N·m]
$\mathbf{M}(\mathbf{q})$	$\in \mathbb{R}^{N \times N}$, generalized inertia matrix, [kg, kg·m^2]
$\mathbf{q}$	$\in \mathbb{R}^N$, vector of generalized coordinates of the skeleton, [m, rad]
$\mathbf{R}(\mathbf{q})$	$\in \mathbb{R}^{N \times J}$, geometric transformation matrix of $\mathbf{F_{mt}}$ to generalized joint forces, [m, -]
$\mathbf{S}(\mathbf{q})$	$\in \mathbb{R}^{N \times 12}$, geometric transformation matrix of $\mathbf{W_{ext}}$ to generalized joint forces, [m, -]
$\mathbf{T_{ext}}(\mathbf{q}, \mathbf{W_{ext}})$	$\in \mathbb{R}^N$, vector of generalized joint forces delivered by the generalized external forces, [N, N·m]
$\mathbf{T_{mt}}(\mathbf{q}, \mathbf{F_{mt}})$	$\in \mathbb{R}^N$, vector of generalized joint forces delivered by the musculotendon forces, [N, N·m]
$\mathbf{T_{pass}}(\mathbf{q}, \dot{\mathbf{q}})$	$\in \mathbb{R}^N$, vector of generalized passive forces, [N, N·m]
$\mathbf{W_{ext}}$	$\in \mathbb{R}^{12}$, vector of generalized external forces acting on the skeleton, [N, N·m]

1. Problem Statement

Walking is a common part of most people's daily activity. Yet, the muscle forces that result in an observed gait pattern are only roughly known. A comprehensive and quantitative picture of these forces would, however, be of great value to clinicians seeking to treat gait pathology (Yamaguchi and Zajac, 1990; Anderson, 1999). Moreover, muscle forces are required for the calculation of joint contact forces, the knowledge of which is indispensable for the improvement of prosthesis design, the investigation of normal and pathological joint function, and the identification of important variables in the etiology of joint diseases (Brand et al., 1994; Bergmann et al., 2001; Kleeman et al., 2003).

Determining the muscle forces that underlie some experimentally observed human motion, is a challenging biomechanical problem, both from an experimental and a computational point of view. The development of a non-invasive method to experimentally *measure* these forces is an active research domain (Drace and Pelc, 1994; Dai et al., 2001; Liu et al., 1999), but at present no method is widely accepted. On the other hand, *computing* the muscle forces corresponding to some observed motion, based on the skeleton's equations of motion, does not yield an unambiguous solution, due to the inherent overactuation of the human body. The latter problem is typically handled by formulating an optimization problem. Optimization approaches are generally classified according to the formalism, inverse or forward, adopted for solving the skeleton's equations of motion. Both formalisms are briefly explained below.

In an *inverse* (or kinetostatic) simulation, the complete motion of the mechanical system is prescribed. The equations of motion of the system are then solved as *algebraic* equations to determine the forces and moments that generate the prescribed motion (the driving forces), as well as the forces in the different mechanical connections (the joint contact forces). On the other hand, in a *forward* dynamic simulation, the driving forces are prescribed. The equations of motion are then solved as a set of *differential* and algebraic equations to determine the resulting motion, as well as the joint contact forces (Haug, 1989). The need to numerically integrate differential equations renders forward dynamic simulations computationally much more expensive than inverse dynamic simulations.

From the above, it follows that inverse optimization approaches are inherently computationally cheaper than forward optimization approaches. On the other hand, inverse approaches are much more sensitive to the quality of the experimentally measured kinematic data, for reasons explained in Section 6.3.

Another major difference between classical inverse and forward approaches concerns the integration of muscle physiology, that is, the dynamic relation between the neural stimulation of a muscle and the actual force produced. Classical inverse approaches (Seireg and Arvikar, 1975; Crowninshield et al., 1978; Pedotti et al., 1978; Patriarco et al., 1981; Crowninshield and Brand, 1981; Brand et al., 1986; Dul et al., 1984; An et al., 1989; Brand et al., 1994; Yamaguchi et al., 1995; Raikova and Prilutsky, 2001; Heller et al., 2001; Bergmann et al., 2001; Stansfield et al., 2003) do not take muscle physiology into account. As a result, the corresponding optimization problem can be decoupled into a set of K (the number of considered time instants) small, usually convex, programs. As explained in Section 3, the inclusion of muscle physiology would turn this favorable, decoupled structure into one, large-scale, nonconvex program. Hence, computational efficiency, the main motivation to actually adopt an inverse approach, is dramatically compromised. Neglecting muscle physiology implies that classical inverse optimization approaches only produce relevant results for slow motions, where muscle physiology plays less significant a role.

Forward approaches (Chow and Jacobson, 1971; Hatze, 1976; Hatze, 1981; Davy and Audu, 1987; Yamaguchi and Zajac, 1990; Pandy et al., 1992; Pandy et al., 1995; Anderson and Pandy, 2001) on the other hand, result in large-scale, nonconvex programs, even if muscle physiology is not considered. Hence, the price to be paid by adding muscle physiology constraints is still significant, but not catastrophic: two to three times more optimization variables, but not the loss of favorable structure (the program has unfavorable structure anyway). Hence, forward approaches generally take into account muscle physiology.

The present paper makes a double contribution. First, it is the first report on a novel inverse approach, called the *physiological inverse approach*, that, as opposed to *classical* inverse approaches, does take into account muscle physiology. Second, it is demonstrated that the corresponding nonconvex program can be solved[1] efficiently by using a hot-start, provided by an approximate convex program. The key element in the approximation is a global linearization of the muscle physiology, based on techniques from experimental system identification.

Section 2 briefly overviews the basic building blocks of a general dynamic musculoskeletal model, that is, the activation dynamics, contraction dynamics and skeleton dynamics model. Section 3 subsequently discusses the classical inverse and forward optimization approaches for dynamic musculoskeletal analysis, and introduces the novel physiolog-

[1] That is, an optimum is found that yields physiologically acceptable results, but of which the global optimality cannot be proven.

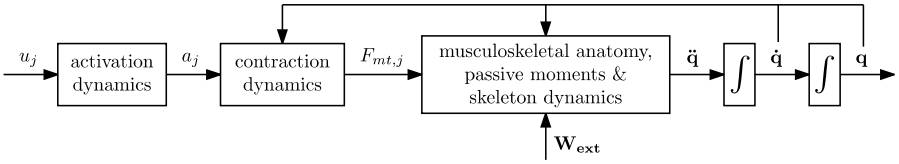

Figure 1. Schematic outline of a dynamic musculoskeletal model.

ical inverse optimization approach. The key issue when formulating an approximate convex program is a global linearization of both the activation and contraction dynamics. This is the subject of Section 4 and 5, respectively. Numerical results are presented and experimentally validated in Section 7, based on a full-scale model of the lower limbs, introduced in Section 6.

2. An Introduction to Dynamic Musculoskeletal Modelling

Figure 1 schematically outlines a dynamic musculoskeletal model, and features three submodels: the activation dynamics, contraction dynamics and skeleton dynamics model. The first two submodels account for muscle physiology, that is, the dynamic relation between the neural stimulation of a muscle and the force it actually generates. The skeleton dynamics model accounts for the dynamic relation between the applied muscle and external forces, and the resulting skeleton motion. The index $(\cdot)_j$ indicates the different muscles in the model: $j = \{1, \ldots, J\}$, and J denotes the total number of muscles included in the model.

2.1. Activation Dynamics

When the central nervous system stimulates a particular muscle, not all of the muscle fibers are equally excited (Lieber, 1992). Therefore, in order to quantify the overall excitation of a particular muscle, the muscle excitation u_j [-] is introduced. It is a dimensionless number between 0 and 1 and measures the relative number of fully excited muscle fibers. $u_j = 0$ implies that the muscle is not excited at all, while all muscle fibers (and hence the overall muscle) are fully excited if $u_j = 1$.

Neurally excited muscle fibers cannot generate force instantaneously. This time delay is modelled by the activation dynamics, which relates the muscle excitation u_j to the muscle activation a_j [-]. The muscle activation a_j corresponds to the "active state" of a muscle, as defined by Ebashi and Endo (1968). It quantifies the instantaneous force generating capacity of the muscle and is again a dimensionless number between 0 and 1.

In motion analysis, the following activation dynamics model, first introduced by He et al. (1991), is frequently used (Anderson and Pandy, 2001; Raasch et al., 1997; Umberger et al., 2003):

$$\frac{da_j}{dt} = (u_j - a_j) \cdot \left(\frac{u_j}{\tau_{act,j}} + \frac{1 - u_j}{\tau_{deact,j}} \right) . \tag{1}$$

In correspondence to Winters and Stark (1988), the activation time constant $\tau_{act,j}$ [s] and deactivation time constant $\tau_{deact,j}$ [s] are set equal, for all muscles j, to 11 ms and 68 ms, respectively. This first-order nonlinear differential equation strongly resembles the model equation of a first-order linear system with unity DC gain and time constant τ [s]:

$$\frac{da_j}{dt} = \frac{u_j - a_j}{\tau} .$$

The nonlinearity of (1) is caused by the time constant of the system being dependent on the instantaneous muscle excitation u_j: it decreases nonlinearly from $\tau_{deact,j}$ at $u_j = 0$ to $\tau_{act,j}$ at $u_j = 1$. That is, the higher the excitation, the faster the response of the model.

2.2. Contraction Dynamics

The contraction dynamics link the muscle activation a_j to the resulting musculotendon force $F_{mt,j}$ [N]. This building block of the dynamic musculoskeletal model takes into account the characteristics of the muscle as an active tissue, the interaction between the muscle fibers and the tendon, as well as the mechanical properties of the tendon tissue.

