Rigid Body Simulations and Related Research

Bogdan Gavrea

Habilitation Thesis

Contents

1 Simulation of Rigid Bodies **6**
1.1 The rigid body constraints 6
1.2 Linear Complementarity Problems 9
1.3 Rigid body DAEs 13
1.4 Measure Differential Inclusions and Rigid Systems 15

2 Contributions **19**
2.1 Linearly implicit schemes 20
2.2 Convergence results 25
2.2.1 Pointedness of the friction cone 26
2.2.2 The reduced MDI and convergence 27
2.2.3 A numerical example 33
2.3 A large rigid-body system 34
2.4 Simulation in a quasi-static setting 44
2.5 Other contributions 54
2.5.1 A problem from optical flow 54
2.5.2 A Chebysev–Grüss type inequality 58
2.5.3 A Hermite-Hadamard type inequality 62

3 Further Research **65**
3.1 A class of stochastic optimization problems 66
3.2 A Newton-Monte Carlo method 69
3.3 Rigid body simulation in autonomous navigation 72
3.4 Conclusions 73

Index **76**

BIBLIOGRAPHY **77**

Abstract

In this habilitation thesis we have described the professional activity conducted by the candidate after obtaining his PhD degree from University of Maryland, Baltimore County, in 2006. The research results presented here are concerned with the simulation of rigid body systems with contact and friction and with mathematical inequalities with applications to probability and statistics.

The thesis is structured in three chapters. In the first chapter, we introduce the terminology and present introductory material related to the simulation of rigid body systems with joints, contacts and Coulomb friction. In the absence of contacts, the rigid body system is modeled by a system of differential algebraic equations (DAEs). The rigid body DAE is an index 3 DAE. The numerical schemes presented here use an equivalent index 2 reformulation.

When frictional contacts are present, unilateral and complementarity constraints become part of the mathematical model giving rise to a *differential complementarity problem (DCP)*. It has been shown, that even when no impacts occur the DCP is not compatible with the Coulomb friction model. This is due to the fact that impulsive forces can act even when no impact occurs and it was pointed out by P. Painlevé in 1895. Therefore a more appropriate continuous-time formulation for rigid body systems experiencing collisions and frictional contacts is given by *measure differential inclusions (MDIs)*. One of the key elements in the MDI formulation is the *friction cone*.

The numerical integration schemes (time-stepping schemes) presented here formulate the integration step as a *mixed linear complementarity problem (MLCP)*. The MLCP is a mixture of a *linear complementarity problem (LCP)* with a system of linear equations. When Coulomb friction is part of the mathematical model the matrix of the underlying LCP is copositive.

In the second chapter of this thesis the main contributions are presented. We start with the contributions related to the simulation of rigid body systems and end with auxiliary results consisting of mathematical inequalities with applications to probability and statistics. First, a class of linearly implicit time-stepping

schemes, [21], was presented here. The schemes solve a mixed linear complementarity problem (MLCP) at each integration step. To obtain the MLCP formulation, the full (nonlinear) frictional cone is replaced by a polyhedral approximation. We addressed the issue of convergence of these time-stepping schemes in the context of measure differential inclusions. One of the key assumptions in proving the convergence results is given by the *uniform pointedness of the total friction cone*. The analysis performed in [21] introduces two new concepts the *reduced friction cone* and the corresponding *reduced measure differential inclusion*.

We continue by describing the results of [37], where a quadratic programming (QP) based model for granular flow simulation inside of a pebble bed reactor was analyzed from a computational point of view. The integration step is formulated as a QP, which is obtained via convex relaxation. Both the primal and the dual formulations are considered. The dual program is a bound-constrained QP, which allows for the use of some bound constrained minimization solvers. Again the pointedness of the friction cone plays an important role and it is shown that together with the feasibility of the primal QP guarantees no duality gap.

Next we described the results obtained in [14], [13] and [17], where rigid body systems in a quasi-static setting were analyzed. In such a framework, Newton' s second law which is specific to a dynamical setting is replaced by an equilibrium equation. The application described in [14] and [13] is known as the peg-in-the-hole problem at the meso-scale. A planar rigid part (peg) is manipulated by pushing operations and the task is to take the peg from an initial sensed configuration A to the goal configuration B. The quasi-static model is motivated by the fact that inertial forces are one order smaller than the frictional forces. There are uncertainties due to sensing, modelling and parameter estimation. Since uncertainties play an important role in such a system, designing controls that will prove robust in the presence of uncertainties is extremely important. In [14] and [13], robust manipulation primitives were designed mostly based on geometric considerations. The robust controls can be used then in the context of randomized planning. Here, we have also shown how the analysis of the LCP structure may lead to the selection of robust controls. More precisely, one can use complementarity matrices to decide whether the applied controls are robust or not. A change of the complementarity matrix will be equivalent to a switching event and the corresponding control will be rejected.

An optimization problem appearing in optical flow estimation is analyzed next. The l_1 minimization problem from optical flow was presented in [23]. The optimization problem is obtained from the discrete version of the classical Horn-Schunck model. The standard l_2 energy functional is replaced by its l_1 counterpart. For the l_1 minimization problem, two linear programming reformulations were analyzed in [23]. Here we have presented the lines of our analysis for the LP with a better structure from the matrix sparsity point of view. Primal dual interior point

methods (IPM) can be used to solve this LP and the sparse structure specific to this optical flow problem can be exploited in the context of parallel algorithms.

The last part of the second chapter is concerned with some mathematical inequalities with applications to probability and statistics. In [18], a Chebysev–Grüss type inequality was given. The results developed here use the modulus of continuity and its least concave majorant, for the case of two linear positive functionals which preserve the constants. The new results of [18] can be used in various probabilistic applications. We have described here how these results can be used in the estimation of covariances for different pairs of random variables. In [19], several new inequalities of the Hermite-Hadamard type were obtained. Here, we presented one such result, that can be immediately used in the estimation of moments of continuous random variables.

In the third and last chapter we present some of the lines that characterizes the present and future research work. Some of the research items which are part of our current and future research are given below.

- Design and implementation of LCP based time-stepping schemes for autonomous navigation. This can be done by allowing virtual contacts. In this context, we plan in using the underlying LCP structure of the integration step in deciding whether switching events occur or not.

- Design of randomized algorithms for solving systems of nonlinear equation.

- Theoretical and computational results related to a class of stochastic optimization problems with mixed expectation and per-scenario constraints (abbreviated by SOESC). From an analytical point of view, we are interested in convergence results for sample average approximations applied to SOESC problems. From a computational point of view, our interest resides in designing parallel algorithms that would exploit the particular structure of these problems in the context of interior point methods.

- Design and analysis of new time-stepping schemes for both index 2 and index 3 rigid body DAEs.

- Convergence results in the measure differential inclusion sense for systems experiencing partially elastic collisions.

Chapter 1

Simulation of RIgid Bodies

1.1 The rigid body constraints

Rigid body systems are subject to bilateral, non–penetration, contact and frictional constraints. We start by introducing the terminology that will be used through out this work. We will then continue by briefly explaining the mathematical models used for each of the rigid body constraints mentioned above.

The state of a rigid-body system is represented at the position level by an array $q = (q_1, ..., q_s)^T$ of generalized coordinates ($s \in \mathbb{N}$). There are many ways in which the generalized coordinates can be selected. For example the configuration of the system can be represented by Cartesian coordinates for position and Euler angles for body centroidal frames. In this setting for each rigid body in the three-dimensional world, its configuration will be given by 6 coordinates: 3 Cartesian coordinates for the $3D$ position of a fixed point on the body and 3 other coordinates (the Euler angles) that set the orientation of the body. The velocity of the system is described by the array of generalized velocities $v = (\dot{q}_1, ..., \dot{q}_s)^T$ and acceleration of the system by the array of generalized accelerations $a = (\ddot{q}_1, ..., \ddot{q}_s)^T$. Here the generalized coordinates are time dependent, i.e., $q := q(t)$ and $\dot{q} := \dot{q}(t)$ represents the time derivative of q.

When analyzing mechanical systems the rigid body assumption considerably simplifies the mathematical model. The deformable body assumption, naturally leads to a system of partial differential equations (PDEs), which depend on material and geometric properties that are not easy to obtain. The rigid body systems analyzed here are subject to joint constraints, non-penetration, contact and frictional constraints. We will briefly present the mathematical representation of these constraints.

In any constrained mechanical system, joints connecting bodies restrict their relative motion and impose constraints on the generalized coordinates. The *joint*

constraints are then formulated as algebraic expressions involving generalized coordinates. For now, we consider holonomic, time-indepent constraints $\Theta^{(i)} : R^s \to R$, $i = 1, ..., m$. The force exerted by joint i on the system is $c_{\nu,i}\nu^{(i)}(q)$, where $\nu^{(i)}(q) = \nabla_q\Theta^{(i)}(q) \in \mathbb{R}^s$ is the gradient of $\Theta^{(i)}(q)$ and $c_{\nu,i}$ is the appropriate Lagrange multiplier [5]. We assume that the equality constraints are not redundant in the sense that the vectors $\nu^{(i)}(q)$, $i = 1, .., m$ are linearly independent. In what follows we denote by $\Theta(q)$ the vector valued function, with components $\Theta^{(i)}(q)$. More precisely, $\Theta : R^s \to R^m$,

$$\Theta(q) = \left(\Theta^{(1)}(q), ..., \Theta^{(m)}(q)\right)^T$$

We also denote by $\nu(q) = \nabla_q\Theta(q)$ the $s \times m$ matrix with columns $\nu^{(i)}(q)$ and by c_ν the vector having as components the Lagrange multipliers $c_{\nu,i}$. In this notation the force exerted by all joints on the system is represented by the vector $\nu(q)c_\nu$.

The *non–penetration constraints* are generated by the rigid body assumption which states that the bodies comprising the system cannot penetrate each other. We assume that for any pair of bodies we can define a continuous signed distance function $\Phi_j(q)$ so that the non–penetration constraints can be written as

$$\Phi^{(j)}(q) \geq 0, \quad j = 1, 2, \ldots, p\,, \tag{1.1.1}$$

where p is the number of pairs of bodies of the system that could get in contact, which in most applications is substantially smaller than the number of all possible pairs of bodies. For now, we will consider only time-independent non-penetration constraints. Although such continuous functions cannot be determined in the most general case, under some weak assumptions it is possible to define them at least in a neighborhood of all contact configurations, see [3, 11].

The *contact constraints* are complementarity constraints.We assume that the function $\Phi^{(j)}(q)$, $j = \overline{1, p}$ is differentiable in a neighborhood of the contact configuration. Let $n^{(j)}(q) = \nabla_q\Phi^{(j)}(q)$ be the gradient of the non-penetration constraint $\Phi^{(j)}(q)$. Then if contact j is active, meaning that the corresponding pair of rigid bodies achieve contact, a *"normal" force* $c_n^{(j)}n^{(j)}(q)$ will act at the contact. To avoid interpenetration the force can be only a compression force, which means that $c_n^{(j)} \geq 0$. We can express the non-penetration and contact constraints by means of the following complementarity condition

$$\Phi^{(j)}(q) \geq 0, \quad c_n^{(j)} \geq 0, \quad \Phi^{(j)}(q)c_n^{(j)} = 0, \quad j = 1, 2, \ldots, p. \tag{1.1.2}$$

Let $\Phi(q) \in \mathbb{R}^p$ denote the vector having as components $\Phi^{(j)}(q)$, $j = \overline{1, p}$ and c_n the vector with components $c_n^{(j)}$, $j = \overline{1, p}$. Then, the above conditions can be written simply as

$$0 \leq \Phi(q) \perp c_n \geq 0, \tag{1.1.3}$$

where "$\perp$" is used to represent complementarity.

The *frictional constraints* connect the tangential force, the normal force, and the velocity at contacts. They are imposed at each contact (j). The set of possible friction forces, for unitary normal force multiplier, is given by

$$FC_0^{(j)}(q) = \left\{ \overline{D}^{(j)}(q)\overline{\beta} \,|\overline{\beta} \in R^{\,d}, \|\overline{\beta}\|_2 \leq \mu \right\},$$

where $\overline{D}^{(j)}(q)$ is a given $s \times d$ matrix and μ is a nonnegative friction coefficient (the friction coefficient can have different values for different pairs of contacting bodies). The total force at contact (j), given a normal force multiplier $c_n^{(j)}$, belongs to the friction cone

$$FC^{(j)}(q) = c_n^{(j)} \left(n^{(j)}(q) + FC_0^{(j)}(q) \right) = \left\{ c_n^{(j)} n^{(j)}(q) + \overline{D}^{(j)}(q)\overline{\beta} \,|\overline{\beta} \in R^{\,d}, \|\overline{\beta}\|_2 \leq \mu c_n^{(j)} \right\}. \tag{1.1.4}$$

If $\overline{D}^{(j)}(q)$ consists of two orthogonal columns that span the tangent plane, then $FC^{(j)}(q)$ becomes the classical circular friction cone. The current representation, however, also covers the representation in global coordinates, where $n(q)$ is not necessarily orthogonal to $\overline{D}^{(j)}(q)$, [3]. The 2–norm appearing in equation (1.1.4) corresponds to isotropic friction. To cover other types of friction one could replace this norm by a function Ψ that, [45], should be convex, positively homogeneous ($\Psi(\alpha\beta) = |\alpha|\Psi(\beta)$, for all $\alpha \in R$) and coercive ($\Psi(\beta) \to \infty$ as $\|\beta\| \to \infty$).

According to the maximal dissipation principle we choose $\overline{\beta}$ to maximize the dissipation rate $-v^T\overline{D}^{(j)}(q)\overline{\beta}$ over $\overline{D}(q)\overline{\beta} \in c_n FC_0(q)$, which defines $\overline{\beta}$ as the solution of the following optimization problem:

$$\min_{\overline{\beta} \in \boldsymbol{R}^{\,d}} v^T\overline{D}^{(j)}(q)\overline{\beta} \quad \text{subject to} \quad \|\overline{\beta}\|_2 \leq \mu c_n^{(j)}. \tag{1.1.5}$$

The numerical schemes presented here use *linear complementarity problems* to impose the frictional constraints, therefore a polyhedral approximation of the friction cone, [6, 45, 46], is needed. Such an approximation is generated by the set

$$\left\{ n^{(j)}(q) + d_i^{(j)}(q),\ i = 1, 2, \ldots, m_C \right\},$$

where $d_i^{(j)}(q)$ is a collection of direction vectors in $FC_0^{(j)}(q)$ and m_C is a nonnegative integer. Using the vectors $d_i^{(j)}(q)$ we construct the matrix

$$D^{(j)}(q) = \left[d_1^{(j)}(q), d_2^{(j)}(q), \ldots, d_{m_C}^{(j)}(q) \right].$$

The columns of $D^{(j)}(q)$ are chosen to be balanced in the sense that for any i there is a k such that $d_k^{(j)}(q) = -d_i^{(j)}(q)$, [46]. This allows for one nonnegative component β_i

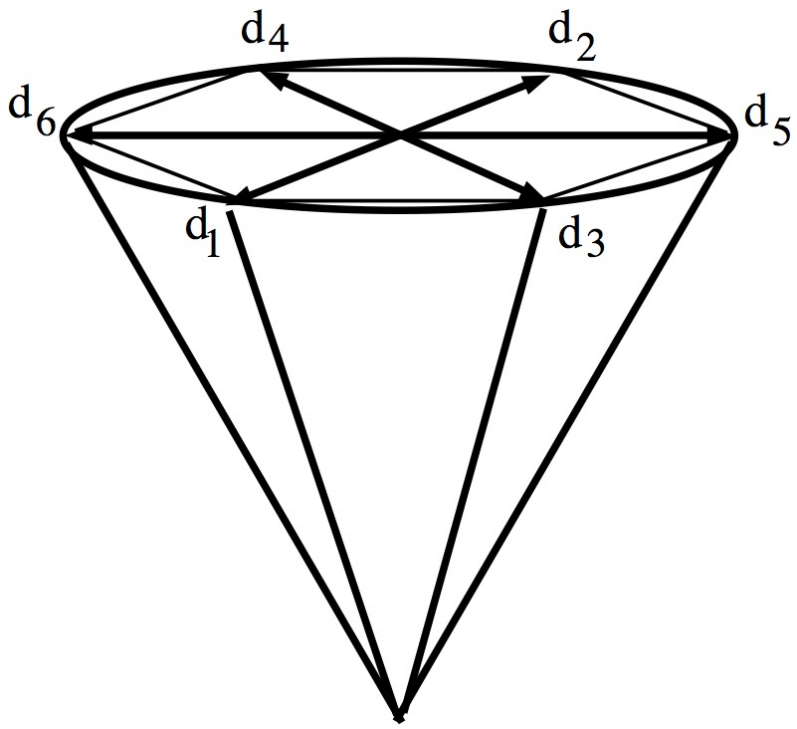


Figure 1.1: Polyhedral approximation of the friction cone. Here $m_C = 6$, and $d^{(j)}_{2k-1} = -d^{(j)}_{2k}$ for $k = 1, 2, 3$.

to be associated with every column $d^{(j)}_i(q)$ of $D^{(j)}(q)$. The (full) friction cone $FC(q)$ is approximated by the polyhedral cone (see Figure 1.1):

$$\widehat{FC}^{(j)}(q) = c^{(j)}_n \left(n^{(j)}(q) + \widehat{FC}^{(j)}_0(q) \right) = \left\{ c^{(j)}_n n^{(j)}(q) + D^{(j)}(q)\beta \,\middle|\, \beta \geq 0, \|\beta\|_1 \leq \mu c^{(j)}_n \right\}. \tag{1.1.6}$$

In the next section, we give a short overview of *linear complementarity problems (LCPs)* which are used to model contact and frictional constraints.

1.2 Linear Complementarity Problems

The time-stepping schemes that we proposed, formulate the integration step in the form of a linear complementarity problem (LCP). The LCP represents a special case of the *nonlinear complementarity problem* which we abbreviate by NCP. We start with a formal definition of the NCP and then continue with the definition of the LCP, where the nonlinear function is replaced with an affine one.

Definition 1.2.1. *Let $f : R^n \to R^n$. The nonlinear complementarity problem is the problem of finding $z \in R^n$ such that*

$$z \geq 0, \quad f(z) \geq 0, \quad \text{and} \quad z^T f(z) = 0. \tag{1.2.1}$$

By taking $f(z)$ to be an affine function one can formulate the linear complementarity problem as follows:

Definition 1.2.2. *The problem of finding $z \in R^n$ such that*

$$z \geq 0, \quad Mz + q \geq 0, \quad \text{and} \quad z^T(Mz + q) = 0. \tag{1.2.2}$$

where $q \in R^n$ and $M \in R^{n \times n}$ is called a linear complementarity problem.

The linear complementarity problem is therefore uniquelly determined by the matrix M and the vector q. We denote the above problem by $LCP(q,\ M)$. In what follows we introduce the notion of a complementarity matrix and complementarity cone. Since the systems that we numerically integrate exhibit a behavior similar to that of hybrid systems, complementarity matrices are used in detecting transitions from one mode to another. We can rewrite $LCP(q, M)$, as the problem of finding a vector pair $(w, z) \in R^{2n}$ such that

$$Iw - Mz = q, \tag{1.2.3}$$

$$w, z \geq 0, \tag{1.2.4}$$

$$w_i z_i = 0, \quad \text{for} \quad i = 1, ..., n, \tag{1.2.5}$$

where $I \in R^{n \times n}$ denotes the identity matrix.

Definition 1.2.3 ([15]). *Given $M \in R^{n \times n}$ and α a subset of $\{1, ..., n\}$ we define the complementarity matrix $C_M(\alpha) \in R^{n \times n}$ as*

$$C_M(\alpha) = \begin{cases} -M_{\cdot,i} & \text{if} \quad i \in \alpha \\ I_{\cdot,i} & \text{if} \quad i \notin \alpha \end{cases} \tag{1.2.6}$$

If $C_M(\alpha)$ is nonsingular then it is called a complementarity basis .

Definition 1.2.4 ([15]). *The cone $pos(C_M(\alpha))$ defined by*

$$pos(C_M(\alpha)) = \{q \in R^n \ : \ q = C_M(\alpha)v \ \text{ for } \ v \in R^n, \ v \geq 0\}$$

is called a complementarity cone relative to the matrix M and the index set α.

The union of all complementarity cones is also a cone that we denote by $K(M)$. The cone $K(M)$ is nothing else but the set of those vectors $q \in R^n$, for which $LCP(q, M)$ has a solution. More precisely:

$$K(M) = \{q \in R^n \ : \ \mathrm{SOL}(LCP(q, M)) \neq \emptyset\}. \tag{1.2.7}$$

It is natural to ask the question of whether for a solvable $LCP(q, M)$ there are solution pairs $(w, z) \in R^{2n}$ that use only linearly independent columns of the matrix $(I \quad -M)$. We are therefore led to the notion of a basic solution, a main concept in the theory of linear programming.

Definition 1.2.5. *A basic solution of the system of equations $Ax = b$ is a solution that uses only linearly independent columns of A.*

There is a strong connection between basic solutions and extreme points of the set

$$X = \{x \ : \ Ax = b, \ x \geq 0\} \tag{1.2.8}$$

We recall that an extreme point x of a convex set $C \in R^n$ is a point of C for which the following implication holds

$$x = \lambda x_1 + (1 - \lambda)x_2, \ x_1, x_2 \in C \ x_1 \neq x_2 \Rightarrow \lambda \in \{0, 1\}$$

It is known that $\overline{x} \in X$ is an extreme point of X if and only if $\overline{x}$ is a basic solution of $Ax = b$ (see [15], for example). We now introduce the feasible set of $LCP(q, M)$, denoted by $FEA(q, M)$, as a subset of R^{2n}, defined by

$$FEA(q, M) = \{(w, z) \in R^{2n} \ : \ Iw - Mz = q, \ w \geq 0, \ z \geq 0\} \tag{1.2.9}$$

Whenever $LCP(q, M)$ is solvable the existence of a solution pair $(\tilde{w}, \tilde{z})$ that is an extreme point of $FEA(q, M)$ is guaranteed by the following result, [15].

Theorem 1.2.6. *If (w, z) is a solution pair of $LCP(q, M)$ then there exists a solution pair $(\tilde{w}, \tilde{z})$ which is an extreme point of $FEA(q, M)$. We also have that the support of the vector pair (w, z) contains the support of $(\tilde{w}, \tilde{z})$, or more precisely,*

$$\mathrm{supp}\ \tilde{w} \times \mathrm{supp}\ \tilde{z} \subseteq \mathrm{supp}\ w \times \mathrm{supp}\ z.$$

Here by the support of a vector u, denoted by $\mathrm{supp}\ u$, we mean the index set given by $\mathrm{supp}\ u = \{i \ : \ u_i \neq 0\}$. Using the above results it follows that any extreme point solution (w, z) of $LCP(q, M)$ is a basic solution of $Au = q$, where A has the partitioned form $A = (I \quad -M)$. Therefore for any such solution there exists a complementarity basis corresponding to it.

The matrices of the underlying LCPs used in the integration schemes described here are copositive matrices. We now introduce this notion as well as an existence result that is essential in proving the solvability of the time-stepping LCPs.

Definition 1.2.7. *A matrix $M \in R^{n\times n}$ is said to be copositive if*

$$x^T M x \geq 0 \ \text{ for all } \ x \in R^n, \ \ x \geq 0.$$

In general a linear complementarity problem with a copositive matrix is not guaranteed to posses a solution. The following result gives a sufficient condition for the solvability of such an LCP.

Theorem 1.2.8 ([15], Th. 3.8.6)**.** *Let $M \in R^{n\times n}$ be a copositive matrix and let $q \in R^n$ be given. If the implication*

$$\left[v \geq 0, \ Mv \geq 0, \ v^T M v = 0\right] \ \Rightarrow \ \left[v^T q \geq 0\right] \tag{1.2.10}$$

holds, then $LCP(q, M)$ has a solution. Lemke's algorithm with precautions taken against cycling will always find a solution of $LCP(q, M)$.

Lemke's algorithm is a pivoting method similar to the simplex method of linear programming. It will produce a solution pair (w^*, z^*) that is an extreme point of $FEA(q, M)$. Cycling here refers to the possibility of using the same basis twice. The drawback of Lemke's algorithm consists in a not very high computational performance especially in the case of large LCPs that appear in the simulation of some rigid body systems such as granular matter.

When equality constraints are coupled with complementarity constraints, we are led to a generalization of the LCP, called the *mixed linear complementarity problem (MLCP)* . This is a mixture of an LCP with a system of linear equations. To be more precise, we follow the definition in [15] and consider the matrices $A \in R^{n\times n}$, $B \in R^{m\times m}$, $C \in R^{n\times m}$ and $D \in R^{m\times n}$. Let $a \in R^n$ and $b \in R^m$ be given.

Definition 1.2.9. *The mixed linear complementarity problem is the problem of finding vectors $u \in R^n$ and $v \in R^m$ such that*

$$\begin{array}{lcl} a + Au + Cv & = & 0 \\ b + Du + Bv & \geq & 0 \\ v & \geq & 0 \\ v^T(b + Du + Bv) & = & 0 \end{array} \tag{1.2.11}$$

If the matrix A is invertible we can write u in terms of v and use this form to reduce the problem to a standard LCP formulation. This observation together with Theorem 1.2.8 leads to the following solvability result, [8]:

Theorem 1.2.10 ([8], Th. 2.1)**.** *Consider a mixed linear complementarity problem of the form*

$$\begin{pmatrix} 0 \\ 0 \\ s \end{pmatrix} = \begin{pmatrix} Q & -F & -H \\ F^T & 0 & 0 \\ H^T & 0 & N \end{pmatrix} \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix} + \begin{pmatrix} -k \\ 0 \\ b \end{pmatrix}, \ \ s \geq 0, \ \ \lambda \geq 0, \ \ \lambda^T s = 0, \tag{1.2.12}$$

where $Q,\ N,\ F,\ H$ are given matrices and $b,\ k$ are given vectors of the appropriate dimension.

