# Mathematical and numerical modeling on the Sweeping process : applications to contact dynamics

Vo Anh Thuong Nguyen

# THÈSE

Pour obtenir le grade de

**Docteur**

Délivré par

**UNIVERSITE DE PERPIGNAN VIA DOMITIA**

Préparée au sein de l'école doctorale :

**Énergie et Environnement ED 305**

Et de l'unité de recherche :

**LAMPS**

Spécialité : **Mathématiques Appliquées**

Présentée par **NGUYEN VO ANH THUONG**

# Mathematical and numerical modeling on the Sweeping process: applications to contact dynamics

Soutenue le 12 Novembre 2024 devant le jury composé de

| | |
|---|---|
| M. Stéphane ABIDE, PR, Université Côte d'Azur | Co-encadrant |
| M. Serge DUMONT, PR, Université de Perpignan via Domitia | Co-encadrant |
| M. Florent NACRY, MCF, Université de Perpignan via Domitia | Co-encadrant |
| M. Samir ADLY, PR, Université de Limoges | Rapporteur |
| M. Mathieu RENOUF, Chargé de recherche CNRS, Université de Montpellier | Rapporteur |
| M. Terence BAYEN, PR, Avignon Université | Examinateur |
| Mme. Aline LEFEBVRE-LEPOT, Directrice de Recherche, CNRS Fédération de Mathématiques de CentraleSupélec | Examinateur |
| M. Mircea SOFONEA, PR, Université de Perpignan via Domitia | Examinateur |
| M. Mikaël BARBOTEU, PR, Université de Perpignan via Domitia | Directeur, Invité |

# Remerciements

Je tiens à exprimer ma profonde gratitude envers toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de cette thèse.

Tout d'abord, je remercie sincèrement mes directeurs de thèse, Mikaël BARBOTEU, Serge DUMONT, Florent NACRY et Stéphane ABIDE, pour leur soutien constant, leurs conseils avisés et leur patience inépuisable tout au long de ce parcours. Leur expertise et leur enthousiasme ont été une source d'inspiration continue.

Je tiens à exprimer ma sincère gratitude à tous les membres du jury de ma thèse, à savoir les rapporteurs Samir ADLY et Mathieu RENOUF, ainsi que les évaluateurs Terence BAYEN, Aline LEFEBVRE-LEPOT et Mircea SOFONEA. Ils ont pris le temps et investi des efforts considérables pour examiner minutieusement mon travail et m'ont apporté des commentaires précieux qui ont grandement contribué à enrichir et à renforcer cette recherche.

Un immense merci à mes collègues et amis du *Laboratoire de Modélisation Pluridisciplinaire et Simulations* (LAMPS) pour leur aide précieuse, leurs discussions stimulantes et leur amitié. Vos encouragements et votre bonne humeur ont rendu ce chemin bien plus agréable.

Je tiens également à remercier mes chers amis Thibault Deverge, Asma Souala, Mélanie Dreina, et bien d'autres. Vous avez été des amis formidables, toujours présents pour m'encourager, partager des moments précieux et m'offrir un soutien indéfectible.

Je souhaite exprimer ma profonde gratitude à mes enseignants et collègues de l'Université de Quy Nhon, qui m'ont toujours soutenu et prodigué de précieux conseils tout au long de mon parcours. De même, un grand merci aux professeurs de l'Université de Limoges, qui m'ont fourni des bases solides avant d'entamer cette thèse.

Je n'oublierai jamais non plus mes amis au Vietnam, notamment Phuoc Tien, Huyen Trang, Tran Duc, ainsi que tous mes collègues, pour leur soutien constant, leur amitié sincère et leurs échanges précieux.

Enfin, je remercie ma famille pour leur soutien inébranlable, leur compréhension et leurs sacrifices. Votre amour et vos encouragements ont été la clé de voûte de ma réussite.

Je remercie également l'École Doctorale 305 - *Énergie Environnement* de l'Université de Perpignan Via Domitia pour leur soutien matériel et financier qui a rendu ce projet possible.

À toutes et à tous, un grand merci.

Nguyen Vo Anh Thuong

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 General context

Mathematically, evolution problems such as contact dynamics are, in most cases, governed by differential inclusions, or more precisely, measure differential inclusions. One of the most famous differential inclusions is known as the **Sweeping Process** and was introduced by J.J. Moreau in 1971 in the famous Convex Analysis Seminar in Montpellier [42]. This evolution problem continues to be widely studied today from both a theoretical and numerical point of view (see [8, 9, 12–22, 29, 30, 44, 47] and references therein). The Sweeping process, which is conditioned by inequality constraints (ivolving problems such as granular media, elastodynamics with contact), is a fundamental tool for mathematical analysis and the design of numerical schemes for these types of applications related to non-regular dynamics. In this subject, we are interested in the Sweeping process with theoretical aspects (mathematical model, solution existence, ...) and numerical aspects (scheme development, system energy analysis, ...) and its application, especially Contact dynamics. The unilateral contact condition in terms of velocity is crucial in the development of Non-Smooth Contact Dynamics methods. The lack of smoothness means that contact laws are not only non-differentiable in the usual sense, but also multivalued. This condition makes it possible to obtain very good energy conservation properties, ensuring physical and numerical stability. In other words, unilateral constraints are written as a relationship between velocity and momentum. In this way, the mechanical power of the system is described in a coherent way. This type of the second-order Sweeping process is a direct extension of Newton's impact law (Moreau's law), written for the first time in a numerically discrete time. The main idea is to extend the mathematical models in the form of measurable differential inclusions for *granular media* and deformable materials to thermodynamics, that is, the transfer of heat in granular media and thermoelasticity for deformable media. Using the previous models and using both the Sweeping process and the Newton-Active set Semi-smooth method, we want to implement a numerical scheme dedicated to the time integration of non-smooth systems with contacts.

In theoretical terms, investigating the properties of the moving set in a sweeping process also plays a crucial role. It helps to expand and develop the understanding of problems, particularly the existence of solutions, in non-convex scenarios, such as prox-regularity or its dual-class, strong convexity. Here, metric regularity serves as a

pivotal tool for studying the weak convexity of specific subsets. This is significant as it provides a robust theoretical foundation for addressing optimization and mechanical problems.

## 1.2 Sweeping process

From a mechanical standpoint, we aim to analyze the movement of a point or particle confined within a convex closed moving set denoted as $C(t)$. Depending on the dynamics of this region, the point will either remain stationary if it avoids colliding with the boundary of the set, or it will be swept towards the interior of the domain while staying within $C(t)$ for all time points $t$. To ensure that the point remains within the region, its velocity must point inwards which is opposite to the velocity's normal direction relative to the boundary of the set. One powerful tool to handle this mechanical problem is Sweping process.

J.J. Moreau [42] introduced the mathematical formulation of the *Sweeping process* which involves *differential inclusions* incorporating the normal cone $N\big(C(t);\cdot\big)$ to the set $C(t)$. The objective is to find a trajectory solution $u : [0,T] \to X$ that satisfies the Cauchy condition, where $T > 0$ and a moving set $C(\cdot)$ exists in the Hilbert space $X$.

**Problem 1. The first order Moreau's sweeping process** is given by

$$\begin{cases} -\dot{u}(t) \in \mathrm{N}(C(t); u(t)); \\ u(0) = u_0 \in C(0); \\ u(t) \in C(t) \quad \forall\, t. \end{cases}$$

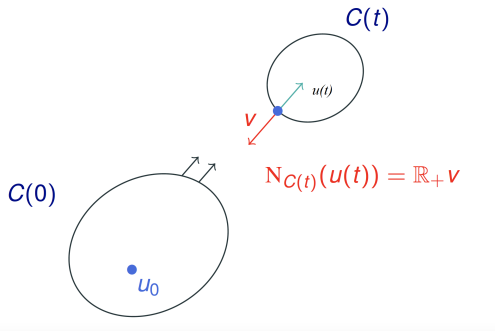



**Figure 1.1:** Mechanical interpretation of Moreau's sweeping process.

The existence and uniqueness of a solution of Problem 1 are established under the assumption that $C(t)$ is a convex closed moving set, and it is Lipschitz continuous in $t$. It is achieved through the *Catching-up Algorithm* (more details in [42]). Within the domain of convexity and mathematical optimization, the investigation of solutions to first and second-order sweeping process problems has engendered a vibrant and profound area of research. Kunze and Monteiro-Marques studied the existence of solutions when the sets have a Lipschitz variation by using a semi-implicit discretization scheme [12]. By using an explicit discretization, Haddad and al. [13] proved the existence of solutions of a perturbed state-dependent sweeping process. In addition, Bounkhel

and Castaing [14] obtained the state-dependent sweeping process in uniformly smooth and uniformly convex in Banach spaces. Later, the authors Jourani and Vilches ([9]) proved the existence of solutions of state-dependent sweeping process by using the Moreau envelope -Yosida regularization technique in the bounded variation continuous with nonregular (subsmooth and positively alpha-far) sets, etc.

In the concept of nonconvex case, it is well-known that Chemetov and Monteiro-Marques [15, 16] established existence of solutions of perturbed state-dependent sweeping process with uniformly prox-regular sets by using Schauder's fixed point theorem or a fixed point argument for the perturbed state-dependent sweeping process. Furthermore, Castaing et al. [17] demonstrated the existence of solutions for the state-dependent sweeping process within uniformly prox-regular sets. This was achieved by employing an extended form of Schauder's theorem in conjunction with a discretization method. Azzam-Laouir et al. [18] and Haddad et al. [19] showed the multivalued perturbed state-dependent sweeping process for uniformly prox-regular case in the finite-dimensional setting. Moreover, the authors Noel [20], Noel and Thibault [21] considered the existence of multivalued perturbed state-dependent sweeping process with equi-uniformly subsmooth and uniformly prox-regularity, etc. On the other hand, the Moreau envelope and Yosida regularization methods [9] have been employed specifically for convex or uniformly prox-regular sets to address the state-dependent sweeping process problem. In addition, related to these problems, both first and second order, notable authors with great contributions to this discourse include Moreau [44], Marques, Venel [29] and Schatzman [8], Acary, Brogliato and Goeleven [22], Maury and Venel [30], etc. and references therein.

Notably, sweeping processes find natural manifestations in a diverse spectrum of practical problems, including crowd motion dynamics, granular media behavior, elastoplastic modeling, contact dynamics, the evolution of sandpiles, and non-regular electrical circuitry, hysteresis, etc. These real-world applications underscore the relevance and ubiquity of sweeping process problems in scientific investigations and practical endeavors.

## 1.3 Solving Granular Media Issues: Computational Methods

### 1.3.1 Granular media problems

A granular media is known as a set of particles that can be small or large (from the grain of sand to a stone block of several meters), more or less complex in shape (sphere or angular polyhedron), and interact with each other through contacts. These materials are found in many industrial sectors with various problems, such as in solid: concrete or the ruin of a monumental building under seismic stress [70], tectonic plates or mixing of pharmaceutical pills [71]; in fluid: avalanche and landslide [72, 74], flow in silos [75], the ballast under cyclic loading model [76], heat production and transmission [78, 79], gluey granular [80], granular flows [81] etc. Researchers in various domains are drawn to granular systems, not only for their practical relevance but also due to their complex behavior.

**Figure 1.2:** Examples of granular materials.
Images: www.freeimages.com

Nowadays, the understanding of the behavior of granular media has become a significant concern for industrial processes or environmental science. One significant hurdle in granular media research is the absence of a comprehensive framework that can universally describe their properties, especially in multi-body, and multi-contact. This complex nature has led to the increasing use of numerical simulations, leveraging advances in computing power to analyze realistic and highly intricate granular systems rapidly, although ensuring the accuracy of the resulting numerical outcomes remains a challenge. Consequently, researchers need to make certain assumptions about the physical states of granular materials to study their behavior effectively.

In summary, granular media represents a fascinating field of study with practical implications spanning various industries and scientific disciplines. Researchers aim to uncover the fundamental principles governing these materials, paving the way for more efficient industrial processes and a deeper understanding of the natural world's intricacies.

### 1.3.2 Numerical resolution methods

In front of these challenges, tools adapted to model granular dynamics have emerged. The Discrete Element Method (DEM), originally called the Distinct Element Method as developed by Cundall [5] in 1971, was designed to simulate granular dynamics by tracking interactions between individual particles. Granular dynamics is then governed by Newton's second law of motion combined with a regular contact model. In particular, DEM simulations can be computationally expensive due to the explicit time scheme with small time steps. Subsequently, these methods have undergone substantial evolution and given rise to various derivatives, such as the Molecular Dynamics method (MD) [35–37]. The behavior of granular materials, though they could benefit from a description within the framework of classical continuum mechanics, remains inadequately addressed by such approaches today. Jean-Jacques Moreau made significant contributions to convex analysis, functions of bounded variation, differential measure theory, and the theory of sweeping processes - key mathematical tools for addressing

nonsmooth dynamics. He transformed these foundational theoretical concepts into an innovative nonsmooth implicit numerical method known as Contact Dynamics (CD) [38, 43, 45, 46]. This method was further extended by M. Jean [48] to accommodate deformable bodies, leading to the development of Non-Smooth Contact Dynamics (NSCD) methods.

NSCD [23–25, 38, 48] focuses on modeling and simulating contact forces and friction between bodies or particles. The core idea of NSCD is to treat the dynamics of the contact forces as an optimization problem, subject to constraints such as conservation of momentum and energy. The method involves introducing a set of unknown contact forces, which are represented as Lagrange multipliers, and then using optimization techniques to solve for the contact forces that satisfy the constraints. These are derived from Signorini's contact law in terms of velocities and impulses and Coulomb's friction law linking frictional force to the tangential component of velocity. Then, M. Jean and J. J. Moreau [38, 46–48] developed an iterative NonLinear Gauss-Seidel solver (NLGS) to determine the contact conditions between rigid bodies. This solver sequentially considers each contact until convergence is achieved. In addition, alternative solvers to NLGS, such as conjugate gradient solvers, have been proposed in [49, 50]. In particular, The nonlinear term of NSCD is computationally expensive compared to DEM, as emphasized in [10] even though NSCD allows for larger time steps than DEM, enabling the resolution of multiple simultaneous contacts. The techniques commonly used in the literature to handle nonlinearities arising from these conditions are based on several types of methods.

- Semi-Augmented Lagrangian (SAL) method [51, 52]
  SAL is used to solve the optimization problem arising from the system's dynamics, governed by the Signorini law - a simplified model assuming contact forces are always perpendicular to contact surfaces. It introduces Lagrange multipliers representing these forces and employs an augmented Lagrangian formulation to ensure force smoothness and compliance with problem's conditions.

- Bipotential method [53, 54]
  The Bipotential method is an extension of the classic Coulomb friction model to address systems with multiple interacting rigid bodies through friction. It utilizes Lagrange multipliers to represent contact forces and employs the bipotential formulation, ensuring energy conservation and stability.

- Projected conjugate gradient (PCG) method [50]
  This method is a numerical technique for solving constrained optimization problems, particularly in contact and granular media contexts. It combines the efficiency of the conjugate gradient method, which minimizes a function by following conjugate gradient directions to avoid the zigzagging of standard gradient descent, with projection techniques that handle constraints by projecting the current point onto the feasible set at each iteration. This ensures that constraints are respected throughout the optimization process, providing a robust and efficient solution framework.

- Primal-Dual Active Set (PDAS) method [55, 56, 58]
  It recently has emerged as promising and relevant approaches for solving de-

formable elastodynamic frictional contact problems [55, 56, 58], owing to their efficiency and ease of implementation. The main principle is: contact and friction conditions play like two elements of nonlinear complementarity functions, the solution of which is directly provided by Newton's Semi-Smooth approach (more details in [55, 62]). In the literature, there are few references to PDAS methods for solving rigid multi-body contact problems. PDAS methods for rigid multi-body systems truly began, by considering PDAS as a method for the local treatment of frictionless contact conditions within the framework of NSCD, with the work of M. Barboteu and S. Dumont [59]. Subsequently, PDAS has been developed and has exhibited a remarkable level of efficiency in comparison to alternative approaches, substantiated by references [60, 61].

Furthermore, Granular Element method (GEM) [32] is a variant of DEM which uses regularization laws and so falls under the category of smooth methods. The main difference between GEM and DEM is its resolution scheme, which resembles a stiffness method that was initially proposed by Kawai [31] and later extended by Kishino for quasi-static simulation of granular materials [32] and dynamics [33]. In Mathieu Renouf's thesis from the University of Montpellier in 2004 [34], GEM is discussed in detail, particularly concerning aspects of numerical integration. Although GEM primarily focuses on modeling the interactions between granular elements, the use of numerical integration methods, such as the Newmark method, is crucial for solving the equations of motion of particles over time. The Newmark method is a commonly used numerical integration scheme for solving ordinary differential equations that arise in the dynamics of structures and continuous media. It is especially popular in structural mechanics for integrating equations of motion due to its stability and its ability to control numerical damping. In the context of GEM, the Newmark method is used to integrate the equations of motion of granular particles over time. This integration scheme is based on an implicit or semi-implicit approach, allowing for stable handling of complex dynamic systems, particularly when multiple and nonlinear contact interactions are present.

The numerical integration scheme described in Mathieu Renouf's thesis involves the techniques used to advance the solutions of particle motion equations over time. The choice of integration scheme is crucial because it affects the accuracy and stability of simulations. Implicit methods like Newmark's are often favored in complex granular media simulations because they allow for larger time steps while maintaining stability, even in the presence of strong nonlinearities due to contact interactions. The Newmark method allows for control over integration accuracy by adjusting specific parameters, such as numerical damping, to avoid non-physical oscillations. In his thesis, Mathieu Renouf discussed the adaptation of the Newmark method for the specific needs of GEM, explaining how this method allows for realistic modeling of the dynamic behavior of granular particles while efficiently managing contact constraints.

## 1.4 Some theoretical aspects in variational analysis

Variational Analysis is a specialized field within mathematics that focuses on optimization, control theory, set-valued analysis, and related problems. It encompasses a broad range of topics, including convex analysis, nonlinear analysis, nonsmooth anal-

ysis, and set-valued theory. This area of study frequently integrates concepts from measure theory, differential geometry, and functional analysis. Variational analysis has a long and deep history which can be found in the books by J.P. Aubin and H. Frankowska [154], R.T. Rockafellar and R.J. B. Wets [120], B.S. Mordukhovich [112], J.P. Penot [116], A.D. Ioffe [107], L. Thibault [124] etc. It can be said that variational analysis is the extension and also the development of the classical Calculus of Variations and Convex Analysis to a more general theory. On the one hand, optimization problems often appear in application sciences. On the other hand, solving problems based on optimization is an effective method in mathematics. This makes Variational Analysis become an area of interest from both theoretical and applied perspectives.

The study of sweeping processes leads to analyzing weak and strong convexity, prox-regularity, and metric regularity, crucial for understanding solutions in non-convex scenarios. In our work, we mainly focus on two concepts in Variational Analysis:

- Weak and strong convexity: They involve sets where local minima are global minima with less stringent curvature requirements than strong convexity, which ensures uniform outward curvature. This is important for granular media and contact dynamics, where sets may not be uniformly convex but can be analyzed using weak convexity. One of the main concepts of weak convexity is prox regularity. It allows non-smooth or irregular sets to be handled in a convex optimization framework, ensuring well-defined normal cone and projection mappings, essential for numerical simulations of granular media. Chemetov and Monteiro-Marques, and Castaing et al., established solutions for perturbed sweeping processes in uniformly prox-regular sets [15, 16].

- Metric regularity: it bridges weak convexity and general set structures, measuring the distance function's behavior to a set, useful in optimization and variational analysis, such as understanding particle behavior in granular systems. Azzam-Laouir et al. [18] and Haddad et al. [19] demonstrated effective multivalued perturbed sweeping processes in uniformly prox-regular sets in finite dimensions.

Integrating these concepts enhances numerical models for complex scenarios, solidifies theoretical foundations, and extends sweeping process applicability in mechanics and optimization. Noel [20] and Thibault [21] explored state-dependent multivalued perturbed sweeping processes with uniform prox-regularity and used Moreau-Yosida regularization for uniformly prox-regular sets [9]. This transition enriches our theoretical and practical understanding of these phenomena, with metric regularity and the farthest distance function playing crucial roles in optimization and mechanics.

### 1.4.1 Weak and strong convexity

The existence of the solution to the sweeping process 1 is established under the assumption that C(t) is a closed convex moving set. A natural extension to achieve the existence of a solution is to consider the set C(t) as weakly convex or its dual, strongly convex. Consequently, investigating the properties of the moving set is also a significant subject of this thesis.

Prox-regularity has been recognized as a fundamental concept in variational analysis which allows to go beyond convexity (see in Poliquin, Rockafellar in [118], Thibault

in [124] and some related papers [127], [84], [125]). It arises numerous interesting applications related to various areas of mathematics such as Optimization, Set-Valued Analysis, Differential Geometry and PDE, etc (see, e.g., [30, 91, 93, 100, 106, 108] and the references therein). One important application is the Sweeping Process, governed by inequality constraints, such as those found in granular media and elastodynamics involving contact. This process is a powerful tool for the mathematical analysis and the development of numerical schemes for these non-regular dynamics. Furthermore, this kind of set is also applied to consider some particularly important properties such as metric (submetric and semimetric) regularity, calmness, transversality, Aubin property, etc. Readers can read more in the monographs [120], [100], [84], ... It is well-established in [115] that the complement of a prox-regular set is nothing but the union of a family of closed balls with a common radius, which are commonly referred to as $R$-strongly convex sets. Consequently, the study of prox-regularity naturally directs our attention towards the exploration of strong convexity, thereby deepening our comprehension of function curvature and optimization landscapes. Noteworthy contributors to this field include H. Frankowska, C. Olech [104], J.P. Vial [125], E.S. Polovinkin [119], M.V. Balashov [88], G. E. Ivanov [105, 108, 110], A. Weber and G. Reibig [126]. For the use of such sets in diverse applied mathematical problems, we refer, e.g., to [42, 96] for the Sweeping Process, [104] for optimal control, [111, 117] for numerical optimization, [108, 110] for differential games, etc.

Distance functions play a fundamental role in mathematical analysis, including convex analysis ([92, 116, 120]), variational analysis ([97, 107, 112, 116, 120, 124]), differential inclusions ([87, 98, 114, 123]), optimal control ([95, 113]), approximation theory ([102, 122]), shape analysis and PDE ([85, 86, 101]), etc. It is well-known (and easily seen) that the convexity of a given closed subset $C$ in a normed space is equivalent to the convexity of its associated (standard/usual) distance function, namely $d_C(x) := \inf_{c\in C} \|x - c\|$. One naturally considered question is what is the set for which the distance is semiconvex? In [88], M.V. Balashov extends in a general Hilbert space such a basic characterization by establishing that the semiconvexity (i.e., convexity up to a square norm) of the distance function $d_C$ (on any convex set of a suitable enlargement of $C$) is equivalent to the uniform prox-regularity of the closed set $C$. Recall that a closed subset $C$ in a Hilbert space $X$ is uniformly prox-regular ([118]) (also known as positively reached, weakly convex, $\varphi-$convex, $\mathcal{O}(2)$-convex, proximally smooth, see, e.g., [94, 99, 103, 121, 125] and the references therein) with constant $r > 0$ provided that the nearest point mapping $\mathrm{proj}_C$ is well defined on a suitable enlargement of $C$ (more precisely on the set $U_r(C) := \{x \in X : d_C(x) < r\}$) and continuous therein.

The distance function to a convex set is known to admit diverse descriptions involving duality relations from convex analysis (see, e.g., [102]). Besides the distance function, we study the farthest points and their corresponding distance functions, which measure the distance from an outside point to the farthest point of a given subset. Early research on these functions was introduced by Jessen ([156], 1940) and further research was performed by Motzkin, Strauss, and Valentine ([158], 1953) in a metric space and by Fitzpatrick ([157], 1980) in Banach space. In convex setting, while the nearest point projection to a convex set is unique, the farthest distance function, namely $\mathrm{dfar}_C(x) := \sup_{c\in C} \|x - c\|$, behaves differently. This operator presents a departure from the conventional notion of the nearest point, highlighting that the existence of

the farthest point is not guaranteed with the classical concepts of convex or weakly compact sets. It is natural to explore alternative sets that are better suited for determining the farthest point. In many situations, it is also of interest to strengthen the convexity property. This can be achieved with $R$-strongly convex sets ([105, 108, 119, 125, 126]) which are basically intersection of closed balls with common radius $R > 0$. In the work [89], it is established that for any closed bounded set in $X$, the strong convexity property is equivalent to the semiconcavity of the (convex) farthest distance function. Therefore, one idea is to investigate the semiconcavity property of the farthest distance function associated with a strongly convex set. Furthermore, we develop the analogs of some properties of convex sets [102] or prox-regular sets [84] for a strongly convex set $C$ when the set $C$ satisfies the strengthened condition of strong convexity.

### 1.4.2 Metric regularity

On the other hand, metric regularity [107, 112, 124, 131, 132, 139, 142, 144] is a concept in mathematical analysis, particularly in the field of optimization and variational analysis. It is closely related to the stability and robustness of solutions to optimization problems and systems of equations. Specifically, metric regularity helps in understanding how sensitive the optimal solutions are to perturbations in the parameters or constraints and also provides error bounds for approximate solutions to equations and inequalities. Moreover, in variation analysis, metric regularity can be a valuable tool for studying the weak convexity of a particular subset. Metric regularity is closely related to other concepts such as: Lipschitz Continuity [124], Clarke's Regularity [124, 161], Robinson-Ursescu Theorem [142, 144], etc.

Let $M : X \rightrightarrows Y$ be a multimapping between two Banach spaces with a closed convex graph $\{(x, y) \in X \times Y : y \in M(x)\} =: \operatorname{gph} M$ and let $(\overline{x}, \overline{y}) \in \operatorname{gph} M$. In 1975-1976, C. Ursescu ([144]) and S.M. Robinson ([142]) independently established that the existence of a real $\gamma \geq 0$ such that

$$d\big(x, M^{-1}(y)\big) \leq \gamma d\big(y, M(x)\big) \quad \text{for all } (x, y) \text{ near } (\overline{x}, \overline{y}) \tag{1.1}$$

is equivalent to the inclusion $\overline{y} \in \operatorname{core} M(X)$. When the above inequality (1.1) holds, one naturally says that the multimapping $M$ is $\gamma$-metrically regular at $\overline{x}$ for $\overline{y}$. Metric regularity property has a long and deep story which goes back to the pioneers works by L.A. Lyusternik ([139]) and L.M. Graves ([132]) and has been developed since in numerous papers and books (see, e.g., [107, 112, 124, 131] and the references therein). Such a property is known to be equivalent either to some Lipschitz behavior of the multimapping $M^{-1}$ or to an openness (with linear rate) type condition, namely the existence of positive constants $\alpha, \beta > 0$ such that

$$B[y, \alpha\beta t] \subset M(B[x, t\alpha]),$$

for all $t \in ]0, 1]$ and all $(x, y) \in \operatorname{gph} M$ near $(\overline{x}, \overline{y})$. Besides the so-called Robinson-Ursescu theorem (which can be viewed as an extension of the famous Banach-Schauder open mapping theorem) metric regularity is strongly involved in subdifferential calculus, estimates of coderivatives and optimality conditions (see, e.g., [112, 116, 124] and the references therein).