The frequently used model of Hill schematically represents a muscle j according to Fig. 2. In this figure, $l_{m,j}$ [m] denotes the length of the muscle fibers, $l_{t,j}$ [m] the length of the tendon, and $l_{mt,j}$ [m] the length of the entire musculotendon unit. The lengthening speed of the muscle fibers is represented by $v_{m,j} = dl_{m,j}/dt$ [m/s]. The angle between the orientation of the muscle fibers and the tendon is the pennation angle α_j [°].

The element CE corresponds to the contractile elements of the muscle. It is the active (that is, energy consuming) force generating element of the model, representing the actin-myosin interaction at sarcomere level (Lieber, 1992). The actively generated force of the muscle fibres depends on both the length $l_{m,j}$ and the lengthening speed $v_{m,j}$ of the fibres. The springs PE and T are passive elements that have a nonlinear spring characteristic as a function of $l_{m,j}$ and $l_{t,j}$, respectively. PE models the connective and support tissue among the muscle fibers. T models the tendon, the passive tissue that attaches the muscle to the skeleton.

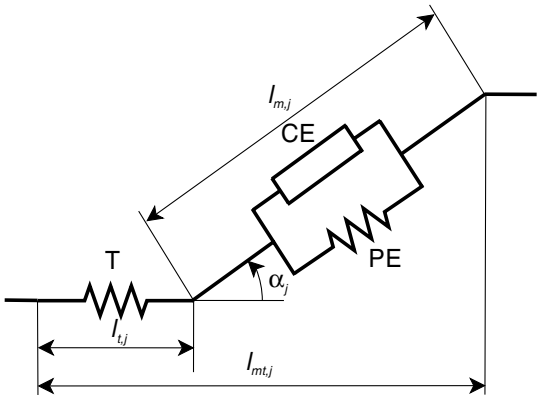

Figure 2. Schematic representation of the muscle model of Hill.

For the details of the muscle model of Hill, the reader is referred to Zajac (1989). The overall model equation is a nonlinear first-order differential equation in $F_{mt,j}$:

$$\frac{dF_{mt,j}}{dt} = f_j(F_{mt,j}, a_j, l_{mt,j}) \; . \tag{2}$$

2.3. Skeleton Dynamics

The relation between the forces and moments (*generalized forces*) acting on the skeleton and its resulting motion is determined by the skeleton's equations of motion. The skeleton motion is described by the generalized coordinates $\mathbf{q} \in \mathbb{R}^N$, corresponding to the N degrees of freedom (dofs) of the skeleton model. Different types of generalized forces act on the skeleton:

Generalized External Forces In the case of gait, the generalized external forces are the ground reaction forces (F) and moments (M). They are grouped into the vector $\mathbf{W_{ext}} \in \mathbb{R}^{12}$:

$$\begin{aligned}
\mathbf{W_{ext}} &= \left[\, \mathbf{W_{ext,r}}^T \;\; \mathbf{W_{ext,l}}^T \, \right]^T \; , \\
\mathbf{W_{ext,r}} &= \left[\, F_{r,x} \;\; F_{r,y} \;\; F_{r,z} \;\; M_{r,x} \;\; M_{r,y} \;\; M_{r,z} \, \right]^T \; , \\
\mathbf{W_{ext,l}} &= \left[\, F_{l,x} \;\; F_{l,y} \;\; F_{l,z} \;\; M_{l,x} \;\; M_{l,y} \;\; M_{l,z} \, \right]^T \; .
\end{aligned}$$

The subscripts $(\cdot)_r$ and $(\cdot)_l$ refer to the right and left foot, respectively. Generally, in gait analysis, the generalized ground reaction forces are the only ones that can be measured directly.

Musculotendon Forces The musculotendon forces $F_{mt,j}$, grouped into

$$\mathbf{F_{mt}} = \left[\, F_{mt,1} \;\; \cdots \;\; F_{mt,J} \, \right]^T \in \mathbb{R}^J \; ,$$

cannot be measured. Like cables, muscles can only exert tensile forces.

Generalized Passive Forces Motion of the joints stretches the ligaments and joint capsules. The resulting joint forces and moments are called the generalized passive forces, grouped into the vector $\mathbf{T_{pass}} \in \mathbb{R}^N$. They are fully determined by the motion of the joints $(\mathbf{q}, \dot{\mathbf{q}})$, see for instance Yamaguchi (2001).

Generalized Joint Contact Forces The generalized forces acting in the joints of the skeleton are the generalized joint contact forces. These forces are for instance relevant for the design of joint prosthesis or the etiology of joint diseases.

Application of the Euler-Lagrange formalism results in the following skeleton's equation of motion (Craig, 1986):

$$\mathbf{M}(\mathbf{q})\cdot\ddot{\mathbf{q}}+\mathbf{c}(\mathbf{q},\dot{\mathbf{q}})+\mathbf{g}(\mathbf{q})-\mathbf{T_{pass}}(\mathbf{q},\dot{\mathbf{q}})-\mathbf{T_{ext}}(\mathbf{q},\mathbf{W_{ext}}) = \mathbf{T_{mt}}(\mathbf{q},\mathbf{F_{mt}}) \; . \tag{3}$$

$\mathbf{M}(\mathbf{q}) \in \mathbb{R}^{N\times N}$ denotes the generalized inertia matrix, $\mathbf{c}(\mathbf{q},\dot{\mathbf{q}}) \in \mathbb{R}^N$ the vector of generalized coriolis and centrifugal forces, and $\mathbf{g}(\mathbf{q}) \in \mathbb{R}^N$ the vector of generalized gravitational forces. The generalized joint forces delivered by the musculotendon forces and the generalized external forces, $\mathbf{T_{mt}}(\mathbf{q},\mathbf{F_{mt}}) \in \mathbb{R}^N$ and $\mathbf{T_{ext}}(\mathbf{q},\mathbf{W_{ext}}) \in \mathbb{R}^N$ respectively, are given by:

$$\mathbf{T_{mt}}(\mathbf{q},\mathbf{F_{mt}}) \;=\; \mathbf{R}(\mathbf{q})\cdot\mathbf{F_{mt}} \; , \tag{4}$$

$$\mathbf{T_{ext}}(\mathbf{q},\mathbf{W_{ext}}) \;=\; \mathbf{S}(\mathbf{q})\cdot\mathbf{W_{ext}} \; . \tag{5}$$

$\mathbf{R}(\mathbf{q}) \in \mathbb{R}^{N\times J}$ denotes the geometric transformation matrix of the musculotendon forces to generalized joint forces, and $\mathbf{S}(\mathbf{q}) \in \mathbb{R}^{N\times 12}$ the geometric transformation matrix of the generalized external forces to generalized joint forces. The resulting equation is

$$\mathbf{M}(\mathbf{q})\cdot\ddot{\mathbf{q}}+\mathbf{c}(\mathbf{q},\dot{\mathbf{q}})+\mathbf{g}(\mathbf{q})-\mathbf{T_{pass}}(\mathbf{q},\dot{\mathbf{q}})-\mathbf{S}(\mathbf{q})\cdot\mathbf{W_{ext}} = \mathbf{R}(\mathbf{q})\cdot\mathbf{F_{mt}} \; . \tag{6}$$

Being internal forces that do not deliver work, the generalized joint contact forces do not appear in this equation. After the second-order nonlinear differential equation (6) is solved, they are readily obtained from the free-body diagram, based on algebraic equations.

3. Dynamic Musculoskeletal Analysis: Computational Approaches

This section illustrates how the building blocks, introduced in the previous section, are used to find the musculotendon forces that actuate an experimentally observed motion. All discussed approaches are based on formulating optimization problems.

3.1. Forward Optimization Approach

Given a time trajectory $u_j(t)$ for the excitation of each of the musculotendon actuators, the resulting motion can be determined by numerical integration of the combined activation, contraction and skeleton dynamics models (Fig. 1).

If, for example, the root-mean-square of the difference between the computed and the measured skeleton motion is determined, a scalar 'goodness-of-fit' measure is obtained. This measure can be used as a goal function for optimizing the excitation trajectories, to obtain the motion that resembles the measured motion as close as possible.

Due to the fact that the excitation trajectories and the resulting skeleton motion are related through a complicated set of nonlinear differential equations, the resulting large-scale optimization problem is nonconvex and very difficult to solve. In 1992, for instance, Anderson and Pandy (Anderson, 1999; Anderson and Pandy, 2001) needed 10000 CPU hours of computation time on an IBM SP-2 machine to simulate human gait, with a forward optimization approach that was based on a musculoskeletal model of a complexity comparable to that of the model used here (see Section 6).

3.2. Classical Inverse Optimization Approach

In an inverse dynamic analysis, the time-trajectories of the generalized coordinates $\mathbf{q}$ and the generalized ground reaction forces $\mathbf{W_{ext}}$ are prescribed, that is, set equal to the experimentally measured trajectories $\tilde{\mathbf{q}}$ and $\tilde{\mathbf{W}}_{\mathbf{ext}}$. Substituting these measured trajectories in Eq. (6) results, at the considered time instant t_k, in the following set of N linear algebraic equations in the J unknown musculotendon forces $F_{mt,j}(t_k)$:

$$\begin{aligned}\mathbf{R}(\tilde{\mathbf{q}}(t_k)) \cdot \mathbf{F_{mt}}(t_k) \;=\; & \mathbf{M}(\tilde{\mathbf{q}}(t_k)) \cdot \ddot{\tilde{\mathbf{q}}}(t_k) + \mathbf{c}(\tilde{\mathbf{q}}(t_k), \dot{\tilde{\mathbf{q}}}(t_k)) + \mathbf{g}(\tilde{\mathbf{q}}(t_k)) \cdots \\ & -\mathbf{T_{pass}}(\tilde{\mathbf{q}}(t_k), \dot{\tilde{\mathbf{q}}}(t_k)) - \mathbf{S}(\tilde{\mathbf{q}}(t_k)) \cdot \tilde{\mathbf{W}}_{\mathbf{ext}}(t_k)\,. \quad (7)\end{aligned}$$

In general, these N equations are linearly dependent: six (in 3D; three in 2D) equations, corresponding to the external equations of motion (Chenut et al., 2002), can be eliminated to obtain a set of $N-6$

linearly independent linear equations in J unknowns. Again in general, $N-6 < J$, and hence additional criteria are needed to obtain an unambiguous solution. Formulating an optimization problem provides a convenient way to do so. Since the optimization approach proposed in this paper is an adaptation of the classical inverse approach, the latter is developed below with full mathematical rigor.