If Q is a positive definite matrix, N a copositive matrix and b a vector having only nonnegative components (in particular all components of b can be 0), then the above MLCP has a solution. Lemke's algorithm, with precautions taken against cycling will always find a solution λ of the LCP obtained by eliminating x and y, and then a solution (x, y, λ) of the original MLCP can be recovered by solving for x and y in the first two rows of (1.2.12).

The theorem above is the main result used in the design of the integration step. The integration steps underlying the time-stepping schemes presented here can all be cast in the same setting of Theorem 1.2.10. This result was the main tool in proving the solvability of the integration step.

1.3 Rigid body DAEs

DAEs stand for *differential algebraic equations* which consist of ordinary differential equations coupled with algebraic equations. The rigid body DAE, is a specific DAE, that is obtained as the solution to a constrained variational problem. We recall that the configuration of the rigid body system is given by $q \in \mathbb{R}^s$, the joints are modeled through the bilateral constraint $\Theta(q) = 0$, with $\Theta : \mathbb{R}^s \to \mathbb{R}^m$ and $\nu(q)$ is the $s \times m$ matrix defined by $\nu(q) = \nabla_q \Theta(q)$. With the joint constraints given by the vector valued function $\Theta(q)$, the equations of motion of an equality constrained rigid body system are given by the following DAE

$$\frac{dq}{dt} = v \tag{1.3.1}$$

$$M(q)\frac{dv}{dt} = \nu(q)c_\nu + k(t, q, v) \tag{1.3.2}$$

$$\Theta(q) = 0\,. \tag{1.3.3}$$

where $M(q)$ is the generalized mass matrix considered to be symmetric positive definite and $k(t, q, v)$ represents the combined external and inertial forces acting on the system. Differentiating (1.3.3) with respect to time gives the velocity kinematic constraint equations:

$$\nu^T(q)v = 0. \tag{1.3.4}$$

In order to ensure that the DAE (1.3.1)–(1.3.3) is solvable, i.e., a solution exists and the solution is uniquely determined by the initial values, the initial position q_0 and initial velocity v_0 have to be consistent in the sense

$$\Theta(q_0) = 0 \tag{1.3.5}$$

$$\nu^T(q_0)v_0 = 0. \tag{1.3.6}$$

Differentiating now (1.3.4) with respect to time leads to the acceleration kinematic constraint equations:

$$\left[\frac{\partial}{\partial q}(\nu^T(q)v)\right] v + \nu^T(q)\frac{dv}{dt} = 0. \tag{1.3.7}$$

The last equation is used to initialize the initial acceleration a^0 and the Lagrange multiplier vector c_ν^0. More precisely, $(a^0,\ c_\nu^0)$ is obtained as the unique solution of the following linear system.

$$\begin{aligned} M(q^0)a^0 &= \nu(q^0)c_\nu^0 + k(t_0, q^0, v^0) \\ \nu(q^0)^T a^0 &= -\left[\frac{\partial}{\partial q}(\nu(q)^T v^0)\right] v^0 \mid_{q=q^0} . \end{aligned} \tag{1.3.8}$$

From (1.3.2) one gets

$$\frac{dv}{dt} = M^{-1}(q)\nu(q)c_\nu + M^{-1}(q)k(t, q, v) \tag{1.3.9}$$

Substituting (1.3.9) in (1.3.7) and differentiating once again with respect to time leads to a differential equation for $c_\nu = c_\nu(t)$. The equation obtained this way together with (1.3.1)–(1.3.2) leads to a system of ordinary differential equations (ODE) in $(q(t), v(t), c_\nu(t))$. We give special attention to the minimum number of differentiations needed to convert the original DAE to an ODE (ODE stands for ordinary differential equations or a system of such equations). This number is known as the *differentiation index*. The higher the index, the more complicated the numerical solution becomes. In what follows we present a formal definition for the *differentiation index* of a DAE.

Definition 1.3.1. *[10] For a general implicit DAE*

$$f(t, y, \dot{y}) = 0 \tag{1.3.10}$$

the minimum number of times that all or part of (1.3.10) has to be differentiated with respect to t in order to determine $\dot{y}$ as a continuous function of t and y is called the differentiation index of the DAE (1.3.10).

Thus, according to definition (1.3.1) the index of the DAE (1.3.1–1.3.3) is 3. An analytically equivalent formulation, which will be exploited by some of the time-stepping schemes, is the index 2 DAE:

$$\frac{dq}{dt} = v \tag{1.3.11}$$

$$M(q)\frac{dv}{dt} = \nu(q)c_\nu + k(t, q, v) \tag{1.3.12}$$

$$\nu(q)^T v = 0\,. \tag{1.3.13}$$

which was obtained by differentiating with respect to time the algebraic constraint $\Theta(q) = 0$. Our work has contributed to the numerical solution of both index 3 rigid body DAEs of the form (1.3.1–1.3.3) and index 2 rigid body DAEs of the form (1.3.11)–(1.3.13). For rigid body systems subject only to bilateral constraints, the numerical schemes of Chapter 2 are applied to the index 2 DAE.

1.4 Measure Differential Inclusions and Rigid Systems

If we combine the Newton equation of dynamics with the joint, non-penetration, contact and frictional constraints, we obtain the following differential complementarity problem (DCP):

$$
\begin{aligned}
M(q)\frac{d^2q}{dt^2} - \sum_{i=1}^{m} \nu^{(i)}(q)c_\nu^{(i)} - \sum_{j=1}^{p} \left(n^{(j)}(q)c_n^{(j)} + D^{(j)}(q)\beta^{(j)}\right) &= k(t,q,\frac{dq}{dt}),\\
\Theta^{(i)}(q) \;&= 0\,, \quad i = 1,2,\ldots,m,\\
0 \le \Phi^{(j)}(q) \;\perp\; c_n^{(j)} \;&\ge 0\,, \quad j = 1,2,\ldots,p,\\
0 \le D^{(j)}(q)^T v + \lambda^{(j)} e^{(j)} \;\perp\; \beta^{(j)} \;&\ge 0\,, \quad j = 1,2,\ldots,p,\\
0 \le \mu^{(j)} c_n^{(j)} - e^{(j)^T}\beta^{(j)} \;\perp\; \lambda^{(j)} \;&\ge 0\,, \quad j = 1,2,\ldots,p,
\end{aligned}
\tag{1.4.1}
$$

where $k(t,q,\frac{dq}{dt}) := k(t,q,v)$ represents the combined external and inertial forces and $M(q)$ is the generalized mass matrix, considered to be symmetric and positive definite. We note that here, we have used polyhedral approximations of the friction cones as given in (1.1.6).

The DCP formulation (1.4.1), which we call the acceleration-force framework, is advantageous in two main ways: first, it depends on a relatively small number of physical properties, and second, its form is easily exploited in numerical integration schemes. The main difficulty consists in the fact that even if no impacts occur the acceleration–force framework may fail to be well defined, as pointed out by P. Painlevé in 1895. The example found by Painlevé shows in fact that Coulomb friction model and the equations of classical rigid body dynamics are incompatible in the sense that models based on them may lead to continuous problems without solution in classical sense. In order to avoid Painlevé–like situations one may allow for the presence of impulsive forces at any time instant. Integration schemes that won't be vulnerable to Painlevé type scenarios will rather consider integrals of the forces appearing in the DCP over small time intervals.

A more appropriate continuous-time formulation for rigid body systems experiencing collisions and frictional contacts is given by *measure differential inclusions*

(MDIs). In what follows we give a short presentation of rigid body MDIs. Before formally introducing the rigid body MDIs, we make the following conventions. Since not all possible contacts of a rigid system are active at a given time, we denote by $\mathcal{A}$, the set of all active contacts. We assume that

$$\mathcal{A} = \{j_1, ..., j_{n_a}\} \subset \{1, ..., p\}, \tag{1.4.2}$$

is the set of active contacts or simply the *active set*. We recall that p is the maximum number of all possible contacts. We also construct the following matrices:

$$\begin{aligned} \widetilde{\nu} &= \left[\nu^{(1)}, \nu^{(2)}, \ldots, \nu^{(m)}\right], \\ \widetilde{D} &= \left[D^{(j_1)}, \ldots, D^{(j_{n_a})}\right], \\ \widetilde{n} &= \left[n^{(j_1)}, n^{(j_1)}, \ldots, n^{(j_{n_a})}\right], \end{aligned}$$

where the dependence on the configuration q was omitted in order to simplify the notation. Now, we give the definition of the *total friction cone*.

Definition 1.4.1. *For a given configuration q, the total friction cone is the portion in the velocity space that can be covered by feasible constraint interaction impulses. More precisely,*

$$\mathcal{FC}(q) = \left\{ z = \tilde{\nu} c_\nu + \tilde{n} c_n + \tilde{D}\tilde{\beta} \,\middle|\, c_n \geq 0,\ \tilde{\beta} \geq 0,\ \|\beta^{(j)}\|_1 \leq \mu^{(j)} c_n^{(j)},\ \forall j \in \mathcal{A} \right\}, \tag{1.4.3}$$

where c_ν, c_n and $\widetilde{\beta}$ are vectors of appropriate dimensions and $\mu^{(j)} \geq 0$ represents the frictional coefficient for contact (j). For $j \in \mathcal{A}$, the j-th components of c_n and $\widetilde{\beta}$ are $c_n^{(j)}$ and $\widetilde{\beta}^{(j)}$, respectively.

We observe that in the absence of bilateral constraints, the above definition is nothing else but the sum of the polyhedral friction cones (1.1.6, taken over the set of active contacts.

In what follows, we use the setup of [44]. Formally, we are looking at complementarity systems of the following form.

$$\frac{dq}{dt} = v \tag{1.4.4}$$

$$M\frac{dv}{dt} = k(q, v) + \rho \tag{1.4.5}$$

$$\Theta^{(i)}(q) = 0, \quad i = 1, 2, ..., m \tag{1.4.6}$$

$$\Phi^{(j)}(q) \geq 0, \quad j = 1, ..., p \tag{1.4.7}$$

$$\rho(t) = \overline{\rho}(t) + \sum_{j=1}^{p} \rho^{(j)}(t) \in \mathcal{FC}(q) \tag{1.4.8}$$

$$\overline{\rho}(t) \in \operatorname{span}\{\nu^{(i)}(q(t)) : \ i = 1, ..., m\} \tag{1.4.9}$$

$$\|\rho^{(j)}\|\Phi^{(j)}(q) = 0, \quad j = 1, 2, ..., p \tag{1.4.10}$$

Here $\rho(t)$ accounts for all constraint forces, the term $\overline{\rho}(t)$ represents the joint forces, while $\rho^{(j)}(t)$ represents the contact and frictional forces at contact (j). Equation (1.4.10) represents the contact constraints, while the membership to $\mathcal{FC}(q)$ is related to Coulomb's friction.

In contact mechanics, measures appear as a result of the presence of impulsive forces,while inclusions appear as a result of the presence of Coulomb friction. Because of possible impulsive forces, the velocity of the system is no longer required to be an absolutely continuous function, but rather a function of bounded variation.

We are going to replace the forces, as they are understood in general, i.e., as functions, by vector measures. A vector measure is defined in terms of its action on a continuous function. Assume now that $v : [0,T] \to R^s$ is a function of bounded variation. That is, the total variation of v, $\bigvee_0^T v(\cdot)$, is finite. Here $\bigvee_0^T v(\cdot)$ is the supremum of the sums $\sum_{i=0}^{N-1} \|v(t_{i+1} - v(t_i)\|$ over all finite partitions $a = t_0 < t_1 < ... < t_{N-1} < t_N = b$. We denote this by $v \in BV([0,T])$. It follows that the measure induced by v can be understood as a linear and continuous operator defined from $C([0,T])$ with values in R^s. More precisely,

$$< dv, \phi >= \int_0^T \phi(t) dv(t), \tag{1.4.11}$$

where $\phi : [0,T] \to R$ is continuous. The Riemann-Stieljes integral in (1.4.11), which exists because of $v(\cdot)$ being of bounded variation, can be approximated by finite Riemann sums:

$$\sum_{i=0}^{N-1} \phi(\tau_i)[v(t_{i+1}) - v(t_i)],$$

where $a = t_0 < \tau_1 < t_1 < ... < \tau_{N-1} < t_N = b$. Discontinuities in the velocity may lead to atoms of the measure dv. Therefore dv is not in general absolutely continuous with respect to the Lebesgue measure dt, and thus $\frac{dv}{dt}(\cdot)$ cannot be defined, in the usual sense, as a Radon–Nykodim derivative. To give a meaning to inclusions of the form

$$\frac{dv}{dt}(t) \in K(t), \quad \text{for } t \in [0,T], \tag{1.4.12}$$

we adopt the following definition [44].

Definition 1.4.2 (Measure Differential Inclusion)**.** *If $v \in BV([0,T])$ and $K(\cdot)$ is a convex-set valued mapping we say that (1.4.12) holds if, for all continuous ϕ :*

$[0,T] \to R$, $\phi \geq 0$ *and* ϕ *not identically zero, we have that*

$$\frac{\int_0^T \phi(t)dv(t)}{\int_0^T \phi(t)dt} \in \bigcup_{\tau:\phi(\tau)\neq 0} K(\tau).$$

Next we define the concept of a weak solution to (1.4.4)-(1.4.10)

Definition 1.4.3. *We say that* $q(t)$, $v(t)$ *is a weak solution of (1.4.4)-(1.4.10) on* $[0,T]$ *if*

1. $v(\cdot)$ *is a function of bounded variation on* $[0,T]$.
2. $q(\cdot)$ *is an absolutely continuous function that satisfies*

$$q(t) = q(0) + \int_0^t v(\tau)d\tau, \quad \text{for} \quad t \in [0,T]. \tag{1.4.13}$$

3. *The measure* $dv(t)$ *must satisfy*

$$M\frac{dv}{dt} - k(q,v) \in \mathcal{FC}(q). \tag{1.4.14}$$

4. $\Theta^{(i)}(q) = 0$, $i = 1, ..., m$
5. $\Phi^{(j)}(q) \geq 0$, $j = 1, ..., p$.

In Chapter 2, we will look at time-stepping schemes used to numerically integrate rigid body systems and analyze the convergence of such schemes to weak solutions of MDIs.

Chapter 2

Contributions

In this chapter we present our research contributions. We focused our attention first on contributions to rigid body simulation and control. We deal with both theoretical as well as practical aspects of rigid body simulation. We present the convergence results obtained in [21] for a class of linearly implicit time-stepping schemes. The convergence analysis, which is presented in detail in [21], follows the same lines as the one given by David Stewart in [44]. The main difference consists in the fact that we allow joint constraints besides contact and frictional constraints. The concept of a *reduced* measure differential inclusion was also introduced.

We continue with a rather different problem, that of simulating large rigid body systems. The LCP formulation is replaced by a quadratic program (QP) in order to circumvent the computational difficulties arising from the copositive LCPs that are used in standard frictional time-stepping schemes.

Simulation and control of a rigid body system in a quasi-static setting is addressed next. The main challenge that we faced while dealing with this problem consisted in performing robust planning and control under uncertainties for a system described at the meso-scale level (the dimensions of the system are at a milimetric scale). The mathematical analysis together with the resulting simulation code were used in a real-life application that is also briefly described here.

We end the chapter with other contributions. These include some contributions to image processing, more precisely to optical flow estimation and mathematical inequalities with applications mainly in probability and statistics.

2.1 A family of linearly implicit schemes

We present a family of time-stepping schemes that was analyzed in [21], mainly from the convergence point of view. These linearly implicit time stepping schemes accommodate methods based on semi-implicit Euler methods [6,43] as well as various instances of the trapezoidal method from [40]. The time-stepping scheme solves at each non-collisional integration step a linear complementarity problem. We will discuss a Poisson collision resolution model, which involves a compression phase followed by a decompression phase. Therefore, for a partially inelastic collision two LCPs need to be solved, one corresponding to the compression phase, the other one to the decompression phase. For the convergence analysis however, only inelastic collisions will be considered.

In what follows, t_0 denotes the initial simulation time, $h > 0$ is the integration step or time-step and for $l = 0, 1, ..., N$, $t_l = t_0 + lh$ are the time instances where approximations of the configuration and the velocity are obtained. More precisely, we denote by q^l the numerical approximation of the generalized position q at time t_l and by v^l the numerical approximation for the generalized velocity at t_l.

To write the integration step as a *mixed linear complementarity* problem, we use the following approximations. The joint constraints are written at the velocity level (1.3.13) and approximated by

$$\left(\nu^{(i)}(q^l)\right)^T \left(\alpha v^{l+1} + (1-\alpha)v^l\right) = 0, \quad i = 1, ..., m,$$

where α is a scalar parameter, $\alpha \in (0, 1]$.

The nonpenetration and frictional constraints are approximated in the same fashion. We can write these as the following complementarity conditions

$$\begin{aligned}
&0 \le \rho^{(j),l+1} := \left(n^{(j)}(q^l)\right)^T \left(\alpha v^{l+1} + (1-\alpha)v^l\right) \perp c_n^{(j),l+1} \ge 0, \quad j \in \mathcal{A},\\
&0 \le \sigma^{(j),l+1} := \lambda^{(j),l+1} e^{(j)} + \left(D^{(j)}(q^l)\right)^T \left(\alpha v^{l+1} + (1-\alpha)v^l\right) \perp \beta^{(j),l+1} \ge 0, \quad j \in \mathcal{A},\\
&0 \le \zeta^{(j),l+1} := \mu^{(j)} c_n^{(j),l+1} - e^{(j)^T} \beta^{(j),l+1} \perp \lambda^{(j),l+1} \ge 0, \quad j \in \mathcal{A}\,.
\end{aligned}$$

Here $e^{(j)}$ is a vector, of dimension $m_C^{(j)}$, whose every entry is 1. We recall that $m_C^{(j)}$ is the number of facets used in the polyhedral approximation of the full friction cone (see (1.1.6) for more details). The equations of motion in implicit form can be written as

$$M\left(v^{l+1} - v^l\right) - z^{l+1} = hk(t_{l+1}, q^{l+1}, v^{l+1}). \tag{2.1.1}$$

Here M is the mass matrix, which is assumed to be a constant symmetric positive definite matrix, z^{l+1} represent the contact and joint impulses, and $k(t_{l+1}, q^{l+1}, v^{l+1})$ are the inertial and applied forces acting at time t_{l+1}. Since the goal is to formulate

the integration step as a linear complementarity problem, we will linearize (2.1.1) as follows. The term

$$z^{l+1} = \widetilde{\nu}(q^{l+1})\widetilde{c}_\nu^{l+1} + \widetilde{n}(q^{l+1})\widetilde{c}_n^{l+1} + \widetilde{D}(q^{l+1})\widetilde{\beta}^{l+1}$$

with the first term accounting for the joint impulses, the second one accounting for normal impulses and the third term accounting for frictional impulses is replaced by

$$z^{l+1} = \widetilde{\nu}^l\widetilde{c}_\nu^{l+1} + \widetilde{n}^l\widetilde{c}_n^{l+1} + \widetilde{D}^l\widetilde{\beta}^{l+1},$$

where $\widetilde{\nu}^l = \widetilde{\nu}(q^l)$, $\widetilde{n}^l = \widetilde{n}(q^l)$ and $\widetilde{D}^l = \widetilde{D}(q^l)$. To linearize the term $k(t_{l+1}, q^{l+1}, v^{l+1})$ in (2.1.1), we first introduce the position update formula. The position update at time t_{l+1} is given by the formula

$$q^{l+1} = q^l + h\left((1-\alpha)v^l + \alpha v^{l+1}\right).$$

For the term $k(t_{l+1}, q^{l+1}, v^{l+1})$ we have

$$\begin{aligned} k(t_{l+1}, q^{l+1}, v^{l+1}) &= f_C(v^{l+1}) + k_1(t_{l+1}, q^{l+1}, v^{l+1}) \\ &= F(v^{l+1})v^{l+1} + k_1(t_{l+1}, q^{l+1}, v^{l+1}), \end{aligned}$$

where $f_C(v^{l+1}) = F(v^{l+1})v^{l+1}$ are the Coriolis forces and $k_1(t_{l+1}, q^{l+1}, v^{l+1})$ are the external forces. The matrix $F(v)$ appearing in the definition of the Coriolis forces $f_c(v)$ is an antisymmetric matrix. We replace the Coriolis term by

$$F(v^{l+1})v^{l+1} \approx F(v^l)\left((1-\alpha)v^l + \alpha v^{l+1}\right) = F(v^l)v^l + \alpha F(v^l)\left(v^{l+1} - v^l\right). \tag{2.1.2}$$

The term $k_1(t_{l+1}, q^{l+1}, v^{l+1})$ is approximated as follows:

$$\begin{aligned} k_1(t_{l+1}, q^{l+1}, v^{l+1}) &\approx (1-\alpha)k_1(t_l, q^l, v^l) + \alpha k_1(t_{l+1}, q^{l+1}, v^{l+1}), \\ &\approx (1-\alpha)k_1(t_l, q^l, v^l) + \alpha k_1(t_{l+1}, q^l, v^l) \\ &+ \alpha\left(\widetilde{k}_{1q}(q^{l+1} - q^l) + \widetilde{k}_{1v}^l(v^{l+1} - v^l)\right), \\ &\approx (1-\alpha)k_1(t_l, q^l, v^l) + \alpha k_1(t_{l+1}, q^l, v^l) + \alpha h\widetilde{k}_{1q}^l v^l \\ &+ \alpha\left(\widetilde{k}_{1v}^l + \alpha h\widetilde{k}_{1q}^l\right)(v^{l+1} - v^l), \end{aligned} \tag{2.1.3}$$

where

$$\widetilde{k}_{1q}^l \approx k_{1q}(t_{l+1}, q^l, v^l) \quad \widetilde{k}_{1v}^l \approx k_{1v}(t_{l+1}, q^l, v^l)$$

are approximations of the Jacobians k_{1q} and k_{1v}. Combining the equations of motion with the joint constraints described at the velocity level and the frictional contact constraints, we obtain the following time-stepping scheme:

$$q^{l+1} = q^l + h\left((1-\alpha)v^l + \alpha v^{l+1}\right) \tag{2.1.4a}$$

$$\widetilde{M}^l v^{l+1} - \sum_{i=1}^{m} \nu^{(i),l} c_\nu^{(i),l+1} - \sum_{j\in\mathcal{A}} (n^{(j),l} c_n^{(j),l+1} + D^{(j),l}\beta^{(j),l+1}) = \widetilde{M}^l v^l + \widetilde{k}^l \tag{2.1.4b}$$

$$\left(\nu^{(i),l}\right)^T \left(\alpha v^{l+1} + (1-\alpha)v^l\right) = 0, \tag{2.1.4c}$$

$$0 \le \rho^{(j),l+1} := \left(n^{(j),l}\right)^T \left(\alpha v^{l+1} + (1-\alpha)v^l\right) \perp c_n^{(j),l+1} \ge 0, \tag{2.1.4d}$$

$$0 \le \sigma^{(j),l+1} := \lambda^{(j),l+1} e^{(j)} + \left(D^{(j),l}\right)^T \left(\alpha v^{l+1} + (1-\alpha)v^l\right) \perp \beta^{(j),l+1} \ge 0, \tag{2.1.4e}$$

$$0 \le \zeta^{(j),l+1} := \mu^{(j)} c_n^{(j),l+1} - {e^{(j)}}^T \beta^{(j),l+1} \perp \lambda^{(j),l+1} \ge 0, \quad , \tag{2.1.4f}$$

where $\nu^{(i),l} = \nu^{(i)}(q^l)$, $n^{(j),l} = n^{(j)}(q^l)$, $D^{(j),l} = D^{(j)}(q^l)$, $j \in \mathcal{A}$, $i = 1, 2, \dots m$ and

$$\begin{aligned} \widetilde{M}^l &= \left(M - \alpha h \left(F(v^l) + \widetilde{k}_{1v}^l\right) - \alpha^2 h^2 \widetilde{k}_{1q}^l\right), \\ \widetilde{k}^l &= h\left((1-\alpha)k_1\left(t_l, q^l, v^l\right) + \alpha k_1\left(t_{l+1}, q^l, v^l\right)\right) + (1-\alpha)hF(v^l)v^l + \alpha h^2 \widetilde{k}_{1q}^l v^l. \end{aligned} \tag{2.1.5}$$

We note that the equations (2.1.4) represent a mixed linear complementarity problem (MLCP). Here we have considered the active set as given in (2.4.17) and assumed that $\widetilde{n}^l$, $\widetilde{c}_n^l$, $\widetilde{D}^l$, and $\widetilde{\beta}^l$ account only for the active constraints. More precisely, we have

$$\begin{array}{ll} \widetilde{n} = [n^{(j_1)}, n^{(j_1)}, \dots, n^{(j_{n_a})}], & \widetilde{c}_n = \lceil c_n^{(j_1)}, c_n^{(j_2)}, \dots, c_n^{(j_{n_a})} \rfloor, \\ \widetilde{D} = [D^{(j_1)}, D^{(j_2)}, \dots, D^{(j_{n_a})}], & \widetilde{\beta} = \lceil \beta^{(j_1)}, \beta^{(j_2)}, \dots, \beta^{(j_{n_a})} \rfloor, \end{array} \tag{2.1.6}$$

for an active set $\mathcal{A} = \{j_1, j_2, \dots, j_{n_a}\}$. In (2.1.6), we have omitted the additional superscript "l", which refers to the fact that all quantities are computed for the configuration q^l, in order to simplify notation. In the same fashion we define the following block-matrices:

$$\begin{array}{llll} \widetilde{\lambda} = [\lambda^{(j_1)}, \lambda^{(j_2)}, \dots, \lambda^{(j_{n_a})}], & \widetilde{\zeta} = [\zeta^{(j_1)}, \zeta^{(j_2)}, \dots, \zeta^{(j_{n_a})}], & \widetilde{E} = \mathbf{diag}(e^{(j_1)}, e^{(j_2)}, \dots, e^{(j_{n_a})}) \\ \widetilde{\sigma} = [\sigma^{(j_1)}, \sigma^{(j_2)}, \dots, \sigma^{(j_{n_a})}], & \widetilde{\rho} = [\rho^{(j_1)}, \rho^{(j_2)}, \dots, \rho^{(j_{n_a})}], & \widetilde{\mu} = \mathbf{diag}(\mu^{(j_1)}, \mu^{(j_2)}, \dots, \mu^{(j_{n_a})}), \end{array} \tag{2.1.7}$$

We can rewrite this MLCP in the matrix form:

$$\begin{bmatrix} \widetilde{M}^l & -\widetilde{\nu}^l & -\widetilde{n}^l & -\widetilde{D}^l & 0 \\ \left(\widetilde{\nu}^l\right)^T & 0 & 0 & 0 & 0 \\ \left(\widetilde{n}^l\right)^T & 0 & 0 & 0 & 0 \\ \left(\widetilde{D}^l\right)^T & 0 & 0 & 0 & \widetilde{E} \\ 0 & 0 & \widetilde{\mu} & -\widetilde{E}^T & 0 \end{bmatrix} \begin{bmatrix} v^{l+1} \\ \widetilde{c}_\nu^{l+1} \\ \widetilde{c}_n^{l+1} \\ \widetilde{\beta}^{l+1} \\ \widetilde{\lambda}^{l+1} \end{bmatrix} - \begin{bmatrix} \widetilde{M}^l v^l + \widetilde{k}^l \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \widetilde{\rho}^{l+1} \\ \widetilde{\sigma}^{l+1} \\ \widetilde{\zeta}^{l+1} \end{bmatrix} \tag{2.1.8}$$

$$0 \le \lceil \widetilde{c}_n^{l+1}, \widetilde{\beta}^{l+1}, \widetilde{\lambda}^{l+1} \rfloor \perp \lceil \widetilde{\rho}^{l+1}, \widetilde{\sigma}^{l+1}, \widetilde{\zeta}^{l+1} \rfloor \ge 0\,. \tag{2.1.9}$$

We denote by

$$\mathcal{L}(q^l, v^l, \widetilde{k}, h, \alpha) \tag{2.1.10}$$

the solution set of this MLCP. It can be noticed that the MLCP (2.1.8) has the same structure as the one given by (1.2.12). In [21], it was shown that for sufficiently small values of the time-step h, the solvability of (2.1.8) is guaranteed. This was obtained by using the result of Theorem 1.2.10. Some of the previously obtained time-stepping schemes can be derived directly from the MLCP (2.1.8) for different values of α. Thus, the choice $\alpha = 1$ results in the scheme from [6, 7], while the choice $\alpha = \frac{1}{2}$ results in a variant of the scheme from [40] .