Over the years, Robinson-Ursescu type theorems for multimappings with possibly nonconvex graph have been provided. A first natural way to go beyond convexity in such a context lies in the concept of paraconvexity. Recall that the (above) multimapping $M$ is said to be $(\theta, C)$-paraconvex ([141]) for a real $\theta > 0$ and a real $C \geq 0$ whenever for all $x_1, x_2 \in X$, for all $t \in [0,1]$,

$$tM(x_1) + (1-t)M(x_2) \subset M(tx_1 + (1-t)x_2) + C \min(t, 1-t) \|x_1 - x_2\|^\theta \mathbb{B}_Y, \quad (1.2)$$

where $\mathbb{B}_Y$ denotes the closed unit ball of $Y$. It is worth pointing out that the latter class of multimappings contains the class of multimappings with convex graphs; in fact (as easily seen) the latter inclusion (1.2) with $C = 0$ characterizes the convexity of gph $M$. A. Jourani proved in 1996 ([136]) under the $\theta$-paraconvexity of the multimapping $M^{-1}$ with $\theta \geq 1$ that the inequality (1.1) is equivalent to the inclusion $\overline{y} \in \operatorname{int} M(X)$. In [138], H. Huang and R.X. Li established that if the inverse of the multimapping $M$, namely $M^{-1}$, is paraconvex, then $M$ is metrically regular at $\overline{x}$ for $\overline{y}$ whenever

$$B(\overline{y}, \beta) \subset M\big(B(\overline{x}, \alpha)\big) \quad (1.3)$$

for some reals $\alpha, \beta > 0$.

In 2012, X.Y. Zheng and K.F. Ng ([145]) showed that prox-regularity of sets ([118]) is also a suitable concept to develop nonconvex versions of Robinson-Ursescu theorem. More precisely, Zheng and Ng established in the Hilbert framework that if gph $M$ is $(r, \delta)$-prox-regular at $(\overline{x}, \overline{y})$, then $M$ is metrically regular at $\overline{x}$ for $\overline{y}$ whenever the inclusion (1.3) holds for some reals $\alpha \in ]0, \frac{\delta}{3}[$, $\beta \in ]0, \delta[$ satisfying the inequality $\beta > \frac{4\alpha^2+\beta^2}{2r}$. Here and below, $N^C(S; x)$ denotes the Clarke normal cone of a set $S \subset X$ at $x \in S$. Two years later, X.Y. Zheng and Q.H. He provided in [146] a Robinson-Ursescu type theorem for multimappings with some variational behavior of order one, namely with a $(\sigma, \delta)$-subsmooth ([129]) graph at $(\overline{x}, \overline{y})$.

As shown by [128], the above results of [145, 146] can be extended to the class of multimappings $M$ with normally $\omega(\cdot)$-regular graph, that is, for some function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$

$$\langle (x^\star, y^\star), (u, v) - (x, y) \rangle \leq \omega(\|(u, v) - (x, y)\|)$$

for appropriate points $u, v, x, y$ and unit normals $x^\star, y^\star$ (see Section 6.2 for the definition and more details). More precisely, it is established in [128] that a multimapping $M$ with a normally $\omega(\cdot)$-regular graph satisfying the openness condition (1.3) for some reals $\alpha, \beta, \rho > 0$ such that

$$\beta > \frac{3\alpha}{\rho} + \big(1 + \frac{1}{\rho}\big)\omega\Big(\sqrt{4\alpha^2 + (\beta - \frac{\alpha}{\rho})^2}\Big). \quad (1.4)$$

is $\gamma$-metrically regular at $\overline{x}$ for $\overline{y}$ for some real $\gamma \leq \rho$. The authors of [128] derive from their study various preservation results for $\omega(\cdot)$-normally regular sets which complement previous works devoted to the stability of prox-regularity and subsmoothness properties [100, 125, 127, 137, 152] (see also the recent paper by G.E. Ivanov [133]).

The Robinson-Ursescu Theorem has inspired various extensions that go beyond the confines of convexity, broadening the scope of metric regularity and its applications. In this thesis, it is interesting to provide normally regular versions of the Robin-Ursescu theorem and also general sufficient conditions ensuring the preservation of normal $\omega$-regularity under metric subregularity.

## 1.5 Thesis objectives and manuscript outline

### 1.5.1 Objectives

**Numerical part**

In this work, we are interested in a discontinuous (but bounded) variations, Moreau second-order sweeping process modeling the contact dynamics of rigid particles ([47] and related references [8, 23, 28, 147]). Recall that sweeping processes [42] are particular differential inclusions governed by the normal cone of a (nonconvex) moving set. The contact law is modeled through the Moreau-Yosida regularization [9, 27] of the unilateral condition. It seems that the Moreau-Yosida regularization with parameter $\alpha \geq 2$ is an appropriate tool to develop a regular contact model (normal compliance) which maintain the system's kinetic energy while ensuring that the contact does not allow objects to pass through each other (non-interpenetration). Based on the work of Hauret and Le Tallec [57], we propose a specific approach within the discrete framework that improves the energy preservation of the system, aligning it with the continuous case: the discrete **Improved Normal Compliance** (INC) which is considered as a suitable method for ensuring energy conservation within the continuous framework. To solve the nonlinearity issue, a combination of the Newmark method [64] and PDAS using complementarity of the various contact models [60] will be employed.

The main aim of the present work is to improve an implicit regularization method for which energy conservation and non-penetration are quite similar to NSCD-NLGS along with a suitable computational cost. Several numerical experiments are reported for verification and validation purposes, and also to evaluate the efficiency and assess the performances of *Newmark-PDAS-INC* method compared to other numerical methods (DEM, NSCD-NLGS).

Moreover, within the scope of our setting, a second pivotal objective involves developing a self-contained code comprehensively employed for the INC method, utilizing the FORTRAN language with the LAPACK library-`Dgesv`. Then, we apply several solvers for solving linearized systems to improve the computational cost, i.e. LU, complete factorization-`SuperLU`, `GMRES` with ILU Preconditioner, Preconditioned Conjugate Gradient (ILU preconditioner-`Pcg` or Jacobi preconditioner-`Jpcg`) by using SPARSKIT library. Besides, we also develop a direct version of the Sparse matrix combining either `SuperLU/GMRES` for friction and `Pcg/Jpcg` for frictionless case.

**Theorical part**

Besides the numerical aspect, we involve the theoretical part of Variational analysis to consider some new properties that are much more useful in mathematics. More precisely, we will focus on two main works.

In the first work, we consider the concepts of **Farthest distance function to strongly convex sets in Hilbert spaces**. The first aim is to give an alternative proof of such a characterization of strongly convex sets. Our approach is in the line of [115] based on the fact that the complement of a prox-regular set is nothing but the union of a family of closed balls with common radius. The second aim of the present

work is to provide in details the analysis of the corresponding features (i), (ii) and (iii) in part 1.4.1 when the set $C$ satisfies the strengthened condition of strong convexity.

The second work is about **Metric subregularity and $\omega(\cdot)$-normal regularity properties**. In the present paper we first show that the metric subregularity property of the multimapping $M$ (that is the inequality (1.1) with $y = \overline{y}$ (see Section 6.2)) is a suitable assumption to get the normal $\omega(\cdot)$-regularity of the inverse image $M^{-1}(\overline{y})$. We also establish the normal regularity of a generalized equation set, say $S := \{x \in X : f(x) \in M(x)\}$ with $f : X \to Y$ a (single-valued) mapping under a metric subregularity inequality, namely

$$d(x, S) \leq \gamma d\Big(\big(x, f(x)\big), \operatorname{gph} M\Big) \quad x \text{ near } \overline{x}.$$

Of course, we require in both cases a normal $\omega(\cdot)$-regularity property on the involved multimapping $M$ (either on the coderivative or on the graph). We then naturally replace (in the line of [128]) the latter metric subregularity assumption by the inclusion (1.3) with $\alpha, \beta > 0$ such that $\beta > \omega\big(\sqrt{\alpha^2 + \beta^2}\big)$. At last but not least, we show that a Lipschitz (with respect to the Hausdorff-Pompeiu distance) multimapping with normal $\omega(\cdot)$-regularity values near $\overline{x}$ enjoys some metric subregularity property at $\overline{x}$ for $\overline{y}$.

### 1.5.2 Outline

Firstly, we start with chapter 2: Notation and preliminaries where we will embark on understanding the elements of Variational Analysis such as convex functions, some classical results in optimization, normal cones and subdifferentials (in the sense of Clarke, Fréchet, and Mordukhovich, proximal), maximal monotone operator and $\alpha$-Moreau Yosida regularization respectively through Section 2.1-2.5.

The thesis consists of two parts.

Part I is **Modeling of the Sweeping Process**.

We start by formulating a dynamic problem to analyze the movement of particles within a predefined region, with a specific focus on the contact law using specific normal compliance conditions, denoted as Improved Normal Compliance (INC). Based on the work of Hauret and Le Tallec [57], we propose a specific approach within the discrete framework that ensures the energy conservation of the system, aligning it with the continuous case. In Chapter 3, we delve into Moreau's second-order sweeping process to model contact dynamics, incorporating the Moreau-Yosida regularization and thus the INC method. This approach aims to ensure energy conservation of the system while preventing penetration between particles.

In Chapter 4, we adopt a time discretization approach using the implicit Newmark scheme for the INC model, combined with the Primal-Dual Active Set (PDAS) method. This integrated method, known as Newmark-PDAS-INC, facilitates effective problem resolution. Additionally, we perform a comparative analysis of the solutions derived from various methods, including INC, Discrete Element Method (DEM), and Non-Smooth Contact Dynamics with Nonlinear Gauss-Seidel (NSCD-NLGS), to evaluate the efficiency of our algorithm. A comparative analysis is also conducted against alternative numerical methods, such as NSCD-NLGS and DEM, specifically in scenarios devoid of friction. Lastly, we elucidate the numerical treatment of the restitution

coefficient and contact conditions involving friction, aiming to offer a robust solution for resolving rigid multi-body contact problems. Most of the results of this chapter, as well as the section on the Moreau-Yosida regularization for Contact Dynamics in Chapter 3, are presented in the following articles [1] and [2].

Part II is **Weak and strong convexity**.

The farthest distance function is a geometric tool used to measure the farthest point in a set from a reference point, with applications in fields ranging from computational geometry to facility location problems. In Chapter 5, we provide some new characterizations of the farthest distance function. On the other hand, the Robinson-Ursescu theorem (which can be viewed as an extension of the famous Banach-Schauder open mapping theorem) metric regularity is strongly involved in subdifferential calculus, estimates of coderivatives and optimality conditions. We spend Chapter 6 providing general sufficient conditions ensuring the preservation of normal $\omega$-regularity for generalized equations and also focus on normally regular versions of Robin-Ursescu theorem.

Finally, in Chapter 7, we summarize the thesis in French (including an overview, objectives, structure, achieved results, and perspectives). Following this, Appendix A presents the code for the Newmark-PDAS-INC method with a direct version of the sparse matrix combined with the `Gmres` solver for the linear system.

# Chapter 2

# Notation and preliminaries

Throughout this part, $X$ stands for a (real) Hilbert space not reduced to $0_X$ endowed with the inner product $\langle \cdot, \cdot \rangle$ and its associated norm $\| \cdot \|$. The set of nonnegative real numbers is denoted $\mathbb{R}_+ := [0, +\infty[$.

We denote by $B[x, r]$ (resp. $B(x, r)$) the closed (resp. open) ball centered at $x \in X$ of radius $r > 0$. The boundary (resp. the interior) of a nonempty set $S \subset X$ is denoted by $\operatorname{bdry} S$ (resp. $\operatorname{int} S$). For the unit balls in $X$ (that is, centered at $0_X$ with radius 1) it will be convenient to set

$$\mathbb{B}_X := B[0_X, 1] \quad \text{and} \quad \mathbb{U}_X := B(0_X, 1).$$

We also set $\mathbb{S}_X := \{x \in X : \|x\| = 1\} = \mathbb{B}_X \setminus \mathbb{U}_X$.

The distance function $d_S$ and the farthest distance function $\operatorname{dfar}_S$ from a nonempty subset $S \subset X$ are defined for every $x \in X$ by

$$d_S(x) := d(x, S) := \inf_{y \in S} \|x - y\|$$

and

$$\operatorname{dfar}_S(x) := \operatorname{dfar}(x, S) := \sup_{y \in S} \|x - y\| .$$

To those functions are naturally associated the multimappings $\operatorname{Proj}_S : X \rightrightarrows X$ of nearest points in $S$ and $\operatorname{Far}_S : X \rightrightarrows X$ of farthest points in $S$ defined for every $x \in X$ by

$$\operatorname{Proj}_S(x) := \operatorname{Proj}(S, x) := \{y \in S : d_S(x) = \|x - y\|\}$$

and

$$\operatorname{Far}_S(x) := \operatorname{Far}(S, x) := \{y \in S : \operatorname{dfar}_S(x) = \|x - y\|\} .$$

Whenever $\operatorname{Proj}_S(\overline{x})$ (resp. $\operatorname{Far}_S(\overline{x})$) is reduced to a singleton for some $\overline{x} \in X$, that is, $\operatorname{Proj}_S(\overline{x}) = \{\overline{y}\}$ (resp. $\operatorname{Far}_S(\overline{x}) = \{\overline{y}\}$) the vector $\overline{y} \in S$ will be denoted by $\operatorname{proj}_S(\overline{x})$ (resp. $\operatorname{far}_S(\overline{x})$).

By convention, $d_S(\cdot) \equiv +\infty$ whenever $S = \emptyset$. It is well-known and not difficult to check that the distance function is 1-Lipschitz on $X$, that is,

$$\|d(x, S) - d(y, S)\| \leq \|x - y\| .$$

The Hausdorff-Pompeiu excess of the set $S$ over another nonempty subset $S' \subset X$ is defined by

$$\operatorname{exc}(S, S') := \sup_{x \in S} d(x, S') = \inf\{r > 0 : S \subset S' + r\mathbb{B}_X\}. \tag{2.1}$$

## 2.1 Convex functions

**Definition 2.1.** Let $f$: $X \to \mathbb{R} \cup \{\infty\}$ and let $r$ be a real, one defines

1. the domain of $f$ is
$$\text{dom f} := \{x \in X \mid f(x) < +\infty\};$$
2. the graph of $f$ is
$$\text{gph f} = \{(x, f(x)) \mid x \in X\};$$
3. the epigraph of $f$ is
$$\text{epi f} = \{(x, r) \in \text{dom} f \times \mathbb{R} \mid f(x) \leq r\}.$$

The function $f$ is proper if dom f $\neq \emptyset$ and $f$ is a lower semi-continuous function on $X$ if and only if epi f is closed.

**Definition 2.2.** Let $f : X \to \mathbb{R} \cup \{\infty\}$ be a function.

1. $f$ is a convex if for all $\alpha \in ]0, 1[$,
$$f\big(\alpha x + (1-\alpha)y\big) \leq \alpha f(x) + (1-\alpha) f(y) \text{ for all } (x, y) \in X \times X.$$
2. $f$ is stricly convex if for all $\alpha \in ]0, 1[$, for all $x \in$ dom f, $y \in$ dom f,
$$x \neq y \Rightarrow f\big(\alpha x + (1-\alpha)y\big) < \alpha f(x) + (1-\alpha) f(y).$$

A function $f$ is convex if and only if its epigraph is convex. In addition, if $f$ is convex, then dom f is convex.

In particular, a convex function $f$ does not have a derivative at point $A$. We observe that the slopes of the lines that pass at this point are under the curve representing function $f$, see Fig. 2.1. These slopes are the subgradients that constitute the *subdifferentiality*.

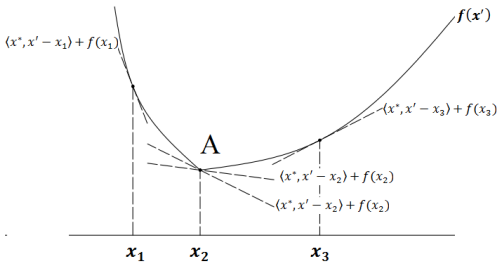



**Figure 2.1:** Subdifferential of convex function.

**Definition 2.3.** Let $f : X \to \mathbb{R} \cup \{\infty\}$ be a proper function. The *Moreau-Rockafellar* subdifferential of $f$, denoted by $\partial f(x)$, is defined by

$$\partial f(x) = \{y \in X : f(x') \geq f(x) + \langle y, x' - x\rangle \ \ \forall x' \in X\}.$$

The vector $y \in \partial f(x)$ is a subgradient of function $f$ at $x$.

We also recall that the subdifferential $\partial$ enjoys Fermat optimality condition, namely

$$0 \in \partial f(x) \quad \text{if } x \text{ is a local minimizer of } f.$$

For the particular case of the convex function $f := \|\cdot\|$, it is known (and not difficult to prove) that

$$\partial\|\cdot\|(0) = \mathbb{B}_X \quad \text{and} \quad \partial\|\cdot\|(x) = \{x^\star \in X : \|x^\star\| = 1, \langle x^\star, x\rangle = \|x\|\} \quad \text{for all } x \neq 0.$$

Let $S$ be a subset of $X$ and let $x \in X$. We recall the indicator function of $S$ is defined by

$$\psi_S(x) = \begin{cases} 0 & \text{if } x \in S, \\ +\infty & \text{if } x \in X \backslash S. \end{cases}$$

Given any subdifferential $\partial$ and its corresponding normal cone $N$, it is well known that

$$\partial\psi_S(x) = N(S; x) \quad \text{for all } x \in X,$$

where (as usual) $\psi_S$ denotes the indicator of the subset $S$ of $X$ (in the sense of variational analysis).

**Definition 2.4.** Let $S$ be a nonempty subset of $X$ and let $x \in X$. The subdifferential of the indicator function of $S$ is the **normal cone** to $S$ at $x$ defined by

$$N(S; x) = N_S(x) = \begin{cases} \{y \in X : \langle y, z - x\rangle \leq 0, \forall z \in S\} & \text{if } x \in S, \\ \emptyset & \text{if } \ x \notin S. \end{cases}$$

If $S$ is convex, then $N_S(x)$ is a closed convex cone for every $x \in S$. If $x \in \text{int}(S)$ then $N_S(x) = \{0\}$.

## 2.2 Classical results in optimization

Let $g : X \to \mathbb{R}$ be a real-valued function on $X$ and let $S \subset X$ be a nonempty set. Consider the constrained optimization problem

$$\min_{x \in S} g(x)$$

and define an extended real-valued function is a function $f : X \to \mathbb{R} \cup \{\infty\}$ by

$$f(x) = \begin{cases} g(x) & \text{if } x \in S, \\ +\infty & \text{if } \ \text{otherwise.} \end{cases} \tag{2.2}$$

It is clear that $f = g + \psi_S$ and the function 2.2 can be successfully used to transform every constrained optimization problem into an unconstrained problem given by

$$\min_{x \in X} f(x).$$

Some classical results in optimization.

1. **Minimizers of a convex function:**

   **Theorem 2.5.** *Let $f : X \to \mathbb{R} \cup \{\infty\}$ be a proper convex function such that $\alpha = \inf f > -\infty$.*

   - *The set $\{x \in X : f(x) = \alpha\}$ is convex.*
   - *Every local minimizer of $f$ is a global minimizer.*
   - *If $f$ is strictly convex, then there exists at most one minimizer.*

   Let $S \subset X$ be a closed convex set and let $f : X \to \mathbb{R} \cup \{\infty\}$ be convex, lower semicontinuous and proper such that dom f $\cap\, S \neq \emptyset$. If $f$ is coercive or $S$ is bounded, then there exists an $x^\star \in S$ such that

   $$f(x^\star) = \inf_{x \in S} f(x).$$

   Moreover, if $f$ is strictly convex, then $x^\star$ is unique.

2. **The first order necessary and sufficient condition:**

   **Theorem 2.6.** *Let $f : X \to \mathbb{R} \cup \{\infty\}$ be convex, lower semi-continuous, proper and continuously differentiable function on $X$. One has $x^\star$ is a global minimizer of function $f$, i.e. $x^\star \in \mathrm{Argmin}_{x \in X} f(x)$, if and only if $\nabla f(x^\star) = 0$.*

   **Theorem 2.7.** *Let $S$ be a nonempty, closed, convex subset of Hilbert space $X$. Let $f : X \to \mathbb{R} \cup \{\infty\}$ be convex, lower semi-continuous, proper and continuously differentiable function on $S$. One has $x^\star$ is a minimizer of function $f$ on $S$, i.e. $x^\star \in \mathrm{Argmin}_{x \in S} f(x)$, if and only if*

   $$\langle \nabla f(x^\star), x - x^\star \rangle \geq 0 \quad \forall x \in S.$$

3. **Second-Order Sufficient and Necessary Conditions for Optimization:**
   • *Second-Order Sufficient Condition*

   **Theorem 2.8.** *Let $f : X \to \mathbb{R}$ be twice continuously differentiable (i.e. $f \in C^2$). A point $x^* \in X$ is a local minimum if:*

   *(i)* $\nabla f(x^*) = 0$,

   *(ii) The Hessian matrix $H_f(x^*)$ of $f$ at $x^*$ is positive definite, i.e., for all nonempty $v \in X$,*

   $$v^T H_f(x^*) v > 0.$$

   • *Second-Order Necessary Condition*
   Assume $f : X \to \mathbb{R}$ is twice continuously differentiable (i.e. assume that $f \in C^2$). A point $x^* \in X$ is a local minimum if:

   (i) $\nabla f(x^*) = 0$,

   (ii) The Hessian matrix $H_f(x^*)$ of $f$ at $x^*$ is positive semidefinite, i.e., for all $v \in X$,

   $$v^T H_f(x^*) v \geq 0.$$

If $H_f(x^*)$ is positive definite, $x^*$ is a strict local minimum.

4. **Karush-Kuhn-Tucker (KKT) Conditions:**
Optimization problem under equality and inequality constraints:

$$x^\star \in \operatorname{Argmin}_{x\in X} f(x)$$
$$\text{subject to } \begin{cases} g_i(x) \leq 0, \forall i \in \{1,\dots m\}, \\ h_j(x) = 0, \forall j \in \{1,\dots p\} \end{cases}$$

where

- $g_i$ is convex, lower semi-continuous, proper for every $i \in \{1,\dots m\}$;
- $h_j$ is convex, lower semi-continuous, proper for every $j \in \{1,\dots p\}$;
- $\text{dom gh} = \left(\bigcap_{i=1}^m \text{dom g}_\text{i}\right) \cap \left(\bigcap_{j=1}^p \text{dom h}_\text{j}\right) \neq \emptyset$.

For all $(x,\lambda,\mu) \in X \times X \times X$, the Lagrangian function is defined by

$$L(x,\lambda,\mu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_i(x)$$

where

- $\text{dom L} = \text{dom gh} \times \mathbb{R}^m \times \mathbb{R}^p$;
- $\lambda = (\lambda_i)_{1\leq i\leq m}$ (*resp.* $\mu = (\mu_j)_{1\leq j\leq p}$) is the Lagrangian multiplier associated with $g_i(x) \leq 0$ for $i = \{1,\dots m\}$ (*resp.* $h_j(x) = 0$ for $j = \{1,\dots p\}$);
- $\lambda$ and $\mu$ are Lagrange multiplier vectors or dual variables.

**Theorem 2.9.** *Let $(\lambda_i)_{1\leq i\leq m}$ and $(\mu_j)_{1\leq j\leq p}$ be continuously differentiable. One has $x^\star$ and couple $(\lambda,\mu)$ are primal and dual solutions if and only if*

| | |
|---|---|
| *Stationarity:* | $\nabla f(x^\star) + \sum_{i=1}^m \lambda_i \nabla g_i(x^\star) + \sum_{i=1}^p \mu_i \nabla h_i(x^\star) = 0;$ |
| *Primal Feasibility:* | $g_i(x^\star) \leq 0 \quad \forall i = 1,\dots,m;$ |
| | $h_j(x^\star) = 0 \quad \forall j = 1,\dots,p;$ |
| *Dual Feasibility:* | $\lambda_i \geq 0 \quad \forall i = 1,\dots,m;$ |
| *Complementary Slackness:* | $\lambda_i g_i(x^\star) = 0 \quad \forall i = 1,\dots,m.$ |

## 2.3 Normal cones, coderivatives and subdifferentials

### 2.3.1 The Clarke tangent and normal cone

**Definition 2.10.** Let $S$ be a set of a Hilbert space $X$ and let $x \in S$. The *Clarke tangent cone* (or *C-tangent cone*) at $x$, denoted by $T^C(S;x)$ or $T^{Cl}(S;x)$, is defined as follows:

A vector $h \in X$ belongs to the Clarke tangent cone $T^C(S;x)$ if and only if it satisfies one of the following equivalent characterizations:

1. For any sequence $(x_n)_n$ of $S$ converging to $x$ and any sequence $(t_n)_n$ with $t_n > 0$ converging to 0, there exists a sequence $(h_n)_n$ in $X$ converging to $h$ such that

$$x_n + t_n h_n \in S \quad \text{for all } n \in \mathbb{N}.$$

2. For any sequence $(x_n)_n$ of $S$ converging to $x$ and any sequence $(t_n)_n$ with $t_n > 0$ converging to 0, there is an increasing function $s : \mathbb{N} \longrightarrow \mathbb{N}$ and a sequence $(h_n)_n$ in $X$ converging to $h$ such that

$$x_{s(n)} + t_{s(n)} h_n \in S \quad \text{for all } n \in \mathbb{N}.$$

**Example 2.11.** Consider a set $S$ in $\mathbb{R}^2$ with $S := \text{hypo}(|\cdot|) = \{(x, y) \in \mathbb{R}^2 : y \leq |x|\}$. The Clarke tangent cone of $S$ at $x = (0, 0)$ is

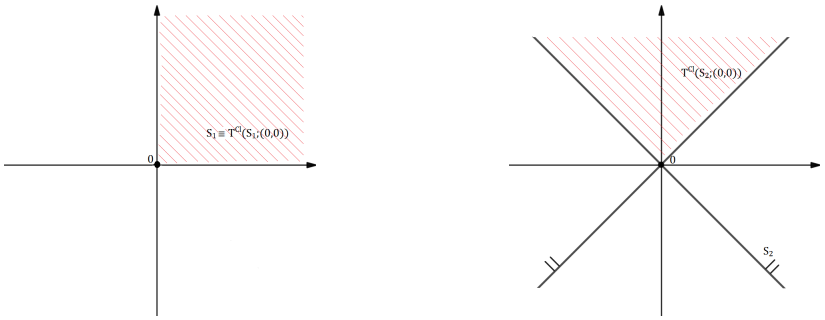
$$T^C\big(S; (0,0)\big) = \text{hypo}(-|\cdot|).$$

**Remark 2.12.** As for the Clarke tangent cone, we should note that, for two sets such that $S_1 \subset S_2$ and let $x \in S_1$, it does not imply that $T^C(S_1; x) \subset T^C(S_2; x)$. It means, the Clarke tangent cone is neither isotone nor antitone. Let consider example to make clearly. Assume that $S_1 = [0, +\infty[ \times [0, +\infty[$ and $S_2 := \text{epi}(-|\cdot|)$. It is obvious that $S_1 \subset S_2$, and get that

$$T^C\big(S_1; (0,0)\big) = [0, +\infty[ \times [0, +\infty[\,,\ T^C\big(S_2; (0,0)\big) = \text{epi}(|\cdot|).$$

We can see that no connection between two sets, that is,

$$T^C\big(S_1; (0,0)\big) \nsubseteq T^C\big(S_2; (0,0)\big) \text{ or } T^C\big(S_2; (0,0)\big) \nsubseteq T^C\big(S_1; (0,0)\big).$$




(a) The Clarke tangent cone of $S_1 = [0, +\infty[ \times [0, +\infty[$ at $(0, 0)$.