As opposed to forward optimization approaches, classical inverse approaches do not take into account muscle physiology. Consequently, the optimization variables are the musculotendon forces themselves. These forces need to be determined for all muscles J, and at each of the K considered time instants $t_k = k \cdot T_s$ [s], $k = \{1, \ldots, K\}$, where T_s [s] denotes the sample period of the simulation. The total number of unknowns hence equals JK.

In an inverse approach, there is no dynamic coupling between the musculotendon forces at different time instants since (i) the excitation and activation dynamics are simply ignored and (ii) the skeleton dynamics reduce to the set of linear algebraic equations (7), which only couple musculotendon forces at the same time instant. As a result, an inverse optimization approach involves solving K decoupled optimization problems with J optimization variables, instead of one large program with JK variables. At each time instant t_k, the decoupled optimization problem has the following structure:

Optimization Variable

$$\mathbf{x_k} = \mathbf{F_{mt}}(t_k) = \left[\; F_{mt,1}(t_k) \;\; \cdots \;\; F_{mt,J}(t_k) \; \right]^T \in \mathbb{R}^J \; . \tag{8}$$

Constraints As for constraints, we firstly have the underdetermined set of linear equations corresponding to Eq. (7), which is concisely written as

$$\mathbf{A_k} \cdot \mathbf{x_k} = \mathbf{B_k} \; , \tag{9}$$

where $\mathbf{A_k} \in \mathbb{R}^{(N-6)\times J}$ has full rank and $\mathbf{B_k} \in \mathbb{R}^{(N-6)}$. Secondly, all musculotendon forces must be nonnegative, but smaller than some muscle-specific maximum force $F^{\max}_{mt,j}$ which depends on $\tilde{\mathbf{q}}(t_k)$. This gives rise to the following linear constraints (bounds on the variables):

$$\mathbf{0} \leq \mathbf{x_k} \leq \mathbf{F^{\max}_{mt,k}}, \tag{10}$$

where

$$\mathbf{F^{\max}_{mt,k}} = \left[\; F^{\max}_{mt,1}(\tilde{\mathbf{q}}(t_k)) \;\; \cdots \;\; F^{\max}_{mt,J}(\tilde{\mathbf{q}}(t_k)) \; \right]^T \; ,$$

and $\leq$ denotes a componentwise inequality.

Goal Function Typical goal functions (Crowninshield and Brand, 1981; Brand et al., 1994; Yamaguchi et al., 1995; Raikova and Prilutsky, 2001; Heller et al., 2001) are of the form

$$f_{goal}(x_k) = \sum_{j=1}^{J} \left(\frac{F_{mt,j}(t_k)}{C_j} \right)^n , \tag{11}$$

with C_j a positive muscle-specific constant, and $n = 2$ or $n = 3$.

The proposed goal function is convex for $n = 2$ and $n = 3$, provided that the musculotendon forces are nonnegative, which is assured by (10). The considered optimization problem is convex, since it has a convex goal function, and equality constraints (9) and inequality constraints (10) that are linear in the optimization variable $\mathbf{x_k}$. The convexity implies that the program can be solved (in the sense that the global minimum is found) numerically very efficiently.

3.3. Physiological Inverse Optimization Approach

As opposed to classical inverse approaches, the inverse approach adopted here does take into account muscle physiology; hence its name *physiological* inverse approach.

The incorporation of muscle physiology affects the corresponding optimization problem in the following way. The number of optimization variables multiplies by three, since at each time instant, the muscle excitation $u_j(t_k)$ and activation $a_j(t_k)$, $j = \{1, \ldots, J\}$, also need to be considered. The optimization variable $\mathbf{y_k}$ at time instant t_k is therefore defined as

$$\mathbf{y_k} = \left[\begin{array}{ccc} \mathbf{u}(t_k)^T & \mathbf{a}(t_k)^T & \mathbf{F_{mt}}(t_k)^T \end{array} \right]^T \in \mathbb{R}^{3J} , \tag{12}$$

where

$$\begin{aligned} \mathbf{u}(t_k) &= \left[\begin{array}{ccc} u_1(t_k) & \cdots & u_J(t_k) \end{array} \right]^T \in \mathbb{R}^J ; \\ \mathbf{a}(t_k) &= \left[\begin{array}{ccc} a_1(t_k) & \cdots & a_J(t_k) \end{array} \right]^T \in \mathbb{R}^J . \end{aligned}$$

At each time instant t_k, the following constraints must be fulfilled:

$$f_{a,j}(a_j(t_{k+1}), a_j(t_k), u_j(t_k)) = 0 ; \tag{13a}$$

$$f_{c,j}(F_{mt,j}(t_{k+1}), F_{mt,j}(t_k), a_j(t_k)) = 0 ; \tag{13b}$$

$$\left[\begin{array}{cc} \mathbf{0}_{(N-6)\times 2J} & \mathbf{A_k} \end{array} \right] \cdot \mathbf{y_k} = \mathbf{B_k} ; \tag{13c}$$

$$\left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array} \right] \leq \mathbf{y_k} \leq \left[\begin{array}{c} \mathbf{1} \\ \mathbf{F}^{\max}_{\mathbf{mt,k}} \end{array} \right] . \tag{13d}$$

Equation (13c) represents the inverse dynamic relationship between the unknown forces and the prescribed skeleton motion. Equation (13d) imposes musculotendon forces between 0 and their respective maxima, and excitation and activation values between 0 and 1. $f_{a,j}$ and $f_{c,j}$ are nonlinear functions corresponding to the difference equations that follow from discretizing the first-order nonlinear differential equations (1) and (2), which govern the activation and contraction dynamics.

Equations (13a–13b) dramatically complicate the optimization problem. Not only do nonlinear equality constraints imply nonconvexity of the problem (Boyd and Vandenberghe, 2004), they also couple the optimization variables at different time instants. Hence, regardless of the structure of the goal function, a large-scale, nonconvex program with $3JK$ variables, grouped into

$$\mathbf{y} = \left[\begin{array}{ccccc} \mathbf{y_1}^T & \cdots & \mathbf{y_k}^T & \cdots & \mathbf{y_K}^T \end{array} \right]^T \in \mathbb{R}^{3JK} \,, \tag{14}$$

needs to be solved, instead of K decoupled, convex programs with $3J$ variables.

As a result, it comes as no surprise that classical inverse approaches neglect muscle physiology, which inherently limits their applicability to slow motions, like gait. This paper claims, however, that inverse approaches that do take into account muscle physiology can still be solved efficiently through *initialization for local optimization* (Boyd and Vandenberghe, 2004). This implies that the nonconvex program is approximated by a convex program, of which the efficiently found global optimum is used as a hot-start for the nonconvex program.

The approximate program must comply with three conditions in order to be convex: (i) the goal function must be convex; (ii) the equality constraints must be linear; (iii) the inequality constraints must define a convex set.

The first condition is met by choosing the following goal function:

$$f = \sum_{k=1}^{K} \left[\sum_{j=1}^{J} \left(\frac{F_{mt,j}(t_k)}{C_j} \right)^2 \right] \,, \tag{15}$$

which sums the convex goal function (11) ($n = 2$) over all considered time instants and hence is convex. Since the inequality constraints (13d) are linear in $\mathbf{y}$, they obviously define a convex set.

The second condition can only be met if the nonlinear constraints (13a–13b) are linearized. *Local linearization*, that is, a first-order Taylor approximation around some stationary point u_j^* or a_j^*, is only useful if $u_j(t)$ and $a_j(t)$ fluctuate around the considered point during the whole time-span. However, in general, both u_j and a_j fluctuate through the

whole of their $[0, 1]$ range. Therefore, a *global linearization* needs to be carried out. This implies that the nonlinear difference equations (13a–13b) are replaced by linear difference equations that give a reasonable approximation for the whole $[0, 1]$ range. For that purpose, techniques from experimental system identification are needed for the activation dynamics (Section 4), while simple linear regression techniques suffice for the contraction dynamics (Section 5).

4. Linearization of the Activation Dynamics

Experimental system identification techniques are used to determine an approximate linear activation dynamics model, sufficiently accurate for the whole [0,1] range of u([2]). Experimental system identification aims at estimating a model of a dynamic system, based on its experimentally measured response to a known system excitation.

Experimental identification typically involves four steps, discussed hereafter: (i) designing the system excitation signal (experiment design), performing the experiment and collecting the input/output data, (ii) choosing a dynamic model structure, (iii) estimating the model parameters based on the measured data, (iv) validating the obtained model.

4.1. Experiment Design

In general, the excitation signal used for the identification should resemble the normal operation of the system as close as possible (Ljung, 1999; Pintelon and Schoukens, 2001). This general guideline is especially important here, since the identification procedure is used to calculate the best linear approximation for the considered nonlinear activation dynamics model. Because of the nonlinear original, this linear approximation depends on both the type of excitation used to generate the identification data, and the goal function used during the parameter estimation step. The linear model obtained will therefore only be the best possible linear activation dynamics model if the input used to generate the identification data resembles the input during the physiological inverse simulation.

A major issue in the present case is the substantial lack of experimental data. The only available experimental input data is provided by the electromyogram (EMG[3]) of twelve superficial muscles, measured

[2] In this section, the subscript j is dropped in order not to overload notation.