We look now at active set selection, collision detection and resolution. We recall that the set of active contacts is denoted by $\mathcal{A}$. The active set at time t_{l+1}, which we denote by $\mathcal{A}^{l+1}$ or $\mathcal{A}(t_{l+1})$, may be defined as

$$\mathcal{A}(t_{l+1}) = \mathcal{A}^{l+1} = \left\{ j : \ \Phi^{(j)}(q^l) \leq 0 \right\} \tag{2.1.11}$$

A slightly different active set selection was used in [40]. This was designed to preserve the second order convergence of the trapezoidal like method presented in this work. For now, we will consider the simpler strategy given by (2.1.11). Whenever a collision is encountered, cubic interpolation is used to determine the pre-collision velocity and the position at which the collision occurs, [40]. We denote by $t* := t^{*,l+1} \in (lh, (l+1)h]$ the detected collision time and by $q^- := q^{-,l+1}$ and $v^- := v^{-,l+1}$ the detected collision position and velocity, respectively.

The collision is modeled through the Poisson restitution model which is briefly described below. The Poisson restitution model is a two-phase model composed of a compression phase and a decompression phase. In the first phase, interpenetration is prevented by normal compression contact impulses, while in the latter a fraction of each normal compression contact impulse is restituted to the system (the Poisson hypothesis [39]).

The detected position q^- and the pre-collision velocity v^- are used in the compression phase. In the compression phase, the only impulses acting on the system are the constraint impulses generated by joints, contacts or friction. We denote the impulses by the same symbols as before but with superscript c. Let v^c be the velocity at the end of the compression phase. At the end of this phase, each contact from the list is either maintained $c_n^c \geq 0, n^T v^c = 0$, or is breaking, $c_n^c = 0, n^T v^c \geq 0$. By adding the complementarity conditions given by joints and friction we can formulate this collision phase in terms of the following MLCP, [6],

$$M(v^c - v^-) - \sum_{i=1}^{m} \nu^{(i)} c_\nu^{c(i)} - \sum_{j \in \mathcal{A}} (n^{(j)} c_n^{c(j)} + D^{(j)} \beta^{c(j)}) = 0 \tag{2.1.12}$$

$$\nu^{(i)T} v^c = 0, \quad i = 1, 2, \ldots, m \tag{2.1.13}$$

$$0 \leq \rho^{c(j)} := n^{(j)T} v^c \perp c_n^{c(j)} \geq 0, \quad j \in \mathcal{A} \tag{2.1.14}$$

$$0 \leq \sigma^{c(j)} := \lambda^{c(j)} e^{(j)} + D^{(j)T} v^c \perp \beta^{c(j)} \geq 0, \quad j \in \mathcal{A} \tag{2.1.15}$$

$$0 \leq \zeta^{c(j)} := \mu^{(j)} c_n^{c(j)} - e^{(j)^T} \beta^{c(j)} \perp \lambda^{c(j)} \geq 0, \quad j \in \mathcal{A} \tag{2.1.16}$$

The joint constraint gradients $\nu^{(j)}$ as well as the contact data $n^{(j)}$, $D^{(j)}$ are evaluated at the pre–collision position q^-. The MLCP (2.1.12)–(2.1.16) has the same structure as the one in Theorem 1.2.10 and therefore its solvability is guaranteed. We can easily see that the structure of the compression phase is similar to the general integration step (2.1.8). It is also clear that the solution set of (2.1.12–2.1.16) can be expressed as $\mathcal{L}(q^-, v^-, 0, 0, 1)$, following the solution set definition (2.1.10).

For each contact (j) the *decompression phase* uses restitution coefficients $e_j \in [0, 1]$ to generate a restitution impulse F^r based on the normal compression impulse $c_n^{c(j)}$. More precisely,

$$F^r = \sum_{j=1}^{p} e_j n^{(j)} c_n^{c(j)}. \tag{2.1.17}$$

Each active contact generates a normal decompression impulse $c_n^{d(j)}$ that is composed of the restituted normal impulse $e_j c_n^{c(j)}$ and an additional impulse $c_n^{x(j)} \geq 0$ needed to prevent interpenetration. That is $c_n^{d(j)} = e_j c_n^{c(j)} + c_n^{x(j)}$. Let $v^+ := v^{+,(l+1)}$ be the velocity after the decompression phase, or the post–collision velocity. At the end of the decompression phase contact j either breaks, and then $n^{(j)T} v^+ \geq 0$, $c_n^{x(j)} = 0$, or it is maintained and then $n^{(j)T} v^+ = 0$, $c_n^{x(j)} \geq 0$. Combining this observation with the constraints given by joints and friction we have, as in the compression phase, that the post–collision velocity and the new constraint impulses are obtained as the solution of a MLCP similar to the compression phase MLCP. More precisely, in the decompression phase, we solve the following MLCP:

$$M(v^+ - v^c) - \sum_{i=1}^{m} \nu^{(i)} c_\nu^{x(i)} - \sum_{j \in \mathcal{A}} (n^{(j)} c_n^{x(j)} + D^{(j)} \beta^{x(j)}) = F^r \tag{2.1.18}$$

$$\nu^{(i)T} v^+ = 0, \quad i = 1, 2, \ldots, m \tag{2.1.19}$$

$$0 \leq \rho^{x(j)} := n^{(j)T} v^+ \perp c_n^{x(j)} \geq 0, \quad j \in \mathcal{A} \tag{2.1.20}$$

$$0 \leq \sigma^{x(j)} := \lambda^{x(j)} e^{(j)} + D^{(j)T} v^+ \perp \beta^{x(j)} \geq 0, \quad j \in \mathcal{A} \tag{2.1.21}$$

$$0 \leq \zeta^{x(j)} := \mu^{(j)} c_n^{x(j)} - e^{(j)^T} \beta^{x(j)} \perp \lambda^{x(j)} \geq 0, \quad j \in \mathcal{A} \tag{2.1.22}$$

Since (2.1.18–2.1.22) has the same structure as the other MLCPs above, its solvability is guaranteed and its solution set is given by $\mathcal{L}^c(q^-, v^c, 0, F^r, 1)$.

In general the post–collision kinetic energy produced following the decompression phase is not guaranteed to be smaller than the pre–collision kinetic energy. A particular case for which the dissipative property holds is given in [6]. Since, we want to obtain convergence results for this class of time-stepping schemes and since the dissipative property is essential in proving this, we will consider only inelastic collisions. In this case, only the compression phase needs to be solved and the post-collision velocity v^+ is equal to the velocity obtained in the compression phase, i.e., $v^+ = v^c$.

Before introducing the convergence results of [21], we make some conventions regarding collisions. From now on we will consider only inelastic collisions (the restitution coefficients are 0) and therefore only the compression phase needs to be solved. It is clear that collision detection usually results in a nonuniform partition of the simulation interval $[0, T]$. More precisely, a collision may be detected at time t^* such that, for a given time-step h, $t^* \neq lh$ for any integer l. When collision is detected at time $t^{*,l+1} \in (lh, (l+1)h]$, the collision is solved, resulting in the collision position $q^{-,l+1}$ and postcollision velocity $v^{+,l+1} \in \mathcal{L}(q^{-,l+1}, v^{-,l+1}, 0, 0, 1)$. Instead of introducing the collision time t^* in the time partition of $[0, T]$ or solving another MLCP in the interval $(t^*, (l+1)h]$, we take

$$t_{l+1} = (l+1)h, \quad q^{l+1} = q^{-,l+1} \text{ and } v^{l+1} = v^{+,l+1}.$$

Assuming that we do this for every collision and that the first integration step is not a collisional one, we have $t_l = lh$, for all l, and the scheme will keep a fixed time-step throughout the integration process.

2.2 Convergence results

In this section, we briefly present the convergence results for the class of linearly implicit time-stepping schemes introduced in the section above. The convergence results were obtained in [21] and we refer the reader to this work for details. The convergence analysis of [21], follows the same lines as the ones given by David Stewart in [44]. The main difference consists in the fact that we allow joint constraints besides contact and frictional constraints. The challenge consists in the fact that impulsive forces may be transmitted to the joints, which means that the joint forces need to be treated in the same way as the contact and frictional forces. We give a simple example of how this can occur.

Consider two particles of equal mass m linked by a massless rod. Assume that the only forces acting on the particles are gravitational forces. As the particles fall they encounter a horizontal frictionless table-top, see Figure 2.1. We assume that the collision between each of the particles and the table-top is inelastic, i.e., no normal velocity is restituted after collision. Then whenever one particle is hitting the table with negative normal velocity a collision takes place, and therefore an impulsive normal reaction force is needed in order to bring the normal velocity from a negative value to zero. If the configuration of the system assumes the form given in Figure 2.1, then the impulsive reaction force will have an impulsive component acting along the joint direction, which results in an impulsive Lagrange multiplier (joint force).

It is conceivable that a proof of convergence for the joint constraint case can be obtained from the one in the jointless case if the system is represented in relative

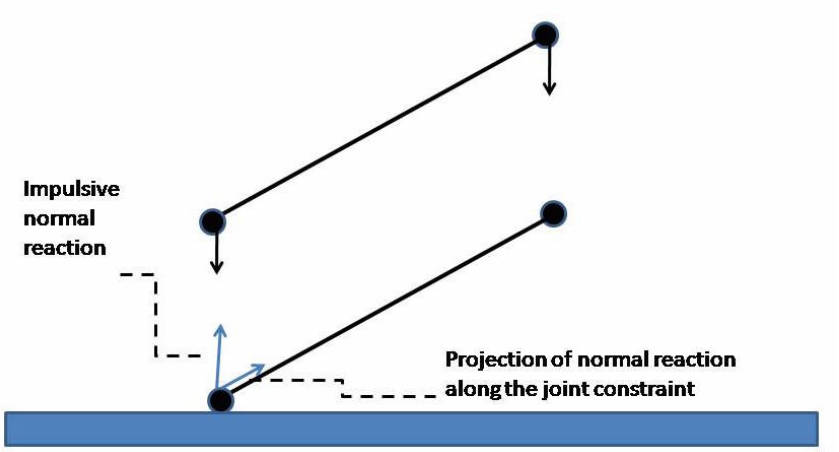


Figure 2.1: Impulsive normal forces are transmitted to the joints.

coordinates which eliminate the joint constraints. Nonetheless, this requires a nonlinear projection at every step, which may be computationally costly. This is the main motivation besides the convergence analysis done in [21].

One of the main ingredients used in the convergence analysis is the pointedness of the total friction cone. To reduce the analysis to a joint-free setting, a *reduced friction cone* is considered, for which the pointedness property can be transferred from the total cone. The reduced friction cone is used to formulate a reduced measure differential inclusion. All of these, together with the assumptions used in obtaining the convergence results are explained next.

2.2.1 Pointedness of the friction cone

The *pointed friction cone* assumption used in [21] is an extension of the one used in [45] for the jointless case. This assumption is one of the main ingredients in proving the convergence results. We recall the definition of the total friction cone which was given in (1.4.1).

Definition 2.2.1. *We say that*

$$\begin{aligned} &\mathcal{FC}(q) \text{ is pointed} \Leftrightarrow \forall \left(c_\nu, c_n \geq 0, \tilde{\beta} \geq 0\right) \neq 0 \text{ such that } \|\beta^{(j)}\|_1 \leq \mu^{(j)} c_n^{(j)}, \\ &\forall j \in \mathcal{A} \text{ we must have that } \tilde{\nu} c_\nu + \tilde{n} c_n + \tilde{D}\tilde{\beta} \neq 0. \end{aligned} \tag{2.2.1}$$

We note that the above definition clearly implies that the joint–constraint matrix $\tilde{\nu}$ is full rank. We also note that the pointed friction cone assumption is weaker than the linear independence of the columns of the matrix $\left(\tilde{\nu}^T,\ \tilde{n}^T,\ \tilde{D}^T\right)^T$. Its

name originates in the fact that, when there are no joint constraints, the condition is equivalent to the cone not containing any proper linear subspace, and thus being "pointed". An equivalent definition of the friction cone assumption is given by the following condition, [2]:

$$\begin{aligned} \mathcal{FC}(q) \text{ is pointed} \Leftrightarrow \text{there exists } c_{\mathcal{FC}} > 0, \text{ such that } \|(c_\nu, c_n, \tilde{\beta})\| \leq c_{\mathcal{FC}}\|z\| \\ \text{with } z = \tilde{\nu} c_\nu + \tilde{n} c_n + \tilde{D}\tilde{\beta} \in \mathcal{FC}(q). \end{aligned} \tag{2.2.2}$$

We will say that the total friction cone is *uniformly pointed* if the constant $c_{\mathcal{FC}}$ in condition (2.2.2) can be taken the same for all possible configurations q. The uniform pointedness of the total friction cone will be used in obtaining a uniform bound for the numerical velocities.

The reduced friction cone $\mathcal{FC}_r(q)$ represents one of the main conceptual novelties of [21]. For the system position q, the reduced friction cone is given by

$$\mathcal{FC}_r(q) = \widetilde{\nu}_\perp^T \mathcal{FC}(q) = \left\{ \widetilde{\nu}_\perp^T z \ : \ z \in \mathcal{FC}(q) \right\}. \tag{2.2.3}$$

Here $\widetilde{\nu}_\perp := \widetilde{\nu}_\perp(q)$ denotes the orthogonal complement of $\widetilde{\nu} = \widetilde{\nu}(q) \in R^{s \times m}$. More precisely, $\widetilde{\nu}_\perp \in R^{s \times (s-m)}$ such that $\widetilde{\nu}_\perp^T \widetilde{\nu} = 0$ and $\widetilde{\nu}_\perp^T \widetilde{\nu}_\perp = I$. It follows that for any $x \in R^s$, there exist unique vectors $u \in R^m$ and $w \in R^{s-m}$ such that the decomposition

$$x = \widetilde{\nu} u + \widetilde{\nu}_\perp w \tag{2.2.4}$$

holds.

2.2.2 The reduced MDI and convergence

In [21], it was shown that the uniform pointedness of the total friction cone $\mathcal{FC}(q)$ implies the uniform pointedness of the reduced cone $\mathcal{FC}_r(q)$. We use the decomposition (2.2.4) to decompose the generalized velocity

$$v = \widetilde{\nu}(q) u + \widetilde{\nu}_\perp(q) w.$$

Since the joint constraints at the velocity level require the velocity v to satisfy $\left(\widetilde{\nu}(q)\right)^T v = 0$ and since $\widetilde{\nu}(q)$ is full rank, one must have $u = 0$ in the above decomposition and therefore,

$$v = \widetilde{\nu}_\perp(q) w, \ \ dv = d\left(\widetilde{\nu}_\perp(q) w\right),$$

which also gives the measure dw. This way, we are led to the concept of a *reduced measure differential inclusion* and *reduced weak solutions*.

Definition 2.2.2 (Reduced Weak Solution of (1.4.4–1.4.10))**.** *We say that $q(t)$, $w(t)$ is a reduced weak solution of (1.4.4–1.4.10) on $[0, T]$ if*

1. $w(\cdot)$ *is a function of bounded variation on* $[0,T]$.

2. $q(\cdot)$ *is an absolutely continuous function that satisfies*

$$q(t) = q(0) + \int_0^t \widetilde{\nu}_\perp(q(\tau))w(\tau)d\tau, \quad \text{for } \ t \in [0,T]. \tag{2.2.5}$$

3. *The measure* $dw(t)$ *must satisfy*

$$\left(\left(\widetilde{\nu}_\perp(q)\right)^T M\widetilde{\nu}_\perp(q)\right)\frac{dw}{dt} - k_{w,\perp}(t,q,w) \in \mathcal{FC}_r(q), \tag{2.2.6}$$

where

$$k_{w,\perp}(t,q,w) = \left(\widetilde{\nu}_\perp(q)\right)^T k_w(t,q,w) \tag{2.2.7}$$

and

$$k_w(t,q,w) = k(t,q,\widetilde{\nu}_\perp(q)w) - M\left(\left(\frac{\partial}{\partial q}\left(\widetilde{\nu}_\perp(q)w\right)\right)\widetilde{\nu}_\perp(q)w\right) \tag{2.2.8}$$

4. $\Phi^{(j)}(q) \geq 0$, $j = 1, ..., p$.

It can be easily noted that the idea behind the above Definition 2.2.2, resides in the decomposition (2.2.4). The connection between the reduced and the full MDIs is established by the following result (for a proof, see [21])

Lemma 2.2.3. *If* (q,w) *is a reduced weak solution of (1.4.4–1.4.10) on* $[0,T]$ *in the sense of Definition 2.2.2 and* $\Theta(q(0)) = 0$*, then* $(q,v) = (q,\widetilde{\nu}_\perp(q)w)$ *is a weak solution of (1.4.4–1.4.10) on* $[0,T]$ *in the sense of Definition 1.4.3*

In order to prove convergence, we need to extend the discrete numerical solution to time instants different from the computational moments t_l, $l = 0, 1, ..., N$. We will define a weighted velocity sequence $v^{h,\alpha}(t)$ and an other sequence $v^h(t)$ where the weighted factor is excluded. The velocity sequence $v^{h,\alpha}(t)$ is defined by

$$v^{h,\alpha}(t) = \begin{cases} v^{l+1,\alpha}, & t \in (lh,(l+1)h] \ \text{ and no collision in } (lh,(l+1)h], \\ v^{l+1} := v^{+,l+1}, & t \in (lh,(l+1)h] \ \text{ and collision in } (lh,(l+1)h], \end{cases} \tag{2.2.9}$$

where $v^{+,l+1}$ denotes the velocity at the end of the compression phase and where

$$v^{l+1,\alpha} = (1-\alpha)v^l + \alpha v^{l+1}. \tag{2.2.10}$$

The velocity function that uses no weighting is denoted by $v^h(\cdot)$ and defined in a similar fashion:

$$v^h(t) = \begin{cases} v^{l+1}, & t \in (lh,(l+1)h] \ \text{ and no collision in } (lh,(l+1)h], \\ v^{l+1} := v^{+,l+1}, & t \in (lh,(l+1)h] \ \text{ and collision in } (lh,(l+1)h]. \end{cases} \tag{2.2.11}$$

For the position sequence, we take $q^{h,\alpha}(t)$ to be

$$q^{h,\alpha}(t) = \frac{1}{h}\left((t-t_l)q^{(l+1)} + (t_{l+1}-t)q^l\right), \text{ whenever } t \in (t_l = lh, t_{l+1} = (l+1)h], \tag{2.2.12a}$$

where

$$q^{l+1} = \begin{cases} q^{(l)} + hv^{l+1,\alpha}, & t \in (lh,(l+1)h] \text{ and no collision in } (lh,(l+1)h], \\ q^{-,l+1}, & t \in (lh,(l+1)h] \text{ and collision in } (lh,(l+1)h]. \end{cases} \tag{2.2.12b}$$

Here q^{l+1} is computed by the position update formula (2.1.4a), except for collisional instants (that is, a collision occurred in the $(lh,(l+1)h]$ interval), in which case $q^{l+1} := q^{-,l+1}$, where $q^{-,l+1}$ is the estimated collision position. Since the collision time $t^{*,l+1}$ is detected by solving

$$\Phi^{(j)}(\widetilde{q}(t)) = 0,$$

where $\widetilde{q} : [lh,(l+1)h] \to R^s$ is the cubic interpolant of the data $\widetilde{q}(lh) = q^l$, $\frac{d\widetilde{q}}{dt}(lh) = v^l$, $\widetilde{q}((l+1)h) = \overline{q}^{l+1}$, $\frac{d\widetilde{q}}{dt}((l+1)h) = \overline{v}^{l+1}$ ($\overline{q}^{l+1}$ and $\overline{v}^{l+1}$ are obtained by applying a regular step with $j \notin \mathcal{A}$) and $q^{l+1} = q^{-,l+1} = \widetilde{q}(t^{*,l+1})$, it can be guaranteed, [21], that

$$\Phi^{(j)}(q^{l+1}) = \Phi^{(j)}(q^{-,l+1}) \geq -C_c h^2.$$

for a fixed constant C_c.

In [21], in order to prove convergence results for the time-stepping scheme introduced in the previous section, a set of assumptions was used. We list these assumptions here and give a brief explanation for every item in the list.

(H1) The nonpenetration constraints are twice-continuously differentiable, and there exists B_H such that

$$\|\nabla_{qq}\Phi^{(j)}(q)\| \leq B_H, \text{ for all } q \text{ and } j = 1,...,p. \tag{2.2.13}$$

(H2) The functions $\Theta^{(i)}(q)$, $i = 1,...,m$ are sufficiently smooth functions.

(H3) The generalized mass matrix, M, is constant, symmetric, and positive definite.

(H4) The total friction cone $\mathcal{FC}(q)$ is uniformly pointed with respect to all configurations q.

(H5) The norm of the external force increases at most linearly with the position and the velocity. That is,

$$\|k_1(t,q,v)\| \leq c_1 + c_2\|q\| + c_3\|v\|. \tag{2.2.14}$$

Here $k_1(q,v)$ denotes the external and inertial forces.

The Coriolis force is given by a bilinear operator

$$[f_C(v)]_i = \sum_{jk} f_{ijk} v_j v_k \ .$$

This is certainly true if the system is described by Newton-Euler equations in body coordinates [33, Section 2.4], where the matrix $F(v)$ of entries

$$[F(v)]_{ij} = \sum_k f_{ijk} v_k$$

is antisymmetric in the sense that

$$u^T F(v) u = 0, \quad \forall u \ .$$

We also assume that the approximations $\tilde{k}_{1q}$ and $\tilde{k}_{1v}$ are bounded. More precisely,

$$\|\widetilde{k}_{1q}\| \leq c_4, \ \ \|\widetilde{k}_{1v}\| \leq c_5. \tag{2.2.15}$$

(H6) The contact data given by $\tilde{n}(q)$, $\tilde{D}(q)$ are globally Lipschitz continuous functions.

(H7) The number of collisions solved by the algorithm is uniformly upper bounded as $h \to 0$.

(H8) The external forces $k_1(t,q,v)$ are linear in v, and the approximation $\widetilde{k}_{1v}$ is constant.

Below we briefly explain the assumption listed here.