(b) The Clarke tangent cone of $S_2 = \text{epi}(-|\cdot|)$ at $(0, 0)$.

**Figure 2.2:** The Clarke tangent cone of $S_1$ and $S_2$ at $(0, 0)$ with $S_1 \subset S_2$.

**Definition 2.13.** Let $S$ is a subset of Hilbert space $X$. The *Clarke normal cone* or *C-normal cone* to $S$ at a point $x \in S$, denoted by $N^{Cl}(S; x)$ (or $N^C(S; x)$), is defined by

$$N^C(S; x) = N^{Cl}(S; x) = \{x^* \in X : \langle x^*, h\rangle \leq 0, \forall h \in T^C(S; x)\}.$$

And note that $N^C(S;x) = \emptyset$ and $T^C(S;x) = \emptyset$ if $x \notin S$. When $S$ is convex, the Clarke tangent and normal cones coincide with the classical tangent and normal cones defined in convex analysis.

The Clarke subdifferential of a function has been defined through the Clarke normal cone of its epigraph.

**Definition 2.14.** Let $X$ be a Hilbert space and let $x \in X$. Let $U$ be a neighborhood of $x$ and $f : U \to \mathbb{R} \cup \{\infty\}$ be an extended real-valued function which is finite at $x$. The *Clarke subdifferential* or *C-subdifferential* of $f$ at $x$, denoted by $\partial_{Cl} f(x)$ or $\partial_C f(x)$, provided by

$$\partial_C f(x) = \partial_{Cl} f(x) := \{x^* \in X : (x^*, -1) \in N^C\big(\text{epi f}; \big(x, f(x)\big)\big)\}.$$

**Example 2.15.** In $\mathbb{R}^2$, given function $f = |x|$, the Clarke tangent cone of $f$ and $-f$ at $(0,0)$ is given by $T^C\big(\text{epi}f; (0,0)\big) = T^C\big(\text{epi}(-f); (0,0)\big) = \{(x,y) \in \mathbb{R}^2 : y \geq |x|\}$. The Clarke subdifferential of $f$ at $(0,0)$ is

$$\begin{aligned}\partial_C f(0) &= \{x^* \in X^* : (x^*, -1) \in N^C(\text{epi}f; (0,0))\} \\ &= \{x^* \in X^* : x^*x - y \leq 0, \forall (x,y), |x| \leq y\} \\ &= \{x^* \in X^* : x^*x \leq |x|\,, \forall x \in \mathbb{R}\} = [-1,1] = \partial_C(-f)(0).\end{aligned}$$

Let $S$ be a subset of $X$. Recall that the Clarke normal cone $N^C$ is linked to its subdifferential $\partial_C$ through the equality

$$0 \in \partial_C f(x) + N^C(S;x).$$

In concept of distance function, we also have

$$\partial_C d_S(x) \subset N^C(S;x) \cap \mathbb{B}_X \quad \text{for all } x \in S.$$

We end this section with fundamental results on subdifferentials. We start with the famous sum rule for the Clarke's subdifferential (see, e.g., [124, Theorem 2.98] (see also the references [116, 130])). These properties still hold when $X$ is a normed space.

**Theorem 2.16. (Sum rule for $C$-subdifferential)** *Let $X$ be a Hilbert space and let $f_1, f_2 : X \to \mathbb{R} \cup \{+\infty\}$ be two functions which are finite at $\overline{x} \in X$. If $f_1(\overline{x}) < +\infty$ and $f_2$ is Lipschitz continuous near $\overline{x}$, then one has*

$$\partial_C(f_1 + f_2)(\overline{x}) \subset \partial_C f_1(\overline{x}) + \partial_C f_2(\overline{x}).$$

The next proposition (see, e.g., [124, Theorem 2.135] (see also the references [116, 130])) gives an estimate for the Clarke subdifferential $\partial_C(g \circ G)(\overline{x})$ for the composition of a locally Lipschitz function $g$ with an inner strictly Hadamard differentiable vector-valued mapping $G$.

**Proposition 2.17.** *Let $G : X \to Y$ be a mapping between two Hilbert spaces $X$ and $Y$ which is strictly Hadamard differentiable at a point $\overline{x} \in X$ and let $g : Y \to \mathbb{R}$ be a function Lipschitz continuous near $G(\overline{x})$. Then, the function $g \circ G$ is Lipschitz continuous near $\overline{x}$ and*

$$\partial_C(g \circ G)(\overline{x}) \subset DG(\overline{x})^\star\big(\partial_C g(G(\overline{x}))\big).$$

The following proposition provides a description of $N^C(\operatorname{gph} M;\cdot)$ for a multimapping $M(\cdot)$ defined as the translation of a fixed set $S$, that is, $M(x) := f(x) + S$ for a given mapping $f$. For the proof, we refer the reader to [124, Proposition 2.129].

**Proposition 2.18.** *Let $S$ be a nonempty set of a Hilbert space $Y$ and let $f : X \to Y$ be a mapping from a Hilbert space $X$ into $Y$. Let $M : X \rightrightarrows Y$ be the multimapping defined by*

$$M(x) := f(x) + S \quad \text{for all } x \in S.$$

*If $f$ is strictly Hadamard differentiable at $\overline{x} \in X$, then for every $y \in M(\overline{x})$*

$$N^C(\operatorname{gph} M;(\overline{x},y)) = \left\{(-Df(\overline{x})^\star(y^\star), y^\star) : y^\star \in N^C\big(S; y - f(\overline{x})\big)\right\}.$$

## 2.3.2 Fréchet and Mordukhovich limiting subdifferential

Firstly, let us introduce the definition of Fréchet subdifferential of a function in Hilbert space.

**Definition 2.19.** Let $X$ be a Hilbert space and let $x \in X$. The Fréchet subdifferential of a function $f : X \to \mathbb{R}$ at a point $x$, denoted $\partial_F f(x)$, is defined as follows:

$$\partial_F f(x) = \{y \in X \mid f(x') \geq f(x) + \langle y, x' - x\rangle \text{ for all } x' \in X\}.$$

Let $S$ be a subset of $X$ and $x \in S$. The Fréchet normal cone (or *F-normal cone* for short) of $S$ at $x$ denoted by $N^F(S;x)$ is the Fréchet subdifferential of the indicator function $\psi_S$ at $x$, that is

$$N^F(S;x) := \partial_F \psi_S(x).$$

In particular, one has $x^* \in N^F(S;x)$ if and only if for any real $\varepsilon > 0$ there exists some neighborhood $V$ of $x$ such that

$$\langle x^*, x' - x\rangle \leq \varepsilon\|x' - x\| \text{ for all } x' \in S \cap V.$$

**Definition 2.20.** Let $(X, \|\cdot\|)$ be an Asplund space and a set $U$ in $X$ be a neighborhood of a point $x$. Given $f : U \to \mathbb{R} \cup \{\infty\}$ be an extended real valued function which is finite at $x \in U$. A continuous linear functional $x^* \in X^*$ is called a *Mordukhovich limiting subgradient* (or *L-subgradient*) of $f$ at $x$ provided that there exists a sequence $\big(x_n, f(x_n)\big)_n$ converging to $\big(x, f(x)\big)$ and a sequence $(x_n^*)_n$ converging weakly star to $x^*$ such that $x_n^* \in \partial_F f(x_n)$.

It is obvious that

$$\partial_F f(x) \subset \partial_L f(x) \subset \partial_C f(x).$$

Let $S$ be a subset of the Asplund space $X$ and $x \in S$. The *Mordukhovich limiting normal cone* (or *L-normal cone*) of $S$ at $x$ denoted by $N^L(S;x)$ is the Mordukhovich limiting subdifferential of the indicator function $\psi_S$ at $x$, that is

$$N^L(S;x) := \partial_L \psi_S(x).$$

In particular, one has $x^* \in N^L(S;x)$ if and only if there exist a sequence $(x_n)_n$ of $S$ converging to $x$ and a sequence $(x_n^*)_n$ converging weakly star to $x^*$ such that

$$x_n^* \in N^F(S;x_n).$$

When $x \notin S$, one convention that

$$N^F(S;x) = N^L(S;x) = N^C(S;x) := \emptyset. \tag{2.3}$$

Let $X$ and $Y$ be two Hilbert spaces. The *coderivative* associated to a concept of normal cone $\mathcal{N}$ in $X \times Y$ of a multimapping $M : X \rightrightarrows Y$ at $(x, y) \in \operatorname{gph} M$ is defined for every $y^\star \in Y^\star$ by

$$D_{\mathcal{N}}M(x,y)(y^\star) := \left\{x^\star \in X^\star : (x^\star, -y^\star) \in \mathcal{N}\big(\operatorname{gph} M; (x,y)\big)\right\}.$$

The *Fréchet* (resp. *Mordukhovich limiting* (resp. *Clarke*)) *subdifferential* of an extended real-valued function $f : X \to \mathbb{R} \cup \{+\infty\}$ finite at $x \in X$ is defined by saying that $x^\star \in X^\star$ belongs to $\partial_F f(x)$ (resp. $\partial_L f(x)$ (resp. $\partial_C f(x)$)) when $(x^\star, -1)$ belongs to the corresponding normal cone of the epigraph of $f$ at $(x, f(x))$, i.e.

$$(x^*, -1) \in \mathcal{N}\big(\operatorname{epi} f; \big(x, f(x)\big)\big).$$

Through to (2.3), we easily see that

$$\partial_F f(x) = \partial_L f(x) = \partial_C f(x) = \emptyset \quad \text{if } f(x) = +\infty.$$

The above normal cones and subdifferentials do not depend on equivalent norms on $X$. In particular, *the subdifferential of a function $f : X \to \mathbb{R} \cup \{+\infty\}$ is always considered for an equivalent norm with respect to the initial one given on $X$.*

Recall that the Fréchet normal cone $N^F$ is linked to its subdifferential $\partial_F$ through the equality

$$\partial_F d_S(x) = N^F(S;x) \cap \mathbb{B}_{X^\star} \text{ and } N^F(S;x) = \mathbb{R}_+ \partial_F d_S(x) \text{ for all } x \in S. \tag{2.4}$$

A similar equality holds for Mordukhovich-limiting normal cone $N^L$ and the subdifferential $\partial_L$.

If the set $S$ and $f$ are convex, the Frechet (resp. Mordukhovich limiting) normal cone $N^F(S;x)$ (resp. $N^L(S;x)$) and subdifferential $\partial_F f(x)$ (resp. $\partial_L f(x)$) coincide respectively with the usual concepts of normal cone and subdifferential in Convex Analysis.

Similar to the Clarke's subdifferential, the sum rule for the Mordukhovich limiting subdifferential can be found in the monographs (see [112, 124]).

**Theorem 2.21. (sum rule for $L$-subdifferential)** *Let $X$ be an Asplund space and let $f_1, f_2 : X \to \mathbb{R} \cup \{+\infty\}$ be two functions which are finite at $\overline{x} \in X$. If $f_1$ is Lipschitz continuous at $\overline{x} \in X$ and $f_2$ is lower semicontinuous near $\overline{x}$, then one has*

$$\partial_L(f_1 + f_2)(\overline{x}) \subset \partial_L f_1(\overline{x}) + \partial_L f_2(\overline{x}).$$

### 2.3.3 Proximal normal cones and subdifferentials

In Variational Analysis, the prox-regular functions are useful and they are motivated by the study of second-order properties of some nonconvex functions. In this part, we will study the concept of proximal normal cones and subdifferentials in Hilbert space based on the research by Poliquin, Rockafellar, Thibault [118] and some related papers.

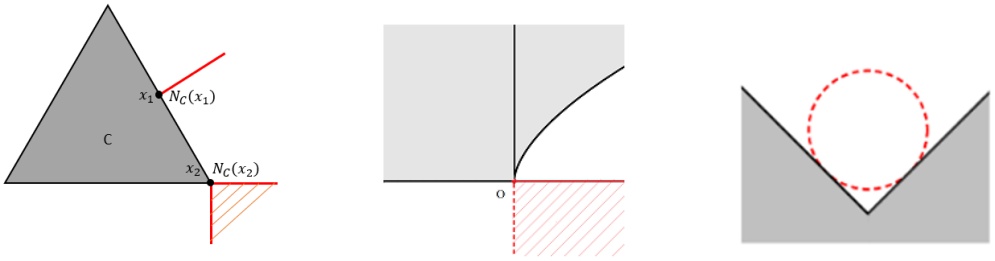
Let $S$ be a nonempty set in Hilbert space $X$. The *proximal normal cone* of $S$ at $x \in S$, denoted by $N^P(S;x)$, is defined as

$$N^P(S;x) := \left\{v \in X : \exists \sigma \geq 0, \forall x' \in S, \langle v, x' - x\rangle \leq \sigma \|x' - x\|^2\right\}.$$

If $S$ is convex, it is known that the proximal normal cone $N^P(S;x)$ coincides with the normal cone *in the sense of convex analysis*, that is,

$$N^P(S;x) = \{v \in X : \langle v, x' - x\rangle \leq 0, \forall x' \in S\} =: N(S;x).$$

Through the equivalences valid for every $x, y \in X$




(a) Normal cone in sense of vonvex analysis

(b) Proximal normal cone

(c) $N^P$ is often reduced to 0

**Figure 2.3:** Illustration the prox-regular cones.

$$y \in \mathrm{Proj}_S(x) \Leftrightarrow y \in S \quad \text{and} \quad \langle x - y, c - y\rangle \leq \frac{1}{2}\|c - y\|^2 \quad \text{for all } c \in S$$

and

$$y \in \mathrm{Far}_S(x) \Leftrightarrow y \in S \quad \text{and} \quad \langle y - x, c - y\rangle \leq -\frac{1}{2}\|c - y\|^2 \quad \text{for all } c \in S, \tag{2.5}$$

we see that

$$x - y \in N^P(S;y) \quad \text{for all } (x,y) \in \mathrm{gph}\,\mathrm{Proj}_S \tag{2.6}$$

and

$$y - x \in N^P(S;y) \quad \text{for all } (x,y) \in \mathrm{gph}\,\mathrm{Far}_S, \tag{2.7}$$

where $\mathrm{gph}\,\mathrm{Proj}_S = \{(x,y) \in X \times X : y \in \mathrm{Proj}_S(x)\}$ and $\mathrm{gph}\,\mathrm{Far}_S = \{(x,y) \in X \times X : y \in \mathrm{Far}_S(x)\}$. It is worth pointing out that for any $(x,y) \in \mathrm{gph}\,\mathrm{Far}_S$

$$\mathrm{Far}_S\big(x + s(x - y)\big) = \{y\} \quad \text{for all } s > 0. \tag{2.8}$$

Given a function $f : X \to \mathbb{R} \cup \{+\infty\}$ and $x \in X$ with $|f(x)| < +\infty$, one defines the *proximal subdifferential* of $f$ at $x$ as the set

$$\partial_P f(x) := \left\{\zeta \in X : (\zeta, -1) \in N^P\big(\mathrm{epi}\, f; (x, f(x))\big)\right\},$$

where as usual $\mathrm{epi}\, f := \{(u,r) \in X \times \mathbb{R} : f(u) \leq r\}$ stands for the *epigraph* of the function $f$. Of course, if $f$ is convex then the latter subdifferential is nothing but the *Moreau-Rockafellar subdifferential*, that is

$$\partial_P f(x) = \{\zeta \in X : \langle \zeta, x' - x\rangle \leq f(x') - f(x) \quad \forall x' \in X\} =: \partial f(x).$$

If $f$ is convex and continuous at $x$, then $f$ is Fréchet differentiable at $x$ whenever $\partial_P(-f)(x) \neq \emptyset$. If $f$ is $C^{1,1}$ near $x$, it is an exercise to check that for a function $g : X \to \mathbb{R} \cup \{+\infty\}$ finite at $x$, one has

$$\partial_P(f+g)(x) = \nabla f(x) + \partial_P g(x). \tag{2.9}$$

For more details on normal cones, coderivatives and subdifferentials, we refer the reader to the books [112, 116, 120, 124, 130] and the references therein.

## 2.4 Maximal Monotone Operator

Next, we dedicate this section to present some knowledge about Maximal Monotone Operator, as it is essential for constructing a regularization framework for the sweeping process.

**Definition 2.22.** Let $X$ be a Hilbert space. An operator $A : X \to X$ is *monotone* whenever

$$\langle x^\star - y^\star, x - y\rangle \geq 0 \tag{2.10}$$

for all $x^\star \in A(x), y^\star \in A(y)$. A is called is *maximal monotone operator* if its graph is not properly contained in the graph of any other monotone operator.

**Definition 2.23.** Let $X$ be a Hilbert space and let A be a maximal monotone operator. For every positive $\alpha$, the *resolvent* $J_\alpha$ and the *Yosida approximation* of $A$, namely $A_\alpha$, defined by

$$J_\alpha = (I + \alpha A)^{-1}, \quad A_\alpha = \frac{1}{\alpha}(I - J_\alpha) \tag{2.11}$$

where $I : X \to X$ be the identity operator.

Instead of representing the Yosida approximation as an operator implicitly dependent on the resolvent through equation 2.11, we can explicitly express the Yosida approximation using the following proposition.

**Proposition 2.24.** *The Yosida approximation of an operator $A$ with parameter $\alpha > 0$ can be presented as*

$$A_\alpha = (\alpha I + A^{-1})^{-1},$$

*where $I : X \to X$ be the identity operator.*

*Proof.* Fix $u \in X$ and let $J_\alpha$ be the resolvent given by $J_\alpha = (I + \alpha A)^{-1}$ where $I$ is the identity matrix. By the definition of the Yosida approximation above combining with the homohenity of the dualtity mapping $J^{-1}$, we obtain that

$$J_\alpha(u) = u - \alpha J^{-1}(A_\alpha(u)).$$

It leads to

$$(I + \alpha A)^{-1}(u) = u - \alpha J^{-1}(A_\alpha(u)).$$

Hence, we get that

$$A_\alpha(u) \in A(u - \alpha J^{-1}(A_\alpha(u)),$$

or it is equivalent to

$$u \in (A^{-1} + \alpha J^{-1})(A_\alpha(u)).$$

Moreover, since $A_\alpha$ is single - valued, we get the conclusion that

$$A_\alpha(u) = (\alpha I + A^{-1})^{-1}(u).$$

□

It is not difficult to check that $A_\alpha$ is a maximal monotone and $\alpha^{-1}-$Lipschitz continuous. We can find more details of the concept of Maximal Monotone in the research of H.Brezis [65].

Related to the first order sweeping process, we consider the evolution equation associated with Maximal monotone operator given by

$$\begin{cases} -\dot{u}(t) \in Au(t), \\ u(0) = u_0. \end{cases}$$

To solve it, one is to use the *Yosida approximation* with parameter $r > 0$ and this problem is wellknown in the theory of optimization in Hilbert space. Hence, we are able to follow the *regularization* system of

$$\begin{cases} -\dot{u}_r(t) = A_r u_r(t), \\ u_r(0) = u_0. \end{cases}$$

For the sweeping process 1 with $A^t(\cdot) = N(C(t), \cdot)$, it can be represented as $-\dot{u}(t) \in A^t u(t)$ with the initial condition $u(0) = u_0$. It is also notable that the subdifferential of the indicator function associated with the set $C(t)$ is a maximal monotone operator. This property, combined with the form of the related evolution equation, leads to the Yosida approximation of $A^t$, denoted as $(A^t)^\alpha$, where

$$A^t(\cdot) = N(C(t), \cdot) = \partial\psi_{C(t)}(\cdot).$$

## 2.5 The $\alpha$-Moreau-Yosida regularization

In this section, we recall the fundamentals of the Moreau-Yosida formulation that will be used in calculating related our mechanical problem in the next chapter.

**Definition 2.25.** Let $X$ be a real Hilbert space, let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a function and let $\alpha \geq 2$. For any real $R > 0$, the $\alpha-$Moreau envelope (or $\alpha$-Moreau-Yosida regularization) with index $R$ of the function $f$ is the function $f_R^\alpha : X \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f_R^\alpha(u) := \inf_{y \in X} \left( f(y) + \frac{1}{\alpha R} \|y - u\|^\alpha \right) \text{ for all } u \in X.$$

It is well known (see, e.g., [120]) that the Yosida regularization of the subdifferential $\partial f$ of a (proper lower semicontinuous) convex function is nothing but the gradient of its usual (i.e., with parameter $\alpha = 2$) Moreau envelope, that is

$$\nabla e_\lambda f = (\lambda I_d + (\partial f)^{-1})^{-1}.$$

Such an equality still holds in the context of $\alpha-$Moreau envelope (or $\alpha$-Moreau-Yosida regularization).

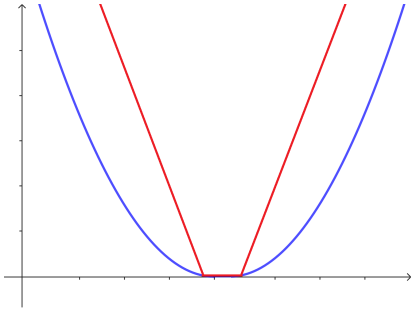


**Figure 2.4:** Moreau envelope of function $f$ in $\mathbb{R}^2$.

**Theorem 2.26.** *[120] Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a function. For any real $R > 0$, the $\alpha$-Moreau-Yosida regularization of $\partial f$ is the gradient $\nabla f_R^\alpha$ associated with the Moreau envelope $f_R^\alpha$.*

In addition, an important identity for computation in next chapter is Moreau's decomposition, which states that

**Theorem 2.27.** *[39]* **(The Moreau decomposition theorem.)** *Let $S$ be a nonempty closed convex cone of Hilbert space $X$. For any $x \in X$,*

$$x = \text{proj}_S(x) + \text{proj}_{S^0}(x) \text{ with } \langle \text{proj}_S(x), \text{proj}_{S^0}(x) \rangle = 0$$

*or equivalently,*

$$x = x_1 + x_2 \text{ where } x_1 = \text{proj}_S(x),\ x_2 = \text{proj}_{S^0}(x) \text{ with } \langle x_1, x_2 \rangle = 0$$

*where $S_0$ is polar cone of $S$ defined by $S_0 = \{x : \langle x, y \rangle \leq 0, \forall y \in X\}$.*

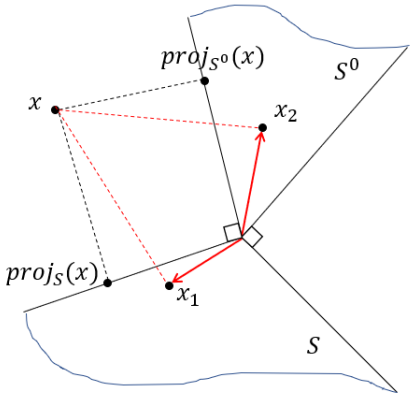



**Figure 2.5:** Moreau decomposition.

Moreau decomposition is a fundamental concept in convex analysis and optimization theory. It involves decomposing a vector in a Hilbert space into the sum of two orthogonal components, each associated with a convex cone and its polar cone. This decomposition has found many applications in optimization and various other areas of applied mathematics (see [164–168]). It is also an important key for computation in mechanical problems, especially in unilateral mechanics [40, 41].

# Part I

# Modeling of the Sweeping Process

# Chapter 3

# Mechanical System

In this chapter, we will study some mathematical models of Mechanics with dynamics of systems given by differential inclusions and their link to the form of the Sweeping Process.

## 3.1 Function of bounded variation

Let $I$ be a nonempty real interval in a Hilbert space $\mathcal{H}$, and let $u : I \to H$. We define the *right-limit* of $u$ at $t$ (resp. *left-limit*) as:

$$u^+(t) = \lim_{s \searrow t} u(s) \quad (resp. \quad u^-(t) = \lim_{s \nearrow t} u(s)). \tag{3.1}$$

If it is defined everywhere in an open interval in $\mathcal{H}$, one says that $u^+$ is the *right-continuous* function (resp. *left-continuous* function).