[3] The EMG is an electrical signal that qualitatively indicates the neural stimulation of the muscle. In the case of human motion, the recording electrodes are usually mounted on the skin. Hence only the EMG of superficial muscles is measurable.

during the gait cycle, discussed in Section 6. Although EMG-signals provide only qualitative indication about the actual muscles excitation u, they are used to represent the actual excitation. Experimental output (activation) data is completely lacking: muscle activation is simply not measurable.

As a result, an indirect approach is adopted by using simulated instead of experimental output data. That is, the data used for the parameter estimation is obtained by simulating the response of the nonlinear model of He et al. (1991) to the measured EMG-signals.

So, according to the above approach, the parameters of the proposed model structure will be estimated based on all the measured EMG signals and their responses, simulated by Eq. (1); in total $I = 780$ input-output couples $\{u(i), a(i)\}$, $i = 1, 2, \ldots, I$.

4.2. Model Structure and Parameter Estimation

Since the aim is to obtain a linear dynamic model, relating sampled input data $u(i)$ to sampled output data $a(i)$, a discrete time linear model structure is proposed:

$$C(q) \cdot a(i) = D(q) \cdot u(i) \; , \tag{16}$$

where

$$\begin{aligned} C(q) &= 1 + c_1 \cdot q^{-1} + c_2 \cdot q^{-2} + \cdots + c_{n-1} \cdot q^{1-n} + c_n \cdot q^{-n} \; ; \\ D(q) &= d_0 \cdot q^{m-n} + d_1 \cdot q^{m-1-n} + \cdots + d_{m-1} \cdot q^{1-n} + d_m \cdot q^{-n} \; . \end{aligned}$$

q^{-1} denotes the one-sample-period delay operator, whereas n indicates the model order and $m \leq n$, so as to obtain a causal model. The unknown model parameters are collected in the parameter vector $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = \left[\; c_1 \; \cdots \; c_n \; \; d_0 \; \cdots \; d_m \; \right]^T \; . \tag{17}$$

The aim of the identification is to derive a linear model that simulates the nonlinear activation dynamics as accurately as possible. Therefore, in the derivation of an approximate linear activation dynamics model, the model parameters $\boldsymbol{\theta}$ should be determined so as to minimize the error between the activation simulated by the original nonlinear model and the activation simulated by the linear model. To obtain this, the sum of the squared simulation errors is chosen as goal function for the parameter estimation (Ljung, 1999):

$$f(\boldsymbol{\theta}) \; = \; \sum_{i=1}^{I} \left(a(i) - \tilde{a}(i, \boldsymbol{\theta}) \right)^2 \; ; \tag{18}$$

$$\begin{aligned} \tilde{a}(i, \boldsymbol{\theta}) \; = \; & d_0 \cdot u(i - n + m) + \cdots + d_m \cdot u(i - n) \cdots \\ & - c_1 \cdot \tilde{a}(i-1) - \cdots - c_n \cdot \tilde{a}(i-n) \; . \end{aligned} \tag{19}$$

$\tilde{a}(i, \boldsymbol{\theta})$ is the output to $u(i)$, simulated by the linear model. This goal function is nonlinear in the model parameters $\boldsymbol{\theta}$. In order to provide a good starting point for this optimization, the linear least squares program corresponding to the minimization of the sum of the squared one-step-ahead prediction errors, is solved first (Ljung, 1999).

Since the nonlinear activation dynamics model has a unity DC-gain, the additional constraint

$$\frac{\sum_{i=0}^{m} d_i}{1 + \sum_{i=1}^{n} c_i} = 1$$

is imposed in both the above optimization problems so as to obtain a model with a unity DC-gain as well.

Since the nonlinear model is of low order, and to prevent overfitting, the order of the linear models is limited to two. The identified first and second order linear systems are given by

$$R_1(q) = \frac{D_1(q)}{C_1(q)} = \frac{0.557}{q - 0.443} , \tag{20}$$

$$R_2(q) = \frac{D_2(q)}{C_2(q)} = \frac{0.623 \cdot q - 0.339}{q^2 - 0.844 \cdot q + 0.129} . \tag{21}$$

The pole of model R_1 corresponds to time constant $\tau = 22$ ms and the poles of model R_2 correspond to time constants $\tau_1 = 12$ ms and $\tau_2 = 39$ ms.

4.3. Model Validation

Proper validation of the identified model implies that the input trajectory for validation is different from the input trajectory used for the estimation of the model parameters. Since all the available EMG signals are used as input for the identification, no different excitation trajectories are available for the validation. Therefore the model validation has to be based on these signals as well.

The root-mean-square value of the remaining simulation errors equals 0.021 for linear approximation R_1 and 0.020 for R_2. Since the increase of the model order from 1 to 2 does not lead to a significant improvement, R_1 is chosen as the global linearization for the nonlinear activation dynamics model of He et al. (1991). In the approximate, convex program the nonlinear equality constraints (13a) are therefore replaced by approximate linear constraints of the form

$$C_1(q) \cdot a(i) = D_1(q) \cdot u(i) . \tag{22}$$

5. Linearization of the Contraction Dynamics

For gait, the influence of the lengthening speed $v_{m,j}$ of the muscle fibers on the active muscle force is generally considered to be negligible. As a result, the nonlinear differential equation (2) can be accurately approximated by a nonlinear *algebraic* equation:

$$F_{mt,j}(t_k) = f_j(a_j(t_k), l_{mt,j}(t_k)) \; . \tag{23}$$

Hence, system identification techniques are not required here for obtaining an approximate linear model: it suffices to fit, e. g. in the least-squares sense, a straight line through a set of points defined by (23):

$$F_{mt,j}(t_k) = v_j(l_{mt,j}(t_k)) + w_j(l_{mt,j}(t_k)) \cdot a_j(t_k) \; . \tag{24}$$

The linear regression coefficients v_j and w_j are muscle-specific — hence the subscript $(\cdot)_j$ — and depend on the musculotendon length $l_{mt,j}$, which is determined by the skeleton motion and hence a priori known in an inverse dynamic analysis. For each muscle j and at each time instant t_k, $v_j(l_{mt,j}(t_k))$ and $w_j(l_{mt,j}(t_k))$ are calculated, based on a set of equidistant values between 0 and 1 for a_j.

For slow motions, Eq. (23) is only slightly nonlinear in a_j, and hence the linear model (24) results in very good approximations. For all muscles and all considered time instants t_k, the correlation coefficient (R^2-value) of the resulting linear regression lies between 0.96 and 1.00.

In faster motions, like running or bicycling, the contraction dynamics no longer simplify to an algebraic equation. As a result, dynamic models have to be estimated using system identification techniques. Since for an inverse simulation, $l_{mt,j}(t)$ is a priori known, Eq. (2) can be written as

$$\frac{dF_{mt,j}}{dt} = f_j(F_{mt,j}, a_j, t) \; . \tag{25}$$

In order to approximate this nonlinear time-varying differential equation by a linear differential equation with time-varying coefficients, recursive identification techniques for linear time-varying systems (Graham and Sin, 1984; Ljung, 1999) can be used. Since Eq. (25) is muscle-specific, this identification has to be performed for each of the muscles separately.

Similarly to the linearization of the activation dynamics, this identification suffers from a substantial lack of experimental input and output data, since muscle activation and musculotendon force are simply not measurable. As a remedy, an indirect approach using simulated instead of experimental data is suggested. The input data for the identification

should resemble the muscle activation trajectory during the observed motion as close as possible. For the muscles for which EMG data are available, the activation trajectory is calculated from the EMG signal by simulating the nonlinear activation dynamics model (Eq. (1)). Whenever no EMG data are available, the muscle activations calculated by the physiological inverse simulation with the static Hill model can be used. The output data for the identification are computed from the selected muscle activation trajectory by simulation of Eq. (25).

6. Musculoskeletal Model and Experimental Data

The developed methodology is applied to the estimation of muscle excitation patterns during gait. First, the musculoskeletal model used for the physiological inverse simulation is detailed. After that the experimental set-up, used to obtain the kinematic data and ground reaction forces, is presented.

6.1. Musculoskeletal Model

The musculoskeletal model used here is generated in *SIMM* (Musculographics Inc., Delp et al. (1990)), a dedicated software package for musculoskeletal modelling. It consists of eight body segments: a HAT segment (head-arms-trunk), the pelvis, left and right thigh, lower leg and foot. The inertial parameters are adopted to reflect the test subject anthropometry (de Leva, 1996). The pelvis is modelled as a single rigid body with six dofs. The HAT-segment actuates with the pelvis via a three-dof spherical joint. Each hip has two rotational degrees of freedom (flexion-extension and ad-adduction, but no endo-exoration, see Section 6.3) and both the knees and ankles are modelled as a one-dof revolute joint. These $N = 17$ degrees of freedom allow the gait motion to be described in the three body planes. $J = 84$ muscles actuate the model. Each leg is actuated by 39 muscles, while six abdominal and back muscles control the relative motion between the HAT-segment and the pelvis. In view of the lack of consistency in the passive moments reported in literature at present (Anderson, 1999; Davy and Audu, 1987; Yamaguchi, 2001), they are not included in the model.

SIMM-Dynamics Pipeline (Musculographics Inc.) converts the musculoskeletal model developed in *SIMM* to *SD/FAST* (Symbolic Dynamics Inc.). *SD/FAST* subsequently calculates, at each time instant t_k, the matrix $\mathbf{A_k}$ and vector $\mathbf{B_k}$ of Eq. (13c), based on the experimentally measured time-trajectories $\tilde{\mathbf{q}}(t)$ and $\tilde{\mathbf{W}}_{\mathbf{ext}}(t)$.