- Assumptions **(H1)**, **(H2)**, **(H5)** and **(H6)** are stronlgy connected to the smoothness of the contact data as well as the properties of the external and Coriolis forces. These are fairly standard assumptions.

- Assumption **(H3)** is satisfied when the Newton-Euler formulation in body coordinates is used, [33].

- Assumption **(H4)** was one of the central points in our analysis. We note, that this implies that $\widetilde{\nu}(q)$ has uniform full rank. More precisely, there exists a constant $\kappa > 0$ such that

$$\sigma_{\min}(\widetilde{\nu}(q)) \geq \kappa \quad \forall\, q,$$

  where $\sigma_{min}(A)$ denotes the smallest singular value of the matrix A.

- Assumption **(H7)** is related to and implied by the assumption of non-Zeno behavior of the system. The non-Zeno behavior means that the number of switching points in the dynamics (collisions and stick-slip transitions) is finite in any bounded time interval. While this is not unreasonable to expect when the restitution coefficient is 0 (as we assume here), it is not hard to find examples with partially elastic collisions (the restitution coefficient belongs to $(0, 1)$), where the non-Zeno property cannot be guaranteed.

- Assumption **(H8)** imposes some restrictions on the dependence on velocity of the external forces. The first part of assumption **(H8)** is quite standard in stability analysis, while the second part is not needed to prove all the results. More precisely, uniform boundedness of the numerical velocities as well as a uniform bound on the variation of the numerical velocities can be obtained without this assumption. We note that this assumption is satisfied when external forces include linear damping terms, by far the prevailing type of external velocity-dependent passive force.

To obtain the convergence result, we first proved a uniform bound on the numerical velocities. The result below is practically proving the existence of such a bound, when the parameter α belongs to the interval $\left[\frac{1}{2}, 1\right]$. For more details, we refer the reader to [21].

Theorem 2.2.4. *If* **(H1)–(H8)** *are satisfied and* $\dfrac{1}{2} \leq \alpha \leq 1$*, then there is a constant* c *such that*

$$(v^l)^T M v^l \leq \max\left\{(v^0)^T M v^0,\ \|q^0\| + 1\right\} e^{ct_l}\,,\ \ l = 0, 1, \ldots, \lfloor T/h \rfloor\,,$$

for all sufficiently small h *(here* $\lfloor x \rfloor$ *denotes the integer part of* x*)* .

The kinetic energy estimate above together with assumption **(H3)** gives a uniform bound for the discrete numerical velocities v^l. We also note that the above theorem implies that both $v^h(\cdot)$ and $v^{h,\alpha}(\cdot)$ are uniformly bounded on $[0, T]$, as $h \to 0$.

Using assumption**(H7)**, which states that the number of collisions solved is uniformly upper bounded as $h \to 0$, we have that

$$q^{h,\alpha}(t) = q^{h,\alpha}(0) + \int_0^t v^{h,\alpha}(\tau)d\tau.$$

This together with the uniform boundedness of the velocities implies that the sequence $\left\{q^{h,\alpha}(\cdot)\right\}$ is equicontinuous and equibounded. Therefore by the Arzela-Ascoli theorem, there exists a uniformly convergent subsequence. Without any loss of generality, we also denote such a subsequence by $q^{h,\alpha}(\cdot)$ and write

$$q^{h,\alpha}(\cdot) \to q(\cdot)$$

uniformly in $[0, T]$.

To obtain a pointwise convergent velocity subsequence and a weak* convergent subsequence for the corresponding measures, we first obtain a uniform bound on the total variation of the numerical velocities $v^{h,\alpha}(\cdot)$. More precisely, we have the following result.

Theorem 2.2.5. *$\bigvee_0^T v^{h,\alpha}(\cdot)$ is uniformly bounded as $h \to 0$, and there exists $v^*(\cdot)$ of bounded variation such that $v^{h,\alpha} \to v^*$ pointwise and $dv^{h,\alpha} \to dv^*$ weakly.*

The proof of Theorem 2.2.5, given in [21], follows mainly the lines given in [44]. The main difference consists in the use of the reduced friction cone and the reduced measure differential inclusion. The uniform pointedness of the reduced friction cone which is induced by the uniform pointedness of the total cone $\mathcal{FC}(q)$ was the main ingredient in obtaining a uniform bound on the sums of normal impulses, $\sum_l \left|\left|\widetilde{c}_n^{l,h}\right|\right|$. This was then used in obtaining similar results for the other constraint impulses. We refer the reader to [21] for more details.

Using the uniform bound on the total variation $\bigvee_0^T v^{h,\alpha}(\cdot)$ given by Theorem (2.2.5) together with Helly's selection theorem the existence of a subsequence $v^{h_k,\alpha}(\cdot)$ of $v^{h,\alpha}(\cdot)$ was obtained.

Since the limiting velocity $v(t)$ may not be well defined for every $t \in [0, T]$, *we assume without loss of generality, [44], that $v(\cdot)$ is right–continuous, i.e. $v(t) = v^+(t)$ for all $t \in [0, T]$.* The corresponding functions $q^{h_k,\alpha}(\cdot)$ converge to the indefinite integral of $v(\cdot)$ by the pointwise convergence theorem for Lebesgue integrals. We assume for simplicity that this is the entire sequence and therefore $q^{h,\alpha}(\cdot) \to q(\cdot)$ uniformly and $v^{h,\alpha}(\cdot) \to v(\cdot)$ pointwise almost everywhere.

To obtain the weak* convergence from Theorem 2.2.5, we used the fact that $\bigvee_0^T v^{h,\alpha}(\cdot)$ is uniformly bounded as $h \to 0$, $v^{h,\alpha}(0) = v(0)$ and $v^{h,\alpha}(\cdot) \to v(\cdot)$ pointwise. These properties imply that $dv^{h,\alpha} \to dv$ weakly*, that is,

$$\int_0^T \phi(t)^T dv^{h,\alpha}(t) \to \int_0^T \phi(t)^T dv(t)$$

for all continuous functions $\phi(t)$. Therefore, $dv^{h,\alpha}(\cdot) \to dv(\cdot)$ weak* as Borel measures.

The final convergence result was obtained by showing that the limits $(q(\cdot), v(\cdot))$ represent a weak solution of the corresponding MDI, in the sense of Definition 1.4.3. This was done by proving that the reduced limiting trajectories $(q(\cdot), w(\cdot))$ represent a reduced weak solution in the sense of Definition 2.2.2 . Here $w(\cdot)$ is obtained from the decomposition

$$v = \widetilde{\nu}(q)u + \widetilde{\nu}_{\perp}(q)w,$$

while the corresponding measure dw is obtained from the resulting decomposition for dv. We end this section by stating the main convergence result, obtained in [21].

Theorem 2.2.6. *Assume that* $\alpha \in \left[\frac{1}{2}, 1\right]$ *and conditions* **(H1)–(H8)** *hold. Then there exists a subsequence* $h_k \to 0$ *such that*

1. $q^{h_k,\alpha}(\cdot) \to q(\cdot)$ *uniformly.*
2. $v^{h_k,\alpha}(\cdot) \to v(\cdot)$ *pointwise a.e.*
3. $dv^{h_k,\alpha}(\cdot) \to dv(\cdot)$ *weak * as Borel measures in [0,T], and every such subsequence converges to a solution* $(q(\cdot), v(\cdot))$ *of the measure differential inclusion (1.4.13–1.4.14).*

Therefore, $(q(t), v(t))$ *is a weak solution of our model.*

2.2.3 A numerical example

In this section we present a numerical example given in [21], which motivates our choice for the velocity sequence. This is a very simple example with stick–slip behavior. The rigid body system consist of a block of mass $m = 1$ subjected to an exterior force $k(t) = 8\cos(t)$ and sliding on a flat table with friction coefficient $\mu = 0.8$. The initial position of the block is $q_0 = (3, 0)^T$ and the initial velocity is $v_0 = (0, 0)^T$. The gravity $G = (0, -mg)^T$ is calculated with $g = 9.81$. We compare the weighted numerical velocity sequence $v^{h,\alpha}(t)$ to the sequence $v^h(t)$, for $\alpha = \frac{1}{2}$. The positions $q^{h,\alpha}(t_l)$ and velocities

$$v^l := v^h(t_l), v^{l,\alpha} := v^{h,\alpha}(t_l)$$

with $\alpha = \frac{1}{2}$ are shown in Figure 2.2.3, and they indicate a typical stick-slip behavior.

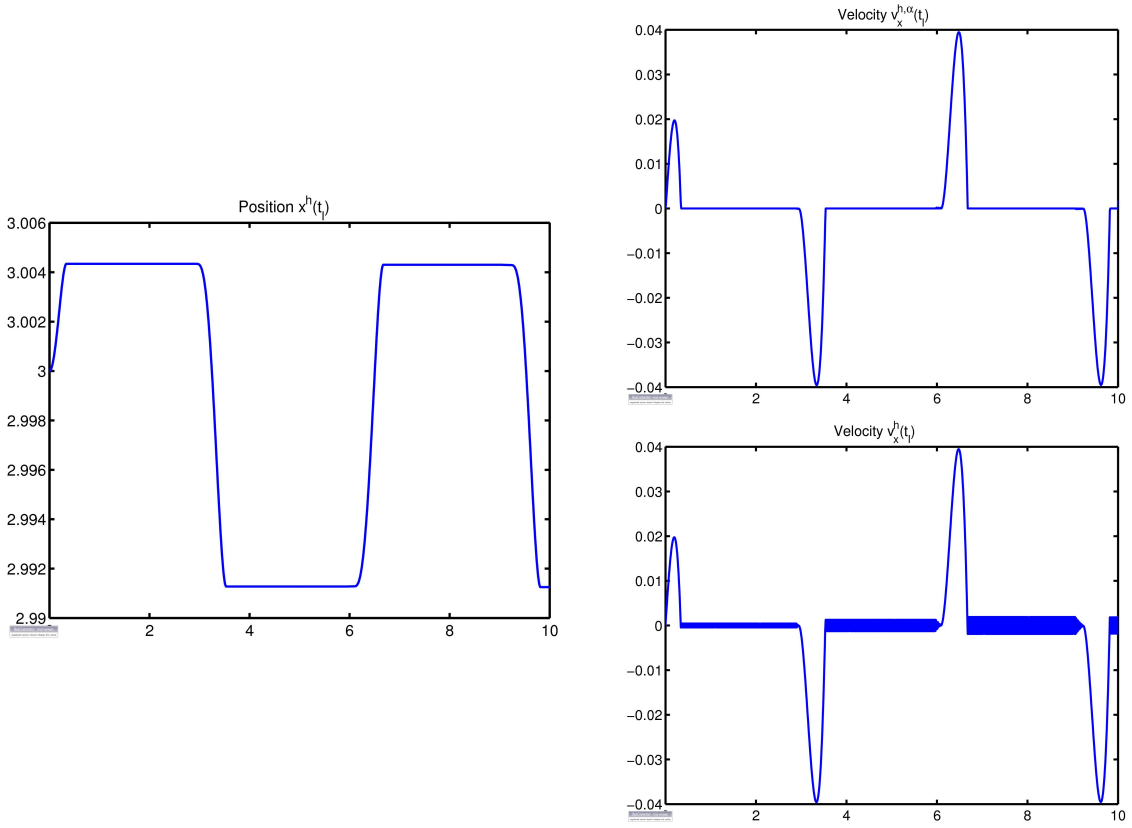


Figure 2.2: Numerical position and velocities for $T = 10$ (s), $\alpha = 1/2$ and $h = 0.01$. The plot on the left shows positions $q_1^{h,\alpha}(t_l)$, while the two plots on the right show the velocity sequences $v_1^{h,\alpha}(t_l)$ at the top and $v_1^h(t_l)$ at the bottom.

We note that the numerical velocities exhibit a quite different behavior. We see that, starting with the onset of sticking, the velocity sequence v^l exhibits oscillations that are not present in the sequence $v^{l,\alpha}$, which has the value 0 during the sticking phase. The example illustrates the difficulty in obtaining a good behavior of the total variation of the velocity solution v^l, as opposed to $v^{l,\alpha}$, and motivates our choice of the latter for our convergence result.

2.3 A large rigid-body system

In this section, we present a slightly different problem that was analyzed in [37]. The main challenge we faced while dealing with the system analyzed in [37] is of a computational nature. As we have seen above, LCP time-stepping schemes for frictional rigid body problems are formulated as LCPs with copositive matrices.

Such LCPs are generally solved by means of Lemke-type algorithms, which are pivotal type algorithms similar to the simplex algorithm from linear programming. For large systems, however, Lemke based solvers or any other pivotal algorithm become impractical from a computational point of view.

The computational difficulties associated with the (copositive) frictional LCP formulation cannot be avoided even for small friction coefficients. In [4], a simple example is used to show that the solution set of the time-stepping LCP fails to be convex for any nonzero friction coefficients. This implies that no polynomial time algorithms are known to exist for solving such problems. A *convex relaxation* which reformulates the integration step as a convex quadratic program (QP) was introduced by Anitescu in [2]. A convergence analysis similar to the one we have presented in the section above was also given in [2]. The advantage of a convex QP formulation of the integration step over the LCP one, consists in the fact that for convex QPs state–of–the–art solvers are available.

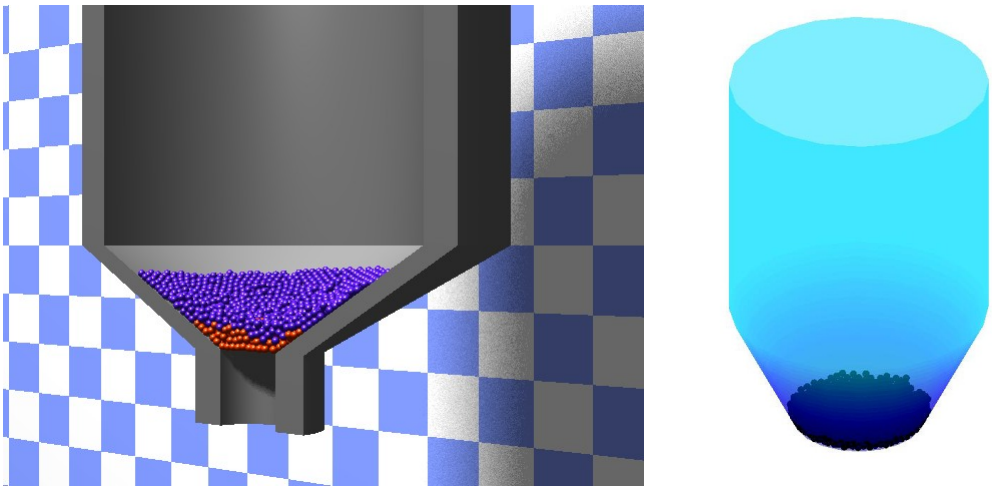
In [37], the QP formulation of [2] was obtained to simulate a large rigid body system that models a pebble bed reactor. Several QP solvers were tested for computational performance on this rigid body system. Here, we will briefly present the application that inspired the analysis from [37], as well as the two (equivalent) QP formulations that were used in obtaining the numerics. Then, we present a summary of the computational performance for the solvers that were tested in [37]. We refer the reader for more details on the implementation as well as resulting computational performance to [37].

The application considered in [37] consists of a pebble bed reactor (PBR), with the reactor vessel being composed of a truncated cone and a cylinder that is opened at both ends, see Figure 2.3. The rigid bodies consist of tennis-ball-sized pebbles which move inside the reactor vessel. The pebbles are extracted from the bottom of the vessel and reinserted through the top [25]. When fully loaded, the reactor has about 400,000 such bodies, but important assessments can be extracted from simulations involving a smaller number [42]. In Figure 2.3 a cross section of the reactor vessel with 3200 pebbles at the end of the simulation is shown on the left and the reactor vessel with 1600 pebbles also at the end of the simulation is shown on the right.

In what follows we briefly describe the mathematical setting used in order to simulate the granular flow inside the PBR. We denote by N the number of pebbles inside the reactor vessel. The position of pebble i is

$$q^{(i)} = (x_i, y_i, z_i, \theta_i, \alpha_i, \gamma_i)^T \,,\ i = 0, ..., N-1,$$

where the first three components are the Cartesian coordinates of the center (in a fixed inertial frame) and the last three are the parameters representing the orientation of a reference frame attached to the sphere. The generalized position



(a) (b)

Figure 2.3: (a) Cross-section of the reactor vessel with 3200 pebbles at the end of the simulation. The coloring for the pebbles was chosen according to their initial position; (b) The reactor vessel with 1600 pebbles at the end of the simulation.

q of the entire system is obtained by concatenating the positions of each body, namely,

$$q = \left(q^{(0)^T}, q^{(2)^T}, ..., q^{(N-1)^T}\right)^T.$$

In a similar fashion, one defines the generalized velocity of the system to be $v \in R^{6N}$ (see [37] for more details). For the PBR example there are two types of possible interactions: *pebble–pebble* and *pebble–wall* interactions. Each pair of interacting bodies was indexed by (i, j), where i, j are indices satisfying $j > i$. The nonpenetration constraints are then written in the form

$$\Phi^{(i,j)}(q) \geq 0,$$

where the $\Phi^{(i,j)}$s are defined to be *signed gap functions* modelling either pebble-pebble nonpenetration or wall-pebble nonpenetration. For example, for two spheres of radius R, indexed by i and j respectively, having the centers of mass at positions x_1, y_1, z_1 and x_2, y_2, z_2, a signed gap function is

$$\Phi^{(i,j)} = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 - 4R^2.$$

The pebble-wall gap functions use the distance from a point to a simple surface and can be easily defined.

In order to give contact specifications, we considered the generic contact depicted in Figure 2.4. We denote by $\overrightarrow{n}$ the unit outward normal for the horizontally

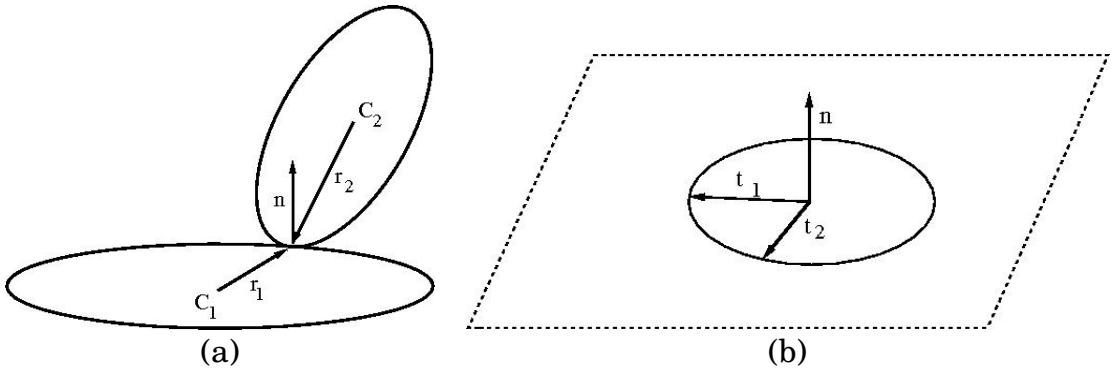


Figure 2.4: (a) A generic contact, (b) Tangent space at contact.

aligned body and by C_1 and C_2 be the centers of mass for each body. We let $\overrightarrow{r}_1$ and $\overrightarrow{r}_2$ be the position vectors of the contact point relative to C_1 and C_2 respectively. Then the 3–dimensional vector $\overrightarrow{n}$ is mapped in generalized coordinates into the 12–dimensional vector n, defined as follows

$$\overrightarrow{n} \longmapsto n := \begin{pmatrix} \overrightarrow{n} \\ \overrightarrow{r}_1 \times \overrightarrow{n} \\ -\overrightarrow{n} \\ -\overrightarrow{r}_2 \times \overrightarrow{n} \end{pmatrix}. \tag{2.3.1}$$

Since the rigid body system is composed of N bodies, the configuration space (the space where the generalized position q lives is $6N$ dimensional, i.e. $q \in \mathbb{R}^{6N}$. Therefore, contact data such as the one described by (2.3.1) must be embedded in the $6N$-dimensional space. For this, first we will use k to index a specific contact (i, j), where $j > i,\ j > 0$. The convention used in [37], was that for a pebble-pebble interaction $i \geq 0$ while for a wall-pebble contact $i < 0$. Taking into consideration (2.3.1), the $6N$-dimensional normal direction for the *pebble–pebble* pair (i, j) has the form

$$n^{(k)} = \left(O_{1,6(i-1)}, \overrightarrow{n}_k^T, (\overrightarrow{r}_i \times \overrightarrow{n}_k)^T, O_{1,6(j-i-1)}, -\overrightarrow{n}_k^T, (-\overrightarrow{r}_j \times \overrightarrow{n}_k)^T, O_{1,6(N-j)}\right)^T, \tag{2.3.2}$$

where $O_{1,\alpha}$ represents a zero row vector of length α, whenever $\alpha > 0$ and the empty vector otherwise. Here n_k is the three-dimensional normal vector at contact k. For a *pebble–wall* interaction (i, j), $i < 0$, only the second nonzero block will contribute to $n^{(k)}$.

As we have seen in the previous section, the Coulomb friction model prescribes that the tangential force is inside a friction disk proportional to the one in Figure 2.4–(b). In order to formulate the integration step as a QP, we use also here a polyhedral approximation of the friction cone. This can be obtained by means of a polygonal approximation of the friction disk.

For a contact indexed by k, let $\overrightarrow{t}_1 := \overrightarrow{t}_{1k}$ and $\overrightarrow{t}_2 := \overrightarrow{t}_{2k}$ be the tangent vectors that span the tangent space at contact k and $t_1 := t_{1k}$, $t_2 := t_{2k}$ be the corresponding directions obtained by the generalized coordinate mapping (2.3.1). We let p_k denote the number of tangent vectors used in the polygonal approximation of the friction disk. In [37], the p_k tangent vectors were defined by

$$\overrightarrow{d}_{sk} := \cos\left(\frac{2\pi s}{p_k}\right)\overrightarrow{t}_1 + \sin\left(\frac{2\pi s}{p_k}\right)\overrightarrow{t}_2, \quad s = 1, ..., p_k. \tag{2.3.3}$$

By mapping equation (2.3.3) into generalized coordinates, one obtains the (generalized) tangent directions

$$d_{sk} := \cos\left(\frac{2\pi s}{p_k}\right) t_1 + \sin\left(\frac{2\pi s}{p_k}\right) t_2, \quad s = 1, ..., p_k.$$

Similar to (2.3.2), the tangential directions $d_s^{(k)} \in R^{6N}$, $s = 1, ..., p_k$ in the $6N$ dimensional space, corresponding to a pebble-pebble contact (k), were defined by

$$d_s^{(k)} = \left(O_{1,6(i-1)}, \overrightarrow{d}_{sk}^T, (\overrightarrow{r}_i \times \overrightarrow{d}_{sk})^T, O_{1,6(j-i-1)}, -\overrightarrow{d}_{sk}^T, (-\overrightarrow{r}_j \times \overrightarrow{d}_{sk})^T, O_{1,6(N-j)}\right)^T. \tag{2.3.4}$$

The matrix associated with the polyhedral approximation of the friction disk at contact (k) is the matrix having its columns the directions $d_s^{(k)}$, i.e., $D^{(k)} \in R^{6N \times p_k}$, $D^{(k)} = \left(d_1^{(k)}, ..., d_{p_k}^{(k)}\right)$, where p_k represents the number of friction generators for contact (k). In [37], the same friction coefficient $\mu \in [0, 1]$ was used for all contacts. The integration step, which was defined by means of a QP, uses the matrices $\widehat{D}^{(k)} \in R^{6N \times p_k}$, which are constructed using the normal directions $n^{(k)}$ and the tangential directions $d_s^{(k)} \in R^{6N}$, $s = 1, ..., p_k$, as follows:

$$\widehat{D}^{(k)} = \left(n^{(k)} + \mu d_1^{(k)}, ..., n^{(k)} + \mu d_{p_k}^{(k)}\right), \tag{2.3.5}$$

In [37] all the p_k were taken to be equal to 3. It was shown that this is the maximal value for p_k such that the matrix $\widehat{D}^{(k)}$ has full column rank.

The generalized matrix (inertia) of the system in a fixed coordinate frame is denoted by $M(q)$. Since we are dealing only with spherical bodies the generalized mass matrix is independent of the configuration q, i.e., $M(q) = M$ where $M \in R^{6N \times 6N}$ is a (constant) diagonal matrix with positive entries. Since the mass matrix is constant, the inertial forces are zero. This implies that, besides contact forces, the only forces acting on the system are external forces.

For PBR, because of the heaviness of the uranium pebble core, the effect of the cooling flow over the dynamics was ignored. This leads to the fact that the applied

external forces, denoted here by k_{app} are made only of gravitational forces. More precisely,

$$k_{app} = \left(u^T, ..., u^T\right)^T,$$

where $u \in R^6$, $u = (0, 0, -g, 0, 0, 0)^T$, for some positive constant g. Here we have also assumed that the mass of each pebble is scaled to 1.

In selecting the active contacts we have used a parameter $\epsilon > 0$. Thus, if the contact pair (i, j) is indexed by k and the corresponding nonpenetration constraint by $\Phi^{(k)}(q) := \Phi^{(i,j)}(q)$, then for a given configuration q, the active set was defined by

$$\mathcal{A}(q, \epsilon) = \left\{k \in \mathbb{Z}_+^2 \mid (k) = (i, j),\ \Phi^{(i,j)}(q) \leq \epsilon\right\}. \tag{2.3.6}$$

If $k \notin \mathcal{A}(q, \epsilon)$, the corresponding nonpenetration constraint is simply ignored by the QP used to formulate the integration step. It has been shown that the choice of ϵ in (2.4.17) does not affect the convergence results, [2]. However, a value of ϵ that is too large results in an exceedingly large QP which becomes computationally expensive. On the other side, a value too small may result in excessive penetration. The choice of ϵ discussed in [37] takes into account the relative velocity at contacts, in such a way that a smaller velocity implies a smaller value for ϵ.