**Definition 3.1.** The *variation of* $u : I := [0, T] \to \mathcal{H}$ can be presented by taking supremum of a partition of $I$ with $0 = t_0 < t_1 < ... < t_N = T$,

$$\mathrm{var}(u) = \mathrm{var}(u, I) = \sup \sum_{i=0}^{N-1} \mid u(t_{i+1}) - u(t_i) \mid .$$

**Remark 3.2.** 1. Assume that $I := [a, b]$, we can write:

$$\mathrm{var}(u) = \mathrm{var}(u; a, b) = \mathrm{var}(u; [a, b]).$$

Moreover, if there is a point $b \in [a, c]$, one has:

$$\mathrm{var}(u; a, c) = \mathrm{var}(u; a, b) + \mathrm{var}(u; b, c).$$

2. Let $t$ be a singleton on $I$, one has:

$$\mathrm{var}(u, t) = 0.$$

3. One has, $u$ be a constant on $I$ if and only if $\mathrm{var}(u, t) = 0$.

**Definition 3.3. Function of bounded variation** Let $I$ be a nonempty real interval in $\mathcal{H}$. A function $u : I \to \mathcal{H}$ is of bounded variation (in short, BV) on $I$, denoted by $u \in \mathrm{BV}(I, \mathcal{H})$ if and only if

$$\mathrm{var}(u) < +\infty.$$

**Lemma 3.4.** *If $u : [0, T] \to \mathcal{H}$ is Lipschitz continuous with constant $L > 0$, then $u$ is BV and* $\mathrm{var}(u) \leq LT$.

*Proof.* Since $u$ is Lipschitz continuous with constant $L > 0$, for any $t, s \in [0, T]$, we have:

$$\mid u(t) - u(s) \mid \leq L \mid t - s \mid .$$

Now, considering a partition of $[0, T]$ with $0 = t_0 < t_1 < ... < t_N = T$, we use Definition 3.3 to obtain:

$$\sum_{i=0}^{N-1} \mid u(t_{i+1}) - u(t_i) \mid \leq L \sum_{i=0}^{N-1} \mid (t_{i+1} - t_i) = L(t_N - t_0) = LT.$$

Taking the supremum, we conclude $\mathrm{var}(u) \leq LT$. □

By definition, we can deduce another property about BV functions as follows.

**Lemma 3.5.** *If $u : [0, T] \to \mathcal{H}$ is an increasing function, then $u$ is BV and* $\mathrm{var}(u) = u(T) - u(0)$.

In general, a BV function $u$ defined on $[0, T]$ has a right limit $u^+$ on $[0, T[$ and left limit $u^-$ on $]0, T]$ (defined by 3.1), and both $u^+$ and $u^-$ are BV functions. Additionally, it is well-known that any continuous BV function $u$ contains a countable set of discontinuous points and is differentiable almost everywhere in $]0, T[$.

On the other hand, let $u : [0, T] \to \mathcal{H}$ be a BV function, it defines the **differential measure** or **Stieltjes measure** $du$ on $[0, T]$ (see more details in [68]) as below.

**Proposition 3.6.** *Let $\phi : [0, T] \to \mathcal{H}$ be a continuous function with compact support. There exists an interval $I$ of $H$, denoted by $I = \int_0^T \phi du$, such that for any positive $\varepsilon$, there is a partition on $[0, T]$ with $0 = t_0 < t_1 < ... < t_N = T$ such that if there is any partition finer $0 = s_0 < s_1 < ... < s_{N'} = T$ and for any $\delta_k$ be a given arbitrarily in $[s_k, s_{k+1}]$ with $k = 0, ..., N' - 1$, one has:*

$$\left| \sum_{k=0}^{N'} \phi(\delta_k)[u(s_{k+1}) - u(s_k)] - I \right| \leq \varepsilon.$$

This property is very useful to prove the existence of solutions to the Mechanical problem that we will introduce in the next section. Regarding the concept of right and left limits of $u$, for any $[a, b] \subset [0, T]$, if $u$ is a BV function, the measure $du$ can be expressed by

$$\int_{[a,b]} du = u^+(b) - u^-(a).$$

Furthermore, for any discontinuous points in $\mathcal{H}$, we have $\int_{\{a\}} du = u^+(a) - u^-(a)$.

## 3.2 Description of the motion of one particle

### 3.2.1 Physical setting in Mechanics

The motivation for this study stems from analyzing a mechanical system with a finite number of degrees of freedom, represented as a point in Euclidean space $\mathbb{R}^2$, and subject to a single smooth constraint, often termed a *unilateral constraint*. Specifically, we examine a system where a particle moves toward a fixed surface.

$$L = \{q \in \mathbb{R}^2 : f(q) \geq 0\}$$

where $f : \mathbb{R}^2 \to \mathbb{R}$ a given sufficiently regular function. The movement of the particle has to meet the boundary of $L$, defined by $S = \{q \in \mathbb{R}^2 : f(q) = 0\}$.

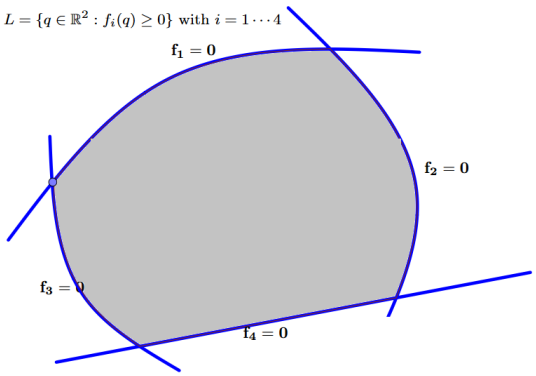



**Figure 3.1:** Illustration of fixed surface in 2D.

Assume the position at time $t$ of the particle denoted by $q = (q^x, q^y) \in \mathbb{R}^2$. The motion $t \to q(t)$ has to remain on $L$, so that $f(q(t)) \geq 0,\ t \in I$ with $I := [0, T]$ be an interval on $\mathbb{R}$. Hence, its the corresponding velocity $v = \dot{q}$ function of *bounded variation*. In contactless problem, for a usual velocity $v \in C^1$ of the motion with

$$v(t) = \lim_{h \to 0} \frac{1}{h}\big(q(t+h) - q(t)\big),$$

we recall the classical integrate

$$q(t) = q_0 + \int_0^t v(s)ds. \tag{3.2}$$

In particular, at the boundary $S$, the velocity of contact often occurs discontinuous. In this case, we prefer to consider the concept of velocity of BV function. That means, for any BV function $u \in \mathrm{BV}(I, \mathbb{R}^2)$, equation (3.2) can be written as

$$q(t) = q_0 + \int_0^t u(s)ds. \tag{3.3}$$

It is well-known that any BV functions $u$ defined on $[0, T]$ has a right limit $u^+$ and left limit $u^-$. Hence, one obtains that $u = u^+ = u^-$ Lebesgue almost everywhere on $[0, T]$ and then

$$q(t) = q_0 + \int_0^t u(s)ds = q_0 + \int_0^t u^+(s)ds = q_0 + \int_0^t u^-(s)ds.$$

That leads to the time-derivative of $q$ at each $t$ in (3.3) or the *right velocity* (resp. the *left velocity*) of the motion $\dot{q}(t)$ can be expressed by the right limit $u^+(t)$ (resp. left limit $u^-(t)$), i.e.

$$\dot{q}^+(t) = u^+(t) = \lim_{h\to 0^+} \frac{1}{h}\left(q(t+h) - q(t)\right).$$

Let us consider the model

$$M\ddot{q} = F(t, q, \dot{q}) + \Lambda,$$

where $M$ is a mass of the particle, $F$ is a given external force and $\Lambda \in \mathbb{R}^2$ is the vector of the contact or "*reaction*" forces. The reaction force is the force acting on the boundary $S$ of the fixed surface. Assume that $M = I$ be the unit matrix, then

$$\Lambda = \ddot{q} - F(t, q, \dot{q}). \tag{3.4}$$

The acceleration of the motion, denoted by $\ddot{q}$, is the derivative of a BV function $u$ with $u = \dot{q}$, we can rewrite (3.4) as

$$\Lambda = \dot{u} - F(t, q, \dot{q}).$$

And thus, since $u$ is a measure, it ensures $\Lambda$ is also a measure

$$d\Lambda = du - F(t, q, \dot{q})dt \tag{3.5}$$

where $dt$ is the Lebesgue on $[0, T]$.

Since, the motion of the particle towards a fixed $L$ with constraint $f(q(t)) \geq 0$. One remark, in the case that $f(q(t)) > 0$, the moving is in the interior of the region, means $d\Lambda = 0$. From (3.5), it follows that

$$\frac{du}{dt} = F(t, q, \dot{q}).$$

In addition, by the smooth property of $u$ and $u^+ = u^-$ be absolutely continuous, it is obvious that there is no reaction here (**contactless**). Otherwise, when $u$ is a discontinuous velocity, there exists an instant $t_k$ on the interval such that $u(t_k)^+ \neq u(t_k)^-$ and its Lebesgue measure is zero. Combining to the fact $du$ coincides with a the Dirac measure, we obtain from (3.5) that

$$d\Lambda(t_k) = du(t_k) = u(t_k)^+ - u(t_k)^- = \int_{t_k} du. \tag{3.6}$$

We observe that there exists a **contact** which can be represented in the form of the reaction force.

On the other hand, since at any points on the boundary $S$, the particle may change the direction of moving (see Fig. 3.2), then this leads to the velocity is discontinuous at these points. A **contact** or **impact** with minimum value occurs at the boundary $S$

where $f(q(t)) = 0$ and obviously the reaction force is nonzero. As mentioned, when $u$ is discontinuous at any time, the right velocity and the left velocity of the motion is basically different. Now, derivating both the right and left velocity, we get that

$$\left\langle \nabla f(q(t_k)), u^+(t_k) \right\rangle \geq 0, \tag{3.7}$$

$$\left\langle \nabla f(q(t_k)), u^-(t_k) \right\rangle \leq 0, \tag{3.8}$$

when $f(q(t_k)) = 0$, we have $\langle \nabla f(q(t_k)), u^-(t_k) \rangle < 0$ and $\langle \nabla f(q(t_k)), u^+(t_k) \rangle > 0$ and obviously $u(t_k)^+ \neq u(t_k)^-$.

Now, let us define

$$V(q) = \begin{cases} \{v \in \mathbb{R}^2 : \langle v, \nabla f(q) \rangle \geq 0\} & \text{if } f(q) \geq 0, \\ \mathbb{R}^2 & \text{if } f(q) < 0, \end{cases} \tag{3.9}$$

to be the set of *admissible right velocities* of $q$ on $L$, also known as the *tangent cone* to $L$ at the point $q$. It is clear that the interior of $V(q)$ is nonempty, and from inequality 3.7), we observe that

$$u^+(t) \in V(q) \quad \text{and} \quad -u^-(t) \in V(q) \quad \text{for all } \ t \in [0, T]. \tag{3.10}$$

With the definition of the admissible velocities of $q$ at $\dot{v}$ on $L$ (defined by 3.9), it is well-known that *the normal cone* $N(V(q), u)$ (in short, $N_q$) is the polar of $V(q)$,

$$N_q = \left\{ w \in \mathbb{R}^2 \mid \langle w, v \rangle \leq 0, \forall v \in V(q) \right\}. \tag{3.11}$$

**Proposition 3.7.** *One has*

$$N(V(q), u) = \{ -\lambda \nabla f(q) \mid \lambda = 0 \text{ if } f(q) > 0 \text{ or } \lambda > 0 \text{ if } f(q) = 0 \}. \tag{3.12}$$

*Proof.* First, we observe that $V(q)$ is nonempty since it is either the whole space or a half-space, and it is also a closed convex cone. In addition, based on the basic result known as the Polar Cone Theorem (see details in [120]), which states that if $S$ is itself a nonempty closed convex set, then $(V(q)^\circ)^\circ = V(q)$. Now, setting

$$K := \{ -\lambda \nabla f(q) \mid \lambda = 0 \text{ if } f(q) > 0 \text{ or } \lambda > 0 \text{ if } f(q) = 0 \},$$

our aim is to show that $V(q) = K^\circ$, and then we can apply the Polar Cone Theorem to conclude. Indeed, fixing $v \in V(q)$, we obtain

$$-\langle \nabla f(q), v \rangle \leq 0.$$

Multiplying by a nonnegative $d$, this is equivalent to

$$-\langle d \cdot \nabla f(q), v \rangle \leq 0. \tag{3.13}$$

This leads to $V(q) = K^\circ$. Therefore, we obtain that

$$\overset{\circ}{N_{V(q)}} = (V(q)^\circ)^\circ = V(q) = K^\circ.$$

Consequently, by taking the polar of both sides, we are able to conclude that the normal cone of $V(q)$ coincides with $K$, i.e.,

$$N(V(q), u) = \{ -\lambda \nabla f(q) \mid \lambda = 0 \text{ if } f(q) > 0 \text{ or } \lambda > 0 \text{ if } f(q) = 0 \}.$$

The proof is complete.

□

In Mechanics, in the case of non-friction (**frictionless**) unilateral contacts, the reaction of the material (particle, body, etc.) can exert a force on the boundary that is normal and pointing outward. Specifically, there exists a constant $\lambda \geq 0$ such that

$$\Lambda = \lambda \nabla f(q) = \lambda \nu,$$

where $\nu$ is the outward normal vector on the boundary $S$ (assuming the function $f$ is sufficiently regular).

Assuming that the duration of the reaction $\Lambda$ after the shock is very small, such that the reaction does not change the direction of motion, from (3.6) we obtain

$$u(t_k)^+ - u(t_k)^- = \alpha \nabla f\big(q(t_k)\big) \quad \text{where } \alpha > 0. \tag{3.14}$$

Building upon the model proposed by Kanze and Marques in 2000 [11], we consider **the second-order sweeping process** problem, formulated as follows.

$$\begin{cases} -du(t) \in N\big(V(q(t)); u(t)\big), \quad u(t) \in V(q(t)), \\ q(0) = q_0, \\ u(0) = u_0, \end{cases}$$

where $V(q)$ is defined by (3.9) and the normal cone $N(V(q), u)$ is defined by (3.12).

We have also introduced the regularization of the Moreau envelope-Yosida approximation of the contact problem based on the evolution equation associated with the Maximal monotone operator (see Section 2.4). This enables us to obtain the *regularization* system as

$$\begin{cases} -du^\alpha(t) = A_r^\alpha u_r(t), \\ q_r(0) = q_0, \ u_r(0) = u_0. \end{cases}$$

Hence, following a similar approach, we derive the Yosida approximation of $A^t$, denoted as $(A^t)_r^\alpha$, where

$$A^t(\cdot) = N(V(q(t)), \cdot) = N_q(\cdot) = \partial \psi_{C_q}(\cdot).$$

Consequently, we find that for any $r > 0$, the relation holds:

$$(A^t)_r^\alpha(u) = \frac{1}{\alpha r}\left(\nabla d^\alpha_{V(q)(t)}(u)\right).$$

The existence of solutions to certain sweeping process problems have been rigorously demonstrated and proven in [8, 22, 29, 30, 42, 44].

### 3.2.2 Impact Laws

In this section, we derive from the contact or impact dynamics in Section 3.2. In particular, it brings meaningful result in Mechanics. To verify this fact, we start with the simply case of a purely inelastic impact.

#### 3.2.2.1 Purely inelastic impact

On the other hand, let us consider when the right velocity $u^+$ at any point on $S$ after shock with the boundary is *tangential* $T(q)$, corresponds to $\langle u^+, \nabla f(q)\rangle = 0$, where

$$T(q) = \left\{w \in \mathbb{R}^2 : \langle w, \nabla f(q)\rangle = 0\right\}.$$

We can easily observe that the right velocity $u^+$ is the orthogonal projection of the left velocity $u^-$ onto $T(q)$. Regarding the normal cone of $N_q$, for $u = u^+ \in N_q$ (see Fig. 3.2-(a)), we have

$$-d\Lambda \in N(V(q), u).$$

#### 3.2.2.2 Purely elastic impact: Conservation of kinetic energy

In Mechanics, the **kinetic energy** can be expressed by

$$\mathcal{E}(\dot{q}) = \frac{1}{2}\left\|\dot{q}\right\|^2. \tag{3.15}$$

Assume that $u$ be tangential with $u = \dfrac{1}{2}(u^+ + u^-) \in T(q) \subset V(q)$ and one has

$$-d\Lambda \in N(V(q), u) = N\left(V(q), \frac{1}{2}(u^+ + u^-)\right).$$

By definition of $N_q$, we obtain

$$\langle d\Lambda, u\rangle = \left\langle d\Lambda, \frac{1}{2}(u^+ + u^-)\right\rangle = \left\langle u^+ - u^-, \frac{1}{2}(u^+ + u^-)\right\rangle = 0 \tag{3.16}$$

by (3.6). It follows that

$$\frac{1}{2}\left\|u^+\right\|^2 - \frac{1}{2}\left\|u^-\right\|^2 = 0 \;\Leftrightarrow\; \frac{1}{2}\left\|u^+\right\|^2 = \frac{1}{2}\left\|u^-\right\|^2.$$

It means that $\|u^+\| = \|u^-\|$ and also corresponds to $\mathcal{E}(\dot{q}^+) = \mathcal{E}(\dot{q}^-)$. Consequently, we observe that the *purely elastic impact* ensures the conservation of kinetic energy in the meaning of $\|u^+\| = \|u^-\|$ (see Fig. 3.2-(b)). It plays an important role in Mechanics, especially in numerical computation.

#### 3.2.2.3 Partially elastic impact

In mechanics, there are scenarios where a body's energy is not conserved. This occurs when the body's velocity exhibits partial elasticity, characterized by $\|u^+\| < \|u^-\|$. The phenomenon of partial elasticity is illustrated in Figure 3.2 (see [66]). When a reaction force acts on the body at point $S$, a dissipative shock arises, leading to the following relationship:

$$u^+ = \mathrm{proj}_{N_q}(u^-)$$

or equivalently,

$$u^+ - u^- \in -N_q(u^+).$$

To generalize further, let the velocity $u$ be a convex combination of its right element $u^+$ and left element $u^-$, such that

$$u = \frac{1}{1+e}u^+ + \frac{e}{1+e}u^-, \tag{3.17}$$

where $e \in [0,1]$ is known as the "restitution coefficient." It is worth noting that when $e = 0$, the combination (3.17) represents purely inelastic impacts, and when $e = 1$, it corresponds to purely elastic impacts.

For any constants $t_k$ at which $u_{t_k}$ becomes discontinuous, it is necessary that $-d\mathbf{\Lambda} \in N(V(q), u)$ to ensure that $u$ belongs to $V_q$. When $u$ lies in the interior of $V_q$, the reaction force is zero since $N_q(u)$ is zero as well. In numerical computations, the problem is often handled through discretization, which implies

$$-d\Lambda \in N_q \quad \text{with } u = \mathrm{Av}(u^-, u^+)$$

where $\mathrm{Av}(u^-, u^+)$ is a type of linear interpolation defined as

$$\mathrm{Av}(u^-, u^+) = u^+ + k(u^- - u^+),\ 0 \le k \le \frac{1}{2}. \tag{3.18}$$

**Remark 3.8.** Note that when $k = 0$, it represents purely inelastic impacts, and when $k = \frac{1}{2}$, it corresponds to purely elastic impacts.

From (3.17) and 3.18, one deduces $k = \frac{e}{1+e}$. Additionally, we expect an extension in the form of a *dissipation index* $\delta$ by putting $e = \frac{1-\delta}{1+\delta}$; $\delta = \frac{1-e}{1+e}$ and $k = \frac{1-\delta}{2}$, which gives

$$u = \frac{1+\delta}{2}u^+ + \frac{1-\delta}{2}u^-,$$

as illustrated in Fig. 3.2-(c,d).

**Remark 3.9.** By using $\langle d\Lambda, u\rangle = 0$, it is easy to see that

$$\frac{1}{2}\left\|u^+\right\|^2 - \frac{1}{2}\left\|u^-\right\|^2 = -\frac{1}{2}\left(\frac{1-e}{1+e}\right)\left\|u^+ - u^-\right\|^2 = -\frac{\delta}{2}\left\|u^+ - u^-\right\|^2,$$

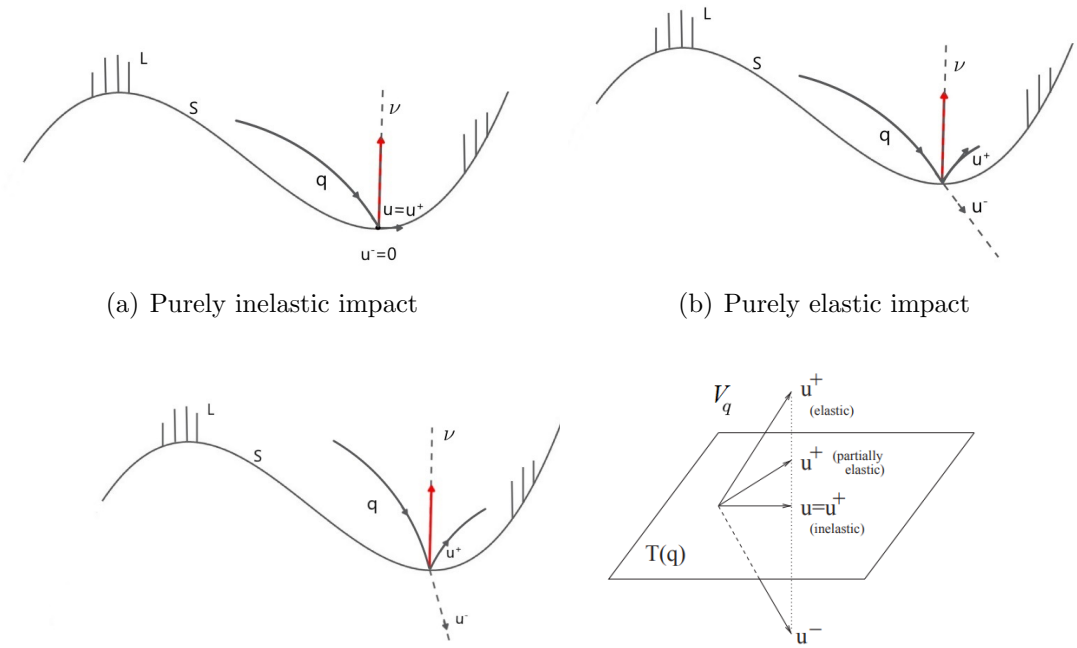
and by definition (3.15), we obtain the dissipation of energy

$$\mathcal{E}(u^+) - \mathcal{E}(u^-) = -\delta\mathcal{E}(u^+ - u^-) \le 0.$$

Clearly, when $\delta = 0$ (resp. $e = 1$), this is the case of kinetic energy conservation, $\mathcal{E}(\dot{q}^+) = \mathcal{E}(\dot{q}^-)$.

### 3.2.3 Problem of impact via the second order sweeping process

From all assumptions above, we need to find $u : I \to \mathbb{R}^2$ : be a BV function with e-averaged of the problem




(a) Purely inelastic impact

(b) Purely elastic impact

(c) Partially elastic impact

(d) Illustration of the general impact laws

**Figure 3.2:** Impact laws concerning velocity.

**Problem 2.** Let $I := [0, T]$ be an interval in Hilbert space. Let $q_0 \in L = \{q : f(q) \geq 0\}$, $u_0 \in V(q_0)$ (defined by (3.9)) and let $p : I \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$. We try to find $u \in \mathrm{BV}(I, \mathbb{R}^2)$ with e-averaged and $u(0) = u_0$ be an initial velocity. Assume that $u(t) \in V_q$ for $t \in I$ with

$$q(t) = q_0 + \int_0^t u(s)ds.$$

Moreover, assume that there is a nonnegative measure $d\mu$ on $I$ relative to its Lebesgue measure dt with $t'_\mu \in L^1(I, d\mu; \mathbb{R})$ and the Stieltjes measure du with $u'_\mu \in L^1(I, d\mu; \mathbb{R}^2)$, one has

$$-d\Lambda = p\big(t, q(t), u(t)\big)dt - du \in N_q(u(t))$$

and the differential inclusion

$$F\big(t, q(t, u(t))\big)t'_\mu(t) - u'_\mu(t) \in N_q(u(t))$$

satisfies for $d\mu$ almost every $t \in I$.

In the subsequent analysis, we shall focus on a simplified scenario involving a moving point or particle directed towards a stationary surface denoted by $L$ (described by (6.14)). We consider a mechanical problem with its motion represented by the variable $q \in \mathbb{R}^2$, and the corresponding velocity is treated as a function of bounded variation. The differential measure $d\dot{q}$ is associated with the velocity $\dot{q}$.

**Problem 3.** The model with velocity contact problem

$$(P_{cv})\begin{cases} Md\dot{q} = F(t,q,\dot{q})dt + d\Lambda, \\ q(0) = q_0, \\ \dot{q}(0) = u_0, \end{cases}$$

where $d\Lambda \in -N_q(\dot{\tilde{q}})$ with $\dot{\tilde{q}} = \frac{1}{2}\big(\dot{q}^+ + \dot{q}^-\big)$.

Problem 3 uses a model with velocity contact because in argument in the normal cone $N_q$ there is $\dot{\tilde{q}}$. On the other hand, we also can consider the model with displacement contact as follow.

**Problem 4.** The model with displacement contact problem

$$(P_{cd})\begin{cases} Md\dot{q} = F(t,q,\dot{q})dt + d\Lambda, \\ q(0) = q_0, \\ \dot{q}(0) = u_0, \end{cases}$$

where $q \in L$ and $d\Lambda \in N_L(q) = \{-\lambda\nabla f(q) \mid \lambda = 0 \text{ if } f(q) > 0 \text{ or } \lambda > 0 \text{ if } f(q) = 0\}$ and

$$L = \begin{cases} \{q \in \mathbb{R}^2 : \langle q, \nabla f(q)\rangle \geq 0\} & \text{if} \quad f(q) \geq 0, \\ \mathbb{R}^2 & \text{if} \quad f(q) < 0. \end{cases}$$

**Proposition 3.10.** *If $d\Lambda \in -N_L(q)$ then there exists a $\lambda \geq 0$ such that*

$$\lambda \in -\partial\psi_{\mathbb{R}^+}\big(\langle \nabla f(q), q\rangle\big)$$

*whenever $\lambda > 0$ if $f(q) = 0$ or $\lambda = 0$ if $f(q) > 0$.*

*Proof.* Assume that $d\Lambda \in -N_L(q)$, we obtain that $d\Lambda = \lambda\nabla f(q)$ with $\lambda \geq 0$ (by Proposition 3.7), i.e $\lambda = 0$ if $f(q) > 0$ or $\lambda > 0$ if $f(q) = 0$. Moreover, since $d\Lambda \in -N_L(q)$, it can be represented by (3.11) in form of $-\langle d\Lambda, q\rangle \leq 0$ or $\langle d\Lambda, q\rangle \geq 0$. Hence,

$$x := \langle \lambda\nabla f(q), q\rangle \geq 0. \tag{3.19}$$

It is clear that when $\lambda = 0$, one has $x > 0$ (since $f(q) > 0$) to hold the inequality (3.19). And also, if $\lambda > 0$, $x = 0$ (since $f(q) = 0$). Thus, we easily to obtain

$$\lambda \in -N_{\mathbb{R}^+}(x),$$

where

$$-N_{\mathbb{R}^+}(x) = \begin{cases} 0 & \text{if } x > 0, \\ [0, +\infty) & \text{if } x = 0, \\ \emptyset & \text{if } x < 0. \end{cases}$$

In addition, we also notice that $N_{\mathbb{R}^+}(x) = \partial\psi_{\mathbb{R}^+}(x)$. Thus, we get that $\lambda \in -\partial\psi_{\mathbb{R}^+}(x)$. □

In addition, by using Karush–Kuhn–Tucker (KKT) condition [120], we have a property which is a unilateral contact displacement $q$ corresponding to $u$ (Fig. 3.3).