An important remark is to be made with respect to the matrix $\mathbf{A_k}$. From an inverse dynamics point of view, different muscles are only

coupled to each other by the degrees of freedom they actuate[4]. In the musculoskeletal model at hand, the muscles of the left and right leg are not coupled to the abdominal and back muscles, nor are they coupled to each other. The block-diagonal structure of $\mathbf{A_k} \in \mathbb{R}^{11\times 84}$ reveals this:

$$\mathbf{A_k} = \begin{bmatrix} \mathbf{A_{r,k}} & \mathbf{0}_{4\times 39} & \mathbf{0}_{4\times 6} \\ \mathbf{0}_{4\times 39} & \mathbf{A_{l,k}} & \mathbf{0}_{4\times 6} \\ \mathbf{0}_{3\times 39} & \mathbf{0}_{3\times 39} & \mathbf{A_{b,k}} \end{bmatrix} , \tag{26}$$

provided that the musculotendon forces $F_{mt,j}$ are numbered as follows: (i) $j = 1, \ldots, 39$: right leg muscles; (ii) $j = 40, \ldots, 78$: left leg muscles; (iii) $j = 79, \ldots 84$: the abdominal and back muscles. $\mathbf{A_{r,k}} \in \mathbb{R}^{4\times 39}$ and $\mathbf{A_{l,k}} \in \mathbb{R}^{4\times 39}$ being nonzero implies that the right and left leg muscles independently actuate the 4 dofs of the right and left leg, respectively. $\mathbf{A_{b,k}} \in \mathbb{R}^{3\times 6}$ being nonzero implies that the 6 abdominal and back muscles control the 3 dofs in the back. This decoupling constitutes a general, additional numerical advantage of inverse approaches, as compared to forward approaches.

As a result of the decoupling, the musculoskeletal analysis decouples into three independent optimization problems: one for each leg, and one for the abdominal and back muscles. The physiological inverse approach is illustrated for the largest problem, namely, (one of) the legs. The right leg is chosen, since EMG-data are only available for the right leg.

6.2. Experimental Data

The required experimental data is obtained during one gait cycle of a test subject. Measured are (i) the skeleton kinematics, (ii) the generalized ground reaction forces, both essential input for the physiological inverse simulation, and (iii) the EMG of the superficial right-leg muscles. The latter signal can be compared qualitatively with the computed muscle excitations, and constitutes the only experimental validation data.

Ground reaction forces and moments are measured at 2400 Hz, using an *AMTI* force plate. The motion of the skeleton, that is, its generalized coordinates $\mathbf{q}$, is determined from the motion of retroreflective markers attached to the skin at well-defined anatomical landmarks. The markers' trajectories are tracked by a six-camera *Motion Analysis* system, sampled at 60 Hz.

[4] This does not hold in a forward approach, for in this case the equations of motion imply a highly coupled set of differential equations to calculate the skeleton motion. So the excitation of one muscle can also affect other degrees of freedom than the ones it directly actuates.

The EMG data is collected using bi-articular surface electrodes; 12 muscle groups of the right leg are examined: m. biceps femoris, m. rectus femoris, m. vastus medialis and lateralis, m. semimembranosus, m. gastrocnemius (lateral and medial head), m. tibialis anterior, m. soleus, m. tensor fasciae latae and m. gluteus medius and maximus. The raw EMG signal, sampled at 2400 Hz, is high-pass filtered, rectified and normalized with respect to the maximal signal amplitude over the gait cycle.

The sample frequency of the simulation is chosen equal to the lowest sample frequency of the experimental data: $f_s = 60$ Hz. The total gait cycle takes 1.08 s, or, equivalently, $K = 65$ time instants. Consequently, the total number of optimization variables amounts to $3JK = 7605$.

6.3. Sensitivity to Measurement and Modelling Errors

Both the classical and the physiological inverse approach are, like any inverse approach, very sensitive to measurement errors, since the measured kinematics and ground reaction forces directly affect $\mathbf{B_k}$ and $\mathbf{A_k}$ in Eq. (9, 13c). Forward approaches suffer less from this problem: the measured kinematics do not affect the constraints, but only the goal function, for instance defined as the root-mean-square of the difference between the computed and the measured kinematics.

The registration of the skeleton kinematics is the dominant source of measurement errors. The motion of the skeleton is determined from the motion of retroreflective markers attached to the skin. Since the skin and the markers attached to it move relative to the underlying bones, this results in large, systematic errors on the generalized coordinates (Cappozzo et al., 1996). Since for the calculation of $\mathbf{B_k}$ and $\mathbf{A_k}$ differentiation of $\mathbf{q}$ is required, appropriate signal processing is required in order not to blow up these measurement errors.

The skin motion errors dominate small motions like hip endo-exorotation. Because of the resulting very low measurement accuracy, this dof is not included in the simulation.

The obtainable accuracy of both forward and inverse approaches is also restricted by the accuracy of the musculoskeletal model (Raikova and Prilutsky, 2001). The development of patient specific musculoskeletal models is a topic of active research, e.g. (Spoor and van Leeuwen, 1992; Arnold et al., 2000; Maganaris et al., 2006), but at present scaling of a specific model according to the patient's anthropometry is the standard procedure, e.g. (de Leva, 1996).

7. Numerical Results

This section discusses the numerical results of the physiological inverse simulation based on the model and data of Section 6. The physiological inverse simulation is an *initialization for local optimization* approach. First an approximate, convex program is solved. This so-called *Phase I* problem determines the optimization variable $\mathbf{y}$, defined by (14), so as to minimize the convex, quadratic goal function (15), subject to (for $k = 1, \ldots, K$) the linear equality constraints (13c), (22) and (24), and the linear inequality constraints (13d). MINOS (Murtagh and Saunders, 1998), a dedicated solver for large-scale nonlinear programs including quadratic programming problems, is chosen to solve[5] this problem.

After that, the global optimum of the Phase I problem is used as a hot-start for the exact, nonconvex optimization problem, in which the approximate, linear activation dynamics constraints (22) are replaced by the exact, nonlinear activation dynamics constraints (13a). Because the linearized contraction model is a very good approximation to the exact nonlinear model of Hill, it was decided not to replace the approximate, linear contraction dynamics constraints (24) by the exact, nonlinear constraints (13b). This so-called *Phase II* problem is solved[6] using SNOPT (Gill et al., 2002), a dedicated solver for large-scale constrained nonlinear programs.

The results of the physiological inverse approach are reported and discussed in the first part of this section. They are compared to the classical inverse approach in Section 7.2. To asses the initialization for local optimization strategy, Section 7.3 compares the results of the Phase II problem, when either the hot-start or randomly selected starting points are used.

7.1. Results of the Physiological Inverse Approach

For use in clinical practice both the computation time and the quality of the obtained excitation and musculotendon force trajectories are important. These aspects of the physiological inverse approach are discussed in the first and second part of this section. After that, the results of the Phase I and Phase II optimization problem are compared to each other in order to asses the quality of the linearization of the activation dynamics.

In figures 3, 4 and 6, the gait cycle is indicated as follows. The time axis starts and ends at right heel contact. The first tick on the time axis

[5] In the sense of determining its global minimum.
[6] In the sense of determining a local minimum.

indicates the start of the left swing phase, which ends at the second tick. The vertical dotted line indicates the stance-swing transition of the right leg.

7.1.1. *Computation time*
Solving the Phase I problem requires 22 CPU seconds on a Pentium IV, 2.4 GHz processor. Based on the resulting hot-start, SNOPT converges after 122 CPU seconds.

Because no results of a forward simulation of the same model on the same PC are available, a direct comparison between the computation times of a forward and a physiological inverse simulation is not possible. However, although it is not really fair to compare an IBM SP-2 from 1992 with a present-day Pentium IV PC, we still believe that the five orders of magnitude difference between 10000 hours (Anderson, 1999; Anderson and Pandy, 2001) and 144 seconds of CPU time, indicate that our inverse approach is significantly faster (despite the incorporation of muscle physiology) than a forward approach. This must of course partly be attributed to the inherently (due to the need for numerical integration in a forward approach) better numerical efficiency of inverse approaches in general, and not only to the use of a hot-start. A more objective assessment of the increase in numerical efficiency due to the use of convex optimization techniques is given in Section 7.3.

7.1.2. *Validation of the results*
The results are not only found quickly, they are also of good quality. The calculated muscle excitations comply generally well with known essential muscle actions during gait (Gage, 1991; Perry, 1992). Consequently, good qualitative correspondence between the measured EMG and the calculated muscle excitations is obtained. To illustrate this, Fig. 3(b) compares both the Phase I and Phase II muscle excitation with the measured EMG of nine muscles. Fig. 3(a) shows the internal joint moments (that is, the right-hand side of Eq. (7)) for the right leg.