The nonpenetration and contact constraints are replaced by

$$\left(n^{(k)}(q^l)\right)^T v + \mu \left(d_s^{(k)}(q^l)\right)^T v \geq -\frac{1}{h}\Phi^{(k)}(q^l), \quad s = 1, 2, \ldots, p_k, \tag{2.3.7}$$

where h is the time step of the numerical scheme and q^l represents the numerical configuration at time t_l. The first term on the left together with the term on the right of (2.3.7) come from the linearization of the nonpenetration constraint $\Phi^{(k)}(q) \geq 0$. The second term on the left of (2.3.7) is unique to the scheme in [2]. Its physical significance is based on a microscopic realization of surface asperities that result in macroscopic friction coefficient μ.

Assume now, that at time t_{l+1} we want to determine the value of the corresponding velocity v^{l+1}, knowing the velocity at time t_l, v^l and the configuration q^l. The discretized version of Newton's second law, at the velocity impulse level is then written as:

$$M\left(v^{l+1} - v^l\right) - z^{l+1} = hk_{app}, \tag{2.3.8}$$

where hk_{app} are the external impulses and z^{l+1} represent the normal and frictional contact impulses:

$$z^{l+1} = \sum_{k \in \mathcal{A}(q^l, \epsilon)} \sum_{s=1}^{p_k} \beta_s^{(k)} \left(n^{(k)}(q^l) + \mu d_s^{(k)}(q^l)\right),$$

with the multipier vectors $\beta^{(k)}$, satisfying

$$\beta^{(k)} = \left(\beta_1^{(k)}, \beta_2^{(k)}, \ldots, \beta_{p_k}^{(k)}\right) \in R^{p_k} \text{ and } \beta^{(k)} \geq 0.$$

The integration step can be formulated using the analysis from [2]

$$\begin{array}{ll} \min & \frac{1}{2}v^T M v + \left(f^l\right)^T v \\ \text{s.t.} & \left(n^{(k)}(q^l)\right)^T v + \mu \left(d_s^{(k)}(q^l)\right)^T v \geq -\frac{1}{h}\Phi^{(k)}(q^l) \\ & k \in \mathcal{A}(q^l, \epsilon),\ s = 1, 2, \ldots, p_k \end{array} \tag{2.3.9}$$

In (2.3.9), f^l is obtained by the following formula:

$$f^{(l)} = -v^l - hk_{app}.$$

We call (2.3.9), the *primal QP* or *the primal formulation*. We note that equations (2.3.8) and (2.3.7) are satisfied as part of the optimality conditions for (2.3.9). The *dual formulation* is easily obtained by using standard duality tehniques. More precisely, in [2], the dual QP was given in the form

$$\begin{array}{ll} \min & \frac{1}{2}\lambda^T P^l \lambda + \left(\kappa^l\right)^T \lambda, \\ \text{s.t.} & \lambda \geq 0 \end{array} \tag{2.3.10}$$

where $P^l = A^l M^{-1} \left(A^l\right)^T$ and $\kappa^l = -b^l - A^l f^l$. Here A^l and b^l are the matrix and the right-hand side, respectively, of the inequality constraints in (2.3.9). If the active contact set is given by $\mathcal{A}(q^l, \epsilon) = \{k_i \mid i = 1, ...p\}$, then, the matrix A^l has the form

$$A^l = \begin{pmatrix} \left(\widehat{D}^{k_1}(q^l)\right)^T \\ \vdots \\ \left(\widehat{D}^{k_p}(q^l)\right)^T \end{pmatrix}.$$

Here, the matrices $\widehat{D}^{k_1}(q^l)$ are given by (2.3.5). The vector b^l is composed of block vectors in R^3, with the block corresponding to contact k_i having all its components equal to $-\frac{1}{h}\Phi^{(k_i)}(q^l)$. The dual formulation (2.3.10) is a bound-constrained quadratic programming problem. This implies that besides general-purpose quadratic programming algorithms, such as interior points, one can use also iterative algorithms of the projected gradient type.

To be able to simulate the same system by both the primal and the dual formulations a *"no duality gap"* result was discussed in [37]. For the "no duality gap" result, the pointedness of the friction cone is again an essential assumption. We defined the ϵ–active friction cone $\mathcal{FC}(q^l, \epsilon)$ by

$$\mathcal{FC}(q^l, \epsilon) = \left\{ \sum_{k \in \mathcal{A}(q^l, \epsilon)} \widetilde{D}^{(k)} \beta^{(k)} \ \Big|\ \beta^{(k)} \in R^{p_k},\ \beta^{(k)} \geq 0, \right\}. \tag{2.3.11}$$

In (2.3.11), the matrices $\widetilde{D}^{(k)}$ are defined by

$$\widetilde{D}^{(k)}(q^l) := \widetilde{D}^{(k)} = \left(d_1^{(k)}, d_2^{(k)}, \ldots, d_{p_k}^{(k)}\right),$$

where the generalized tangential directions $d_s^{(k)} := d_s^{(k)}(q^l), s = 1, 2, \ldots, p_k$ are given in (2.3.4). In order to have no duality gap, the program (2.3.9) must be feasible and the ϵ–active friction cone must be pointed. We note that for the friction cone defined by (2.3.11), pointedness is equivalent to the fact that $\mathcal{FC}(q^l, \epsilon)$ does not contain any proper subspaces. As pointed out in [37], loss of pointedness corresponds, from a physical point of view, to jamming, which is a phenomenon that was not encountered in the simulation of the granular flow in the PBR.

We now briefly present the software packages used in [37] for solving the QPs presented above. Some of these packages solve the primal–dual formulation (2.3.9), while others solve only the dual formulation (2.3.10). Two of the first software packages, OOQP and MOSEK, use interior-point algorithms and solve the primal-dual formulation of both (2.3.10) and (2.3.9). The other two packages tested are TRON and BLVM, which use projected gradient algorithms on the dual problem (2.3.10). A short description of each of these solvers is given in the list below:

- OOQP (Object-Oriented software for Quadratic Programming), [24], is a C++ package for solving convex quadratic programming problems. It is based on primal-dual interior-point methods. In the simulations we generate for different sizes of the system, OOQP's general sparse formulation was used to solve the primal form (2.3.9). Two distinct linear algebra packages MA27 and CHOLMOD were used for solving the inner OOQP linear systems. For the second one, CHOLMOD, which implements a modified Cholesky factorization, we developed an interface and reformulated the linear systems. We refer the reader to [37] for a detailed description of the implementation.

- TRON , [30], is a trust region Newton method for bound constrained optimization problems and it is used to solve the dual formulation (2.3.10). The algorithm uses a quadratic model function, projected searches during the subspace minimization phase and a preconditioned conjugate gradient method to determine the minor iterates. The limited memory preconditioner used is the incomplete Cholesky factorization of [31]. More details about the Cauchy step as well as some other inner workings of this software package are given in [37].

- BLMVM [9] is a projected gradient solver for nonlinear bound-constrained optimization problems and also this one is applied to the dual formulation (2.3.10). In order to reduce the cost of storing the inverse Hessian approximation, BLMVM uses the limited memory BFGS method (L-BFGS), [35].

- The MOSEK Optimization Software (www.mosek.com) is a collection of tools for solution of large-scale optimization problems. When solving convex quadratic problems subject to linear constraints, which is the type of QPs we are dealing with in the PBR simulation, MOSEK employs an homogeneous interior-point algorithm for monotone complementarity problems [1]. The algorithm can start at a feasible or infeasible point as well. In [37], we have used MOSEK 4.0 to solve the primal formulation (2.3.9).

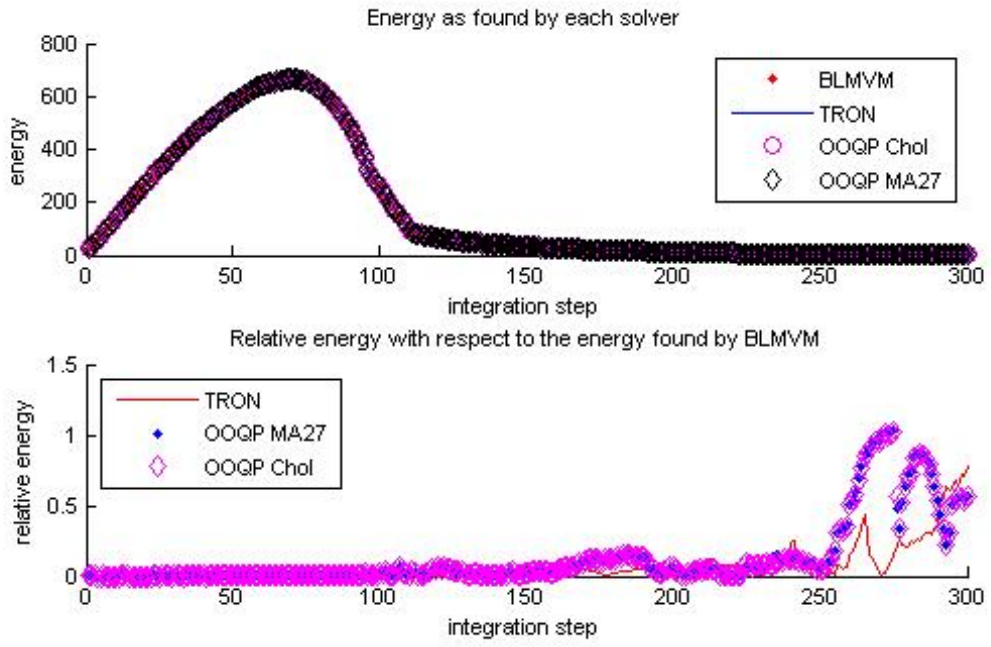


Figure 2.5: Top: total kinetic energy obtained with each of the four solvers; bottom: relative energy E_{rel}.

In [37], the software packages were tested for computational performance on the simulation of the granular flow in the PBR. For a detailed analysis of these computational results we refer the reader to [37]. Here, we restrict our attention to a different issue, that of correctness for the simulation results.

For isolated QPs one can compare the correctness of the results obtained by using the software packages mentioned above. Things change when running an entire simulation. The use of different solvers to obtain an entire simulation (the end of the simulation corresponds to the moment when the pebbles are at rest) will likely cause totally different individual configurations. This is motivated by the fact that, by nature, the system is chaotic. Therefore, to measure the correctness

of an entire simulation, at least partially, some ensemble properties must be used. In [37] we used the total kinetic energy to test full simulations. We present these results also here.

For the simulation experiments the pebbles are initially randomly arranged in horizontal planes. On each horizontal plane the pebbles are distributed in several inner circles. In Figure 2.5, we look at how the total kinetic energy of a system consisting of 800 pebbles changes in time. The kinetic energy defined to be $E(t) = \frac{1}{2}v(t)^T M v(t)$, where M is the (whole) system matrix and v the system velocity. was found by simulating the same configuration with the four solvers. The second plot in Figure 2.5 represents the relative energy, E_{rel} defined as

$$E_{rel} = \frac{|E_s - E_b|}{E_b},$$

where E_b and E_s are the total kinetic energies found by making use of BLMVM and one of the remaining solvers, respectively.

It can be seen from the top plot of Figure 2.5 that the total kinetic energy obtained with the four solvers is fairly similar. The differences can be better analyzed from the bottom plot which gives the relative energy E_{rel}. It can be noted from this plot that he relative error in energy is insignificant, *in physical interpretation terms*, except after a large amount of simulation time. It can also be noted that this error occurs only at a very small value of the kinetic energy (essentially, around the time the pebbles have stopped). This is because when the pebbles are almost at rest the denominator in the definition of E_{rel} becomes very small in absolute value.

We end this section by summarizing the analysis of the computational performance of the solvers presented above. For more details, as well as complete data tables regarding the computational performance, we refer the reader to [37]. Simulations that were reported and analyzed from a computational perspective, consider 800, 1000, 1600 or 3200 pebbles. The values for the time-steps used to produce these simulations were $h = 0.05$ or $h = 0.01$. There were two types of tests: *"the simulation test"* and *"the optimization test"*. In the simulation test, the solvers were analyzed for an "entire simulation". More precisely each solver was used in the numerical integration process until the kinetic energy fell under a specific value. It is important to note that in the simulation test, the solvers may solve different problems at each step since the system trajectories may be different due to accumulation of the numerical errors. In the *optimization test*, the performance of all solvers for the same QP problem was tested. It was concluded that for the PBR simulation the OOQP-Chol implementation was the fastest of all packages tested. It used only about three times more memory than BLMVM, while achieving much higher precision levels. For more details on our analysis as well as technical data

that supports this conclusion we refer the reader to [37].

2.4 Simulation in a quasi-static setting

In this section we look at a slightly different problem that was treated in the papers [14], [13] and [17]. The dynamics setting analyzed in the previous sections is replaced by a quasi-static setting. In the quasi-static setting, Newton's second law is replaced by an equilibrium equation. One real-life application that fits this model is given in [14] and [13]. The problem treated in these works is the canonical problem of the peg-in-the-hole at the mesoscale. This was defined in [14] and [13] as the problem of assembling a planar rectangular part (peg) into a planar, rectangular slot by means of pushing operations. This problem is graphically described in Figure 2.6.

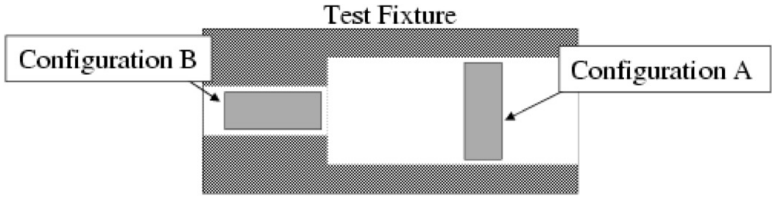


Figure 2.6: Typical meso-scale assembly task: move peg from configuration A to configuration B.

To move the peg from the sensed configuration A to the goal configuration B two types of probes were used in [14] and [13]. The first type was a passive Single-Tip Probe (STP), while the second one was an active Dual-Tip Probe (DTP). The STP is passive and it can be positioned at the beginning of any manipulation sequel, its motion is not controlled during the manipulation process.

The quasi-static model is motivated by the fact that inertial forces are one order smaller than the frictional forces. More precisely, inertial forces are of the order of nano-newtons for the accelerations involved, while the frictional forces are of the order of micro-newtons. We assumed Coulomb friction for peg-probe contacts and peg-fixture contacts. We also assume the support plane to be uniform, and all pushing motions of the probes to be parallel to this plane. Since in experimentation, we coated the support surface with oil, we assumed viscous damping between peg and the support plane. The corresponding damping matrix was estimated based on experimental data. The probe-peg and peg-fixture type contacts were considered to be Coulomb frictional contacts. In [14] and [13] only contacts between the probe and the peg were allowed, but contacts between peg and the fixture can be treated in the same fashion, [17].

We mention just a few of the difficulties associated with peg-in-the-hole tasks at the mesoscale level :

- uncertainties in sensing, actuation, manufacturing and control,
- identifying the exact position of the peg and/or the hole,
- gripping or manipulating the peg,
- position control.

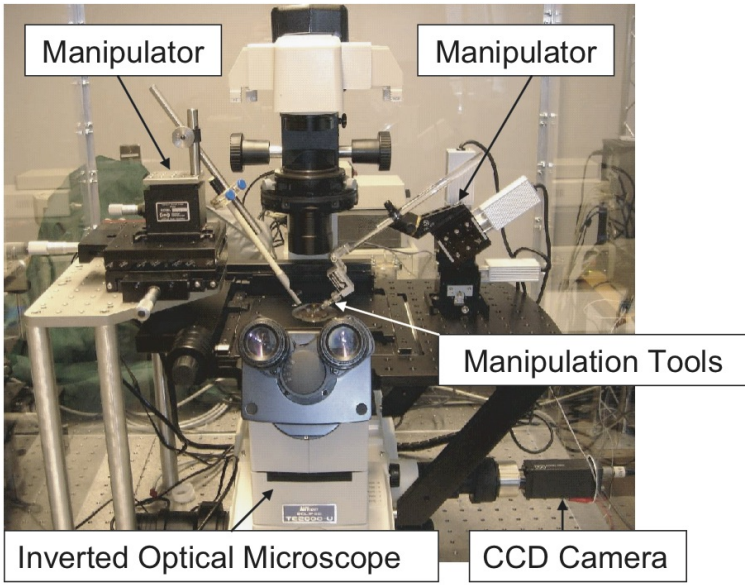


Figure 2.7: The experimental setup used in [14] and [13]

The manipulation system used in [14] and [13] is shown in Figure 2.7 and consists of an inverted optical microscope and charge-coupled device (CCD) camera which was used for sensing the configuration, a computer controlled micromanipulator, controller, a manual micromanipulator, 5 μm and 25 μm diameter tip tungsten probes, a motorized XY stage, and control computer. The 4X microscope objective used in this application produced a field of view (FOV) of 3.37 mm x 2.52 mm. The CCD camera records the images in the FOV and sends them to the control computer.

The fixture has the geometry presented in Figure 2.6 and the state of the peg is given by $q = (q_x, q_y, q_\theta)$, where (q_x, q_y) are the Cartesian coordinates of the centroid of the peg and q_θ denotes the peg orientation. As we mentioned above, in [14]

and [13], two probes were used in manipulating the peg. The STP is a passive probe and its position doesn't change during a manipulation task. The control of the DTP was fully characterized by a vector $u = (d_2, v_p, p^t)$ denoting a push in x-direction with relative distance d_2, constant speed v_p and duration p^t. Here d_2 refers to the relative position of the DTP on one side of the peg. We refer the reader to [13] for more details. In the more general model which was presented in [17] and will be given also here in the next pages, the control for each of the probes is not restricted by the pushing direction and it will assume a more general form.

Since the support surface was lubricated, the interaction between the peg and this surface was modeled by means of a damping force. This damping force is fully characterized by a 3×3 damping matrix E, which is assumed to be symmetric positive definite. The matrix E together with the friction coefficients need to be estimated based on a matching between experimental data and numerical simulation. In [14] and [13], we have used a three-point support model for the interaction between the peg and the supporting surface. For parameter estimation, an "L_∞–fit" was formulated in [13] and the Nelder-Mead algorithm, [34], was used.

In what follows we present the quasi-static differential complementarity problem (DCP) that was used in designing the time-stepping schemes used for the simulation of the peg-in-the-hole problem. Once again the state of the peg is given by $q(t) := q = (q_x, q_y, q_\theta)$, while its velocity is $v = \dot{q}$. The damping matrix E is assumed to be symmetric and positive definite. Let

$$u = \left(u_1^T, u_2^T, ..., u_{n_u}^T\right)^T \tag{2.4.1}$$

denote the partitioned vector of controls applied in the time interval $[0, T]$. In the time interval $[0, T]$, we assume that only one manipulation task is performed. The partition (2.4.1) implies that n_u probes are used to manipulate the peg on $[0, T]$. One way of setting these controls is by encoding the position, orientation and velocity of probe i in the control vector

$$u_i \in R^4,\ u_i = (u_{xi}, u_{yi}, u_{\theta i}, v_{ui})^T.$$

Here (u_{xi}, u_{yi}) are the Cartesian coordinates of the tip of probe i, and v_{ui} is the *constant* speed of the tip along the probe direction $d_{ui} = (\cos(u_{\theta i}), \sin(u_{\theta i}))^T$. Since we are talking about a single manipulation task, this control is applied for a fixed period of time $[0, T]$.

Before giving the DCP that models this type of applications we introduce the contact data. Let $\Psi_{nk}(q, u, t)$ denote the normal displacement function corresponding to contact k, $k \in \{1, ..., n_c\}$. Here we assume n_c possible contacts between the peg and the probes and contacts between the peg and the fixture. In the same fashion we define $\Psi_{tk}(q, u, t)$ to be the tangential displacement function corresponding

to contact k. The normal and tangential wrench vectors corresponding to contact k are denoted by $W_{nk}(q,u,t)$ and $W_{tk}(q,u,t)$ respectively and satisfy

$$W_{nk}(q,u,t) = \frac{\partial \Psi_{nk}}{\partial q}(q,u,t) \quad \text{and} \quad W_{tk}(q,u,t) = \frac{\partial \Psi_{tk}}{\partial q}(q,u,t).$$

We end the description of the contact data by mentioning that Coulomb's friction coefficient corresponding to contact k, is denoted by μ_k and satisfies $\mu_k \in [0,1]$.

The continuous model under the rigid body assumption is given by the following differential complementarity problem (DCP):

$$\dot{q}(t) = v(t), \tag{2.4.2}$$

$$Ev(t) - W_n(q,u,t)\lambda_n(t) - W_t(q,u,t)\lambda_t(t) = 0, \tag{2.4.3}$$

$$0 \leq \Psi_n(q,u,t) \quad \perp \quad \lambda_n(t) \geq 0, \tag{2.4.4}$$

$$\dot{s}^+_{tk}(t) - \dot{s}^-_{tk}(t) = \left(W_{tk}(q,u,t)\right)^T v(t) + \frac{\partial \Psi_{tk}}{\partial t}(q,u,t), \quad k = 1,...,n_c, \tag{2.4.5}$$

$$0 \leq \dot{s}^+_{tk}(t) \quad \perp \quad \mu_k \lambda_{nk}(t) + \lambda_{tk}(t) \geq 0, \quad k = 1,...,n_c, \tag{2.4.6}$$

$$0 \leq \dot{s}^-_{tk}(t) \quad \perp \quad \mu_k \lambda_{nk}(t) - \lambda_{tk}(t) \geq 0, \quad k = 1,...,n_c. \tag{2.4.7}$$

The quasi-static assumption is reflected by the equilibrium equation (2.4.3). Here $\lambda_n(t) \in R^{n_c}$ and $\lambda_t(t) \in R^{n_c}$ are vectors encoding all normal and tangential forces (for eg., $\lambda_{nk}(t)$ represents the normal force corresponding to the k–th contact), while $W_n(q,u,t)$ and $W_t(q,u,t)$ are the normal and tangential wrench matrices. More precisely, the k-th column of $W_n(q,u,t)$ ($W_t(q,u,t)$) is the normal (tangential) wrench vector $W_{nk}(q,u,t)$ ($W_{tk}(q,u,t)$).

Equation (2.4.4) gives the non-penetration and contact constraints in the form of a *complementarity* condition . Here, $\Psi_n(q,u,t)$ represents the vector of normal displacements and it is constructed in the same fashion as $\lambda_n(t)$ was built above. Since both quantities in (2.4.4) are vector functions the inequality signs are to be understood component–wise, while the complementarity condition ("$\perp$") may be written as $\left(\lambda_n(t)\right)^T \Psi_n(q,u,t) = 0$. Therefore condition (2.4.4) is encoding non-penetration (rigid bodies) constraints ($\Psi_n(q,u,t) \geq 0$, $\lambda_n(t) \geq 0$) and contact constraints (if no contact than the normal force is 0, otherwise the normal force must be non-negative in order to avoid subsequent penetration).

Equation (2.4.5) gives the definition of the positive, $\dot{s}^+_{tk}(t) \geq 0$, and negative, $\dot{s}^-_{tk}(t) \geq 0$, sliding velocities at contact k. In other words, the (overall) sliding velocity defined by (2.4.5) is seen decomposed in its negative part and its positive part. The right-hand side of (2.4.5) represents the (overall) sliding velocity $\dot{s}_{tk}(t) := \dot{\Psi}_{tk}(q,u,t) = \left(W_{tk}(q,u,t)\right)^T v(t) + \frac{\partial \Psi_{tk}}{\partial t}(q,u,t)$ at contact k. The last two equations, namely (2.4.6) and (2.4.7), represent Coulomb's friction law at contact k.

To obtain the time-stepping scheme we used an equivalent formulation of the DCP (2.4.2)–(2.4.7). This new DCP formulation can be obtained by replacing the positive and negative sliding velocities with a different set of signed restricted variables. The technique is not new and it is frequently used to approximate circular (non-polyhedral) friction cones by polyhedral ones.

For each contact k we define the 3×2 matrix $W_{fk}(q,u,t)$ by joining the column vectors $W_{tk}(q,u,t)$ and $-W_{tk}(q,u,t)$. That is,

$$W_{fk}(q,u,t) = \left[W_{tk}(q,u,t) \quad -W_{tk}(q,u,t)\right].$$

The tangential wrench $W_t(q,u,t)$ matrix will be replaced by $W_f(q,u,t)$ with its partitioned form:

$$W_f(q,u,t) = \left[W_{f1}(q,u,t) \;\; ... \;\; W_{fn_c}(q,u,t)\right].$$

The tangential displacement functions $\Psi_{tk}(q,u,t)$ are replaced by $\Psi_{fk}(q,u,t) \in R^{\,2}$,

$$\Psi_{fk}(q,u,t) = \left(\Psi_{tk}(q,u,t), -\Psi_{tk}(q,u,t)\right)^T.$$

The friction forces $\lambda_{tk}(t)$ are decomposed into their negative and positive parts. More precisely, $\lambda_{fk}(t) = \left(\lambda^+_{tk}(t), \lambda^-_{tk}(t)\right)^T$ and $\lambda_f(t) = \left(\lambda^T_{f1}(t), ..., \lambda^T_{fn_c}(t)\right)^T \in R^{\,2n_c}$.