**Proposition 3.11.** *From all setting above, the following statement holds*

$$\lambda \in -\partial\psi_{\mathbb{R}^+}(x) \Leftrightarrow \begin{cases} \lambda \geq 0, \\ x \geq 0, \\ \langle \lambda, x \rangle = 0. \end{cases} \tag{3.20}$$

*Proof.* Based on the proof of Proposition 3.10, it is well-known that for any $\lambda \in -\partial\psi_{\mathbb{R}^+}(x)$, we have

$$\lambda = 0 \text{ and } x \geq 0, \text{ or } \lambda \geq 0 \text{ and } x = 0. \tag{3.21}$$

Let $-\lambda \in \partial\psi_{\mathbb{R}^+}(x)$, by using definition of subdifferential, we obtain that

$$\begin{aligned} & \psi_{\mathbb{R}^+}(x) + \psi^\star_{\mathbb{R}^+}(-\lambda) + \langle x, -\lambda \rangle \leq 0 \\ \Leftrightarrow\ & \psi_{\mathbb{R}^+}(x) + \psi_{\mathbb{R}^-}(-\lambda) + \langle x, \lambda \rangle \leq 0 \\ \Leftrightarrow\ & \langle x, \lambda \rangle \leq 0 \end{aligned}$$

due to $\psi_{\mathbb{R}^+}(x) = 0$. In addition, it is clear that $x \geq 0$ and $\lambda \geq 0$, then

$$\langle x, \lambda \rangle \geq 0.$$

We get the conclusion $\langle x, \lambda \rangle = 0$. □

In the context of optimization, the classical expression of the *Complementarity Condition* (also known as the Karush-Kuhn-Tucker condition in [120]) for pairs of complementary variables involves a nonnegative variable $\lambda$ and the inner product $\langle \nabla f(q), q \rangle$ or $D^\gamma_\nu$, the signed distance from particle $i$ to particle $j$ at contact $\gamma$. This condition requires that at least one of the two inequalities reaches its bound. Alternatively, it can be formulated as

$$0 \leq \lambda \perp \langle \nabla f(q), q \rangle \geq 0. \tag{3.22}$$

In this expression, the symbol $\perp$ represents the orthogonality relationship between the variables $\lambda$ and $\langle \nabla f(q), q \rangle$, indicating that they are complementary and cannot both be strictly positive. At least one of them must be at its lower or upper bound, while the other is nonnegative.

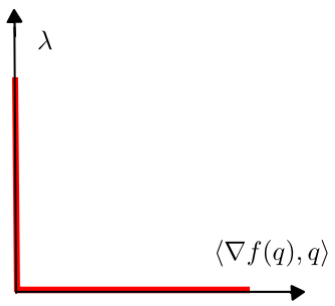



**Figure 3.3:** Unilateral contact displacement.

## 3.3 Description of the collective motion of multiple particles - Granular media problem

To describe the motion of a multi-contact system between rigid bodies (see Fig. 3.4), we use the following writing conventions. Assuming that a particle $q$ among $N_q$ particles is described by the position of its center and its rotation, we will denote by $\boldsymbol{q}$ the generalized coordinate describing its position in space of the entire multibody system ($\boldsymbol{q} \in \mathbb{R}^{d \times N_q}$, where $d = 6$ for a 3D problem and $d = 3$ for a 2D problem). As a consequence of the possible collisions between particles, we introduce the generalized velocity denoted by $\dot{\boldsymbol{q}}$ as a function of bounded variations, and its associated differential $d\dot{\boldsymbol{q}}$. According to the fundamental principle of rigid dynamics, the equations of motion formulated in terms of differential measures can be written as follows

$$\mathbb{M} d\dot{\boldsymbol{q}} = \boldsymbol{F}(t, \boldsymbol{q}, \dot{\boldsymbol{q}}) dt + d\boldsymbol{\Lambda} \tag{3.23}$$

where $\mathbb{M}$ represents the generalized mass matrix; $\boldsymbol{F}$ represents external forces; $dt$ is the Lebesgue measure and $d\boldsymbol{\Lambda}$ is a non-negative real measure that represents the reaction forces and impulses occurring between particles in contact.

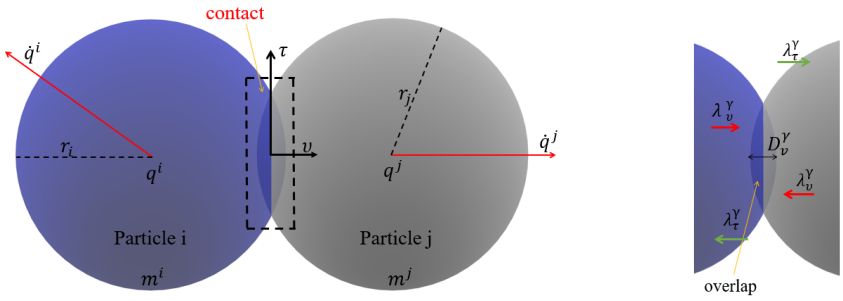



**Figure 3.4:** The multi-contact problem.

We now establish the mechanical interaction arising from a potential collision between a rigid particle $i$ and its neighbor $j$, involving a contact point $\gamma$. As the contact laws are expressed using the contact variables, we have to express these local variables from the generalized velocity and impulses. Thus, we define a local-global mapping from the global to the local frame for each contact node $\gamma \in [1, N_\gamma]$, with $N_\gamma$ denoting the total number of possible contacts between two rigid particles to write the corresponding local contact law.

For a potential contact $\gamma$ ($\gamma \in [1, N_\gamma]$), the local relative velocity $\boldsymbol{v}^\gamma$ and the local contact impulse $\boldsymbol{\lambda}^\gamma$ can be respectively decomposed into a sum of normal and tangential components $\boldsymbol{v}^\gamma = \boldsymbol{v}_\nu^\gamma \cdot \nu + \boldsymbol{v}_\tau^\gamma$ and $\boldsymbol{\lambda}^\gamma = \boldsymbol{\lambda}_\nu^\gamma \cdot \nu + \boldsymbol{\lambda}_\tau^\gamma$ where $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$ are respectively the inward unit normal and tangential vector.

The shock/contact law can be written as follows.

$$\text{If} \quad D_\nu^\gamma > 0 \quad \text{then} \quad \lambda_\nu^\gamma = 0, \tag{3.24}$$

$$\text{if} \quad D_\nu^\gamma \leq 0 \quad \text{then} \quad \begin{cases} v_\nu^\gamma \geq 0, \\ \lambda_\nu^\gamma \geq 0, \\ v_\nu^\gamma \lambda_\nu^\gamma = 0 \end{cases} \tag{3.25}$$

where $D_\nu^\gamma = D_\nu^\gamma(q_i, q_j)$ represents the signed distance from particle $i$ to particle $j$ at contact $\gamma$.

The Coulomb friction law is written as:

$$\boldsymbol{\lambda}_\tau^\gamma \neq \mathbf{0}. \tag{3.26}$$

$$\begin{cases} \lambda_\nu^\gamma = 0: \ v_\nu^\gamma \geq 0, \\ \lambda_\nu^\gamma > 0 \quad \text{and} \quad ||\boldsymbol{\lambda}_\tau^\gamma|| < \mu|\lambda_\nu^\gamma| : \ \boldsymbol{v}_\tau^\gamma = 0, \\ \lambda_\nu^\gamma > 0 \quad \text{and} \quad \|\boldsymbol{\lambda}_\tau^\gamma|| = \mu|\lambda_\nu^\gamma| : \ \exists \beta \geq 0, \boldsymbol{v}_\tau^\gamma = \beta \dfrac{\boldsymbol{\lambda}_\tau^\gamma}{||\boldsymbol{\lambda}_\tau^\gamma||}, \end{cases} \tag{3.27}$$

where $\mu$ is the friction coefficient.

In the case of a collision, the non-smooth movement implies sudden changes in velocity (*velocity jumps*). They are determined under the non-interpenetrability conditions. The time derivative $\dot{u}$ is not unique and then, the left-limit and the right-limit velocity, denoted as $v_\nu^-$ $v_\nu^+$ respectively, have to be distinguished separately. By analogy with a binary collision between two rigid bodies, we compute $v_\nu^+$ based on $v_\nu^-$ using a time-stepping scheme. This leads us to compute $v_\nu$ in Signorini's conditions by taking a weighted average of $v_\nu^-$ and $v_\nu^+$, expressed as $v_\nu = \eta v_\nu^- + (1-\eta) v_\nu^+$ where $\eta$ characterizes the contact properties. This concept is detailed in references such as [24, 25, 46, 48]. For $\eta \neq 0$, the ratio $\frac{-v_\nu^+}{v_\nu^-} = \frac{\eta}{(1-\eta)}$ in a collision between rigid particles is identified with the normal restitution coefficient $e_\nu$, leading to $\eta = \frac{e_\nu}{(1+e_\nu)}$. Consequently, we define

$$v_\nu = \frac{v_\nu^+ + e_\nu v_\nu^-}{(1+e_\nu)}.$$

The computation of tangential velocity $\mathbf{v}_\tau$, governed by Coulomb's frictional conditions, follows a similar approach. The tangential force $\boldsymbol{\Lambda}_\tau$ represents the combined effect of static and repulsive forces during the time interval $\Delta t$. The tangential velocity is thus determined as an average velocity, as explained in [24, 25, 46, 48]. Similar to normal forces, a consistent model incorporating the tangential restitution coefficient $e_\tau$ is defined by

$$\boldsymbol{v}_\tau = \frac{\boldsymbol{v}_\tau^+ + e_\tau \boldsymbol{v}_\tau^-}{(1+e_\tau)}.$$

In the following, two couples $(v_\nu^\gamma, \lambda_\nu^\gamma)$ and $(\boldsymbol{v}_\tau^\gamma, \boldsymbol{\lambda}_\tau^\gamma)$ verifying these set of conditions for a potential contact node $\gamma$ between two particles are denoted by

$$\lambda_\nu^\gamma \in -\partial \psi_{\mathbb{R}^+}(v_\nu^\gamma), \tag{3.28}$$

$$\boldsymbol{\lambda}_\tau^\gamma \in \partial \psi^\star_{C(\boldsymbol{\lambda}_\tau^\gamma)}(\boldsymbol{v}_\tau^\gamma) \tag{3.29}$$

where $\psi_S$ is the indicator function of a set $S \subset X$ defined by $\psi_S(x) = 0$ if $x \in S$ and $\psi_S(x) = +\infty$ if othervise; $\psi_S^\star(\cdot)$ is the Fenchel conjugate of $\psi_S(\cdot)$ of $S$. In the latter formulation, we notice that (3.28) is related to the the inelastic shock law's condtions (3.24) – (3.25), while (3.29) describes the Coulomb's friction law (3.26) – (3.27).

## 3.4 The displacement contact problem

Firstly, we consider mechanical problem described by its motion $\boldsymbol{q} \in \mathbb{R}^{d\times N_q}$, its corresponding velocity $\dot{\mathbf{q}}$ as a function of bounded variation and $d\dot{\boldsymbol{q}}$, the differential measure associated, given by

**Problem 5.** The model with velocity contact problem

$$\begin{cases} \mathbb{M}d\dot{\boldsymbol{q}} = \boldsymbol{F}dt + d\boldsymbol{\Lambda}, \\ \boldsymbol{q}(0) = \boldsymbol{q}_0, \\ \dot{\boldsymbol{q}}(0) = \boldsymbol{v}_0 \end{cases}$$

where, $\lambda_\nu^\gamma \in -\partial\psi_{\mathbb{R}^+}(v_\nu^\gamma)$ and $\boldsymbol{\lambda}_\tau^\gamma \in \partial\psi^\star_{C(\boldsymbol{\lambda}_\tau^\gamma)}(\boldsymbol{v}_\tau^\gamma)$ for each contact $\gamma \in [1, N_\gamma]$.

In particular, the velocity contact problem corresponds to the case of the *purely elastic impact*, which ensures the conservation of kinetic energy in the sense that $|v_\nu^+| = |v_\nu^-|$. It plays an important role in contact mechanics, especially in numerical computation.

Now, for one contact $\gamma$ between two particles, we will omit $\gamma$ in the following for the sake of simplicity. On the other hand, it leads to the model with displacement contact as follow.

**Problem 6.** The model with displacement contact problem via subdifferential form

$$\begin{cases} \mathbb{M}d\dot{\boldsymbol{q}} = \boldsymbol{F}dt + d\boldsymbol{\Lambda}, \\ \boldsymbol{q}(0) = \boldsymbol{q}_0, \\ \dot{\boldsymbol{q}}(0) = \boldsymbol{v}_0 \end{cases}$$

where $\lambda_\nu^\gamma \in \partial\psi_{\mathbb{R}^-}(D_\nu^\gamma)$ and $\boldsymbol{\lambda}_\tau^\gamma \in \partial\psi^\star_{C(\boldsymbol{\lambda}_\tau^\gamma)}(\boldsymbol{v}_\tau^\gamma)$ for each contact $\gamma \in [1, N_\gamma]$.

This model cannot ensure the conservation of the energy of the system since the contact law does not allow to determine the velocity $\boldsymbol{v}^+$ from $\boldsymbol{v}^-$. In the following, we will approximate this law in order to be able to ensure the conservation of the system.

### 3.4.1 The Moreau-Yosida regularization for Contact Dynamics

In this section, we introduce a regularization framework for contact conditions. We begin by outlining the fundamentals of Moreau-Yosida regularization (see Section 2.5) and then present the regularized formulations in both continuous and discrete contexts, accompanied by an energy analysis. Related to Definition (2.25) and Theorem (2.26), when $f$ coincides to the indicator function of the set C, for any $R > 0$ and with $\alpha \geq 2$,

one has

$$\big(\psi_C\big)_R^{\alpha}(u) = \inf_{y\in\mathcal{H}}\left(\psi_C(y) + \frac{1}{\alpha R}\left\|y-u\right\|^{\alpha}\right) = \inf_{y\in C}\frac{1}{\alpha R}\left\|y-u\right\|^{\alpha} = \frac{1}{\alpha R}d_C^{\alpha}(u). \quad (3.30)$$

Upon computing the gradient of $(\psi_C)_R^{\alpha}$, we derive the following proposition.

**Proposition 3.12.** *With $\alpha \geq 2$, one has*

$$\nabla\big(\psi_C\big)_R^{\alpha}(u) = \nabla\big(\frac{1}{\alpha R}d_C^{\alpha}(u)\big) = \frac{1}{R}\left\|\mathrm{proj}_{C_0}(u)\right\|^{\alpha-2}\mathrm{proj}_{C_0}(u).$$

To prove it, first let us recall the lemma below.

**Lemma 3.13.** *One has*

$$\nabla\left\|x\right\|^{\alpha} = \alpha\left\|x\right\|^{\alpha-2}x.$$

*Proof.* Indeed, it is not difficult to check that

$$\nabla\left\|x\right\|^{\alpha} = \nabla\left[\big(\left\|x\right\|^2\big)^{\frac{\alpha}{2}}\right] = \frac{\alpha}{2}\left(2x\right)\big(\left\|x\right\|^2\big)^{\frac{\alpha}{2}-1} = \alpha\left\|x\right\|^{\alpha-2}x.$$

□

*Proof of Proposition 3.12.* Applying directly the definition of distance function and the lemma above, for any $R>0$ and $\alpha\geq 2$, we get that

$$\begin{aligned}\nabla(\psi_C)_R^{\alpha}(u) &= \nabla\big(\frac{1}{\alpha R}d_C^{\alpha}(u)\big) = \nabla\big(\frac{1}{\alpha R}\inf_{u\in C}\left\|y-u\right\|^{\alpha}\big)\\ &= \frac{1}{\alpha R}\nabla\big(\left\|y-\mathrm{proj}_C(y)\right\|\big)^{\alpha}\\ &= \frac{1}{R}\left\|y-\mathrm{proj}_C(y)\right\|^{\alpha-2}\big(y-\mathrm{proj}_C(y)\big)\\ &= \frac{1}{R}\left\|\mathrm{proj}_{C_0}(u)\right\|^{\alpha-2}\mathrm{proj}_{C_0}(u).\end{aligned}$$

□

We are now in position to find the $\alpha$-*Moreau-Yosida* regularization of $\partial\psi_{\mathbb{R}^-}\big(u\big)$ by taking $C\equiv\mathbb{R}^-$.

**Corollary 3.14.** For any $R>0$ and $\alpha\geq 2$, one has

$$\lambda_\nu = (\partial\psi_{\mathbb{R}^-})_R^{\alpha}(u) \Leftrightarrow \lambda = \frac{1}{R}\left\|\mathrm{proj}_{\mathbb{R}^+}(u)\right\|^{\alpha-2}\mathrm{proj}_{\mathbb{R}^+}(u) \text{ for any } u\in\mathcal{H}.$$

We derive from Corollary 3.14 to apply to the displacement contact problem 6 to get that, for any $R>0$ and $\alpha\geq 2$,

$$\lambda_\nu = \frac{1}{R}\left\|\mathrm{proj}_{\mathbb{R}^+}\big(D_\nu\big)\right\|^{\alpha-2}\mathrm{proj}_{\mathbb{R}^+}\big(D_\nu\big) = \frac{1}{R}\mathrm{proj}_{\mathbb{R}^+}\big(D_\nu\big)^{\alpha-1} =: \frac{1}{R}\big(\left[D_\nu\right]_+\big)^{\alpha-1}.$$

Set $c_\nu := 1/R$ and for simplicity, we omit the contact exponent $\gamma$ for each contact $\gamma\in[1,N_\gamma]$.

**Problem 7.** The $\alpha$-Moreau-Yosida regularization of displacement contact problem

$$\begin{cases} \mathbb{M}d\dot{\boldsymbol{q}} = \boldsymbol{F}dt + d\boldsymbol{\Lambda}, \\ \boldsymbol{q}(0) = \boldsymbol{q}_0, \\ \dot{\boldsymbol{q}}(0) = \boldsymbol{v}_0, \end{cases}$$

where $\lambda_\nu = c_\nu\big([D_\nu]_+\big)^{\alpha-1}$, $c_\nu > 0$, $\alpha \geq 2$, and

$$\begin{cases} \text{if} \quad \|\boldsymbol{\lambda}_\tau\| < \mu\,\lambda_\nu : \quad \boldsymbol{\lambda}_\tau = c_\tau \dot{\boldsymbol{q}}_\tau, \\ \text{if} \quad \|\boldsymbol{\lambda}_\tau\| \geq \mu\,\lambda_\nu : \quad \boldsymbol{\lambda}_\tau = \mu\,\lambda_\nu \dfrac{\dot{\boldsymbol{q}}_\tau}{\|\dot{\boldsymbol{q}}_\tau\|}. \end{cases}$$

Based on the Normal compliance method, we will provide balance energy over the interval $[0, T]$. Let us consider the work of contact force

$$W_C = \Big\langle \lambda_\nu, \dot{D}_\nu \Big\rangle \quad \text{with } \lambda_\nu = c_\nu\big([D_\nu]_+\big)^{\alpha-1}.$$

It is equivalent to

$$W_C = \Big\langle c_\nu\big([D_\nu]_+\big)^{\alpha-1}, \dot{D}_\nu \Big\rangle = \frac{c_\nu}{\alpha}\frac{d}{dt}\big([D_\nu]_+\big)^{\alpha}.$$

When moving continuously over the interval $[0, T]$, the energy holds this equation:

$$E_C^{[0,T]} = E_C(T) - E_C(0) = \int_0^T W_C \; ds.$$

From the framework of the work of contact force $W_C$, one has

$$\begin{aligned} E_C(T) - E_C(0) &= \int_0^T W_C \; ds \\ &= \frac{c_\nu}{\alpha}\int_0^T \frac{d}{ds}\big([D_\nu]_+\big)^{\alpha} ds \\ &= \frac{c_\nu}{\alpha}\big([D_\nu]_+\big)^{\alpha}(T) - \frac{c_\nu}{\alpha}\big([D_\nu]_+\big)^{\alpha}(0). \end{aligned} \tag{3.31}$$

It is easy to see that when $\alpha$ is large enough, the energy balance is guaranteed due to low penetrations from the Normal compliance regularization method.

### 3.4.2 Discrete formulation of the regularized contact problems

Let $T > 0$ be a real and fix any integer $N \geq 1$. We consider the time discretization of the interval $[0, T]$ given by $0 = t^0 < t^1 < ... < t^N = T$, i.e. $[0, T] = \cup_{n=0}^{N}[t^n, t^{n+1}]$ with $t^{n+1} = t^n + \Delta t$ where $\Delta t := T/N$.

#### 3.4.2.1 Standard normal compliance

Problem 7 then can be discretized in time in a classical form of Standard Normal Compliance (SNC) in the following via the Moreau-Yosida regularization with $\alpha = 2$.

**Problem 8.** Standard Normal Compliance problem

$$(P_{cd}^{SNC}) \quad \begin{cases} \mathbb{M}\dfrac{\dot{\boldsymbol{q}}^{n+1} - \dot{\boldsymbol{q}}^{n}}{\Delta t} = \boldsymbol{F}^{n+1} + \boldsymbol{\Lambda}^{n+1}, \\ \boldsymbol{q}(0) = \boldsymbol{q}_0, \\ \dot{\boldsymbol{q}}(0) = \boldsymbol{v}_0, \end{cases}$$

where $\lambda_\nu^{n+1} = c_\nu [D_\nu^{n+1}]_+$, $c_\nu > 0$.

Notice that the standard law of normal compliance does not ensure the balance energy under the continuous formalism see equation (3.31) unlike the Improved Normal Compliance model in the next section.

Based on the paper [57], we present an improved energy conservation method for granular media, extending it to account for dissipation phenomena with friction. This method enforces the standard normal compliance law at each time step, achieving minimal contact penetrations and maintaining conservation properties that closely respect energy.

#### 3.4.2.2 Improved normal compliance

In order to ensure the energy balance coming from the continuous case, we replace the normal displacement $[D_\nu^{n+1}]_+$ by $\tilde{D}_\nu^{n+1}$ in the normal stress of contact given by

$$\tilde{D}_\nu^{n+1} = \frac{([D_\nu^{n+1}]_+)^\alpha - ([D_\nu^{n}]_+)^\alpha}{\alpha(D_\nu^{n+1} - D_\nu^{n})}.$$

**Proposition 3.15.** *One has*

$$E_C^{n+1} - E_C^{n} = \frac{c_\nu}{\alpha}\big( \left[D_\nu^{n+1}\right]_+ \big)^\alpha - \frac{c_\nu}{\alpha}\big( [D_\nu^{n}]_+ \big)^\alpha.$$

*Proof.* By setting above, we have

$$\begin{aligned} E_C^{n+1} - E_C^{n} &= \int_{t^n}^{t^{n+1}} W_C \, ds = \int_{t^n}^{t^{n+1}} \left\langle \lambda_\nu^{n+1}, \dot{D}_\nu^{n+1} \right\rangle ds \\ &= c_\nu \Delta t \; \tilde{D}_\nu^{n+1} \frac{D_\nu^{n+1} - D_\nu^{n}}{\Delta t} \\ &= c_\nu \frac{\big( [D_\nu^{n+1}]_+ \big)^\alpha - \big( [D_\nu^{n}]_+ \big)^\alpha}{D_\nu^{n+1} - D_\nu^{n}} (D_\nu^{n+1} - D_\nu^{n}) \\ &= \frac{c_\nu}{\alpha}\big( \left[D_\nu^{n+1}\right]_+ \big)^\alpha - \frac{c_\nu}{\alpha}\big( [D_\nu^{n}]_+ \big)^\alpha. \end{aligned}$$

□

This discrete formulation of Improved Normal Compliance (INC) based on the specific normal stress of contact $\lambda_\nu^{n+1} = c_\nu \tilde{D}_\nu^{n+1}$ makes it possible to ensure the balance energy under the continuous formalism, see equation (3.31).

Now, we provide the time-dependent discretization of the implicit system of INC.

**Problem 9.** Improved Normal Compliance problem

$$(P_{cd}^{INC}) \quad \begin{cases} \mathbb{M}\dfrac{\dot{\boldsymbol{q}}^{n+1} - \dot{\boldsymbol{q}}^{n}}{\Delta t} = \boldsymbol{F}^{n+1} + \boldsymbol{\Lambda}^{n+1}, \\ \boldsymbol{q}(0) = \boldsymbol{q}_0, \\ \dot{\boldsymbol{q}}(0) = \boldsymbol{v}_0, \end{cases}$$

where $\lambda_\nu^{n+1} = c_\nu [\tilde{D}_\nu^{n+1}]_+$, $c_\nu > 0$, $\alpha \geq 2$ and

$$\tilde{D}_\nu^{n+1} = \frac{([D_\nu^{n+1}]_+)^\alpha - ([D_\nu^n]_+)^\alpha}{\alpha(D_\nu^{n+1} - D_\nu^n)}. \tag{3.32}$$

In the following chapter, we will provide a comprehensive and detailed presentation of this model. This will encompass both the theoretical framework and a series of numerical experiments.

# Chapter 4

# Numerical resolution of the Sweeping processes with impact laws.

In this chapter, we first provide a description of the contact and friction laws within the framework of the NSCD via NLGS and DEM method with penalization. Then, we study the discrete INC in which, a combination of the Newmark method and PDAS is in use in order to solve the nonlinearity issue. Furthermore, it entails conducting numerical simulations to rigorously assess the efficacy and validity of the Newmark-INC-PDAS approach. A comparative analysis is also conducted against alternative numerical methods, such as NSCD-NLGS and DEM, specifically in scenarios devoid of friction. Lastly, we elucidate the numerical treatment of the restitution coefficient and contact conditions involving friction, aiming to offer a robust solution for resolving rigid multi-body contact problems.