Generally, only qualitative correspondence can be expected between the EMG signals and the predicted muscle excitations. This is due to the inherent inaccuracy of the EMG, which is related to the following effects. First, the surface EMG is a local registration of the electrical activity of a muscle. This local information is extrapolated to approximate the overall muscle excitation. Second, excitation of nearby muscles may disturb the EMG, so called muscle cross-talk. Third, there are many ways in which the EMG-signal can be processed in order to get a signal representative to muscle excitation (Hermens et al., 2000). For these reasons, care has to be taken when comparing the EMG data to the computed excitation trajectories. To facilitate the comparison,

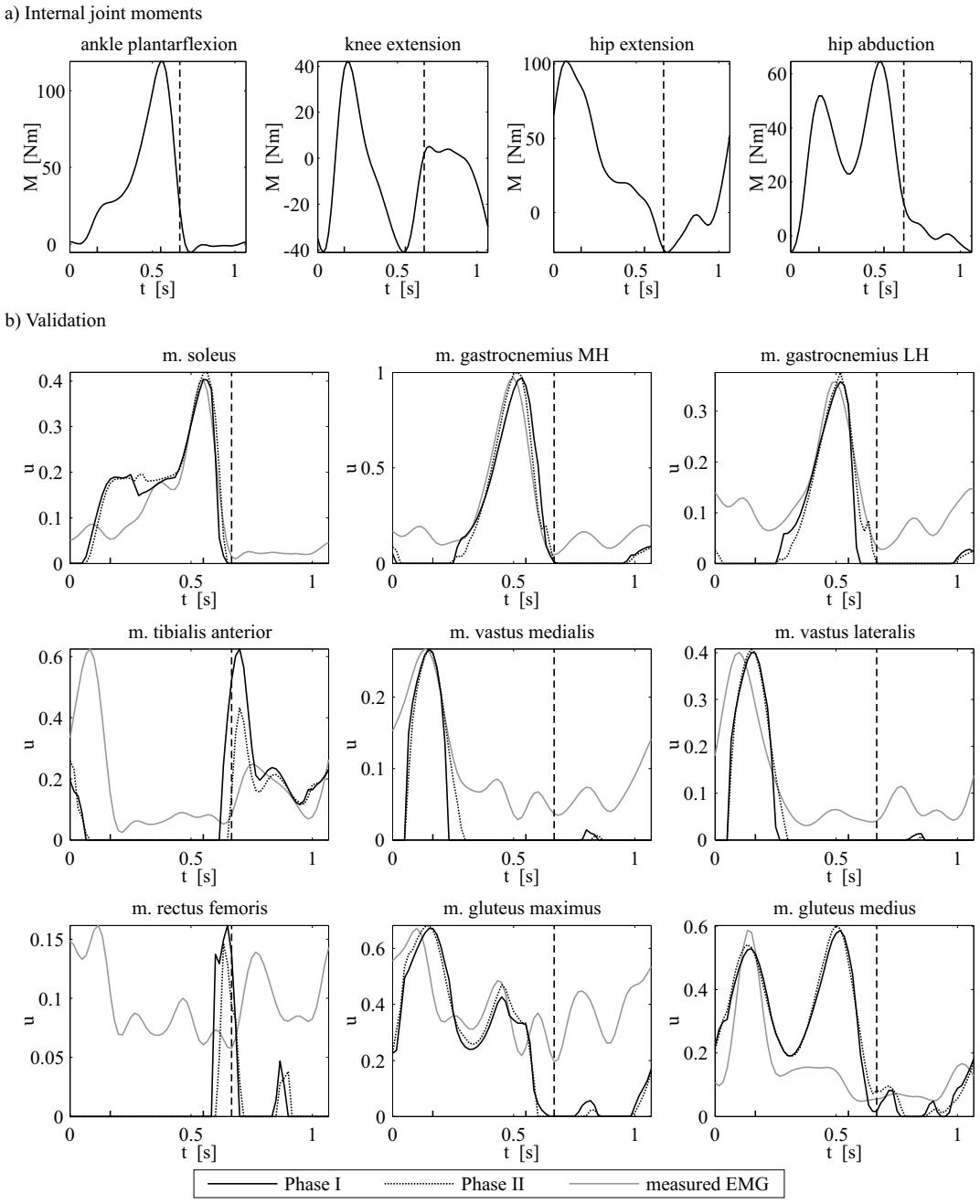

Figure 3. Initialization for local optimization: validation of the results for nine muscles: (a) internal joint moments; (b) qualitative comparison of both the Phase I and Phase II muscle excitation with the measured EMG. The time axis starts and ends at right heel contact. The first tick indicates the start of the left swing phase, which ends at the second tick. The vertical dotted line indicates the stance-swing transition of the right leg.

the EMG in Fig. 3(b), is scaled to the maximum amplitude of the computed excitation signals.

For the m. soleus, the m. gastrocnemius lateral (LH) and medial (MH) head, the m. vastus medialis and lateralis, and the m. gluteus maximus, a good qualitative correspondence between EMG and excitation is observed. The slight timing difference observed for some of these muscles, may be attributed to the fact that the activation dynamics of particular muscles are slower (or faster) than the values $\tau_{act,j} = 11$ ms and $\tau_{deact,j} = 68$ ms, suggested by Winters and Stark (1988).

The physiological inverse simulation predicts a negligible excitation of the m. rectus femoris, a bi-articular muscle that generates a combination of knee extension and hip flexion moment. This contradicts the surface EMG recordings, which indicate an equal distribution of the knee extension moment among the vasti and the m. rectus femoris. Since excitation of the m. rectus femoris would result in an additional hip flexion moment, whereas a hip extension moment is required in the stance phase, the optimization prefers not to excite this muscle. The differential nature of the excitation pattern of m. rectus femoris and vasti, although contradicting the EMG, is in close agreement with findings reported by Nene et al. (1999).

Similar to the EMG, the physiological inverse simulation predicts excitation of the m. tibialis anterior at the beginning of both the stance and the swing phase. The high excitation at the stance-swing transition is related to a relative high ankle dorsiflexion moment at these time instants and, as explained in Section 7.2, antagonist muscle action.

The computed excitation trajectory of the m. gluteus medius corresponds well to the measured EMG except in the second part of the stance phase around $t = 0.5$ s. Since the hip endorotation dof is not included in the optimization, excitation of the m. gluteus medius is not restricted at these time instants.

7.1.3. *Comparison Phase I and Phase II Results*

The close agreement between the Phase I and Phase II results in Fig. 3(b) illustrates that the linearized activation model of the Phase I problem is a good approximation to the exact nonlinear model of the Phase II problem.

7.2. Physiological Versus Classical Inverse Approach

Muscle physiology limits the speed at which muscles can generate and cut back force. Therefore, the inclusion of muscle physiology in an inverse approach will have a larger effect on the obtained muscle force pat-

terns when simulating faster motions. But even for gait, the inclusion of muscle physiology affects the calculated muscle force trajectories.

To asses this effect, Fig. 4 compares the musculotendon force of the classical (no muscle physiology) and the physiological inverse approach for the m. soleus, m. gastrocnemius and m. tibialis anterior. Although both methods result in similar profiles around the maximal force production, differences at the onset and decay of the muscle force can be observed. The fast decrease in the musculotendon force of the m. gastrocnemius and the m. soleus in the beginning of the swing phase ($t = 0.7$ s) predicted by the classical inverse approach does not comply with muscle physiology. At these time instants the physiological inverse approach predicts a higher excitation of the m. tibialis anterior in order to achieve sufficient ankle dorsiflexion moment in the presence of the decaying muscle force of the m. gastrocnemius and the m. soleus, both ankle plantarflexors. This phenomenon of antagonistic muscle action arising from premature force production and force decay can be predicted by the physiological inverse approach, while the classical inverse approach cannot. This constitutes a major advantage of the former approach over the latter.

Another advantage of the physiological inverse approach is that it facilitates the evaluation of the motion pattern by providing the clinicians with the corresponding muscle excitations. These signals can be compared to the measured EMG signals. Contrary, for the classical inverse approach no muscle excitation signals can be computed, since in this case the predicted musculotendon forces are inconsistent with muscle physiology, even for slow motions like gait. Using the classical inverse approach, the best the clinician can do is to compare the EMG signals to the computed musculo-tendon forces.

These two advantages are combined with a required CPU time of 144 seconds, which is certainly acceptable in a clinical environment.

7.3. Random Simulations

This section assesses the increase in numerical efficiency arising from the use of convex optimization techniques, by comparing the computation time and optimum obtained for the Phase II problem, when either the hot-start or randomly selected starting points are used.

100 Phase II optimizations are carried out starting from random starting points that only meet the bound constraints (13d). Based on the solver's exit flag, these Phase II optimizations are grouped into four different classes, as illustrated in Fig. 5(a). All the optimizations that converged to an optimum are grouped in the class 'optimal solution'. 'crash' indicates that either the simulation was interrupted after 12

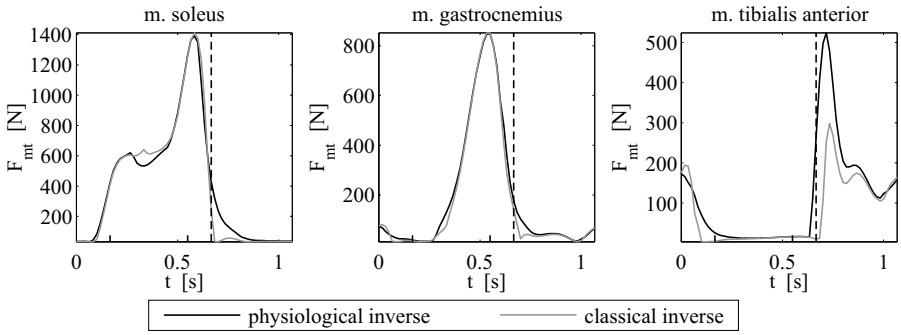

Figure 4. Comparison of the results of the classical and the physiological inverse approach for m. soleus, m. gastrocnemius and m. tibialis anterior. The time axis starts and ends at right heel contact. The first tick indicates the start of the left swing phase, which ends at the second tick. The vertical dotted line indicates the stance-swing transition of the right leg.

hours of calculation, or that Matlab (from where the solver was called) crashed during the optimization. The class 'ill cond. basis' contains the experiments that reached a point where no well conditioned basis could be found to eliminate the linearized constraints. 'infeasible' indicates that SNOPT was not able to find a feasible point. Only 55 of the 100 optimizations reached an optimum: starting a Phase II optimization from a randomly selected point does not guarantee convergence to an optimum.

For the converged experiments, both the obtained minimum f^*_{sim} [-] (Fig. 5(b)) and the computation time CPU_{sim} [s] (Fig. 5(c)) are compared to the results of the optimization starting from the hot-start: $CPU_{hot} = 122$ s and $f^*_{hot} = 74.56$. These figures prove that, when starting from a random starting point, the Phase II optimization takes a lot *more time* (a factor 2 to 54) to converge to a *worse optimum* (2.5 to 14.5 %), compared to starting from the hot-start.