Using these decompositions, we are led to the alternative formulation

$$\dot{q}(t) = v(t), \tag{2.4.8}$$

$$Ev(t) - W_n(q,u,t)\lambda_n(t) - W_t(q,u,t)\lambda_t(t) = 0, \tag{2.4.9}$$

$$0 \le \Psi_n(q,u,t) \quad \perp \quad \lambda_n(t) \ge 0, \tag{2.4.10}$$

$$0 \le \left(W_{fk}(q,u,t)\right)^T v(t) + \frac{\partial \Psi_{fk}}{\partial t}(q,u,t) + \sigma_k(t)e \quad \perp \quad \lambda_{fk}(t) \ge 0, \quad k = 1, ..., n_c, \tag{2.4.11}$$

$$0 \le \sigma_k(t) \quad \perp \quad \mu_k \lambda_{nk}(t) - e^T \lambda_{fk}(t) \ge 0, \quad k = 1, ..., n_c \tag{2.4.12}$$

Here $e \in R^{\,2}$ is a column vector with both components equal to 1, i.e., $e = (1,1)^T$ and $\sigma_k(t)$ is the sliding speed for the k–th contact. We note that the DCP (2.4.2)–(2.4.7) or its equivalent formulation (2.4.8)–(2.4.12) covers more general situations than the ones encountered in [14] and [13]. More precisely it is easy to see that with the above DCPs, the number of probes, the direction of pushing and the passive/active status of each probe are not restricted.

We present now the time-stepping scheme that was used in [14] and [13]. Let $h > 0$ denote the integration step (time-step) and $t_l = lh$ the discrete integration times. We approximate the new configuration q^{l+1} using a backward Euler formula, as follows

$$q^{l+1} = q^l + hv^{l+1}, \tag{2.4.13}$$

where v^{l+1} is an estimate for the velocity at time t_{l+1} and will be found by solving a mixed linear complementarity problem.We approximate the nonlinear vector

function $\Psi_n(q^{l+1}, u, t_{l+1})$ by its linearization, so that the complementarity condition (2.4.10) at t_{l+1} is written as:

$$0 \leq \Psi_n(q^l, u, t_l) + \big(W_n(q^l, u, t_l)\big)^T \big(hv^{l+1}\big) + \frac{\partial \Psi_n}{\partial q}(q^l, u, t_l) \quad \perp \quad \lambda_n^{l+1} \geq 0. \tag{2.4.14}$$

Finally, the functions W_n, W_f, W_{fk} and $\frac{\partial \Psi_{fk}}{\partial t}$ ($k = 1, ..., n_c$) in (2.4.9), (2.4.11) and (2.2.3) are evaluated at (q^l, u, t_l). To shorten the notation we use the following conventions:

$$W_n^l := W_n(q^l, u, t_l),\ W_f^l := W_f(q^l, u, t_l),\ \frac{\partial \Psi_n}{\partial t}^l := \frac{\partial \Psi_n}{\partial t}(q^l, u, t_l) \text{ and } \frac{\partial \Psi_f}{\partial t}^l := \frac{\partial \Psi_f}{\partial t}(q^l, u, t_l). \tag{2.4.15}$$

It follows that the unknowns $\big(hv^{l+1}, h\lambda_n^{l+1}, h\lambda_f^{l+1}, h\sigma^{l+1}\big)$ may be obtained as the solution of the following MLCP:

$$\begin{pmatrix} 0 \\ \rho_n^{l+1} \\ \rho_f^{l+1} \\ s^{l+1} \end{pmatrix} = \begin{pmatrix} E & -W_n^l & -W_f^l & 0 \\ (W_n^l)^T & 0 & 0 & 0 \\ (W_f^l)^T & 0 & 0 & E_f \\ 0 & U_f & -E_f^T & 0 \end{pmatrix} \begin{pmatrix} hv^{l+1} \\ h\lambda_n^{l+1} \\ h\lambda_f^{l+1} \\ h\sigma^{l+1} \end{pmatrix} + \begin{pmatrix} 0 \\ \Psi_n^l + h\frac{\partial \Psi_n}{\partial t}^l \\ h\frac{\partial \Psi_f}{\partial t}^l \\ 0 \end{pmatrix} \tag{2.4.16}$$

with $0 \leq \big[\rho_n^{l+1}, \rho_f^{l+1}, s^{l+1}\big] \quad \perp \quad \big[h\lambda_n^{l+1}, h\lambda_f^{l+1}, h\sigma^{l+1}\big] \geq 0$. Here $U_f \in R^{n_c \times n_c}$, $E_f \in R^{2n_c \times n_c}$ with U_f a diagonal matrix with elements on its diagonal equal to μ_k, $k = 1, ..., n_c$ and E_f a block diagonal matrix, with diagonal blocks given by the vector e. That is,

$$U_f = \begin{pmatrix} \mu_1 & ... & 0 \\ \vdots & & \vdots \\ 0 & ... & \mu_{n_c} \end{pmatrix}, \quad E_f = \begin{pmatrix} 1 & ... & 0 \\ 1 & ... & 0 \\ \vdots & & \vdots \\ 0 & ... & 1 \\ 0 & ... & 1 \end{pmatrix}.$$

As done also in the previous sections, we use a parameter $\epsilon > 0$ to define the set of active contacts corresponding to time t_l, configuration q^l and control u. This contact set which we denote by $\mathcal{A}(q^l, u, t_l, \epsilon)$ or briefly by $\mathcal{A}_\epsilon^l$ is given by

$$\mathcal{A}_\epsilon^l = \big\{k \in \{1, ..., n_c\} \mid \Psi_{nk}(q^l, u, t_l) \leq \epsilon\big\} \tag{2.4.17}$$

Also here the value of ϵ can be used to avoid numerical interpenetration and can be chosen as a quantity dependent on the time step h. For an active contact $k \in \mathcal{A}_\epsilon^l$, we define the friction cone corresponding to that contact by

$$FC_k^\epsilon(q^l, u, t_l) = \big\{z = W_{nk}^l \lambda_{nk} + W_{fk} \lambda_{fk}^l \mid \lambda_{nk} \geq 0,\ \lambda_{fk} \geq 0 \text{ and } e^T \lambda_{fk} \leq \mu_k \lambda_{nk}\big\} \tag{2.4.18}$$

Taking into account all active friction cones defined by (2.4.18), we are led to the *total active friction cone* $FC^{\epsilon}(q^l, u, t_l)$, given by

$$FC^{\epsilon}(q^l, u, t_l) = \sum_{k \in \mathcal{A}_{\epsilon}^l} FC_k^{\epsilon}(q^l, u, t_l). \tag{2.4.19}$$

Solvability and pointdeness of the friction cone. The solvability of the MLCP (2.4.16), which together with (2.4.13) characterizes the integration step depends on the same regularity assumption that we have seen in the sections above, namely *pointedness of the friction cone*. We recall that pointedness of the friction cone given by (2.4.19) reduces to the fact that the cone doesn't contain any proper linear subspace. As pointed out in [17], the solvability of (2.4.16) can be guaranteed by the following theorem.

Theorem 2.4.1. *Assume that the friction cone $FC^{\epsilon}(q^l, u, t_l)$ is pointed and that only ϵ–active contact constraints are encoded in (2.4.16). Then a solution of the MLCP (2.4.16) exists and Lemke's algorithm, with precautions taken against cycling will always return a solution.*

In [17], we have pointed out that the proof of the result given by Theorem 2.4.1 can be easily obtained by first reducing the MLCP (2.4.16) to a standard LCP and then using an existence result for LCPs with copositive matrices. The reduction of the MLCP (2.4.16) to a standard LCP can be obtained if we eliminate hv^{l+1} from the first equation. The resulting LCP has the form

$$\begin{pmatrix} \rho_n^{l+1} \\ \rho_f^{l+1} \\ s^{l+1} \end{pmatrix} = \begin{pmatrix} \left(W_n^l\right)^T E^{-1} W_n^l & \left(W_n^l\right)^T E^{-1} W_f^l & 0 \\ \left(W_f^l\right)^T E^{-1} W_n^l & \left(W_f^l\right)^T E^{-1} W_f^l & E_f \\ U_f & -E_f^T & 0 \end{pmatrix} \begin{pmatrix} h\lambda_n^{l+1} \\ h\lambda_f^{l+1} \\ h\sigma^{l+1} \end{pmatrix} + \begin{pmatrix} \Psi_n^l + h\frac{\partial \Psi_n}{t}^l \\ h\frac{\partial \Psi_f}{t}^l \\ 0 \end{pmatrix}$$

$$0 \leq \begin{pmatrix} \rho_n^{l+1} \\ \rho_f^{l+1} \\ s^{l+1} \end{pmatrix} \perp \begin{pmatrix} h\lambda_n^{l+1} \\ h\lambda_f^{l+1} \\ h\sigma^{l+1} \end{pmatrix} \geq 0 \tag{2.4.20}$$

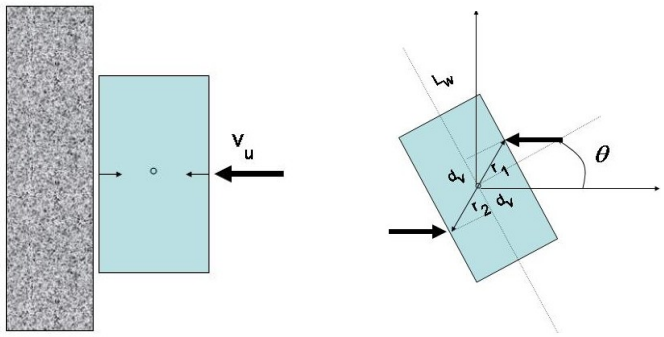
It is not hard to prove that the matrix on the right hand-side in the first equation of (2.4.20) is a copositive matrix. Now this fact, together with the fact that the matrix E is positive definite and the friction cone $FC^{\epsilon}(q^l, u, t_l)$ is pointed can be used to obtain the solvability for the integration step. This is done by applying the standard solvability result, given by Theorem 1.2.8.

As pointed out also in the previous sections the pointedness assumption is weaker than the linear independence of the contact wrenches. This is important especially for the application discussed here, since there are perfectly valid situations (from a physical point of view), when the contact wrenches are linearly

dependent. The next question that can be addressed is related to the *lack of pointedness*. Since *lack of pointedness* will lead to a time-step LCP with empty solution set, we would like to know, whether *lack of pointedness* corresponds also to a physically unfeasible configuration. The short discussion below addresses this question.

In Figure 2.8, there are two situations when the friction cone fails to be pointed. The first situation corresponds to a stuck configuration, while in the second one the applied controls will damage the peg/probes.

For Figure 2.8(a), the normal wrench W_{n1} corresponding to contact 1 (peg–probe contact), $W_{n1} = (0, -1, 0)^T$, while $W_{n2} = (0, 1, 0)^T$. Clearly we can write 0 as a positive linear combination of W_{n1}, W_{n2}: $0 = W_{n1}\lambda_{n1} + W_{n2}\lambda_{n2}$, for $\lambda_{n1} = \lambda_{n2} > 0$. Therefore the pointedness assumption doesn't hold.



(a) (b)

Figure 2.8: (a) A stuck configuration, (b) Controls that damage the peg.

In Figure 2.8(b), L_w denotes half the length of the short side while d_v denotes half the separation between the probes (along the common sliding direction). If we apply the two probes as shown in this picture it can be easily observed that the normal wrenches corresponding to the two contacts become linear dependent. This alone does not imply lack of pointedness. However, it can be shown that for $d_v < L_w$ and certain values of the friction coefficient μ (both contacts have the same friction coefficient), the friction cone fails to be pointed. More precisely, the situation under which this failure occurs is given by the following conditions:

$$d_v < L_w, \quad \frac{d_v}{L_w} \leq \mu.$$

The physical situation explained above is somehow intuitive. We can imagine that if we push the peg with two probes acting in entirely opposite directions very close to the middle of the corresponding sides (as depicted in Figure 2.8(b)), will result in a damage to the peg and/or the probes. This type of situation is undesirable and should be avoided by any manipulation plan. Therefore whenever the LCP

solver fails to return a solution, we can simply reject the applied control due to "stuck"–configurations or "damage"–situations.

In [14] and [13] robust pushing primitives were designed using a lot of geometrical insight. The presence of uncertainty which clearly characterizes the application described here, imposes manipulation plans based on robust pushing operations. Small perturbations should not affect in a significant way the original trajectory. Since the pushing operations are determined by the controls applied, we will identify a robust pushing operation by a corresponding *robust control*. Deciding whether a control u is robust or not can be done by analyzing the structure of the LCP integration step. An easy algorithm for deciding whether an applied control is robust or not was given in [16] for the dynamic setting. We briefly present this algorithm here in the quasi-static setting. Lack of robustness, will mean a change of mode,as for example a change from a sliding contact to sticking one. Detecting whether an applied control results in a mode change can be done by looking at the complementarity matrices used by the solutions of subsequent integration steps. This is shown next.

For this, assume that the time-step MLCP is reduced to a standard LCP as the one given in (2.4.20). Given a control vector u (which is kept constant), we are interested to see whether this control will result in a mode change during a given number of integration steps. If this happens, the applied control will not be robust and it will be rejected. For simplicity, we can view LCP (2.4.20) in the form

$$z \geq 0,\ A(t,h)z + b(t,h) \geq 0,\ z^T(A(t,h)z + b(t,h)) = 0, \tag{2.4.21}$$

where $A(t,h)$ and $b(t,h)$ are matrices and vectors of appropriate dimensions. We denote the above LCP, by $LCP(b(t,h), A(t,h))$. Since the control u is kept constant, the explicit dependence on u is ommited from (2.4.21). The matrix $A(t,h)$ from (2.4.21) is coposiitive and assuming that the conditions for solvability are met, Lemke's algorithm will return a solution which is an extreme point of the feasible set. Therefore Lemke's algorithm will return an extreme point solution $z^* = z^*(t,h)$ that can be recovered using a complementarity basis $C_{A(h)}(\alpha)$, where α is the index set determined by z^* and $C_{A(t,h)}(\alpha)$ is obtained. Determining whether an applied control is robust or not is based on detecting whether a change of complementarity basis occurs or not.

Let N be a positive integer. We are interested in detecting whether a switching event (change of complementarity basis) occurred in the time interval $[0, Nh]$. If such a switching event occurs in the given time interval the applied control will not be considered robust and it will be rejected. Otherwise the control is said to be robust and may be used in manipulation plans. The algorithm is given below and it is adapted from the one presented in [16] for the dynamic setting.

1. Set $n := 0$, $T := Nh$, $t := 0$.
2. Solve $LCP(b(0,h), A(0,h))$ by means of Lemke's method.
3. Determine the index set α_0 such that $C_{A(0,h)}(\alpha_0)$ is a complementarity basis for the solution obtained in Step 2 above.
4. Set $t := t + h$ and $IsSwitch := 0$.

WHILE ($t < T$ and $IsSwitch = 0$)

5. Solve for the unknown vector w the linear system $C_{A(t,h)}(\alpha_0)w = b(t,h)$.

IF ($\exists i,$ such that $w_i < 0$)

6. $IsSwitch := 1$.

ENDIF

7. t:=t+h.

ENDWHILE

We note that the algorithm solves only one LCP in step 2 above. In the while loop, only linear systems are solved and the resulting solutions are checked for non-negativity. If they keep remaining non-negative (the sign is checked componentwise) for the time interval $[0, Nh]$, then the applied control is considered robust and the algorithm exits with $IsSwitch = 0$. If at any point inside the loop the non-negativity is violated, it follows that a solution to the corresponding LCP cannot be obtained using the *"initial"* complementarity basis and therefore $IsSwitch = 0$ implying that a mode switch has occurred. Since this corresponds to a control u which is not robust, u will be rejected.

We end this section by pointing out that robust controls may be used in the context of randomized planning. A sampling-based motion planning inspired by the Rapidly-exploring Random Trees (RRT), [29], was used in [13] to produce relatively short manipulation paths. While in [13], a *"geometric"* approach was used to determine apriori robust motion primitives, it is important to note that the robustness analysis based on the underlying LCP structure may be extremely useful in the context of randomized planning for both the dynamic and the quasi-static settings. This is mainly because it has a relatively small dependence on the details of the contact problem (such as the geometry of the fixture, number of probes, pushing directions, etc.) and the computational effort is not extremely big since only one LCP needs to be solved in the Algorithm presented above. For more details as well as experimental and numerical proofs we refer the reader to the works [14], [13], [17] and [16].

2.5 Other contributions

In this section we present some other contributions which one can say that they are not directly connected to the simulation of rigid body systems. However many of these results can be used in the numerical integration, parameter estimation, as well as planning and control of such systems.

The first work presented here deals with an optimization problem that comes from optical flow estimation. We discuss several formulations for the optimization problem used in the estimation of the optical flow. We also briefly present how the structure of the optimization problem can be exploited from a computational point of view.

The second part of this section is dedicated to mathematical inequalities with applications to probability and statistics. The first inequalities discussed give an improvement of some inequalities of Chebysev-Grüss type and lead to a probabilistic inequality that can be used in the estimation of the covariance matrix . The second result refers to a Hermite-Hadamard type inequality that can be used in estimating moments of continuous random variables.

2.5.1 A problem from optical flow

In [23], we have given a study of some linear programming formulations (LP) used in the estimation of optical flow. We focused on a version of the Horn-Schunck model [26] with the l_1 norm in place of the classical l_2 norm. We analyzed two linear programming reformulations of the l_1 minimization problem and addressed issues related to the linear structure induced by the optical flow problem in the context of primal-dual interior point methods. Here we briefly present the (LP) formulations given in [23] and discuss about the sparse structure of these programs, a structure that can be easily exploited by parallel solvers.

In [23], we considered the discrete version of the classical Horn-Schunck model (HS) [26] with the optical flow energy functional being expressed as:

$$E_2^{(HS)}(u,v) = \sum_{\substack{i=\overline{1,W}\\ j=\overline{1,H}}} \left(I_{i,j}^x \cdot u_{i,j} + I_{i,j}^y \cdot v_{i,j} + I_{i,j}^t\right)^2 + \lambda \left(|\nabla u_{i,j}|^2 + |\nabla v_{i,j}|^2\right)$$

where W and H are the width and height (in pixels) of the images, respectively; $I_{i,j}^x$, $I_{i,j}^y$ and $I_{i,j}^t$ are the first-order partial derivatives with respect to the spatial coordinates x, y and time t, respectively, of the image intensity I at some fixed time t_0 at pixel (i,j). The image intensity is a time-dependent matrix of the form

$$I(t) = (I_{i,j}(t))_{j=\overline{1,H}}^{i=\overline{1,W}}.$$

The vectors $u = (u_{i,j})_{j=\overline{1,H}}^{i=\overline{1,W}}$ and $v = (v_{i,j})_{j=\overline{1,H}}^{i=\overline{1,W}}$ are the (unknown) horizontal and vertical components, respectively, of the optical flow. The parameter λ is a regularization parameter that controls the influence of the spatial term in the energy functional, while $\nabla u = (\nabla u_{i,j})_{j=\overline{1,H}}^{i=\overline{1,W}}$ and $\nabla v = (\nabla v_{i,j})_{j=\overline{1,H}}^{i=\overline{1,W}}$ are discrete linear approximations of the gradient of the optical flow components. The problem is to find u, v that minimize $E_2^{(HS)}$.

The minimization of $E_2^{(HS)}(u, v)$ is a linear least-squares problem of the form:

$$\min_{x \in \mathbb{R}^n} \|Ax + b\|_2^2 \tag{2.5.1}$$

for an appropriate choice of the matrix A and the vector b. The matrix A is a structured sparse matrix of size $(m, n) = (5N, 2N)$ which depends on the approximations $\nabla u = (\nabla u_{i,j})_{j=\overline{1,H}}^{i=\overline{1,W}}$ and $\nabla v = (\nabla v_{i,j})_{j=\overline{1,H}}^{i=\overline{1,W}}$.

Since least squares approach (2.5.1) is sensitive to outliers, an l_1 functional of the form

$$E_1^{(HS)}(u, v) = \sum_{\substack{i=\overline{1,W} \\ j=\overline{1,H}}} \left| I_{i,j}^x \cdot u_{i,j} + I_{i,j}^y \cdot v_{i,j} + I_{i,j}^t \right| + \lambda(\left|u_{i,j}^x\right| + \left|v_{i,j}^x\right| + \left|u_{i,j}^y\right| + \left|v_{i,j}^y\right|)$$

could be more useful in many situations. The minimization of $E_1^{(HS)}$ can be stated as the following l_1 optimization problem

$$\min_{x \in \mathbb{R}^n} \|Ax + b\|_1 , \tag{2.5.2}$$

where the matrix A is dependent on the approximations used for the gradient components $u_{i,j}^x$, $u_{i,j}^y$, $v_{i,j}^x$ and $v_{i,j}^y$.

In [23], we have used a five-point approximation formula for the first order derivates and obtained the matrix A with the form:

$$A = \begin{pmatrix} \delta(I_x) & \delta(I_y) \\ R & O_{N,N} \\ O_{N,N} & R \\ \widetilde{S} & O_{N,N} \\ O_{N,N} & \widetilde{S} \end{pmatrix} \in M_{m,n}(\mathbb{R}) \tag{2.5.3}$$

$$b = \begin{pmatrix} I_t \\ O_{4N,1} \end{pmatrix} \in \mathbb{R}^m$$

In (2.5.3), $\delta(x)$ denotes a diagonal matrix with the elements on its diagonal equal to the components of the vector x. The matrix $O_{N,N}$ is the $N \times N$ zero matrix and the matrices R and $\widetilde{S}$ are matrices of appropriate dimensions with a sparse diagonal

structure. We refer the reader to [23] for more details on how these matrices were constructed.

To obtain a linear programming (LP) formulation of the l_1–minimization problem (2.5.2), we considered a decomposition of the vector $Ax + b$ into its negative and positive parts. More precisely, let $w \in \mathbb{R}^m$, $w := Ax+b$ and consider the decomposition $w = w^+ - w^-$, where $w^+ \in {\mathbb{R}_+}^m$ and $w^- \in {\mathbb{R}_+}^m$ are the nonnegative and negative parts of w, respectively. Using the new variables w^+ and w^-, the objective function of (2.5.2) is written in the form

$$\begin{array}{rcl} ||Ax+b||_1 & = & ||w||_1 \\ & = & (e_m)^T w^+ + (e_m)^T w^- \end{array} \tag{2.5.4}$$

where $e_m \in \mathbb{R}^m$ is a vector of all ones. The vectors $\overline{u} \in \mathbb{R}^{n+2m}$ and $c_1 \in \mathbb{R}^{n+2m}$ are constructed such that they to have the block-partitioned form:

$$\overline{u} = \left[x^T, (w^+)^T, (w^-)^T\right]^T \text{ and } c_1 = \left[0, (e_m)^T (e_m)^T\right]^T.$$

Let D be the $(2m)\times(n+2m)$ matrix with the partitioned form $D = [0 \;\; I_{2m}]$, where I_{2m} denotes the $2m \times 2m$ identity matrix. We also define B to be the $m\times(n+2m)$ matrix given by $B = [A \;\; -I_m \;\; I_m]$. This partition led us to the following LP reformulation of problem (2.5.2):

$$\begin{array}{ll} \min\limits_{\overline{u}\in\mathbb{R}^{n+2m}} & c_1^T \overline{u} \\ \text{such that} & D\overline{u} \geq 0 \\ & B\overline{u} + b = 0 \end{array} \tag{2.5.5}$$

As it can be easily observed the matrix D above, introduces only lower bound constraints on a subset of the variables and therefore it is sometimes more useful to rewrite problem (2.5.5) in the following form

$$\begin{array}{ll} \min\limits_{\overline{u}\in\Omega} & c_1^T \overline{u} \\ \text{such that} & B\overline{u} + b = 0, \end{array} \tag{2.5.6}$$

where $\Omega = \left\{\overline{u} \in \mathbb{R}^{n+2m} \mid \overline{u}_{n+i} \geq 0, i = \overline{1, 2m}\right\}$. The formulation (2.5.6) can be exploited by bound constrained minimization solvers.

If one eliminates the x-variable from (2.5.5), an equivalent LP formulation, exploited by M. Rus in [41], is obtained. Here we recall that the x-variable is given by the first (unrestricted by sign) block of the unkonown vector $\overline{u}$. The LP used in [41] can be obtained by considering a basis of the null space of A^T to be represented by the columns of $A^\perp$ and by setting $\overline{v} = \left[(w^+)^T, (w^-)^T\right]^T$. This way, the equivalent LP formulation of [41] is given by

$$\begin{array}{ll} \min\limits_{\overline{v}\in\mathbb{R}^{2m}} & (e_{2m})^T \overline{v} \\ \text{such that} & \overline{v} \geq 0 \\ & \left[\left(A^\perp\right)^T \;\; \left(-A^\perp\right)^T\right] \overline{v} = \left(A^\perp\right)^T b \end{array} \tag{2.5.7}$$

We note that reaching the LP formulation (2.5.7) carries the additional cost of computing the matrix $A^{\perp}$. Since the sparsity of matrix A is also lost by $A^{\perp}$, formulation (2.5.5) is more attractive from a computational point of view. We will focus here on this formulation and briefly describe the linear system that are solved by typical primal-dual interior point methods.