## 4.1 Some usual non-smooth methods in Contact Mechanic

Regarding the rigid body model and contact dynamics, which gave rise to the Contact Dynamics method (CD) [38, 43, 45, 46], and later to Non-Smooth Contact Dynamics (NSCD) [48], was originally developed for simulating granular materials and emerged from the mathematical formulation of nonregular dynamic systems.. The resolution of multi-contact problems largely depends on the mechanical behavior including deformable or rigid bodies involved. Within the framework of granular modeling of rigid bodies, we have to consider two factors: one is for all the bodies, namely *global*, to describe the dynamics and solve the equations of motion, and the other one is linked to the contact, namely *local*, to describe the contact interactions of a particle related to the others. The local contact conditions between rigid bodies are ensured using the iterative Nonlinear Gauss-Seidel (NLGS) solver. In addition, alternative solvers to NLGS have been proposed, such as conjugate gradient solvers [49, 50], Semi-Augmented Lagrangian (SAL) method [51, 52], Bipotential method [53, 54], Primal-Dual Active Set (PDAS) method [55, 56, 58]. PDAS methods initially initiated in a deformable

environment, are widely used because of their effectiveness and the simplicity of their implementation because they determine the constraints that will influence the final result of the optimization (see [55, 56, 58, 62]). With these advantages, Barboteu and Dumont [59] developed a PDAS-type method for the local treatment of contact conditions using the NSCD formalism and it is extended in [60] in case of friction conditions with high effectiveness in terms of precision, numerical convergence, and gain of calculation time.

In order to describe the behavior of granular media, it is necessary to consider certain non-restrictive hypotheses about these media and to formulate the appropriate behavioral law for this type of problem. The complex nature of the interactions governing these environments is all the more restrictive as it requires the use of digital simulation. In this sense, simulation strategies from Discrete Element Methods (DEM) (more details in [5–7] and the references therein) make it possible to overcome the difficulties linked to the divided nature of these environments.

Leveraging these characteristics in conjunction with the effectiveness of the PDAS approach, we aim to propose an implicit regularization technique, specially Moreau-Yosida approximation by using Newmark methods which is capable of describing the behavior of granular materials as well as ensuring energy conservation similar to the NSCD-NLGS method while maintaining a reasonable computational cost. It is called Improved Normal Compliance (INC) with Newmark-PDAS-INC approach.

The approach integrates the Newmark implicit scheme with the Semi-smooth Newton method and the Primal-Dual Active Set (PDAS) strategy to reformulate contact and friction conditions into non-linear complementarity functions, which are then solved using the semi-smooth Newton iterative method. This method categorizes potentially contacting nodes into active and inactive subsets, aiming to accurately identify those in active contact (subset $\mathcal{A}$) as opposed to inactive ones (subset $\mathcal{I}$). It eliminates the need for Lagrange multipliers by directly enforcing boundary conditions on subsets $\mathcal{A}$ and $\mathcal{I}$ through the semi-smooth Newton method, making implementation straightforward. Additionally, PDAS methods process each contact condition numerically, solving the laws of contact and friction that describe the dynamics of discrete systems, thereby streamlining and improving the efficiency of these computational processes.

## 4.1.1 Nonsmooth Contact Dynamic (NSCD) via Non-Linear Gauss Seidel (NLGS)

### 4.1.1.1 Time discretization and global-local scheme

Let $T > 0$ be a real number, and fix any integer $N \geq 1$. We consider the time discretization of the interval $[0, T]$ defined by $0 = t^0 < t^1 < \ldots < t^N = T$, i.e., $[0, T] = \bigcup_{n=0}^{N-1} [t^n, t^{n+1}]$ with $t^{n+1} = t^n + \Delta t$ where $\Delta t := \frac{T}{N}$.

The dynamic (3.23) can be reformulated as a time-stepping scheme ([24, 46, 48])

$$\begin{cases} \mathbb{M}\dfrac{\dot{\boldsymbol{q}}^{n+1} - \dot{\boldsymbol{q}}^n}{\Delta t} = \boldsymbol{F}^{n+1} + \boldsymbol{\Lambda}^{n+1}, \\ \boldsymbol{q}^{n+1} = \boldsymbol{q}^n + \Delta t \cdot \dot{\boldsymbol{q}}^{n+1}. \end{cases}$$

Here, $\boldsymbol{\Lambda}^{n+1}$ represents the total impulse at time $t_{n+1}$. We denote the free velocity of the particle as $\dot{\boldsymbol{q}}^{n,free} := \dot{\boldsymbol{q}}^n + \mathbb{M}^{-1}\Delta t \boldsymbol{F}^{n+1}$, which is the velocity without contact impulses.

The motion can then be represented by

$$\dot{\boldsymbol{q}}^{n+1} = \dot{\boldsymbol{q}}^{n,free} + \mathbb{M}^{-1}\boldsymbol{\Lambda}^{n+1}. \tag{4.1}$$

It is well-known that the contact impulses are unknown. Therefore, impulses and velocities must be determined at the same time. A simple transition from the global coordinate system to the local coordinate system at each contact, designated by $\gamma \in [1, N_\gamma]$ allows us to write contact law which is relative to it.

From the global variables $\dot{\boldsymbol{q}}$ and $\boldsymbol{\Lambda}$, it is possible to obtain the local ones at a contact $\gamma$ using the local-global mapping:

$$\boldsymbol{v}^\gamma = H^\star(\boldsymbol{q}, \gamma)\dot{\boldsymbol{q}} \quad \text{and} \quad \boldsymbol{\Lambda} = H(\boldsymbol{q}, \gamma)\boldsymbol{\lambda}^\gamma$$

where $H(\boldsymbol{q}, \gamma)$ is the local-global matrix of size $dN_\gamma \times dN_q$ carrying the information on the geometry of the contact network and $H^\star(\boldsymbol{q}, \gamma)$ is the transpose of matrix $H(\boldsymbol{q}, \gamma)$.

Thereby, setting $\tilde{\boldsymbol{v}}^{n,free} = \mathbb{H}^\star(\boldsymbol{q}, \gamma)\dot{\boldsymbol{q}}^{n,free}$ be the relative velocity without contact, we derive from 4.1 that $\tilde{\boldsymbol{v}}^{n+1,free} = \tilde{\boldsymbol{v}}^{n,free} + \mathbb{H}^\star(\boldsymbol{q}, \gamma)\mathbb{M}^{-1}\mathbb{H}(\boldsymbol{q}, \gamma)\boldsymbol{\lambda}^{n+1}$. Then, the discretization of the motion of a multi-contact system between rigid bodies can be written as follows:

$$\tilde{\boldsymbol{v}}^{n+1,free} = \tilde{\boldsymbol{v}}^{n,free} + \mathbb{W}^{n+1}\boldsymbol{\lambda}^{n+1} \tag{4.2}$$

where $\mathbb{W} := \mathbb{H}^\star(\boldsymbol{q}, \gamma)\mathbb{M}^{-1}\mathbb{H}(\boldsymbol{q}, \gamma)$. The procedure can be shown in the following figure. The local resolution is carried out by NLGS algorithm [38, 46–48]. In general, if the

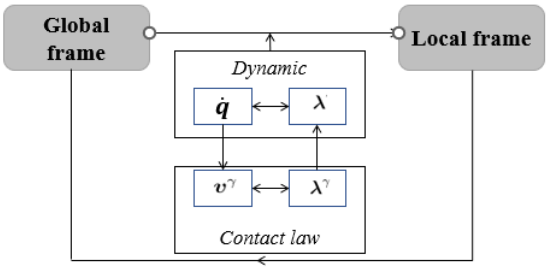



**Figure 4.1:** The global-local scheme of particles at contact points.

contact $\gamma$ is considered, and the pulses at the other contacts $\beta$ are assumed to be fixed, then the iterative scheme is then defined as follows (at the iteration)

$$\begin{cases} \boldsymbol{v}_\gamma^{k+1} = \boldsymbol{v}_\gamma^{free} + \mathbb{W}_{\gamma\gamma}\boldsymbol{\lambda}_\gamma^{k+1} + \sum_{\beta<\alpha} \mathbb{W}_{\gamma\beta}\boldsymbol{\lambda}_\gamma^{k+1} + \sum_{\beta>\gamma} \mathbb{W}_{\gamma\beta}\boldsymbol{\lambda}_\gamma^{k}, \\ \text{Conditions } \big((3.24) - (3.25)\big) \text{ and } \big((3.26) - (3.27)\big) \text{ hold.} \end{cases}$$

The problem consists of finding the couple $(\boldsymbol{v}_\gamma^{n+1}, \boldsymbol{\lambda}_\gamma^{n+1})$. Noticing that for each contact $\gamma$, combine $\mathbb{W}_{\gamma\gamma} = \sum_i \mathbb{H}_{i\gamma}^\star \mathbb{M}_i^{-1} \mathbb{H}_{i\gamma}$ where the sum is made over all the grains $i^{th}$ linked to the contact $\gamma$.

#### 4.1.1.2 Non-Linear Gauss Seidel method (NLGS)

The algorithm used at the global level to solve the multi-rigid body dynamic contact problems can be described as follows. We would use NLGS algorithm, which involves treating each contact successively until convergence is achieved. This method combines the time-stepping approach with the NLGS algorithm to efficiently handle the dynamic contact interactions.

- Loop of discretization in time $n$

  1. Calculate the prediction of position (local - global mapping)

$$\boldsymbol{q}_{n+\frac{1}{2}} = \boldsymbol{q}_n + \frac{\Delta t}{2}\dot{\boldsymbol{q}}_n. \tag{4.3}$$

  2. Choose initial conditions of motion $\dot{\boldsymbol{q}}^0_{n+1} = \dot{\boldsymbol{q}}^{free}_n$ and the contact impulses $\boldsymbol{\Lambda} = 0$.
  3. Loop on $k \geq 0$ (NLGS iterations):
     (a) Computation of the local-global mapping

$$\text{Pre-velocity: } \boldsymbol{v}^- = \mathbb{H}^\star(\boldsymbol{q}_{n+\frac{1}{2}}, \boldsymbol{\lambda})\dot{\boldsymbol{q}}_n, \tag{4.4}$$

$$\text{Pos-velocity at contact point: } \boldsymbol{v}^{\gamma,k^+} = \mathbb{H}^\star(\boldsymbol{q}_{n+\frac{1}{2}}, \boldsymbol{\lambda})\dot{\boldsymbol{q}}^k_{n+1}. \tag{4.5}$$

     (b) Calculating the Moreau velocity

$$v_\nu^{\gamma,k^+} = \frac{v_\nu^{k^+} + e_\nu v_\nu^-}{1+e_\nu}, \tag{4.6}$$

$$\boldsymbol{v}_\tau^{\gamma,k^+} = \frac{\boldsymbol{v}_\tau^{k^+} + e_\tau \boldsymbol{v}_\tau^-}{1+e_\tau}. \tag{4.7}$$

     (c) Check the contact law 3.24 and 3.25.
     (d) Set $\mathbb{W} = \mathbb{H}^\star\mathbb{M}^{-1}\mathbb{H}$ and update the global velocity of motion

$$\dot{\boldsymbol{q}}^{k+1}_{n+1} = \dot{\boldsymbol{q}}^k_{n+1} + \mathbb{W}\boldsymbol{\lambda}^{k+1}. \tag{4.8}$$

  4. End loop on $k$. When the convergence is reached, the velocity is

$$\dot{\boldsymbol{q}}_{n+1} = \dot{\boldsymbol{q}}^{k+1}_{n+1}.$$

  5. Update the actualization of the generalized motion: $\boldsymbol{q}_{n+1} = \boldsymbol{q}_{n+\frac{1}{2}} + \frac{1}{2}\dot{\boldsymbol{q}}_{n+1}$.

- End time-loop $n$.

Alternatively, based on the established contact conditions, the calculation of the local contact step within the NLGS iteration loop at index $j$ can be replaced by the Primal–Dual Active Set Algorithm as described in [60, 61].

### 4.1.2 Discrete Element Method (DEM) with penalisation

The Discrete Element Method (DEM) [5] is a key tool for simulating the behavior of granular materials. Initially introduced by Cundall in 1979 under the name "Distinct Element Method," it has evolved into a powerful technique with an explicit scheme. Its application of DEM across various engineering and scientific fields contributes considerably to a deeper understanding of the mechanics of granular materials.

#### 4.1.2.1 Modeling Particles

The Discrete Element Method (DEM) models each particle as a discrete element with distinct physical properties, such as mass, shape, and stiffness. While particles can be spherical or non-spherical, spherical particles are commonly used in basic models to simplify calculations. DEM has various applications across multiple fields. In geotechnical engineering, it is used to simulate the behavior of soil and granular materials. In mineral processing, DEM models the crushing and fragmentation of rocks. Additionally, in mechanical engineering, it is employed to study friction, wear, and contact phenomena between rough surfaces.

#### 4.1.2.2 The regular contact law

In DEM, the regularization of contact involves simplifying and stabilizing the calculations of contact forces between discrete particles. When particles come into contact, the calculation of these forces can be complex due to the irregular shapes and varying interactions. Cundall introduced several methods to handle these interactions effectively and one of them is the Soft-Sphere model:

- The force is typically modeled using a *Spring-Dashpot* model (more details in [151]), where the spring component represents the elastic behavior, and the dashpot represents the damping (energy dissipation).
  - Normal Forces: Normal forces are perpendicular to the contact surface between two particles. Cundall's regularized contact law proposes that the normal force is proportional to the normal deformation, similar to a spring law (force = stiffness*deformation).
  - Tangential Forces: Tangential forces are parallel to the contact surface. They are modeled using an approach similar to Coulomb's friction law, with the tangential force proportional to the tangential displacement up to a sliding threshold.
  - Deformation and stiffness: The regularized contact law uses normal stiffness $c_\nu$ and tangential stiffness $c_\tau$ to model forces based on normal and tangential deformations.
  - Energy Dissipation: To model energy dissipation during impacts, a damping component is often added to the contact forces. This helps simulate the attenuation of vibrations and oscillations post-impact.
- The *soft contact* approach [7]: particles are allowed to *overlap* slightly at contact points. Then, the contact forces are directly calculated based on this overlap. Contact forces are updated at each time step based on the relative positions, velocities, and deformations of the contacting particles.

More precisely, let's consider the contact between two 2D spherical particles $i$ and $j$. We define the signed distance $D_\nu$ between them by

$$D_\nu = (r^i + r^j) - |\boldsymbol{q}^i - \boldsymbol{q}^j|, \tag{4.9}$$

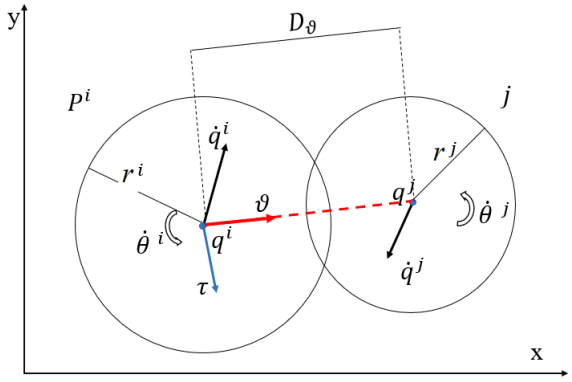



**Figure 4.2:** The regular contact law

where $r^i$ and $r^j$ are the radii of particles $i$ and $j$ respectively, and $\boldsymbol{q}^i, \boldsymbol{q}^j \in \mathbb{R}^3$ are their positions. When the overlap or interpenetration $D_\nu < 0$, two particles are to be in contact and then the normal reaction force at contact is $\lambda_\nu = c_\nu D_\nu$ where $c_\nu$ is the normal stiffness.

The unit vector $\nu$ along the contact line pointing from $i$ to $j$ and the unit vector $\tau$, which is the clockwise rotation of $\nu$, are defined respectively as follows:

$$\nu^{ij} = \frac{\boldsymbol{q}^i - \boldsymbol{q}^j}{|\boldsymbol{q}^i - \boldsymbol{q}^j|} \quad \text{and} \quad \tau^{ij} = (\nu_2^{ij}, -\nu_1^{ij}).$$

Then, the relative velotiy at contact is calculated by

$$\dot{\boldsymbol{q}}^{ij} = (\dot{\boldsymbol{q}}^i - \dot{\boldsymbol{q}}^j) - (\dot{\theta}^i r^i + \dot{\theta}^j r^j) \cdot \tau^{ij}.$$

The normal and tangential components of the contact velocity, $\dot{\boldsymbol{q}}_\nu^{ij}$ and $\dot{\boldsymbol{q}}_\tau^{ij}$ respectively, are

$$\dot{\boldsymbol{q}}_\nu^{ij} = (\dot{\boldsymbol{q}}^i - \dot{\boldsymbol{q}}^j) \cdot \nu^{ij} \nu^{ij} \quad \text{and} \quad \dot{\boldsymbol{q}}_\tau^{ij} = \dot{\boldsymbol{q}}^{ij} - \dot{\boldsymbol{q}}_\nu^{ij} \cdot \nu^{ij} \nu^{ij}.$$

#### 4.1.2.3 Simulation Algorithm

1. Contact Detection: At each time step, particle positions are checked for contacts based on (4.9).
2. Force Calculation: For each pair of contacting particles, normal and tangential forces are calculated using the regularized contact law in 4.1.2.2.
3. Position and Velocity Update: The calculated forces are used to update the particle' accelerations, velocities, and positions using equations of motion:
   - Position Update:
   $$\boldsymbol{q}^{n+1} = \boldsymbol{q}^n + \Delta t\, \dot{\boldsymbol{q}}^n + \frac{\Delta t^2}{2} \ddot{\boldsymbol{q}}^n.$$
   - Velocity Update:
   $$\dot{\boldsymbol{q}}^{n+1} = \dot{\boldsymbol{q}}^n + \Delta t \ddot{\boldsymbol{q}}^n.$$

   where $\boldsymbol{q}^n$ and $\dot{\boldsymbol{q}}^n$ are the position and velocity of particle at time $t^n$, $\ddot{\boldsymbol{q}}^n$ is the acceleration due to the forces acting on the particle, and $\Delta t$ is the time step.

### 4.1.3 Some numerical simulations

In this paragraph, we will present some numerical simulations based on the exactitude of the implementation of the NCSD-NLGS program in Fortran90.

#### NSCD via NLGS

Some simple cases of the frontal collision between 2 particles $P_1$ and $P_2$ are presented (assume that 2 rigid bodies are punctual and without rotation).

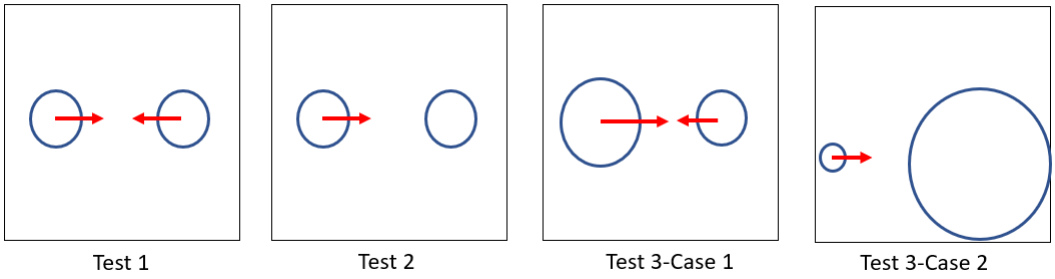



**Figure 4.3:** Physical setting of some simualtions - NSCD-NLGS

• **Test 1: Two balls with same masses and velocities.** This test is for two particles with same masses and velocities (Tab. 4.1 and Fig. 4.4).

| | Value |
|---|---|
| Mass | $m_1 = m_2 = 1$ |
| Radius | $r_1 = r_2 = 0.1$ |
| Initial velocity | $v_1^0 = (1, 0), v_2^0 = (-1, 0)$ |
| Initial position | $q_1^0 = (0.5, 1), q_2^0 = (1.5, 1)$ |
| Time-step | $\Delta t = 10^{-4}$ |
| Total time | $T = 1.5$ |
| Box | $[0; 2] \times [0; 2]$ |

**Tab 4.1:** Initial conditions for collision of two particles with identical masses and velocities.

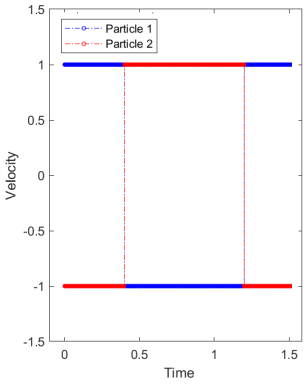



**Figure 4.4:** Velocity of two particles with the same masses and velocities.

After the collision, the motion of the particle 1 is transmitted to the particle 2, and otherwise, i.e. $v_1' = -v_2$ and $v_2' = -v_1$.

• **Test 2: Two balls with same masses and different velocities.** Next, the simulation of two particles with same masses and different velocities is considered (see Fig. 4.5 and Tab. 4.2 ).

After the collision, the motion of the particle 1 is transmitted to the particle 2, and otherwise, i.e. $v_1' = -v_2$ and $v_2' = -v_1$.

• **Test 3: Two balls with different masses and velocities.** We study the case of two particles with different masses and velocities, one is assumed that $m_1 >$

| | Value |
|---|---|
| Mass | $m_1 = m_2 = 1$ |
| Radius | $r_1 = r_2 = 0.1$ |
| Initial velocity | $v_1^0 = (1,0), v_2^0 = (0,0)$ |
| Initial position | $q_1^0 = (1,1), q_2^0 = (1.5,1)$ |
| Time-step | $\Delta t = 10^{-4}$ |
| Total time | $T = 1.5$ (s) |
| Box | $[0;2] \times [0;2]$ |

**Tab 4.2:** Input of collision of two particles with same masses and different velocities.

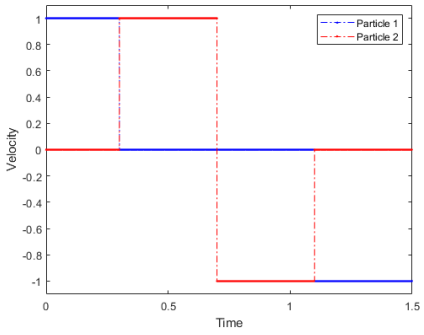
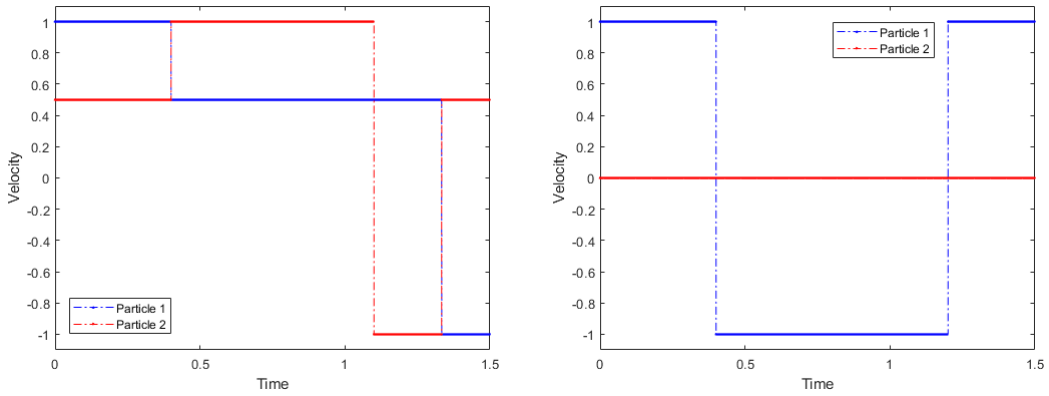



**Figure 4.5:** Simulation the collision of two particles with same masses and different velocities.

$m_2, v_1 > v_2$ (Tab. 4.3-(case 1) and Fig. 4.6-(a)) and another is $m_2 >> m_1$ and $v_2 = 0$ (Tab. 4.3-(case 2) and Fig. 4.6-(b)).

| Parameter | Case 1 | Case 2 |
|---|---|---|
| Mass | $m_1 = m_2 = 1$ | $m = m_2 = 1$ |
| Radius | $r_1 = 0.2, r_2 = 0.1$ | $r_1 = 0.1, r_2 = 1$ |
| Initial velocity | $v_1^0 = (1,0), v_2^0 = (0.5,0)$ | $v_1^0 = (1,0), v_2^0 = (0,0)$ |
| Initial position | $q_1^0 = (0.5,1), q_2^0 = (1,1)$ | $q_1^0 = (0.5,1), q_2^0 = (2,1)$ |
| Time step | $\Delta t = 10^{-4}$ | $\Delta t = 10^{-4}$ |
| Total time | $T = 1.5$ | $T = 1.5$ (s) |
| Box | $[0;2] \times [0;2]$ | $[0;3] \times [0;3]$ |

**Tab 4.3:** Input of two particles with different masses and velocities.




(a) Velocity of 2 particles (Case 1)

(b) Velocity of 2 particles (Case 2)

**Figure 4.6:** Simulation the collision of two particles with different masses and velocities.

After the collision the particle 2 remains where it is and the particle 1 bounces back which is opposite to the initial velocity and has values approximately the initial one.

| | Initial $\mathcal{E}_c$ | $\mathcal{E}_c$ | $\Delta\mathcal{E}_{c\ relative}$ |
|---|---|---|---|
| Test 1 | 1.0000 | 0.9999 | $3.7253{\cdot}10^{-9}$ |
| Test 2 | 0.5000 | 0.4999 | $5.5879{\cdot}10^{-9}$ |
| Test 3 - Case 1 | 0.6250 | 0.6249 | $3.9116{\cdot}10^{-9}$ |
| Test 3 - Case 2 | 0.5000 | 0.4999 | $1.9156{\cdot}10^{-8}$ |

**Tab 4.4:** System's energy - Test 1, 2, 3.

**Conclusion:** With the velocity contact problem combining with NSCD via NLGS, the energy of the system is conserved (Tab. 4.4). This problem coresponds to the *purely elastic impacts*.