There are no two random starting points from which the optimization converges to the same optimum. Compared to the optimum reached from the hot-start, the envelopes of the resulting excitation trajectories are roughly the same, but the excitations switch very nervously to 0 and back. As a result, the corresponding activations and musculotendon forces exhibit high-frequency perturbations. This is illustrated for a particular random starting point in Fig. 6 for the m. gluteus medius 1. These results are representative for all the random simulations and all the muscles. Hence, the optimum reached from the hot-start has not only a lower goal function value, but is also characterized by physiologically sounder excitation, activation and musculotendon force trajectories.

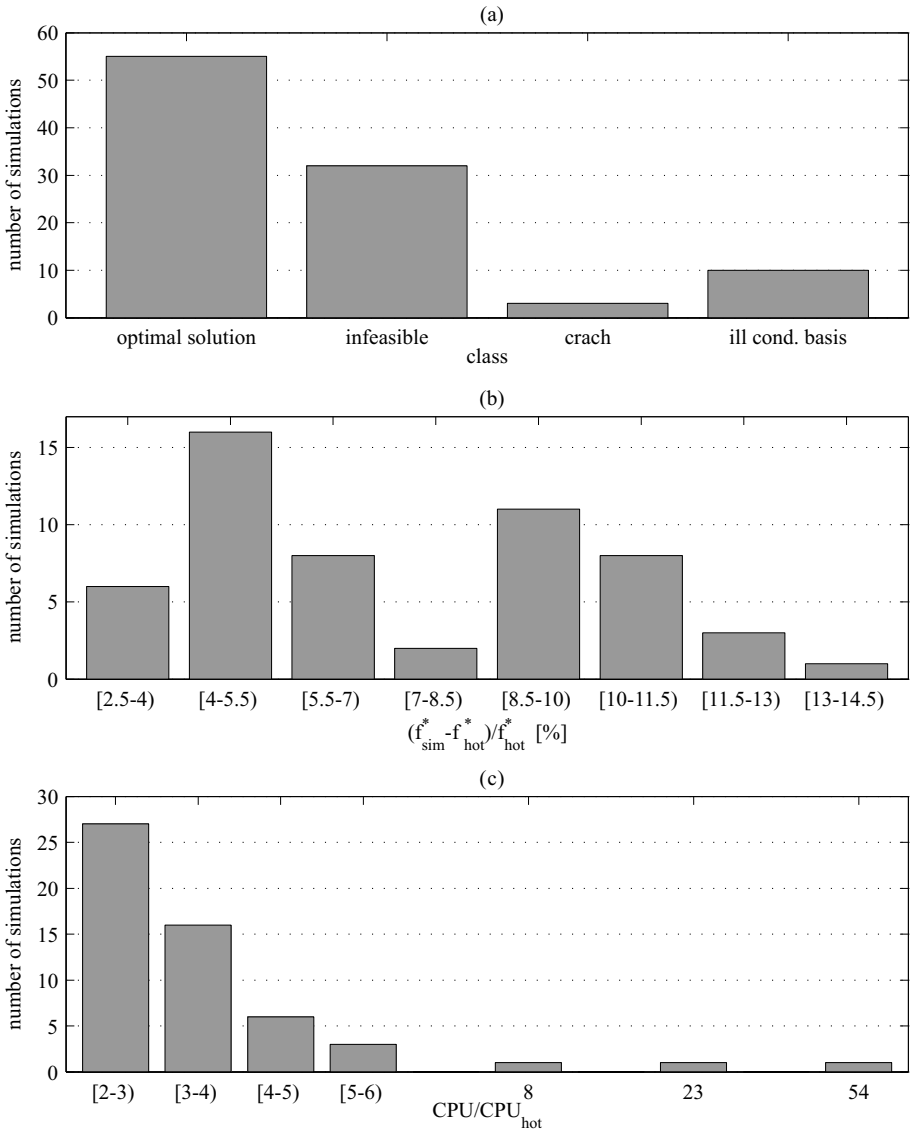

Figure 5. Results of 100 Phase II optimizations, starting from a randomly selected starting point: (a) exit conditions of the simulations; (b) relative goal function value $(f^*_{sim} - f^*_{hot})/f^*_{hot}$ of the converged experiments; (c) relative computation time CPU_{sim}/CPU_{hot} of the converged experiments.

8. Conclusions

This paper introduces the physiological inverse approach, a new approach to calculating the muscle forces that give rise to an experimentally measured motion pattern. This new approach combines the

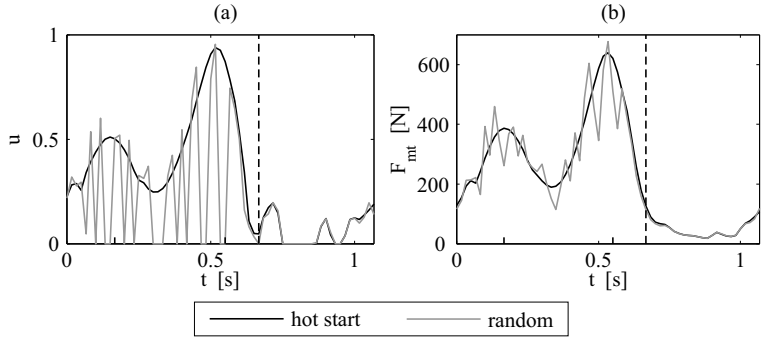

Figure 6. Comparison of the results of the Phase II optimization for the m. gluteus medius 1 when either started from the hot-start or from a particular random starting point: (a) muscle excitation; (b) corresponding musculotendon force. The time axis starts and ends at right heel contact. The first tick indicates the start of the left swing phase, which ends at the second tick. The vertical dotted line indicates the stance-swing transition of the right leg.

numerical efficiency of classical inverse approaches, with the possibility to consider muscle physiology (muscle activation and contraction dynamics). Muscle physiology was previously only included in forward approaches, which are however extremely costly from a computational point of view.

The numerical efficiency is preserved, despite the nonconvexity introduced by the muscle physiology, through initialization for local optimization, a well-known technique in convex optimization. This implies that the nonconvex program is first approximated by a convex program, whose global optimum is used as a hot-start for the original, nonconvex program. Formulating the approximate convex program requires globally linearizing the nonlinear activation and contraction dynamics. This is done based on experimental identification and simple curve fitting techniques, respectively. The numerical results indicate that these approximations are quite accurate.

The applicability of this approach is numerically demonstrated by applying it to human gait. It is shown that (i) this novel inverse approach is potentially much faster than forward approaches, despite the inclusion of muscle physiology; (ii) the hot-start yields a local optimum that is (iii-a) found faster than when the Phase II optimization is started from a random staring point, (iii-b) physiologically sound, since it is confirmed qualitatively by the measured EMG signals and essential muscle actions known from literature, and (iii-c) definitely physiologically sounder than when the Phase II optimization is started from a random starting point.

Although the results for gait are promising, experimental study of faster motions is needed to demonstrate the full power and advantages of the proposed methodology, and therefore is the subject of subsequent research. In faster motions, like running or bicycling, the contraction dynamics no longer simplify to an algebraic equation, such that system identification techniques, similar to those applied for the activation dynamics, are needed. Moreover, the nonlinearity of the muscle physiology will prevail more.

In addition, the implementation of more physiological goal functions like muscle fatigue or metabolic energy consumption is to be explored. The fact that convex functions constitute a much broader function class than linear or quadratic functions, should allow the incorporation of these physiological phenomena without significantly compromising the numerical efficiency of the proposed framework.

Future research is needed to investigate how the constraints and the goal function of the optimization can be adapted in order to enforce neurophysiologically imposed control aspects which may affect the observed co-contraction patterns (Larsen et al., 2006).

To conclude, it will be investigated how the physiological inverse approach can be combined with the classical forward approach. An interesting experiment would, for instance, be to use the results of the physiological inverse approach as a starting point for the forward approach.

Acknowledgements

The authors thank Prof. Lieven Vandenberghe (UCLA) for an interesting discussion on the topic of this paper and several pertinent hints for future research. The authors also thank the Neuromuscular Biomechanics Lab of the Stanford University for providing the experimental gait-data. They also wish to thank the anonymous reviewers, who have made several pertinent comments that have enhanced the quality of this paper.

Goele Pipeleers is a Research Assistant of the Research Foundation - Flanders (FWO–Vlaanderen). Bram Demeulenaere is a Postdoctoral Fellow of the Research Foundation - Flanders and has carried out part of this research as a visiting scholar (2004-2005) at the Electrical Engineering Dept. of the University of California, Los Angeles. Ilse Jonkers is a Postdoctoral Fellow of the Research Foundation - Flanders and has carried out part of this research as a visiting scholar at the Neuromuscular Biomechanics Lab of the Stanford University.

She receives additional funding from the Belgian Educational Foundation and the Koning Boudewijn Fonds. The support of the following projects is gratefully acknowledged: project G.0462.05 'Development of a powerful dynamic optimization framework for mechanical and biomechanical systems based on convex optimization techniques' of the Research Foundation - Flanders and K.U.Leuven-BOF EF/05/006 Center-of-Excellence Optimization in Engineering.

References

An, K. N., K. R. Kaufman, and E. Y. Chao: 1989, 'Physiological considerations of muscle force through the elbow joint'. *J. Biomech.* **22**(11-12), 1249–1256.

Anderson, F. C.: 1999, 'A Dynamic Optimization Solution for a Complete Cycle of Normal Gait'. Ph.D. thesis, University of Texas.

Anderson, F. C. and M. G. Pandy: 2001, 'Dynamic optimization of human walking'. *J. Biomech. Eng.* **123**(5), 381–390.

Arnold, A. S., S. Salinas, D. J. Asakawa, and S. L. Delp: 2000, 'Accuracy of muscle moment arms estimated from MRI-based musculoskeletal models of the lower extremity.'. *Comput. Aided Surg.* **5**(2), 108–119.

Bergmann, G., G. Deuretzbacher, M. Heller, F. Graichen, A. Rohlmann, J. Strauss, and G. N. Duda: 2001, 'Hip contact forces and gait patterns from routine activities'. *J. Biomech.* **34**(7), 859–871.