We briefly present here how the structure of the linear systems specific to primal-dual interior problems applied to (2.5.5) can be exploited from a computational point of view. For this, in [23], we followed the general framework for primal-dual interior point methods of [35], and considered the underlying linear system to be solved at iteration $(k+1)$ of the form

$$\begin{pmatrix} 0 & 0 & A^T & 0 \\ 0 & 0 & \widetilde{B}^T & I_m \\ A & \widetilde{B} & 0 & 0 \\ 0 & S & 0 & V \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \overline{v} \\ \Delta \lambda \\ \Delta s \end{pmatrix} = - \begin{pmatrix} r_x \\ r_{\overline{v}} \\ r_\lambda \\ r_s \end{pmatrix}, \tag{2.5.8}$$

where $\widetilde{B} \in \mathcal{M}_{m,2m}(\mathbb{R})$ is given by

$$\widetilde{B} = (-I_m \; I_m)\,.$$

The vectors λ and s represent Lagrange multipliers associated with the equality constraints and the non-negativity constraints respectively. This implies that s and $\overline{v}$ must satisfiy the complementarity condition $0 \leq s \perp \overline{v} \geq 0$. In primal-dual interior point methods, the iterates $s^{(k)}$, $\overline{v}^{(k)}$ are kept positive, i.e., $s^{(k)} > 0$ and $\overline{v}^{(k)} > 0$, where the inequalities are to be understood componentwise. The $\Delta\lambda$ and the Δs components of the solution to the linear system (2.5.8) represent the directions used in obtaining the new iterates $\lambda^{(k+1)}$ and $s^{(k+1)}$, respectively. The right-hand side of (2.5.8) is a vector of residuals, [35]. The matrices S and V are diagonal matrices of the form:

$$S = \text{diag}\left(s_1^{(k)}, ..., s_{2m}^{(k)}\right),\; V = \text{diag}\left(\overline{v}_1^{(k)}, ..., \overline{v}_{2m}^{(k)}\right).$$

By subsequent elimination one can obtain expression for $\Delta\lambda$, $\Delta\overline{v}$ and Δs as affine functions of Δx. The quantity Δx can be obtained by solving a linear system of the form

$$\left(A^T \widehat{D}^{-1} A\right) \Delta x = \widetilde{b}, \tag{2.5.9}$$

where a complete form of $\widetilde{b}$ is given in [23]. Here $\widehat{D}$ is a diagonal matrix with positive entries on the diagonal, that was obtained during the elimination process.

Since A is a sparse matrix with the block diagonal structure (2.5.3) and D is a diagonal matrix, the sparse block structure of $\left(A^T \widehat{D}^{-1} A\right)$ can be exploited by doing parallel Cholesky factorizations for sub-matrices of A, similar to what was

done in [12]. Therefore the formulation (2.5.5) can be easily exploited by linear algebra parallel algorithms in the context of interior point methods.

The numerical solution to the second LP formulation (2.5.6) is also discussed in [23], and we refer the reader to this paper for more details. It was noted however that by computing the matrix $A^{\perp}$ the sparsity of matrix A is lost and the advantage of using *parallel* linear algebra algorithms becomes less obvious.

2.5.2 A Chebysev–Grüss type inequality

In this section we briefly present some of the results obtained in [18]. In this paper we designed a Grüss type inequality based on the concave majorant of the classical modulus of continuity, for the case of two linear positive functionals which preserve the constants. Applications to Bernstein and Bernstein-Stancu operators were given in [18] as well as a probabilistic interpretation of the new results. Here, we briefly describe the results and emphasize their probabilistic interpretation.

The classical Grüss inequality can be stated as follows: if we have two integrable functions $f, g : [a, b] \to \mathbb{R}$, satisfying $m_f \leq f(x) \leq M_f$, $m_g \leq g(x) \leq M_g$, $\forall x \in [a, b]$, then the following inequality holds

$$\left| \frac{1}{b-a} \int_a^b f(x)g(x)dx - \frac{1}{(b-a)^2} \left(\int_a^b f(x)dx \right) \left(\int_a^b g(x)dx \right) \right| \leq \frac{1}{4} (M_f - m_f)(M_g - m_g). \tag{2.5.10}$$

If (X, d) is a compact metric space, $C(X)$ is the space of all continuous real valued functions defined on X and $A : C(X) \to \mathbb{R}$ is a linear positive functional which preserves the constants, then the classical Grüss inequality (2.5.10) can be restated as

$$|A(fg) - A(f)A(g)| \leq \frac{1}{4}(M_f - m_f)(M_g - m_g), \tag{2.5.11}$$

where f and g are functions in $C(X)$ satisfying $m_f \leq f(x) \leq M_f$, $m_g \leq g(x) \leq M_g$, $\forall x \in X$.

The results in [18] were motivated by the following *probabilistic* inequality which involves the covariance of two random variables. More precisely, let (U, V) be a two dimensional continuous random vector with the joint probability density function (joint pdf) $\rho_{U,V} : [a, b] \times [a, b] \to \mathbb{R}$ and the marginal pdfs $\rho_U : [a, b] \to \mathbb{R}$, $\rho_V : [a, b] \to \mathbb{R}$ given by

$$\rho_U(u) = \int_a^b \rho_{U,V}(u, v)dv, \ \rho_V(v) = \int_a^b \rho_{U,V}(u, v)du.$$

Assume that $m_U \leq U \leq M_U$ and $m_V \leq V \leq M_V$. Let $X = [a,b] \times [a,b]$ and $A : C(X) \to \mathbb{R}$ be the linear positive functional given by

$$A(h) = \int_a^b \int_a^b h(u,v)\rho_{U,V}(u,v)dudv. \tag{2.5.12}$$

It is straightforward to see that A preserves the constants and therefore by applying (2.5.11) with A given by (2.5.12) and $f, g : [a,b] \times [a,b] \to \mathbb{R}$ given by

$$f(u,v) = u,\ g(u,v) = v,$$

we obtain

$$\left| \int_a^b \int_a^b uv\rho_{U,V}(u,v)dudv - \left(\int_a^b u\rho_U(u)du \right) \left(\int_a^b v\rho_V(v)dv \right) \right| \leq \frac{1}{4} (M_U - m_U)(M_V - m_V),$$

or

$$|\mathrm{COV}[U,V]| \leq \frac{1}{4} (M_U - m_U)(M_V - m_V), \tag{2.5.13}$$

where $\mathrm{COV}[U,V]$ denotes the covariance of the random variables U and V. Inequality (2.5.13) is known as the *probabilistic* Grüss inequality and can be used to estimate the covariance $\mathrm{COV}[U,V]$.

The inequalities obtained in [18] use the least concave majorant of the modulus of continuity. We give here the definitions for the modulus of continuity and its least concave majorant and then we present the main results of [18]. For this, let (X,d) be a compact metric space and $C(X)$ the Banach lattice of continuous real-valued functions defined on the compact metric space X. Let $f \in C(X)$. For $t \in [0,\infty)$, the usual modulus of continuity of the function f in the point t is defined by

$$\omega_d(f;t) := \sup\{|f(x) - f(y)|,\ d(x,y) \leq t\}$$

and its least concave majorant is given by

$$\widetilde{\omega}_d(f;t) = \begin{cases} \sup_{0 \leq \alpha \leq t \leq \beta \leq d(X), \alpha \neq \beta} \frac{(t-\alpha)\omega_d(f;\beta) + (\beta - t)\omega_d(f;\alpha)}{\beta - \alpha}, & \text{for } 0 \leq t \leq d(X), \\ \omega_d(f;d(X)), & \text{if } t > d(X). \end{cases}$$

When X is a compact subset of $\mathbb{R}^n$ and d is the Euclidean metric, we will use the short notations $\omega(f;t)$ and $\widetilde{\omega}(f;t)$ instead of $\omega_d(f;t)$ and $\widetilde{\omega}_d(f;t)$, respectively.

We will say that a functional $A : C(X) \to \mathbb{R}$ is positive if $A(f) \geq 0$, for all $f \geq 0$ (all functions with nonnegative values). Let $A, B : C(X) \to \mathbb{R}$ be two positive linear functional satisfying $A(e_0) = B(e_0) = 1$. Here $e_0 : X \to \mathbb{R}$, $e_0(x) = 1,\ x \in X$. In [18], bounds of the quantity $D_{A,B}(f,g)$ defined by

$$D_{A,B}(f,g) = A(fg) + B(fg) - A(f)B(g) - B(f)A(g)$$

were obtained by using the least concave majorants of the moduli of continuity for the functions that define this quantity. It is easy to observe that if $A = B$, then

$$D_{A,A}(f,g) = 2(A(fg) - A(f)A(g)),$$

which except for a multiplicative constant is equal to the left-hand side of (2.5.11). In what follows we give the main results of [18] and then we focus on their probabilistic interpretation.

The first result from [18], that we present here, is bounding the quantity $D_{A,B}(f,g)$, where the functions f and g are continuous over a compact metric space (X,d).

Theorem 2.5.1. *Let $f, g \in C(X)$ be two continuous functions on the compact metric space (X,d). If A, B are two positive linear functionals, $A, B : C(X) \to \mathbb{R}$ reproducing constants ($A(e_0) = B(e_0) = 1$), then the following inequality*

$$|D_{A,B}(f,g)| \leq \widetilde{\omega}_d\left(f; \sqrt{A_x B_y(d^2(x,y))}\right) \widetilde{\omega}_d\left(g; \sqrt{A_x B_y(d^2(x,y))}\right) \tag{2.5.14}$$

holds.

It is not hard to see that the inequality (2.5.14) is sharp, since for $X = [a,b]$ and $f = g = e_1$, an equality is obtained. The second result is concerned with the case $X = [0,1]$ and it is given by the following theorem

Theorem 2.5.2. *If $A, B : C[0,1] \to \mathbb{R}$ are two linear positive functionals satisfying $A(e_0) = B(e_0) = 1$, then the following inequality*

$$|D_{A,B}(f,g)| \leq \widetilde{\omega}\left(f; \sqrt{D_{A,B}(e_1,e_1)}\right) \widetilde{\omega}\left(g; \sqrt{D_{A,B}(e_1,e_1)}\right) \tag{2.5.15}$$

holds. Moreover, if $A(e_1) = B(e_1)$, then

$$D_{A,B}(e_1,e_1) \leq \frac{1}{2}, \tag{2.5.16}$$

with equality if and only if $A = B = \frac{1}{2}(\delta_0 + \delta_1)$, where $\delta_x(f) = f(x)$, $x \in [0,1]$ is the Dirac δ-functional.

In the last part of this section we present how these Grüss type inequalities can be used to obtain some probabilistic results. For the proofs of the theorems above as well as for other connected results, we refer the reader to [18]. Now, let (U,V) and $(\widetilde{U},\widetilde{V})$ be two dimensional continuous vectors with joint probability density functions (pdfs) given by $\rho_{U,V} : [a,b] \times [a,b] \to \mathbb{R}_+$ and $\rho_{\widetilde{U},\widetilde{V}} : [a,b] \times [a,b] \to \mathbb{R}_+$

respectively. Let $X = [a,b] \times [a,b]$ and consider the linear positive functionals $A, B : C(X) \to \mathbb{R}$ induced by the joint pdfs of (U,V) and $(\widetilde{U},\widetilde{V})$, i.e.,

$$A(h) = \int_a^b \int_a^b h(u,v)\rho_{U,V}(u,v)dudv, \quad B(h) = \int_a^b \int_a^b h(u,v)\rho_{\widetilde{U},\widetilde{V}}(u,v)dudv. \quad (2.5.17)$$

Let $f, g \in C(X)$. Then, with A and B defined by (2.5.17), we have

$$\begin{aligned} D_{A,B}(f,g) &= COV\left[f(U), g(V)\right] + COV\left[f(\widetilde{U}), g(\widetilde{V})\right] \\ &+ \left(E\left[g(V)\right] - E[g(\widetilde{V})]\right)\left(E\left[f(U)\right] - E[f(\widetilde{U})]\right), \end{aligned}$$

where $E[W]$ denotes the expectation of the random variable W. For $d(\cdot,\cdot)$ being the Euclidean distance in $\mathbb{R}^2$ and $u, v \in \mathbb{R}^2$, $u = (u_1, u_2),\ v = (v_1, v_2)$, after some calculations (see [18] for details), we obtain

$$\begin{aligned} A_u B_v\left(d^2(u,v)\right) &= \sigma_U^2 + \sigma_V^2 + \sigma_{\widetilde{U}}^2 + \sigma_{\widetilde{V}}^2 \\ &+ \left(E[U] - E[\widetilde{U}]\right)^2 + \left(E[V] - E[\widetilde{V}]\right)^2, \end{aligned}$$

where $\sigma_W^2 := VAR[W]$ denotes the variance (dispersion) of the random variable W. Here the subscript in A_u refers to the fact that the linear functional A acts on the variable u. If Theorem 2.5.1 is used now, then the following inequality, involving covariances, is obtained

$$\begin{aligned} &\Big|COV\left[f(U), g(V)\right] + COV\left[f(\widetilde{U}), g(\widetilde{V})\right] \\ &+ \left(E\left[g(V)\right] - E[g(\widetilde{V})]\right)\left(E\left[f(U)\right] - E[f(\widetilde{U})]\right)\Big| \\ &\leq \widetilde{\omega}\left(f; \sqrt{\tau}\right)\widetilde{\omega}\left(g; \sqrt{\tau}\right), \end{aligned} \quad (2.5.18)$$

where $\tau := \tau_{U,V,\widetilde{U},\widetilde{V}}$ is given by

$$\tau_{U,V,\widetilde{U},\widetilde{V}} = \sigma_U^2 + \sigma_V^2 + \sigma_{\widetilde{U}}^2 + \sigma_{\widetilde{V}}^2 + \left(E[U] - E[\widetilde{U}]\right)^2 + \left(E[V] - E[\widetilde{V}]\right)^2. \quad (2.5.19)$$

We have the following special cases of the inequality (2.5.18).

- If a single random vector (U,V) is considered, i.e., $\rho_{U,V} \equiv \rho_{\widetilde{U},\widetilde{V}}$, then (2.5.18–2.5.19) becomes

$$|COV(f(U), g(V))| \leq \frac{1}{2}\widetilde{\omega}\left(f; \sqrt{2\left(\sigma_U^2 + \sigma_V^2\right)}\right)\widetilde{\omega}\left(g; \sqrt{2\left(\sigma_U^2 + \sigma_V^2\right)}\right). \quad (2.5.20)$$

  We note that inequality (2.5.20) establishes bounds for the covariance using the concave majorant of the modulus of continuity.

- For the case when f and g are Lipschitz functions with Lipschitz constants L_f and L_g respectively, one can bound $\widetilde{\omega}\left(f;\tau_{U,V,\widetilde{U},\widetilde{V}}\right)$ and $\widetilde{\omega}\left(g;\tau_{U,V,\widetilde{U},\widetilde{V}}\right)$ by $L_f\tau_{U,V,\widetilde{U},\widetilde{V}}$ and $L_g\tau_{U,V,\widetilde{U},\widetilde{V}}$ respectively, to obtain

$$\begin{aligned}&\Big|COV[f(U),g(V)]+COV[f(\widetilde{U}),g(\widetilde{V})]\\&+(E[g(V)]-E[g(\widetilde{V})])(E[f(U)]-E[f(\widetilde{U})])\Big|\\&\leq L_f\ L_g\ \tau_{U,V,\widetilde{U},\widetilde{V}}.\end{aligned}$$

  In particular, for a single random vector, inequality (2.5.20) can be relaxed to

$$|COV(f(U),g(V))|\leq L_fL_g\left(\sigma_U^2+\sigma_V^2\right).$$

2.5.3 A Hermite-Hadamard type inequality

In [19], several new inequalities of the Hermite-Hadamard type were obtained. Here, we present one such result that can be immediately used in the estimation of moments of continuous random variables. We first introduce the classical Hermite-Hadamard inequality. For this, let $f:I\to\mathbb{R}$ be a convex function on the interval $I\subseteq\mathbb{R}$ and let $a,b\in I$ satisfying $a<b$.The inequality

$$f\left(\frac{a+b}{2}\right)\leq\frac{1}{b-a}\int_a^b f(x)dx\leq\frac{f(a)+f(b)}{2}\tag{2.5.21}$$

is well known in the literature as Hermite-Hadamard's inequality. One of the important issues that comes up (see for example [27,28,36,47]) consists in obtaining sharp bounds for

$$\left|f\left(\frac{a+b}{2}\right)-\frac{1}{b-a}\int_a^b f(x)dx\right|\ \text{ and }\ \left|\frac{f(a)+f(b)}{2}-\frac{1}{b-a}\int_a^b f(x)dx\right|.$$

In [19], several such bounds were obtained. We focus here on one such inequality that has a direct application to the estimation of moments of continuous random variables. For this, we let $w:[a,b]\to[0,\infty)$ be a continuous function such that

$$\int_a^b w(x)dx=1.$$

The first-order moment of w is denoted by a_1 and defined by

$$a_1=\int_a^b xw(x)dx.$$

The following result was obtained in [19] for differentiable functions f with convex $|f'|$. We list here this result and then discuss its implications in the estimation of moments of continuous random variables.

Theorem 2.5.3. *Let f be a differentiable mapping on $[a,b]$. If $|f'|$ is convex on $[a,b]$, then the following inequality holds:*

$$\left|\int_a^b w(x)f(x)dx - f(a_1)\right| \leq \frac{|f'(a)|\,A(a,b) + |f'(b)|\,B(a,b)}{b-a}, \tag{2.5.22}$$

where

$$\begin{aligned} A(a,b) &= \frac{1}{2}\int_a^b (x-a_1)^2 w(x)dx - \int_a^{a_1}(x-a_1)^2 w(x)dx \\ &+ 2(b-a_1)\int_{a_1}^b (x-a_1)w(x)dx, \\ B(a,b) &= \frac{1}{2}\int_a^b (x-a_1)^2 w(x)dx - \int_a^{a_1}(x-a_1)^2 w(x)dx \\ &+ 2(a_1-a)\int_{a_1}^b (x-a_1)w(x)dx. \end{aligned}$$

If we assume the hypothesis of Theorem 2.5.3 and the weight function $w(x)$ is symmetric with respect to the midpoint $\dfrac{a+b}{2}$, then the following inequality was obtained in [19].

$$\left|\int_a^b f(x)w(x)dx - f\left(\frac{a+b}{2}\right)\right| \leq \frac{|f'(a)|+|f'(b)|}{b-a}\int_{\frac{a+b}{2}}^b \left(x-\frac{a+b}{2}\right)w(x)dx$$

We conclude the short note on the Hermite -Hadamard inequalities obtained in [19] with an application to random variables. Let $0<a<b$, $r\in\mathbb{R}$ and let X to be a continuous random variable with probability density function $w:[a,b]\to[0,\infty)$. We denote the r-th moment of X the following integral

$$E_r(X) := \int_a^b t^r w(t)dt,$$

which is assumed to be finite. Then, Theorem 2.5.3 can be used to obtain the following estimates for the moments.

Theorem 2.5.4. *For $r\geq 2$, the following inequality holds*

$$|E_r(X) - (E(X))^r| \leq \frac{r}{b-a}\left(|a|^{r-1}A(a,b) + |b|^{r-1}B(a,b)\right), \tag{2.5.23}$$

where

$$A(a,b) = \frac{1}{2}\int_a^b (x-E(X))^2 w(x) - \int_a^{E(X)} (x-E(X))^2 w(x)dx$$
$$+2(b-E(X))\int_{E(X)}^b (x-E(X))w(x)dx,$$
$$B(a,b) = \frac{1}{2}\int_a^b (x-E(X))^2 w(x) - \int_{E(X)}^b (x-E(X))^2 w(x)dx$$
$$+2(E(X)-a)\int_{E(X)}^b (x-E(X))w(x)dx,$$

and $E(X)$ *denotes the first-order moment (expectation) of the random variable* X.

Some similar results were obtained in [22] and [20]. In [20], a mean value result was given for the Chebyshev functional. The mean value theorem, together with Chebysev-Grüss type inequalities can be used in statistics, if the weight function is associated with a probability density function. In [22], the inequalities obtained can be used to generalize some Hadamard–type inequalities. The results obtained here can be applied successfully to some generalized means. We refer the reader to [22] and [20] fo more details.

Chapter 3

Further research

In this last chapter we present some of the lines that characterizes the present and future work. There are several directions that we describe here. First we will discuss about a class of stochastic optimization problems and numerical algorithms that can be used to solve these problems. We are interested both in theoretical issues such as convergence of the numerical schemes as well as practical issues such as computational performance. This is joint work with Mihai Anitescu, from Argonne National Laboratory, USA.

The next item is part of the current work and deals with a simple modification of Newton's method for solving scalar equations. The modification gives rise to a family of methods which can be used in a randomized setting. The advantage of using such a randomized environment consists in the fact that the resulting Newton-Monte-Carlo method converges for cases where the standard Newton method does not.

An other line of work that we follow is related to the simulation and control of rigid body systems. The goal is to use LCP-based time-stepping schemes for obtaining the numerical trajectory of some systems such as the ones appearing in autonomous navigation. Besides obtaining the numerical trajectories, we plan to use the LCP structure in designing robust controls in the presence of uncertainty. Auxiliary results that we plan to obtain in the future are related to mathematical inequalities with applications to probability and statistics. Some of these results could be used in any of the problems described above, since all of them treat uncertainty as part of their model.

3.1 A class of stochastic optimization problems

In this section we look at a class of stochastic optimization problems. These problems are known as stochastic optimization problems with mixed expectations and per-scenario constraints or briefly SOESCs. More precisely, one can write such problems in the following form:

$$\begin{array}{rl} \min\limits_{x\in K_x, y(\omega)\in K_y} & E_\omega\left[\phi(x,y(\omega))\right] \\ \text{such that} & 0 = E_\omega\left[\psi(x,y(\omega))\right]. \\ & 0 = \Gamma(x,y(\omega)),\ \forall \omega\in\Omega, \end{array} \tag{3.1.1}$$

Here

$$\begin{array}{l} \phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, \\ \psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p, \\ \Gamma : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q \end{array}$$

are differentiable functions. The sets K_x and K_y are considered to be convex and closed. The mapping $E_\omega[\cdot]$ is the expectation operator. The stochasticity is given by ω, where $\omega \in \Omega$ and $(\Omega, \mathcal{A}, P)$ is a probability space.

Applications formulated as SOESC include portfolio optimization and stochastic receding horizon control of constrained systems. We note that both inequality and complementarity constraints can be accommodated for appropriate choices of K_x and K_y.

In order to write problem (3.1.1) in a more compact form, we define the following quantities

$$\begin{array}{l} z := (x, y(\cdot)),\ K_z := K_x \times K_W, \\ \widetilde{f}(z) := \mathbb{E}_\omega\left[\phi(x,y(\omega))\right],\ \widetilde{g}(z) := \Gamma(x,y(\omega)), \\ \widetilde{h}(z) := \mathbb{E}_\omega\left[\psi(x,y(\omega))\right]. \end{array} \tag{3.1.2}$$

Here we assumed that the functions $y(\cdot)$ belong to a functional space W and we denote by K_W, the subset of W containing those functions that have their images included in K_y. We note that the objective function $\widetilde{f}(z)$ and the function $\widetilde{h}(z)$ are constant with respect to ω since the expectation operator was used, while the function $\widetilde{g}(z)$ giving the per-scenario constraints is ω–dependent. Here $\widetilde{f} : Z \to \mathbb{R}$, $\widetilde{g} : Z \to Y$ and $\widetilde{h} : Z \to \mathbb{R}^m$. We assume that $(Y, \langle\cdot,\cdot\rangle_Y)$, $(Z, \langle\cdot,\cdot\rangle_Z)$ are Hilbert spaces (even though this condition could be relaxed) and $K_z \subseteq Z$ is a closed convex cone with non-empty interior.

With the notation introduced above, we can rewrite problem (3.1.1) in the more compact form

$$\begin{array}{rl} \min\limits_{z\in K_z} & \widetilde{f}(z) \\ \text{such that} & 0 = \widetilde{g}(z). \\ & 0 = \widetilde{h}(z). \end{array} \tag{3.1.3}$$

By considering a sample of size N and replacing the expectation operator by the sample mean, we obtain the sampling average approximation, SAA(N). More precisely in the process of solving the SOESC (3.1.1), we solve a sequence (indexed by N) of finite dimensional programs of the form

$$\begin{array}{ll} \min\limits_{x\in K_x, y(\omega_i)\in K_y,\ i=\overline{1,n}} & \frac{1}{N}\sum_{i=1}^{N}\Phi(x,y(\omega_i)) \\ \text{such that} & 0=\frac{1}{N}\sum_{i=1}^{N}\psi(x,y(\omega_i)). \\ & 0=\Gamma(x,y(\omega_i)),\ \forall i\in\{1,...,N\}, \end{array} \tag{3.1.4}$$

One of the theoretical issues that appear here is related to the convergence of the approximations given by $SAA(N)$. For this, one can rewrite (3.1.4) in a compact form similar to (3.1.3) and perform a sensitivity analysis by considering parametric variational inequalities associated with the optimality conditions of the programs above.