• **Test 4: Superposition of 4 particles.** We investigate the case of 4 particles superimposed on each other with the same mass (Tab. 4.5 and Fig. 4.7).

| | **Value** |
|---|---|
| Mass density | $m_i = 1\ (i = 1 \dots 4)$ |
| Radius | $r_i = 0.1$ |
| Initial velocity | $v_i^0 = (0, 0)$ |
| Initial position | $q_i^0 = (0.5, (2i-1)r_i)$ |
| Time step | $\Delta t = 10^{-4}$ (s) |
| Total time | $T = 2$ (s) |
| Gravity | $g = -9.80665$ (s) |
| Domain of movement | $[0; 1] \times [0; 1]$ |

**Tab 4.5:** Input of collision of four particles with same masses.

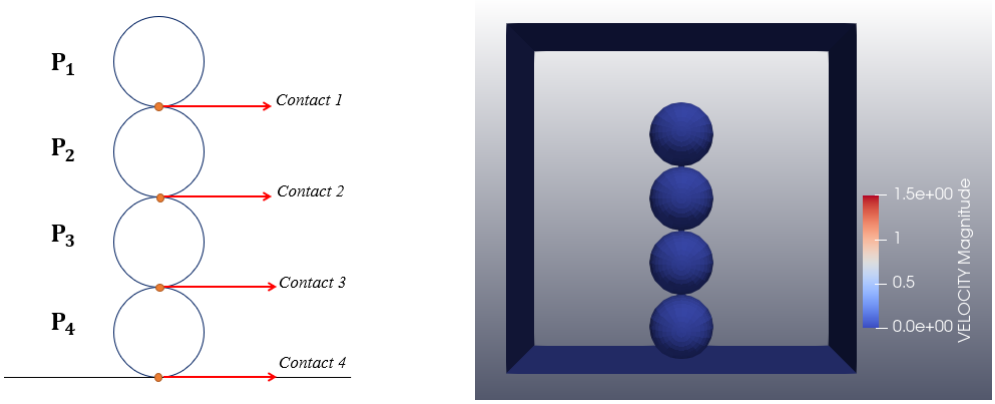



**Figure 4.7:** Simulation the collision of 4 particles - NSCD-NLGS.

Theoretically, it is capable by using the fundamental principle of dynamics to determine the impulses of each contact of the system: $\sum F = m_i g$ with $i = 1 \dots 4$ and $g$ is the gravity force. Then, by caculating we obtain the analysis solution of the equilibrium problem as $\lambda_i = im_i g$ with $i = 1 \dots 4$ and the system reaches the equilibrium state.

Using NSCD via NLGS, the system's energy converges to approximately $5 \cdot 10^{-7}$, closely with the analytical value.

• **Test 5: 81 particles in a box** (Fig. 4.8) . For this method, we also test the case of a system consisting of 81 particles moving in a box in two-dimensional space, with a random initial position (see Fig. 4.8) and which moves under the influence of gravity according to verticality axis (can be shown in Tab. 4.6 and Fig. 4.8).

| | **Value** |
|---|---|
| Mass density | $\rho_i = 2600$ $(i = 1 \dots 81)$ |
| Radius | $r_i = 0.08$ (m) |
| Initial velocity | $v_i^0 = (0, 0)$ (m/s) |
| Initial position | $q_i^0$ arbitrary |
| Time step | $\Delta t = 10^{-4}$ (s) |
| Total time | $T = 2.5$ (s) |
| Gravity | $g = -9.80665$ (s) |
| Domain of movement | $[0; 1] \times [0; 1]$ |

**Tab 4.6:** Input of simulation of 81 particles in a box.

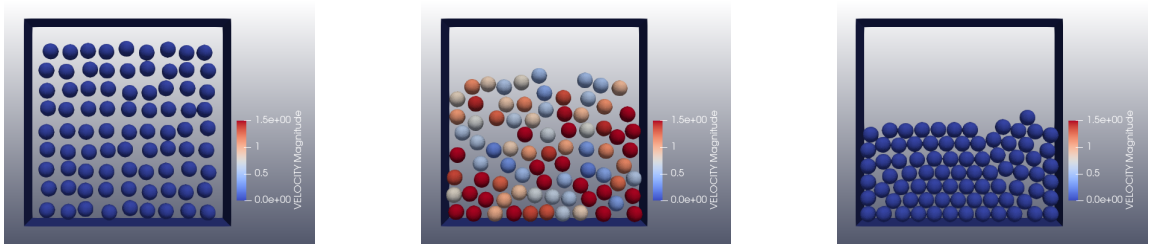


**Figure 4.8:** Simulation of 81 particles in a box

### DEM method with penalization - explicit scheme

• **Test 1: Collision between 2 particles.**

Now, we simulate a simple case of the frontal collision between 2 particles $P_1$ and $P_2$ (see Fig. 4.9) with same masses, velocities (input in Tab. 4.1) with different chosen parameters $R$ and $\Delta t$. Recall that the color gray represents non-convergence of the method.

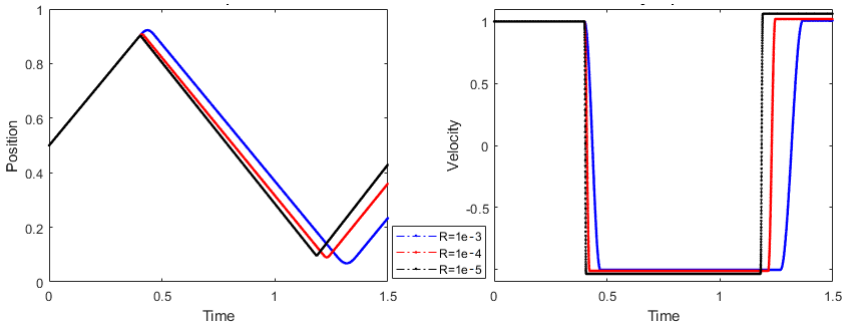
• **Test 2:** Similarly, we simulate the model of 4 particles superimposed on each other with the same mass (Tab. 4.5. The latter figure is about the kinetic energy of the system with $R = 10^{-3}, 10^{-4}, 10^{-5}$.

**Remark 4.1.** Through numerical experiments, there are some notable observations below.

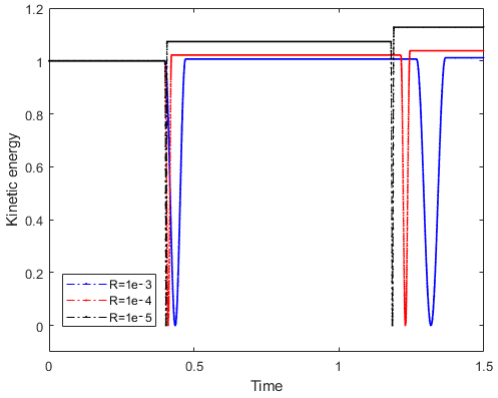
1. It is advisable to use a very small time step in DEM methods ($\Delta t = 10^{-4}, 10^{-5}$) to accurately depict the variations in velocity both prior to and following a collision, see Tab. 4.7. Failure to do so may result in the omission of overlap and compromise the correctness of the solution.

| Time step | $R$ | $v_1 = -v_2$ | $\Delta\mathcal{E}_{c\ relative}$ | Penetration |
|---|---|---|---|---|
| $10^{-3}$ | $10^{-3}$ | 1.0619 | $1.2767{\cdot}10^{-1}$ | $2.2756{\cdot}10^{-1}$ |
| | $10^{-4}$ | | | |
| | $10^{-5}$ | | | |
| $10^{-4}$ | $10^{-3}$ | 1.006 | $1.2068{\cdot}10^{-2}$ | $2.2390{\cdot}10^{-1}$ |
| | $10^{-4}$ | 1.0192 | $3.8667{\cdot}10^{-2}$ | $7.1105{\cdot}10^{-2}$ |
| | $10^{-5}$ | 1.0619 | $1.2767{\cdot}10^{-1}$ | $2.2756{\cdot}10^{-2}$ |
| $10^{-5}$ | $10^{-3}$ | 1.0006 | $1.1999{\cdot}10^{-3}$ | $2.2365{\cdot}10^{-1}$ |
| | $10^{-4}$ | 1.0019 | $3.7998{\cdot}10^{-3}$ | $7.0749{\cdot}10^{-2}$ |
| | $10^{-5}$ | 1.0060 | $1.2069{\cdot}10^{-2}$ | $2.2399{\cdot}10^{-2}$ |

**Tab 4.7:** Collision of two particles with the same masses and velocities - DEM.




(a) Position (left) and velocity (right) of particle 1




(b) Kinetic energy of the system

**Figure 4.9:** Simulation of two particles with same masses and velocities - DEM.

2. Opting for a larger parameter value $R$, (i.e., a smaller coefficient, represented by $c_\nu = 1/R$), enhances the conservation of the system's energy (Tab. 4.7).

3. Kinetic energy conservation is ensured for small values of R due to their greater accuracy in penetration and impulse after contact. Conversely, with larger R values, the system's energy seems amplified, as depicted in Fig. 4.9, 4.10.

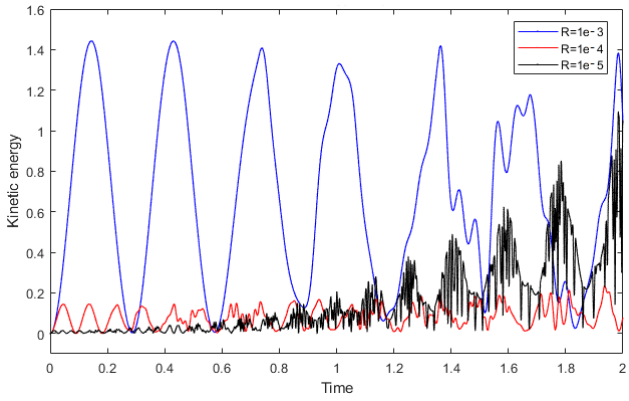



**Figure 4.10:** Simulation of four particles - DEM method.

### Bouncing ball - Comparison between NSCD-NLGS and DEM

We proceed to simulate the bouncing ball, observe Fig. 4.11, with input in Tab. 4.8 by using NSCD-NLGS and DEM method. The kinetic energy of the bouncing ball is represented in figure below with NSCD-NLGS and DEM with different parameter $R$ (or $c_\nu = 1/R$).

| Parameter | Value |
|---|---|
| Mass | $m = 1$ |
| Radius | $r = 0.5$ |
| Initial velocity | $v^0 = (0, 0)$ |
| Initial position | $q^0 = (0, 2)$ |
| Duration of a time step | $\Delta t = 10^{-4}$ (s) |
| Total time | $T = 2.5$ (s) |
| Gravity force | $g = 9.80665$ |

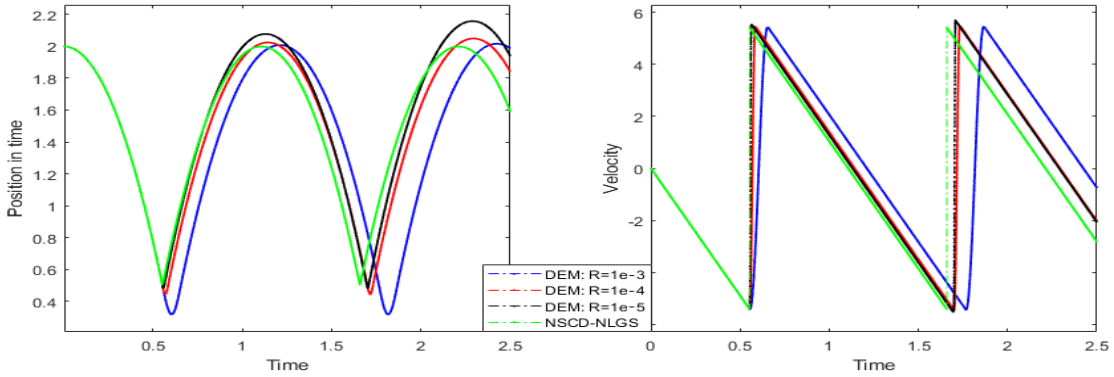
**Tab 4.8:** Bouncing ball with different restitution coefficients.

**Remark 4.2.**

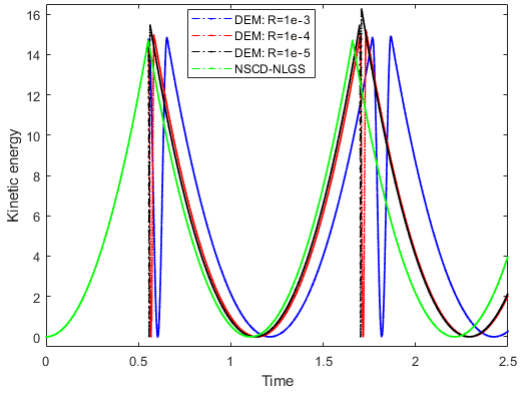
- The preservation of kinetic energy by using NSCD-NLGS is better than DEM with small $R$.
- With DEM, the behavior of individual particles during contact is effectively demonstrated (see Fig. 4.11 - velocity) since DEM models the behavior of discrete particles in the system while models the behavior of the system as a whole, treating the bodies as continuous objects rather than individual particles.

## 4.2 Newmark method

Newmark method is a numerical integration scheme used to solve ordinary differential equations in the context of structural dynamics and time-dependent problems [64]. It is an implicit time integration method that provides stable and accurate solutions




(a) Position (left) and velocity (right) of particle 1




(b) Kinetic energy of the system

**Figure 4.11:** Bouncing ball - Comparison NSCD-NLGS and DEM method.

for both linear and nonlinear systems. The aim of this section is to propose an innovative method based on the combination of Newmark methods, Newton semi-smooth methods and Improved Normal compliance conditions. This integrated methodology aims to guarantee the preservation of the system's kinetic energy.

In this section, our initial focus is to study the time discretization of the implicit scheme applied to the displacement contact problem. Subsequently, we delve into the application of the Semi-smooth Newton method, in the context of handling Standard Normal Compliance (SNC), Improved Normal Compliance (INC), and Compliance conditions relevant to friction. Lastly, we proceed to discuss the application of the primal-dual active set method in the subsequent section.

### 4.2.1 Time descretization

The idea is to use the generalized Newton's method [63] (classic and semi-smooth) to solve both the contact dynamic problems via Normal Compliance condition. The Newmark method introduces two parameters $\beta$ and $\gamma$, which control the numerical

damping and accuracy of the method, its discretization in time are

$$\begin{aligned}\boldsymbol{q}^{n+1} &= \boldsymbol{q}^n + \Delta t\ \dot{\boldsymbol{q}}^n + \frac{\Delta t^2}{2}\left[(1-2\beta)\ \ddot{\boldsymbol{q}}^n + 2\beta\ \ddot{\boldsymbol{q}}^{n+1}\right],\\ \dot{\boldsymbol{q}}^{n+1} &= \dot{\boldsymbol{q}}^n + \Delta t\left[(1-\gamma)\ \ddot{\boldsymbol{q}}^n + \gamma\ \ddot{\boldsymbol{q}}^{n+1}\right].\end{aligned}$$

For $\gamma = 1$ and $\beta = 1/2$, the Newmark method [26, 64] is implicit and unconditionally stable with second order accuracy, denoted midpoint scheme

$$\boldsymbol{q}^{n+1} = \boldsymbol{q}^n + \Delta t\ \dot{\boldsymbol{q}}^n + \frac{\Delta t^2}{2}\ddot{\boldsymbol{q}}^{n+1}, \quad \dot{\boldsymbol{q}}^{n+1} = \dot{\boldsymbol{q}}^n + \Delta t\ddot{\boldsymbol{q}}^{n+1}. \tag{4.10}$$

equivalently,

$$\begin{aligned}\ddot{\boldsymbol{q}}^{n+1} &= \frac{2}{\Delta t^2}(\boldsymbol{q}^{n+1} - \boldsymbol{q}^n) - \frac{2}{\Delta t}\dot{q}^n,\\ \dot{\boldsymbol{q}}^{n+1} &= -\dot{\boldsymbol{q}}^n + \frac{2}{\Delta t}(\boldsymbol{q}^{n+1} - \boldsymbol{q}^n).\end{aligned}$$

The dynamic of Problem 9 can be rewritten in form of

$$\frac{2\mathbb{M}}{\Delta t^2}\boldsymbol{q}^{n+1} - \boldsymbol{\Lambda}(\boldsymbol{q}^{n+1}) = \frac{2\mathbb{M}}{\Delta t}\dot{\boldsymbol{q}}^n + \frac{2\mathbb{M}}{\Delta t^2}\boldsymbol{q}^n.$$

and keep in mind that with SNC: $\boldsymbol{\Lambda} = c_\nu[D_\nu^{n+1}]_+\boldsymbol{\nu}^{n+1}$ or INC: $\boldsymbol{\Lambda} = c_\nu[\tilde{D}_\nu^{n+1}]_+\boldsymbol{\nu}^{n+1}$.

Asusme that $F = 0$, the dynamic is rewriten by

$$\begin{aligned}&\frac{\mathbb{M}}{\Delta t}\left(-\dot{\boldsymbol{q}}^n + \frac{2}{\Delta t}(\boldsymbol{q}^{n+1} - \boldsymbol{q}^n) - \dot{\boldsymbol{q}}^n\right) = \boldsymbol{\Lambda}(\boldsymbol{q}^{n+1})\\ \Leftrightarrow\ &2\frac{\mathbb{M}}{\Delta t}\left(-\dot{\boldsymbol{q}}^n + \frac{1}{\Delta t}(\boldsymbol{q}^{n+1} - \boldsymbol{q}^n)\right) = \boldsymbol{\Lambda}(\boldsymbol{q}^{n+1})\\ \Leftrightarrow\ &\frac{2\mathbb{M}}{\Delta t^2}\boldsymbol{q}^{n+1} - \boldsymbol{\Lambda}(\boldsymbol{q}^{n+1}) = \frac{2\mathbb{M}}{\Delta t}\dot{\boldsymbol{q}}^n + \frac{2\mathbb{M}}{\Delta t^2}\boldsymbol{q}^n.\end{aligned}$$

Then, we deduce the nonlinear system with time-discretization form

$$\mathcal{R}(\mathbf{q}^{n+1}) \ = \ \frac{1}{\Delta t}\mathbb{M}(\dot{\boldsymbol{q}}^{n+1} - \dot{\boldsymbol{q}}^n) \ - \ \boldsymbol{F}^{n+1} - \boldsymbol{\Lambda}(\boldsymbol{q}^{n+1}) = \mathbf{0}, \tag{4.11}$$

Using the midpoint scheme (4.10), this method leads to the solution of linearized systems

$$\begin{aligned}&\mathcal{K}^{k,n+1}\Delta\boldsymbol{q}^{k,n+1} = -\frac{1}{\Delta t}\boldsymbol{M}(\dot{\mathbf{q}}^{k,n+1} - \dot{\boldsymbol{q}}^n) - \mathbf{F}(\boldsymbol{q}^{n+1}) - \boldsymbol{\Lambda}(\boldsymbol{q}^{k,n+1}) \qquad (4.12)\\ &\text{with} \quad \mathcal{K}^{k,n+1} = \frac{2}{\Delta t^2}\mathbf{M} - \mathbf{K}_c^{k,n+1} \text{ and } \quad \Delta\boldsymbol{q}^{k,n+1} = \mathbf{q}^{k+1,n+1} - \boldsymbol{q}^{k,n+1}\end{aligned}$$

where $\mathbf{M} = \partial_{\boldsymbol{q}^{n+1}}\mathbb{M}$ represents the mass matrix and $\mathbf{K}_c^{k,n+1} = \partial_{\boldsymbol{q}^{n+1}}\boldsymbol{\Lambda}(\boldsymbol{q}^{k,n+1})$ which denotes the frictional contact tangent matrix.

The nonlinear system (4.11) can be solved by a generalized Newton method. This method leads to the following iterative scheme (indexed by $k$)

$$\boldsymbol{q}^{k+1,n+1} = \boldsymbol{q}^{k,n+1} - \left(\partial_{\mathbf{q}^{n+1}}\mathcal{R}(\boldsymbol{q}^{k,n+1})\right)^{-1}\mathcal{R}(\boldsymbol{q}^{k,n+1}).$$

### 4.2.2 Semi-smooth Newton method for the normal contact conditions

In this section, we aim to provide a comprehensive overview of the Semi-smooth Newton method to Normal Compliance conditions. We begin by considering the fundamental principles, including complementary function, its generalized derivation, and fixed point condition, firstly of SNC and then focus on INC. Lastly, we navigate the intricacies of compliance conditions within the context of friction.

#### 4.2.2.1 Complemetary function

**4.2.2.1.1 Standard Compliance Normal** The standard normal compliance contact conditions can be formulated from the following non-linear complementary function $\mathcal{C}_\nu^\lambda(D_\nu^{n+1}, \lambda_\nu^{n+1})$

$$\mathcal{C}_\nu^\lambda(D_\nu^{n+1}, \lambda_\nu^{n+1}) = \lambda_\nu^{n+1} - [c_\nu D_\nu^{n+1}]_+. \tag{4.13}$$

For the sake of simplicity, in the following we use this equation:

$$\mathcal{C}_\nu^\lambda(D_\nu, \lambda_\nu) = \lambda_\nu - [c_\nu D_\nu]_+.$$

**4.2.2.1.2 Improved Compliance Normal** The improved normal compliance contact conditions can be formulated from the following non-linear complementary function $\mathcal{C}_\nu^\lambda(D_\nu^{n+1}, \lambda_\nu^{n+1})$

$$\mathcal{C}_\nu^\lambda(D_\nu^{n+1}, \lambda_\nu^{n+1}) = \lambda_\nu^{n+1} - [c_\nu \tilde{D}_\nu^{n+1}]_+, \tag{4.14}$$

For the sake of simplicity, in the following we use this equation:

$$\mathcal{C}_\nu^\lambda(D_\nu, \lambda_\nu) = \lambda_\nu - [c_\nu \tilde{D}_\nu^{n+1}]_+.$$

**4.2.2.1.3 Compliance for friction condition** In addition to the complementary conditions of persistence and non-interpenetrability due to the non-regular impact motion between two particles, the Coulomb friction law is added to consider the behaviors of contacts. The compliance for friction conditions (3.26, 3.27) can be formulated from the following non-linear complementary function $\mathcal{C}_\tau^\lambda(D_\nu^{n+1}, \dot{\boldsymbol{q}}_\tau^{n+1}, \lambda_\nu^{n+1}, \boldsymbol{\lambda}_\tau^{n+1})$ with

$$\mathcal{C}_\tau^\lambda(D_\nu^{n+1}, \dot{\boldsymbol{q}}_\tau^{n+1}, \lambda_\nu^{n+1}, \boldsymbol{\lambda}_\tau^{n+1}) = \max(\mu\lambda_\nu^{n+1}, \|c_\tau \dot{\boldsymbol{q}}_\tau^{n+1}\|)\boldsymbol{\lambda}_\tau^{n+1} - \mu\lambda_\nu^{n+1}(c_\tau \dot{\boldsymbol{q}}_\tau^{n+1}). \tag{4.15}$$

For the sake of simplicity, in the following we use this equation:

$$\mathcal{C}_\tau^{\boldsymbol{\lambda}}(D_\nu, \dot{\boldsymbol{q}}_\tau, \lambda_\nu, \boldsymbol{\lambda}_\tau) = \max(\mu\lambda_\nu, \|c_\tau \dot{\boldsymbol{q}}_\tau\|)\boldsymbol{\lambda}_\tau - \mu\lambda_\nu(c_\tau \dot{\boldsymbol{q}}_\tau).$$

#### 4.2.2.2 Generalized derivative of complementary functions

**4.2.2.2.1 Standard Compliance Normal** We provide the directional generalized derivative of the complementary functions in the gap and contact cases.

- Gap case : $D_\nu \leq 0$

Given the complementary function $\mathcal{C}_\nu^\lambda(D_\nu, \lambda_\nu) = \lambda_\nu$, we have the following derivative

$$d_{D_\nu}\mathcal{C}_\nu^\lambda = 0, \tag{4.16}$$
$$d_{\lambda_\nu}\mathcal{C}_\nu^\lambda = d\lambda_\nu. \tag{4.17}$$

- Contact case : $D_\nu > 0$

Given the complementary functions $\mathcal{C}_\nu^\lambda(D_\nu, \lambda_\nu) = \lambda_\nu - c_\nu D_\nu$, we have

$$d_{D_\nu}\mathcal{C}_\nu^\lambda = -c_\nu dD_\nu, \tag{4.18}$$
$$d_{\lambda_\nu}\mathcal{C}_\nu^\lambda = d\lambda_\nu. \tag{4.19}$$

By combining (4.16)–(4.17), with $\mathcal{D}_{\mathcal{C}_\nu^\lambda}$ the generalized derivative of $\mathcal{C}_\nu^\lambda$, we obtain

$$\mathcal{D}_{\mathcal{C}_\nu^\lambda}(D_\nu, \lambda_\nu)(\delta D_\nu, \delta\lambda_\nu) = -c_\nu(1_{Contact})\delta D_\nu + \delta\lambda_\nu, \tag{4.20}$$

where

$$1_{Gap} = 1, 1_{Contact} = 0 \text{ if } D_\nu \leq 0,$$
$$1_{Gap} = 0, 1_{Contact} = 1 \text{ if } D_\nu > 0.$$

**4.2.2.2.2 Improved Compliance Normal** Now, we provide the generalized derivative of the complementary functions in the gap and contact cases.

- Gap case : $D_\nu^{n+1} \leq 0$

Given the complementary function $\mathcal{C}_\nu^\lambda(D_\nu, \lambda_\nu) = \lambda_\nu$ we have the following derivative

$$d_{D_\nu}\mathcal{C}_\nu^\lambda = 0, \tag{4.21}$$
$$d_{\lambda_\nu}\mathcal{C}_\nu^\lambda = d\lambda_\nu. \tag{4.22}$$

- Contact case : $D_\nu^{n+1} > 0$

Given the complementary functions $\mathcal{C}_\nu^\lambda(D_\nu, \lambda_\nu) = \lambda_\nu - c_\nu\tilde{D}_\nu$, we have

$$d_{D_\nu}\mathcal{C}_\nu^\lambda = -c_\nu\frac{([D_\nu^{n+1}]_+)^{\alpha-1} - \tilde{D}_\nu^{n+1}}{D_\nu^{n+1} - D_\nu^n}dD_\nu, \tag{4.23}$$
$$d_{\lambda_\nu}\mathcal{C}_\nu^\lambda = d\lambda_\nu. \tag{4.24}$$

By combining (4.21)–(4.22), with $\mathcal{D}_{\mathcal{C}_\nu^\lambda}$ the generalized derivative of $\mathcal{C}_\nu^\lambda$, we obtain

$$\mathcal{D}_{\mathcal{C}_\nu^\lambda}(D_\nu, \lambda_\nu)(\delta D_\nu, \delta\lambda_\nu) = -c_\nu\frac{([D_\nu^{n+1}]_+)^{\alpha-1} - \tilde{D}_\nu^{n+1}}{D_\nu^{n+1} - D_\nu^n}(1_{Contact})\delta D_\nu + \delta\lambda_\nu, \tag{4.25}$$

where

$$1_{Gap} = 1, 1_{Contact} = 0 \text{ if } \tilde{D}_\nu^{n+1} \leq 0,$$
$$1_{Gap} = 0, 1_{Contact} = 1 \text{ if } \tilde{D}_\nu^{n+1} > 0.$$

**4.2.2.2.3 Compliance for friction condition** Now, we provide the generalized derivative of the complementary function in the gap and friction cases.