Boyd, S. and L. Vandenberghe: 2004, *Convex Optimization*. Cambridge University Press.

Brand, R. A., D. R. Pedersen, D. T. Davy, G. M. Kotzar, K. G. Heiple, and V. M. Goldberg: 1994, 'Comparison of hip force calculations and measurements in the same patient'. *J. Arthroplasty* **9**(1), 45–51.

Brand, R. A., D. R. Pedersen, and J. A. Friederich: 1986, 'The sensitivity of muscle force predictions to changes in physiologic cross-sectional area'. *J. Biomech.* **19**(8), 589–596.

Cappozzo, A., F. Cantani, A. Leardini, M. G. Benedetti, and U. D. Croce: 1996, 'Position and orientation in space of bones during movement: experimental artefacts'. *Clin. Biomech.* **11**(2), 90–100.

Chenut, X., P. Fisette, and J.-C. Samin: 2002, 'Recursive formalism with a minimal dynamic parametrization for the identification and simulation of multibody systems: application to the human body.'. *Multibody System Dynamics* **8**(2), 117–140.

Chow, C. K. and D. H. Jacobson: 1971, 'Studies of human locomotion via optimal programming'. *Math. Biosci.* **10**, 239–306.

Craig, J. J.: 1986, *Introduction to Robotics: Mechanics and Control*. Addison-Wesley, Reading, MA.

Crowninshield, R. D. and R. A. Brand: 1981, 'A physiologically based criterion of muscle force prediction in locomotion'. *J. Biomech.* **14**(11), 793–801.

Crowninshield, R. D., R. C. Johnston, J. G. Andrews, and R. A. Brand: 1978, 'A biomechanical investigation of the human hip'. *J. Biomech.* **11**(1-2), 75–85.

Dai, T. H., J. Z. Liu, V. Saghal, R. W. Brown, and G. H. Yue: 2001, 'Relationship between muscle output and functional MRI-measured brain activation'. *Exp. Brain Res.* **140**(3), 290–300.

Davy, D. T. and M. L. Audu: 1987, 'A dynamic optimization technique for predicting muscle forces in the swing phase of gait'. *J. Biomech.* **20**(2), 187–201.

de Leva, P.: 1996, 'Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters'. *J. Biomech.* **29**(9), 1223–1230.

Delp, S. L., J. P. Loan, M. G. Hoy, F. E. Zajac, E. L. Topp, and J. M. Rosen: 1990, 'An interactive graphics based model of the lower extremity to study orthopaedic surgical procedures'. *IEEE Trans. Biomed. Eng.* **37**(8), 757–767.

Drace, J. E. and N. J. Pelc: 1994, 'Skeletal muscle contraction: analysis with use of velocity distributions from phase-contrast MR imaging'. *Radiology* **193**(2), 423–429.

Dul, J., G. E. Johnson, R. Shiavi, and M. A. Townsend: 1984, 'Muscular synergism 2: A minimum-fatigue criterion for load sharing between synergistic muscles'. *J. Biomech.* **17**(9), 675–684.

Ebashi, S. and M. Endo: 1968, 'Calcium ion and muscular contraction'. *Prog. Biophys. Mol. Biol.* **18**, 123–183.

Gage, J. R.: 1991, *Gait Analysis in Cerebral Palsy.* Mac Keith Press.

Gill, P. E., W. Murray, and M. A. Saunders: 2002, *User's Guide for SNOPT Version 6, A Fortran Package for Large-Scale Nonlinear Programming.*

Graham, G. C. and K. S. Sin: 1984, *Adaptive Filtering, Prediction and Control.* Prentice-Hall, Inc.

Hatze, H.: 1976, 'The complete optimization of a human motion'. *Math. Biosci.* **28**, 99–135.

Hatze, H.: 1981, 'A comprehensive model for human motion simulation and its application to the take-off phase of the long jump'. *J. Biomech.* **14**(3), 135–142.

Haug, E. J.: 1989, *Computer Aided Kinematics and Dynamics of Mechanical Systems - Vol.I: Basic Methods.* Allyn and Bacon.

He, J. P., W. S. Levine, and G. E. Loeb: 1991, 'Feedback gains for correcting small perturbations to standing posture'. *IEEE Trans. Automatic Control* **36**(3), 322–332.

Heller, M. O., G. Bergmann, G. Deuretzbacher, L. Durselen, M. Pohl, L. Claes, N. P. Haas, and G. N. Duda: 2001, 'Musculo-skeletal loading conditions at the hip during walking and stair climbing'. *J. Biomech.* **34**(7), 883–893.

Hermens, H. J., B. Freriks, C. Disselhorst-Klug, and G. Rau: 2000, 'Development of recommendations for SEMG sensors and sensor placement procedures'. *J. Electromyogr. Kinesiol.* **10**(5), 361–374.

Kleeman, R. U., M. O. Heller, U. Stoeckle, W. R. Taylor, and G. N. Duda: 2003, 'THA loading arising from increased femoral anteversion and offset may lead to critical cement stresses'. *J. Orthop. Res.* **21**(5), 767–774.

Larsen, B., N. Mrachacz-Kersting, B. A. Lavoie, and M. Voigt: 2006, 'The amplitude modulation of the Quadriceps H-reflex in relation to the knee joint action during walking'. *Exp. Brain Res.* (Epub. ahead of print).

Lieber, R. L.: 1992, *Skeletal Muscle Structure and Function: Implications for Rehabilitation and Sports Medicine.* Williams and Wilkins.

Liu, M. M., W. Herzog, and H. Savelberg: 1999, 'Dynamic muscle force predictions from EMG: an artificial neural network approach'. *J. Electromyogr. Kinesiol.* **9**(6), 391–400.

Ljung, L.: 1999, *System Identification - Theory for the User.* PTR Prentice Hall.

Maganaris, C. N., V. Baltzopoulos, and D. Tsaopoulos: 2006, 'Muscle fibre length-to-moment arm ratios in the human lower limb determined in vivo.'. *J. Biomech.* (Epub. ahead of print).

Murtagh, B. A. and M. A. Saunders: 1998, *MINOS 5.5 User's Guide.*

Nene, A., R. Mayagoitia, and P. Veltink: 1999, 'Assessment of rectus femoris function during initial swing phase'. *Gait and Posture* **9**(1), 1–9.

Pandy, M. G., F. C. Anderson, and D. G. Hull: 1992, 'A parameter optimization approach for the optimal control of large-scale musculoskeletal systems'. *J. Biomech. Eng.* **114**(4), 450–460.

Pandy, M. G., B. A. Garner, and F. C. Anderson: 1995, 'Optimal control of non-ballistic muscular movements: a contraint-based performance criterion for rising from a chair'. *J. Biomech. Eng.* **117**(1), 15–26.

Patriarco, A. G., R. W. Mann, S. R. Simon, and J. M. Mansour: 1981, 'An evaluation of the approaches of optimization models in the prediction of muscle forces during human gait'. *J. Biomech.* **14**(8), 513–525.

Pedotti, A., V. V. Krishnan, and L. Stark: 1978, 'Optimization of muscle-force sequencing in human locomotion'. *Math. Biosci.* **38**(1-2), 57–76.

Perry, J.: 1992, *Gait Analysis: Normal and Pathological Function.* SLACK Incorporated.

Pintelon, R. and J. Schoukens: 2001, *System Identification: a Frequency Domain Approach.* Wiley-IEEE Press.

Raasch, C. C., F. E. Zajac, B. Ma, and W. S. Levine: 1997, 'Muscle coordination of maximum-speed pedaling'. *J. Biomech.* **30**(6), 595–602.

Raikova, R. T. and B. I. Prilutsky: 2001, 'Sensitivity of predicted muscle forces to parameters of the optimization-based human leg model revealed by analytical and numerical analyses'. *J. Biomech.* **34**(10), 1243–1255.

Seireg, A. and R. J. Arvikar: 1975, 'Prediction of muscular load sharing and joint forces in the lower extremities during walking'. *J. Biomech.* **8**(2), 89–102.

Spoor, C. W. and J. L. van Leeuwen: 1992, 'Knee muscle moment arms from MRI and from tendon travel.'. *J. Biomech.* **25**(2), 201–206.

Stansfield, B. W., A. C. Nicol, J. P. Paul, I. G. Kelly, F. Graichen, and G. Bergmann: 2003, 'Direct comparison of calculated hip joint contact forces with those measured using instrumented implants: An evaluation of a three-dimensional mathematical model of the lower limb'. *J. Biomech.* **36**(7), 929–936.

Umberger, B. R., K. G. Gerritsen, and P. E. Martin: 2003, 'A model of human muscle energy expenditure'. *Comput. Methods. Biomech. Biomed. Engin.* **6**(2), 99–111.

Winters, J. M. and L. Stark: 1988, 'Estimated mechanical properties of synergistic muscle involved in movements of a variety of human joints'. *J. Biomech.* **21**(12), 1027–1041.

Yamaguchi, G. T.: 2001, *Dynamic Modelling of Musculoskeletal Motion - A Vectorized Approach for Biomechanical Analysis in Three Dimensions.* Kluwer Academic Publishers.

Yamaguchi, G. T., D. W. Moran, and J. Si: 1995, 'A computationally efficient method for solving the redundant problem in biomechanics'. *J. Biomech.* **28**(8), 999–1005.

Yamaguchi, G. T. and F. E. Zajac: 1990, 'Restoring unassisted natural gait to paraplegics via functional neuromuscular stimulation: a computer simulation study'. *IEEE Trans. Biomed. Eng.* **37**(9), 886–902.

Zajac, F. E.: 1989, 'Muscle and tendon: properties, models, scaling and application to biomechanics and motor control'. *Crit. Rev. Biomed. Eng.* **17**(4), 359–411.