A different problem, which we are interested in, consists in obtaining computational efficient algorithms for solving quadratic SAA-type programs. For this, let us assume that the SOESC, objective function is quadratic and the constraints are affine. More precisely,

$$\begin{array}{rcl} \phi(x,y(\omega),\omega) & = & \frac{1}{2}\left(x^TQ_0x+y(\omega)^T\widetilde{Q}(\omega)y(\omega)\right)+c_0^Tx+\widetilde{c}(\omega)^Ty(\omega), \\ \psi(x,y(\omega),\omega) & = & T_0x-b_0+\widetilde{T}(\omega)y(\omega)-\widetilde{b}(\omega), \\ \Gamma(x,y(\omega),\omega) & = & W_0x-d_0+\widetilde{W}(\omega)y(\omega)-\widetilde{d}(\omega). \end{array}$$

If we allow only nonnegative values for x and $y(\omega)$, then for a sample of size N, the QP obtained by sampling average approximations has the form

$$\begin{array}{ll} \min\limits_{x\geq 0, y_i\geq 0, i=\overline{1,N}} & \frac{1}{2}x^TQ_0x+c_0^Tx+\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}y_i^T\widetilde{Q}_iy_i+\widetilde{c}_i^Ty_i\right) \\ \text{such that} & T_0x+\frac{1}{N}\sum_{i=1}^{N}\widetilde{T}_iy_i=\overline{b}+b_0, \\ & W_0x+\widetilde{W}_iy_i=\widetilde{d}_i+d_0,\ i=\overline{1,N}, \end{array} \tag{3.1.5}$$

where $\widetilde{Q}_i := \widetilde{Q}(\omega_i)$, $\widetilde{T}_i := \widetilde{T}(\omega_i)$, $\widetilde{W}_i := \widetilde{W}(\omega_i)$. In a similar fashion one defines the vectors $\widetilde{c}_i$ and $\widetilde{d}_i$. The vector $\overline{b}$ represents the sample mean of the selection $\{\widetilde{b}(\omega_1),...,\widetilde{b}(\omega_N)\}$, i.e.,

$$\overline{b}=\frac{1}{N}\sum_{i=1}^{N}\widetilde{b}(\omega_i).$$

We rescale some of these quantities as follows:

$$Q_i:=\frac{1}{N}\widetilde{Q}_i,\ T_i:=\frac{1}{N}\widetilde{T}_i \text{ and } c_i:=\frac{1}{N}\widetilde{c}_i.$$

If we put now the unknowns in the vector u, so that $u = \left(x^T, y_1^T, ..., y_N^T\right)^T$, we obtain the following reformulation of (3.1.5):

$$\begin{array}{ll} \min\limits_{u \geq 0} & \frac{1}{2} u^T Q u + c^T u \\ \text{such that} & Au = b, \end{array} \tag{3.1.6}$$

where the matrices A and Q have the block partitioned form

$$Q = \begin{pmatrix} Q_0 & & & \\ & Q_1 & & \\ & & \ddots & \\ & & & Q_N \end{pmatrix}, \; A = \begin{pmatrix} T_0 & T_1 & T_2 & ... & T_N \\ W_0 & \widetilde{W}_1 & 0 & ... & 0 \\ W_0 & 0 & \widetilde{W}_2 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ W_0 & 0 & 0 & ... & \widetilde{W}_N \end{pmatrix}$$

and the vectors c and b are given by

$$\begin{aligned} c &= \left(c_0^T, c_1^T, ..., c_N^T\right)^T \\ b &= \left(\overline{b}^T + b_0^T, \widetilde{d}_1^T + d_0^T, ..., \widetilde{d}_N^T + d_0^T\right)^T \end{aligned}$$

We are interested in obtaining computational efficient algorithms for the numerical solution of (3.1.6) by exploiting its linear structure. If an analysis similar to the one performed in [38] is performed in the context of interior point methods (IPM), then the inner linear systems that need to be solved at each IPM iteration, will be determined by matrices of the form

$$K := \begin{pmatrix} K_1 & & & B_1 \\ & \ddots & & \\ & & K_N & B_N \\ B_1^T & ... & B_N^T & K_0 \end{pmatrix}, \tag{3.1.7}$$

Here the matrices K_0, K_i and B_i for $i = \overline{1, N}$ are given by

$$K_0 := \begin{pmatrix} Q_0 + D_0 & T_0^T \\ T_0 & 0 \end{pmatrix}, K_i := \begin{pmatrix} Q_i + D_i & \widetilde{W}_i^T \\ \widetilde{W}_i & 0 \end{pmatrix}, B_i := \begin{pmatrix} 0 & T_i^T \\ W_0 & 0 \end{pmatrix},$$

and $D_0, D_1, ..., D_N$ are diagonal matrices with positive entries.

One can use the particular block-angular structure of the matrix K from (3.1.7), to design efficient parallel algorithms. The steps we plan to take in order to obtain such an algorithm are similar to the ones given in [38], namely:

- *Schur complement calculation*. In this step, the matrix C, representing the Schur complement of K_0 is computed.

- *Schur linear systems.* Here we compute the solution of the corresponding Schur linear system.
- *Solving K_i subsystems.* In this step we solve some linear systems with K_i as the underlying matrices and depending on the solution of the Schur linear systems above.

Following the steps above we plan to design and implement parallel algorithms for solving quadratic SOESC as the ones described here.

3.2 A Newton-Monte Carlo method for solving nonlinear scalar equations

The research we intend here is related to a modified Newton-Raphson method for solving nonlinear scalar equations. The method was developed by Trevor J. McDougall and Simon J. Wotherspoon in [32] and it is a predictor-corrector algorithm with $1+\sqrt{2}$ order of convergence.

We consider the nonlinear scalar equation

$$f(x) = 0, \tag{3.2.1}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a function of class C^2 in a neigborhood of a root r that satisfies $f(r) = 0$ and $f'(r) \neq 0$. The method of [32] initiated with the simple observation that if $f(x)$ is quadratic , then

$$r = x_0 - \frac{f(x_0)}{f'\left(\frac{1}{2}(x_0 + r)\right)}. \tag{3.2.2}$$

The identity (3.2.2) was then used in designing both the predictor and corrector steps in the modified Newton's method of [32]. More precisely, the authors in [32] showed that the method defined by

$$\begin{aligned}
x_0^* &= x_0, \\
x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)}, \\
x_k^* &= x_k - \frac{f(x_k)}{f'\left(\frac{1}{2}(x_{k-1} + x_{k-1}^*)\right)} \text{ (predictor), } k \geq 1, \\
x_{k+1} &= x_k - \frac{f(x_k)}{f'\left(\frac{1}{2}(x_k + x_k^*)\right)} \text{ (corrector), } k \geq 1
\end{aligned}$$

has order of convergence $1+\sqrt{2}$ when x_0 is chosen sufficiently close to r.

The method we are designing uses a parameter $\gamma_k \in (0,1)$, so that at each step k, $k \geq 1$, the derivative in the predictor and corrector step is evaluated at $(1-\gamma_k)x_{k-1}+\gamma_k x^*_{k-1}$ and $(1-\gamma_k)x_k+\gamma_k x^*_k$, respectively. More precisely the iterations can be defined as follows,

$$x^*_0 = x_0, \tag{3.2.3}$$

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \tag{3.2.4}$$

$$x^*_k = x_k - \frac{f(x_k)}{f'\left((1-\gamma_k)x_{k-1}+\gamma_k x^*_{k-1}\right)} \text{ (predictor), } k \geq 1, \tag{3.2.5}$$

$$x_{k+1} = x_k - \frac{f(x_k)}{f'\left((1-\gamma_k)x_k+\gamma_k x^*_k\right)} \text{ (corrector), } k \geq 1 \tag{3.2.6}$$

By looking at (3.2.3–3.2.6), we can make the following remarks:

- We note that if γ_k is set to $1/2$ for all values of k, then the method of McDougall and Wotherspoon is recovered.

- One may generalize the method above even more by introducing different weighting factors for the predictor and corrector steps above. More precisely, a different parameter $\widetilde{\gamma}_k$ may be used in the corrector step. For the sake of simplicity we, will use the same parameter γ_k for both predictor and corrector steps.

- One can choose the parameters γ_k randomly, which adds the Monte-Carlo flavor to our algorithm. More precisely, γ_k will be considered a realization of a random variable $\widetilde{\Gamma}$, that follows the uniform distribution on the interval $[0,1]$, i.e., $\widetilde{\Gamma} \sim Unif(0,1)$. We note that the expectation of $\widetilde{\Gamma}$ equals $1/2$, i.e., $E[\widetilde{\Gamma}] = \frac{1}{2}$ and we can argue that "in the mean" the method performs with an order of convergence of $1+\sqrt{2}$, i.e., equal to the one characterizing the method of McDougall and Wotherspoon.

Our asymptotic analysis shows that the proposed method has an (local) order of convergence equal to 2. Preliminary numerical experiments show an order of convergence slightly better than 2, when the γ_k–s are selected according to the uniform distribution on $[0,1]$. But this can be explained by the fact that the mean of the γ_k–s is $\overline{\gamma} = \frac{1}{2}$, i.e., the value found in the method of McDougall and Wotherspoon, which has the order of convergence $1+\sqrt{2} > 2$.

The Newton-Monte Carlo algorithm implied by the last observation above can now be described. We will consider N runs, each run starting from a given x_0 and from a deterministically computed x_1. The algorithm described below depends on

two user defined parameters: `maxIter` and `tolFun`. The first one gives the maximum iterations allowed in every run, while the second one is used as a stopping criteria, i.e., if $|f(x_k)| \leq$ `tolFun`, then we consider that the approximation x_k is acceptable.

Newton-Monte-Carlo Algorithm.

>User defined parameters:
  >N:= number of runs
  >maxIter:= maximum number of iterations in every run
  >tolFun:= tolerance used on the function values
>Let $x_0^* = x_0$ and compute x_1 from a pure Newton step
>for i=1:N
  >k=1;
  >repeat
    >randomly select γ_k from the uniform distribution on [0,1];
    >compute x_k^* according to (3.2.5);
    >compute x_{k+1} according to (3.2.6);
    >$k := k + 1$;
  >until $k >$ maxIter or $|f(x_k)| \leq$ tolFun
>endfor

Related to the algorithm presented above we study two questions that relate to the practical performance of this algorithm. These questions are

- How well does the method perform when the test functions are "well behaved"?

- Is the method able to numerically compute solutions for functions with oscillatory behavior?

Here, by "well behaved" function we mean a function (together with a starting point) for which the classical Newton method converges, while "badly behaved " situations will refer to those functions and starting points for which the classical Newton-Raphson method fails to converge. Preliminary numerical results show that the Newton-Monte Carlo algorithm proposed here, converges in both situations, even when the classical Newton-Raphson method fails to converge. We intend to adress some other issues such as *error estimation*, any correlation between the computational effort (number of runs, average number of iterations,etc.) and some properties of the function giving the nonlinear scalar equation (3.2.1). We also intend to address the issue of efficiently extending this algorithm to systems of nonlinear equations.

3.3 Rigid body simulation in autonomous navigation

The research intended here is based on developing LCP integration schemes that will lead to the selection of robust controls in autonomous navigation. In our opinion, the LCP–based strategy can be useful even in the case when friction is neglected. More precisely, for the application discussed here, collisional events or contacts between two or more autonomous vehicles are not desirable, in general. Therefore, in the absence of friction between the autonomous vehicle and the supporting surface on which this vehicle moves, there is *no real contact problem* and therefore no apparent need for complementarity formulations.

Virtual contacts may be created by embedding each element from the processed scene in a *"safety cage"* or *"safety box"*. Additional virtual elements such as "guidance elements" may be artificially introduced in the scene. Now, between the "guidance elements" and the "safety cages" contacts are allowed. This gives rise to differential complementarity formulations (DCPs) that describe the dynamics. Time-stepping schemes similar to the ones presented in the previous sections may be used to numerically integrate these DCPs and generate a (numerical) reference trajectory. To control the autonomous system, these reference trajectories are used.

Algorithms similar to the ones presented in Section 2.4 for a quasi-static setting can be used in a dynamic setting. More precisely, one can use complementarity matrices to determine if a trajectory piece or a vector of applied controls is robust or not. We recall that lack of robustness was corresponding to a change of complementarity matrix. This is similar to a mode switch in the hybrid systems context. This way, we can determine and use only robust controls in a randomized planning setting.

Some other items of interest are:

- Design and analysis of new time-stepping schemes with higher order of convergence for index 2 and index 3 rigid body DAEs,

- Design and analysis of time-stepping schemes for rigid body systems, based on *cone complementarity problems*,

- Obtaining convergence results in the measure differential inclusion sense for rigid body systems that experience partially elastic contacts (so far we have considered only inelastic collisions).

3.4 Conclusions

In the present work we have described results related to the simulation of rigid body systems with contact and friction. Both theoretical and computational issues were addressed. Since uncertainty is an important part in the problems studied here, auxiliary results such as mathematical inequalities with applications to probability and statistics were also given.

A class of linearly implicit time-stepping schemes, [21], was presented here. The schemes solve a mixed linear complementarity problem (MLCP) at each integration step. To obtain the MLCP formulation, the full (nonlinear) frictional cone is replaced by a polyhedral approximation. We addressed the issue of convergence of these time-stepping schemes in the context of measure differential inclusions. One of the key assumptions in proving the convergence results is given by the *uniform pointedness of the total friction cone*. The analysis performed in [21] introduces two new concepts the *reduced friction cone* and the corresponding *reduced measure differential inclusion*.

In [37], a quadratic programming (QP) based model for granular flow simulation inside of a pebble bed reactor was analyzed from a computational point of view. The integration step is formulated as a QP, which is obtained via the convex relaxation of [2]. Both the primal and the dual formulations are considered. The dual program is a bound-constrained QP, which allows for the use of some bound constrained minimization solvers. Again the pointedness of the friction cone plays an important role and it is shown that together with the feasibility of the primal QP guarantees no duality gap.

In [14], [13] and [17] rigid body systems in a quasi-static setting are analyzed. In such a framework, Newton' s second law which is specific to a dynamical setting is replaced by an equilibrium equation. The application described in [14] and [13] is known as the peg-in-the-hole problem at the meso-scale. A planar rigid part (peg) is manipulated by pushing operations and the task is to take the peg from an initial sensed configuration A to the goal configuration B. The quasi-static model is motivated by the fact that inertial forces are one order smaller than the frictional forces. There are uncertainties due to sensing, modelling and parameter estimation. Since uncertainties play an important role in such a system, designing controls that will prove robust in the presence of uncertainties is extremely important. In [14] and [13], robust manipulation primitives were designed mostly based on geometric considerations. The robust controls can be used then in the context of randomized planning. Here, we have also shown how the analysis of the LCP structure may lead to the selection of robust controls. More precisely, one can use complementarity matrices to decide whether the applied controls are robust or not. A change of the complementarity matrix will be equivalent to a switching event and the corresponding control will be rejected.

In [23], an l_1 minimization problem from optical flow was presented. The optimization problem is obtained from the discrete version of the classical Horn-Schunck model. The standard l_2 energy functional is replaced by its l_1 counterpart. For the l_1 minimization problem, two linear programming reformulations were analyzed in [23]. Here we have presented the lines of our analysis for the LP with a better structure from the matrix sparsity point of view. Primal dual interior point methods (IPM) can be used to solve this LP and the sparse structure specific to this optical flow problem can be exploited in the context of parallel algorithms.

In [18], a Chebysev–Grüss type inequality was given. The results developed here use the modulus of continuity and its least concave majorant, for the case of two linear positive functionals which preserve the constants. The new results of [18] can be used in various probabilistic applications. We have described here how these results can be used in the estimation of covariances for different pairs of random variables. In [19], several new inequalities of the Hermite-Hadamard type were obtained. Here, we presented one such result, that can be immediately used in the estimation of moments of continuous random variables.

We conclude with a list of items that are part of our current and future research. These research lines will be coupled with auxiliary results such as mathematical inequalities with applications to probability and statistics. We summarize the list of our current and future research interests as follows:

- Design and implementation of LCP based time-stepping schemes for autonomous navigation. As it was explained above, this can be done by allowing virtual contacts. In this context, we plan in using the underlying LCP structure of the integration step in deciding whether switching events occur or not. No switching events will mean ”robust” controls.

- Design of randomized algorithms for solving systems of nonlinear equation. A brief description of a Newton-Monte Carlo algorithm for the scalar case was presented here.

- Theoretical and computational results related to a class of stochastic optimization problems with mixed expectation and per-scenario constraints (abbreviated by SOESC). From an analytical point of view, we are interested in convergence results for sample average approximations applied to SOESC problems. From a computational point of view, our interest resides in designing parallel algorithms that would exploit the particular structure of these problems in the context of interior point methods. Some preliminary results were given in Section 3.1.

- Design and analysis of new time-stepping schemes for both index 2 and index 3 rigid body DAEs.

- Design and analysis of higher order time-stepping schemes for rigid body systems with contact and friction, based on cone-complementarity problems.
- Convergence results in the measure differential inclusion sense for systems experiencing partially elastic collisions.

Index

ϵ–active friction cone, 40
l_1 energy functional, 55
l_2–energy functional, 55

basic solution, 11
BLMVM, 41

complementarity basis, 10
complementarity cone, 10
complementarity matrix, 10
compression phase, 23
copositive matrix, 12

decompression phase, 24
differentiation index of a DAE, 14
dual QP, 40

friction cone, 8

Grüss inequality, 58

Hermite-Hadamard inequality, 62
Horn-Schunck model, 54

interior point methods, 68
IPM, 68

lcp, 10
least concave majorant, 59
linear complementarity problem, 10

MDI, 17
measure differential inclusion, 17
mixed LCP, 12
MLCP, 12
modulus of continuity, 59
MOSEK, 42

NCP, 9
nonlinear complementarity problem, 9

OOQP, 41

pointedness of the friction cone, 26
Poisson restitution model, 23
primal QP, 40

TRON, 41

vector measure, 17

weak solution of an MDI, 18

Bibliography

[1] E. Andersen and Y. Ye. On a homogeneous algorithm for the monotone complementarity problem. *Mathematical Programming*, 84(2):375–399, 1999.

[2] M. Anitescu. Optimization-based simulation of nonsmooth dynamics. *Mathematical Programming*, 105:113–143, 2006.

[3] M. Anitescu, J. F. Cremer, and F. A. Potra. On the existence of solutions to complementarity formulations of contact problems with friction. In Michael C. Ferris and Jong-Shi Pang, editors, *Complementarity and Variational Problems: State of the Art*, pages 12–21, Philadelphia, 1997. SIAM Publications.

[4] M. Anitescu and G. D. Hart. Solving nonconvex problems of multibody dynamics with joints, contact and small friction by sequential convex relaxation. *Mechanics Based Design of Machines and Structures*, 31(3):335–356, 2003.

[5] M. Anitescu and G. D. Hart. A constraint-stabilized time-stepping approach for rigid multibody dynamics with joints, contact and friction. *International Journal for Numerical Methods in Engineering*, page to be determined by the publisher, 2004.

[6] M. Anitescu and F. A. Potra. Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems. *Nonlinear Dynamics*, 14:231–247, 1997.

[7] M. Anitescu and F. A. Potra. A time-stepping method for stiff multibody dynamics with contact and friction. *International Journal for Numerical Methods in Engineering*, 55(7):753–784, 2002.

[8] M. Anitescu and F. A. Potra. Time-stepping schemes for stiff multi-rigid-body dynamics with contact and friction. *International Journal for Numerical Methods in Engineering*, 55(7):753–784, 2002.

[9] S. J. Benson and J. Moré. A limited-memory variable-metric algorithm for bound-constrained minimization. Technical Report ANL/MCS-P909-0901, Mathematics and Computer Science Division, Argonne National Laboratory, 2001.

[10] K. E. Brenan, S. L. Campbell, and L. R. Petzold. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. North-Holland, New York, 1989.

[11] J. C. and G. Vaněček. Isaac: Building simulations for virtual environments. In *Workshop IFIP IC 5 WG 5.10 on Virtual Environments*, Coimbra, Portugal, 1994.

[12] T. D. Cao, J. F. Hall, and R. A. Van De Geijn. Parallel Cholesky factorization of a block tridiagonal matrix. *Proceedings. International Conference on Parallel Processing Workshops*, pages 327–335, 2002.

[13] D.J. Cappelleri, Peng Cheng, J. Fink, B. Gavrea, and V. Kumar. Automated assembly for mesoscale parts. *Automation Science and Engineering, IEEE Transactions on*, 8(3):598–613, 2011.

[14] Peng Cheng, D. J. Cappelleri, B. Gavrea, and V. Kumar. Planning and control of meso-scale manipulation tasks with uncertainties. In *Robotics: Science and Systems III, June 27-30, 2007, Georgia Institute of Technology, Atlanta, Georgia, USA*, 2007.

[15] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. Academic Press, Boston, MA, 1992.

[16] B. Gavrea. Switching events in rigid-body time-stepping schemes. *Automation Computers Applied Mathematics*, 19(1):121 – 126, 2010.

[17] B. Gavrea. Rigid body time-stepping schemes in a quasi-static setting. *Stud. Univ. Babeş-Bolyai Math.*, 56(2):1–11, 2011.

[18] B. Gavrea. Improvement of some inequalities of Chebysev-Grüss type. *Computers and Mathematics with Applications*, 64(6):2003 – 2010, 2012.

[19] B. Gavrea. A Hermite-Hadamard type inequality with applications to the estimation of moments of continuous random variables. *Applied Mathematics and Computation*, 254:92 – 98, 2015.

[20] B. Gavrea. A mean value theorem for the Chebysev functional. *Mathematical Inequalities and Applications*, 18(2):751 – 757, 2015.

[21] B. Gavrea, M. Anitescu, and F. Potra. Convergence of a class of semi-implicit time-stepping schemes for nonsmooth rigid multibody dynamics. *SIAM J. Optim*, 19(2):969–1001, 2008.

[22] B. Gavrea, J. Jakšetić, and J. Pečarić. On a global upper bound for Jessen's inequality. *ANZIAM*, 50:246 – 257, 2008.

[23] B. Gavrea and M.D. Rus. On an l_1–minimization problem from optical flow. *Automation Computers Applied Mathematics*, 22(1):327–335, 2013.

[24] E. M. Gertz and S. J. Wright. Object-oriented software for quadratic programming. *ACM Trans. Math. Softw.*, 29(1):58–81, 2003.

[25] H. D. Gougar. *Advanced core design and fuel management for pebble-bed reactors*. Ph.D in Nuclear Engineering, Department of Nuclear Engineering, Penn State University, 2004.

[26] B. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[27] U.S. Kirmaci. Inequalities for differentiable mappings and applications to special means of real numbers and to midpoint formula. *Applied Mathematics and Computation*, 147:137–146, 2004.

[28] U.S. Kirmaci and M.E. Ozdemir. On some inequalities for differentiable mappings and applications to special means of real numbers and to midpoint formula. *Applied Mathematics and Computation*, 153:361–368, 2004.

[29] S. LaValle. Rapidly-exploring random trees: A new tool for path planning. online, 1998.

[30] C-J. Lin and J. J. Moré. Newton's method for large bound-constrained optimization problems. *SIAM Journal on Optimization*, 9(4):1100–1127, 1999.

[31] C-J. Lin and J.J. Moré. Incomplete Cholesky Factorizations with Limited Memory. *SIAM Journal on Scientific Computing*, 21(1):24–45, 1999.

[32] T. J. McDougall and S. J. Wotherspon. A simple modification of Newton s method to achieve convergence of order $1 + \sqrt{2}$. *Applied Mathematics Letters*, 29:20–25, 2014.

[33] R. M. Murray, Z. Li, and S. S. Sastry. *A mathematical introduction to robotic manipulation*. CRC Press, Boca Raton, FL, 1993.

[34] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[35] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[36] C.E.M. Pearce and J. Pecaric. Inequalities for differentiable mappings with applications to quadrature formulae. *Appl. Math. Lett.*, 13:51 – 55, 2000.

[37] C. Petra, B. Gavrea, M. Anitescu, and F. Potra. A computational study of the use of an optimization-based method for simulating large multibody systems. *Optimization Methods and Software*, 24(6):871–894, 2009.

[38] C. G. Petra, O. Schenk, M. Lubin, and K. Gaertner. An augmented incomplete factorization approach for computing the schur complement in stochastic optimization. *SIAM Journal on Scientific Computing*, 36(2):C139–C162, 2014.

[39] F. Pfeiffer and Ch. Glocker. *Multibody dynamics with unilateral contacts*. Wiley Series in Nonlinear Science. John Wiley & Sons Inc., New York, 1996. A Wiley-Interscience Publication.

[40] F. Potra, M. Anitescu, B. Gavrea, and J. Trinkle. A linearly implicit trapezoidal method for integrating stiff multibody dynamics with contact, joints and friction. *International Journal for Numerical Methods in Engineering*, 66(7):1079–1124, 2005.

[41] M. D. Rus. Optimization methods for l_1-energy minimization in the estimation of optical flow. *Carpathian J. Math.*, 29(1):109–117, 2013.

[42] C.H. Rycroft, M.Z. Bazant, G.S. Grest, and J.W. Landry. Dynamics of random packings in granular flow. *Physical Review E*, 73(5):51306, 2006.

[43] D. E. Stewart. Existence of solutions to rigig body dynamics and the Painlevé paradoxes. *C. R. Acad. Sci. Paris*, 325:689–693, 1997.

[44] D. E. Stewart. Convergence of a time-stepping scheme for rigid body dynamics and resolution of Painlevé's problems. *Archive Rational Mechanics and Analysis*, 145(3):215–260, 1998.

[45] D. E. Stewart. Rigid-body dynamics with friction and impact. *SIAM Review*, 42(1):3–39, 2000.

[46] D. E. Stewart and J. C. Trinkle. An implicit time-stepping scheme for rigid-body dynamics with inelastic collisions and Coulomb friction. *International Journal for Numerical Methods in Engineering*, 39:2673–2691, 1996.

[47] G-S. Yang, D-Y. Hwang, and K l. Tseng. Some inequalities for differentiable convex and concave mappings. *Comput. Math. Appl.*, 47:207–216, 2004.