• Gap case: $\boldsymbol{\lambda}_\tau = \underline{0}$

• Stick case: $\|\boldsymbol{\lambda}_\tau\| < \mu\lambda_\nu$

$$\mathcal{C}_\tau^{\boldsymbol{\lambda}}(D_\nu, \dot{\boldsymbol{q}}_\tau, \lambda_\nu, \boldsymbol{\lambda}_\tau) = \mu\lambda_\nu\boldsymbol{\lambda}_\tau - \mu c_\tau\lambda_\nu\dot{\boldsymbol{q}}_\tau.$$

Then

$$d_{D_\nu}\mathcal{C}_\nu^{\boldsymbol{\lambda}} = 0, \tag{4.26}$$
$$d_{\dot{\boldsymbol{q}}_\tau}\mathcal{C}_\tau^{\boldsymbol{\lambda}} = -\mu c_\tau\lambda_\nu d\dot{\boldsymbol{q}}_\tau, \tag{4.27}$$
$$d_{\lambda_\nu}\mathcal{C}_\tau^{\boldsymbol{\lambda}} = (\mu\boldsymbol{\lambda}_\tau - \mu c_\tau\dot{\boldsymbol{q}}_\tau)d\lambda_\nu, \tag{4.28}$$
$$d_{\boldsymbol{\lambda}_\tau}\mathcal{C}_\tau^{\boldsymbol{\lambda}} = \mu\lambda_\nu d\boldsymbol{\lambda}_\tau. \tag{4.29}$$

• Slip case: $\|\boldsymbol{\lambda}_\tau\| \geq \mu\lambda_\nu$

$$\mathcal{C}_\tau^{\boldsymbol{\lambda}}(D_\nu, \dot{\boldsymbol{q}}_\tau, \lambda_\nu, \boldsymbol{\lambda}_\tau) = \|c_\tau\dot{\boldsymbol{q}}_\tau\|\boldsymbol{\lambda}_\tau - \mu\lambda_\nu(c_\tau\dot{\boldsymbol{q}}_\tau).$$

Then

$$d_{D_\nu}\mathcal{C}_\tau^{\boldsymbol{\lambda}} = 0, \tag{4.30}$$
$$d_{\dot{\boldsymbol{q}}_\tau}\mathcal{C}_\tau^{\boldsymbol{\lambda}} = \Big(c_\tau\boldsymbol{\lambda}_\tau\frac{(\dot{\boldsymbol{q}}_\tau)^T}{\|\dot{\boldsymbol{q}}_\tau\|} - \mu c_\tau\lambda_\nu\boldsymbol{I}_2\Big)d\dot{\boldsymbol{q}}_\tau, \tag{4.31}$$
$$d_{\lambda_\nu}\mathcal{C}_\tau^{\boldsymbol{\lambda}} = -\mu c_\tau\dot{\boldsymbol{q}}_\tau d\lambda_\nu, \tag{4.32}$$
$$d_{\boldsymbol{\lambda}_\tau}\mathcal{C}_\tau^{\boldsymbol{\lambda}} = \|c_\tau\dot{\boldsymbol{q}}_\tau\|d\boldsymbol{\lambda}_\tau \tag{4.33}$$

where $\boldsymbol{I}_2 = \boldsymbol{\nu}\boldsymbol{\nu}^T + \boldsymbol{\tau}\boldsymbol{\tau}^T$.
By combining (4.26)–(4.33), with $\mathcal{G}_{\mathcal{C}_\tau^{\boldsymbol{\lambda}}}$ the generalized derivative of $\mathcal{C}_\tau^{\boldsymbol{\lambda}}$, respectively, we obtain

$$\begin{aligned}\mathcal{G}_{\mathcal{C}_\tau^{\boldsymbol{\lambda}}}(D_\nu, \dot{\boldsymbol{q}}_\tau, \lambda_\nu, \boldsymbol{\lambda}_\tau)(\delta D_\nu, \delta\dot{\boldsymbol{q}}_\tau, \delta\lambda_\nu, \delta\boldsymbol{\lambda}_\tau) &= 1_{Stick}\Big(\mu\boldsymbol{\lambda}_\tau\Big)\delta\lambda_\nu \\ &+ 1_{Slip}\Big(c_\tau\boldsymbol{\lambda}_\tau\frac{(\dot{\boldsymbol{q}}_\tau)^T}{\|\dot{\boldsymbol{q}}_\tau\|}\Big)\delta\dot{\boldsymbol{q}}_\tau - \mu c_\tau\dot{\boldsymbol{q}}_\tau\delta\lambda_\nu - \mu c_\tau\lambda_\nu\delta\dot{\boldsymbol{q}}_\tau \\ &+ 1_{Stick}\Big(\mu\lambda_\nu\Big)\delta\boldsymbol{\lambda}_\tau + 1_{Slip}\Big(\|c_\tau\dot{\boldsymbol{q}}_\tau\|\Big)\delta\boldsymbol{\lambda}_\tau\end{aligned}$$

where

$$\begin{aligned}1_{Gap} = 0, 1_{Stick} = 1, 1_{Slip} = 0 \text{ if } \|\boldsymbol{\lambda}_\tau\| < \mu\lambda_\nu, \\ 1_{Gap} = 0, 1_{Stick} = 0, 1_{Slip} = 1 \text{ if } \|\boldsymbol{\lambda}_\tau\| \geq \mu\lambda_\nu.\end{aligned}$$

#### 4.2.2.3 Fixed point conditions from Newton's Semi-Smooth approach

**4.2.2.3.1 Standard Compliance Normal** Using now the semi-smooth Newton formalism (indexed by the subscript $k$) at the current fixed point $(D_\nu^k, \lambda_\nu^k)$ of the complementary functions $\mathcal{C}_\nu^\lambda$, one can derive the new iterate $(D_\nu^{k+1}, \lambda_\nu^{k+1})$

$$\mathcal{D}_{\mathcal{C}_\nu^\lambda}(D_\nu^k, \lambda_\nu^k)(\delta D_\nu^{k+1}, \delta\lambda_\nu^{k+1}) = -\mathcal{C}_\nu^\lambda(D_\nu^k, \lambda_\nu^k), \tag{4.34}$$

$$(D_\nu^{k+1}, \lambda_\nu^{k+1}) = (D_\nu^k, \lambda_\nu^k) + (\delta D_\nu^{k+1}, \delta\lambda_\nu^{k+1})$$

where $\delta D_\nu^{k+1} = D_\nu^{k+1} - D_\nu^k$.

- Gap case: $1_{Gap} = 1, 1_{Contact} = 0$

From the equations (4.34) we have

$$\lambda_\nu^{k+1} - \lambda_\nu^k = -\lambda_\nu^k.$$

Next, the gap conditions by using the semi-smooth Newton approach are

$$\lambda_\nu^{k+1} = 0.$$

- Contact case : $1_{Gap} = 0, 1_{Contact} = 1$

From the equations (4.34) we have

$$-c_\nu(D_\nu^{k+1} - D_\nu^k) + (\lambda_\nu^{k+1} - \lambda_\nu^k) = -\lambda_\nu^k + c_\nu D_\nu^k.$$

Next,

$$\lambda_\nu^{k+1} = c_\nu D_\nu^{k+1}.$$

**4.2.2.3.2 Improved Compliance Normal** Using now the semi-smooth Newton formalism (indexed by the subscript $k$) at the current fixed point $(D_\nu^k, \lambda_\nu^k)$ of the complementary functions $\mathcal{C}_\nu^\lambda$, one can derive the new iterate $(D_\nu^{k+1}, \lambda_\nu^{k+1})$

$$\mathcal{D}_{\mathcal{C}_\nu^\lambda}(D_\nu^k, \lambda_\nu^k)(\delta D_\nu^{k+1}, \delta\lambda_\nu^{k+1}) = -\mathcal{C}_\nu^\lambda(D_\nu^k, \lambda_\nu^k), \tag{4.35}$$

$$(D_\nu^{k+1}, \lambda_\nu^{k+1}) = (D_\nu^k, \lambda_\nu^k) + (\delta D_\nu^{k+1}, \delta\lambda_\nu^{k+1}).$$

- Gap case: $1_{Gap} = 1, 1_{Contact} = 0$

From the equations (4.35) we have

$$\lambda_\nu^{k+1} - \lambda_\nu^k = -\lambda_\nu^k.$$

Next, the gap conditions by using the Semi-smooth Newton approach are

$$\lambda_\nu^{k+1} = 0.$$

- Contact case : $1_{Gap} = 0, 1_{Contact} = 1$

From the equations (4.35) we have

$$-c_\nu \frac{([D_\nu^{k,n+1}]_+)^{\alpha-1} - \tilde{D}_\nu^{k,n+1}}{D_\nu^{k,n+1} - D_\nu^n}(D_\nu^{k+1,n+1} - D_\nu^{k,n+1}) + (\lambda_\nu^{k+1} - \lambda_\nu^k) = -\lambda_\nu^k + c_\nu \tilde{D}_\nu^{k,n+1}.$$

Next,

$$\lambda_\nu^{k+1} = c_\nu \left[ \tilde{D}_\nu^{k,n+1} + \frac{([D_\nu^{k,n+1}]_+)^{\alpha-1} - \tilde{D}_\nu^{k,n+1}}{D_\nu^{k,n+1} - D_\nu^n}(D_\nu^{k+1,n+1} - D_\nu^{k,n+1}) \right].$$

**4.2.2.3.3 Compliance for friction condition** Using now the semi-smooth Newton formalism at the current $(D_\nu^k, \dot{\boldsymbol{q}}_\tau^k, \lambda_\nu^k, \boldsymbol{\lambda}_\tau^k)$, one can derive the new iterate $(D_\nu^{k+1}, \dot{\boldsymbol{q}}_\tau^{k+1}, \lambda_\nu^{k+1}, \boldsymbol{\lambda}_\tau^{k+1})$

$$\mathcal{G}_{\mathcal{C}_\tau^{\boldsymbol{\lambda}}}(D_\nu^k, \dot{\boldsymbol{q}}_\tau^k, \lambda_\nu^k, \boldsymbol{\lambda}_\tau^k)(\delta D_\nu^{k+1}, \delta\dot{\boldsymbol{q}}_\tau^{k+1}, \delta\lambda_\nu^{k+1}, \delta\boldsymbol{\lambda}_\tau^{k+1}) = -\mathcal{C}_\tau^{\boldsymbol{\lambda}}(D_\nu^k, \dot{\boldsymbol{q}}_\tau^k, \lambda_\nu^k, \boldsymbol{\lambda}_\tau^k),$$

$$(D_\nu^{k+1}, \dot{\boldsymbol{q}}_\tau^{k+1}, \lambda_\nu^{k+1}, \boldsymbol{\lambda}_\tau^{k+1}) = (D_\nu^k, \dot{\boldsymbol{q}}_\tau^k, \lambda_\nu^k, \boldsymbol{\lambda}_\tau^k) + (\delta D_\nu^{k+1}, \delta\dot{\boldsymbol{q}}_\tau^{k+1}, \delta\lambda_\nu^{k+1}, \delta\boldsymbol{\lambda}_\tau^{k+1}).$$

Stick case: $1_{Stick} = 1, 1_{Slip} = 0$

We have

$$\begin{aligned}&-\mu c_\tau \lambda_\nu^k(\dot{\boldsymbol{q}}_\tau^{k+1} - \dot{\boldsymbol{q}}_\tau^k) + \mu\lambda_\nu^k(\boldsymbol{\lambda}_\tau^{k+1} - \boldsymbol{\lambda}_\tau^k) + (\mu\boldsymbol{\lambda}_\tau^k - \mu c_\tau\dot{\boldsymbol{q}}_\tau^k)(\lambda_\nu^{k+1} - \lambda_\nu^k)\\ &= -\mu\lambda_\nu^k\boldsymbol{\lambda}_\tau^k + \mu c_\tau\lambda_\nu^k\dot{\boldsymbol{q}}_\tau^k.\end{aligned}$$

Next, supposing $\boldsymbol{\lambda}_\tau^k = c_\tau\dot{\boldsymbol{q}}_\tau^k$, we obtain

$$\boldsymbol{\lambda}_\tau^{k+1} = c_\tau\dot{\boldsymbol{q}}_\tau^{k+1}.$$

Slip case: $1_{Stick} = 0, 1_{Slip} = 1$

We obtain

$$\begin{aligned}&\Big(c_\tau\boldsymbol{\lambda}_\tau^k\frac{(\dot{\boldsymbol{q}}_\tau^k)^T}{\|\dot{\boldsymbol{q}}_\tau^k\|} - \mu c_\tau\lambda_\nu^k\boldsymbol{I}_2\Big)(\dot{\boldsymbol{q}}_\tau^{k+1} - \dot{\boldsymbol{q}}_\tau^k)\\ &- \mu(c_\tau\dot{\boldsymbol{q}}_\tau^k)(\lambda_\nu^{k+1} - \lambda_\nu^k) + \Big(\|c_\tau\dot{\boldsymbol{q}}_\tau^k\|\Big)(\boldsymbol{\lambda}_\tau^{k+1} - \boldsymbol{\lambda}_\tau^k)\\ &= -\|c_\tau\dot{\boldsymbol{q}}_\tau^k\|\boldsymbol{\lambda}_\tau^k + \mu\lambda_\nu^k(c_\tau\dot{\boldsymbol{q}}_\tau^k).\end{aligned}$$

Therefore, after an elementary computation, supposing $\boldsymbol{\lambda}_\tau^k = \mu\lambda_\nu^k\frac{\dot{\boldsymbol{q}}_\tau^k}{\|\dot{\boldsymbol{q}}_\tau^k\|}$ we have

$$\boldsymbol{\lambda}_\tau^{k+1} = \mu\lambda_\nu^{k+1}\frac{\dot{\boldsymbol{q}}_\tau^k}{\|\dot{\boldsymbol{q}}_\tau^k\|} + \Big(\boldsymbol{\lambda}_\tau^k\frac{(\dot{\boldsymbol{q}}_\tau^k)^T}{\|\dot{\boldsymbol{q}}_\tau^k\|^2} - \mu\lambda_\nu^k\frac{\boldsymbol{I}_2}{\|\dot{\boldsymbol{q}}_\tau^k\|}\Big)(\dot{\boldsymbol{q}}_\tau^{k+1} - \dot{\boldsymbol{q}}_\tau^k).$$

In a two-dimensional scenario, a simplified version of the preceding condition can be derived, keeping in mind that $\boldsymbol{I}_2 = \boldsymbol{\nu}\boldsymbol{\nu}^T + \boldsymbol{\tau}\boldsymbol{\tau}^T$

$$\boldsymbol{\lambda}_\tau^{k+1} = \mu\lambda_\nu^{k+1}\frac{\dot{\boldsymbol{q}}_\tau^k}{\|\dot{\boldsymbol{q}}_\tau^k\|}.$$

### 4.2.3 Primal-Dual Active Set method via Newton Semi-Smooth approach

This part focuses on the numerical handling of contact conditions through the Primal-Dual Active Set method within the context of dynamic contact problems. First, we identify the active and inactive subsets of all nodes currently in contact. Then, we compute the contact conditions for each subset solely based on contact reactions or impulses, utilizing the local general equations of motion.

#### 4.2.3.1 Standard Normal Compliance

We denote by $\mathcal{S}$ the set of potential contact and $\gamma$ a potential contact node (where $\gamma \in [1, N_\gamma]$) belonging to $\mathcal{S}$. The active and inactive sets are defined by

$$\mathcal{A}_\nu^k = \{\gamma \in \mathcal{S} : D_\nu^{k,n+1} \geq 0\},$$
$$\mathcal{I}_\nu^k = \{\gamma \in \mathcal{S} : D_\nu^{k,n+1} < 0\}.$$

The contact tangent matrix $\mathbf{K}_c^{k,n+1}$ is

$$\mathbf{K}_c^{k,n+1} = \mathbb{H}_c^{k,n+1}\mathbb{K}_c{\mathbb{H}_c^{k,n+1}}^T$$
$$\text{with} \quad \mathbb{K}_c = \text{diag}(\mathbb{K}_{SNC},\ \gamma = 1, n_c) \quad \text{and} \quad \mathbb{K}_{SNC} = \begin{bmatrix} c_\nu & 0 \\ 0 & c_\tau = 0 \end{bmatrix}$$

where $n_c$ is the number of contact ; $n_p$ is the number of particules and $nd$ is the space dimension. The matrix $\mathbb{H}_c \in \mathbb{R}^{nd \times n_c \times n_p}$ is a matrix assembly of $\mathbb{H}_p^\gamma$ and $\mathbb{H}_q^\gamma$ ($\gamma$ is contact between particle $p$ and $q$) defined by

$$\text{If } \gamma \in \mathcal{I}_\nu^k \text{ then } \mathbb{H}_p^\gamma = \mathbf{0} \text{ and } \mathbb{H}_q^\gamma = \mathbf{0};$$
$$\text{If } \gamma \in \mathcal{A}_\nu^k \text{ then } \mathbb{H}_p^\gamma = \begin{bmatrix} \boldsymbol{\nu} & \boldsymbol{\tau} \\ 0 & 1 \end{bmatrix} \text{ and } \mathbb{H}_q^\gamma = \begin{bmatrix} -\boldsymbol{\nu} & -\boldsymbol{\tau} \\ 0 & 1 \end{bmatrix}.$$

The general form of the iterative active set algorithm of the index $k$ can be represented as follow.

(i) Choose $(\boldsymbol{q}^{(0)}, \boldsymbol{\lambda}^{(0)})$, $c_\nu > 0$, $c_\tau > 0$, very small $\epsilon > 0$ and set $k = 0$.
(ii) Compute: $\tau_\nu^\gamma = D_\nu^{\gamma,k}$ for each $\gamma \in \mathcal{S}$.
(iii) Set

$$\mathcal{A}_\nu^{k+1} = \{\gamma \in \mathcal{S} : \tau_\nu^\gamma \geq 0\},$$
$$\mathcal{I}_\nu^{k+1} = \mathcal{S} \setminus \mathcal{A}_\nu^{k+1}.$$

(iv) Find $(\boldsymbol{q}^{\gamma,k+1}, \boldsymbol{\lambda}^{\gamma,k+1})$ such that

$$\lambda_\nu^{\gamma,k+1} = 0, \qquad \text{for all } \gamma \in \mathcal{I}_\nu^{k+1},$$
$$\lambda_\nu^{\gamma,k+1} = c_\nu D_\nu^{\gamma,k+1} \qquad \text{for all } \gamma \in \mathcal{A}_\nu^{k+1},$$

(v) If $\|(\boldsymbol{q}^{\gamma,k+1}, \boldsymbol{\lambda}^{\gamma,k+1}) - (\boldsymbol{q}^{\gamma,k}, \boldsymbol{\lambda}^{\gamma,k})\| \leq \epsilon$, $\mathcal{A}_\nu^{k+1} = \mathcal{A}_\nu^k$ stop, else goto (ii).

#### 4.2.3.2 Improved Normal Compliance

The active and inactive sets are defined by

$$\mathcal{A}_\nu^k = \{\gamma \in \mathcal{S} : \tilde{D}_\nu^{k,n+1} \geq 0\},$$
$$\mathcal{I}_\nu^k = \{\gamma \in \mathcal{S} : \tilde{D}_\nu^{k,n+1} < 0\}.$$

The contact tangent matrix $\mathbf{K}_c^{k,n+1}$ is

$$\mathbf{K}_c^{k,n+1} = \mathbb{H}_c^{k,n+1}\mathbb{K}_c{\mathbb{H}_c^{k,n+1}}^T$$

$$\text{with } \mathbb{K}_c \ = \text{diag}(\mathbb{K}_{INC},\ \gamma = 1, n_c),\ \mathbb{K}_{INC} = \begin{bmatrix} c_\nu \dfrac{([D_\nu^{k,n+1}]_+)^{\alpha-1} - \tilde{D}_\nu^{k,n+1}}{D_\nu^{k,n+1} - D_\nu^{k,n}} & 0 \\ 0 & c_\tau = 0 \end{bmatrix}$$

where $n_c$ is the number of contact ; $n_p$ is the number of particules and $nd$ is the space dimension. The matrix $\mathbb{H}_c \in \mathbb{R}^{nd \times n_c \times n_p}$ is a matrix assembly of $\mathbb{H}_p^\gamma$ and $\mathbb{H}_q^\gamma$ ($\gamma$ is contact between particle $p$ and $q$) defined by

$$\text{If } \gamma \in \mathcal{I}_\nu^k, \text{ then } \mathbb{H}_p^\gamma = \mathbf{0} \text{ and } \mathbb{H}_q^\gamma = \mathbf{0};$$
$$\text{If } \gamma \in \mathcal{A}_\nu^k, \text{ then } \mathbb{H}_p^\gamma = \begin{bmatrix} \boldsymbol{\nu} & \boldsymbol{\tau} \\ 0 & 1 \end{bmatrix} \text{ and } \mathbb{H}_q^\gamma = \begin{bmatrix} -\boldsymbol{\nu} & -\boldsymbol{\tau} \\ 0 & 1 \end{bmatrix}.$$

The general form of the iterative active set algorithm of the index $k$ can be represented as follow.

(i) Choose $(\boldsymbol{q}^{(0)}, \boldsymbol{\lambda}^{(0)})$, $c_\nu > 0$, very small $\epsilon > 0$ and set $k = 0$.
(ii) Compute: $\tau_\nu^\gamma = \tilde{D}_\nu^{\gamma,k,n+1}$ for each $\gamma \in \mathcal{S}$.
(iii) Set the active and inactive sets:

$$\mathcal{A}_\nu^{k+1} = \{\gamma \in \mathcal{S} : \tau_\nu^\gamma \geq 0\},$$
$$\mathcal{I}_\nu^{k+1} = \mathcal{S} \setminus \mathcal{A}_\nu^{k+1}.$$

(iv) Find $(\boldsymbol{q}^{\gamma,k+1}, \boldsymbol{\lambda}^{\gamma,k+1})$ such that

$$\lambda_\nu^{\gamma,k+1} = 0, \quad \text{for all } \gamma \in \mathcal{I}_\nu^{k+1},$$
$$\lambda_\nu^{\gamma,k+1} = c_\nu \tilde{D}_\nu^{\gamma,k,n+1} + c_\nu \frac{([D_\nu^{\gamma,k,n+1}]_+)^{\alpha-1} - \tilde{D}_\nu^{\gamma,k,n+1}}{D_\nu^{\gamma,k,n+1} - D_\nu^{\gamma,k,n}} (D_\nu^{\gamma,k+1,n+1} - D_\nu^{\gamma,k,n+1})$$
$$\text{for all } \gamma \in \mathcal{A}_\nu^{k+1},$$

(v) If $\|(\boldsymbol{q}^{\gamma,k+1}, \boldsymbol{\lambda}^{\gamma,k+1}) - (\boldsymbol{q}^{\gamma,k}, \boldsymbol{\lambda}^{\gamma,k})\| \leq \epsilon$, $\mathcal{A}_\nu^{k+1} = \mathcal{A}_\nu^k$ stop, else goto (ii).

#### 4.2.3.3 Compliance for friction condition

The improved contact conditions are achieved by implementing an active set strategy, which directly results from computing the fixed point on the non-linear complementary functions $\mathcal{C}_\nu^\lambda$ and $\mathcal{C}_\tau^\lambda$ using the Newton semi-smooth approach. The active and inactive sets are defined as follows

$$\mathcal{A}_\nu^{k+1} = \{\gamma \in \mathcal{S} : \tilde{D}_\nu^{\gamma,k,n+1} \geq 0\},$$
$$\mathcal{I}_\nu^{k+1} = \{\gamma \in \mathcal{S} : \tilde{D}_\nu^{\gamma,k,n+1} < 0\},$$
$$\mathcal{A}_\tau^{k+1} = \{\gamma \in \mathcal{S} : \|\boldsymbol{\lambda}_\tau^{\gamma,k}\| < \mu\lambda_\nu^{\gamma,k}\},$$
$$\mathcal{I}_\tau^{k+1} = \{\gamma \in \mathcal{S} : \|\boldsymbol{\lambda}_\tau^{\gamma,k}\| \geq \mu\lambda_\nu^{\gamma,k}\}.$$

The status of a given potential contact $\gamma$ at the non-linear iteration $k$ depends on the set it belongs to. It can be either in the non-contact or frictional contact status.

(i) Choose $(\boldsymbol{q}^{(0)}, \boldsymbol{\lambda}^{(0)})$, $c_\nu > 0$, $c_\tau > 0$, very small $\epsilon > 0$ and set $k = 0$.
(ii) Compute: $\tau_\nu^\gamma = \tilde{D}_\nu^{\gamma,k,n+1}$ and $\tau_\tau^\gamma = -\|\boldsymbol{\lambda}_\tau^{\gamma,k}\| + \mu\lambda_\nu^{\gamma,k}$ for each $\gamma \in \mathcal{S}$.
(iii) Set the active and inactive sets:

$$\begin{aligned}
\mathcal{A}_\nu^{k+1} &= \{\gamma \in \mathcal{S} : \tau_\nu^\gamma \geq 0\},\\
\mathcal{I}_\nu^{k+1} &= \mathcal{S} \setminus \mathcal{A}_\nu^{k+1},\\
\mathcal{A}_\tau^{k+1} &= \{\gamma \in \mathcal{S} : \tau_\tau^\gamma > 0\},\\
\mathcal{I}_\tau^{k+1} &= \mathcal{S} \setminus \mathcal{A}_\tau^{k+1}.
\end{aligned}$$

(iv) Find $(\boldsymbol{q}^{\gamma,k+1}, \boldsymbol{\lambda}^{\gamma,k+1})$ such that

$$\lambda_\nu^{\gamma,k+1} = 0, \qquad \boldsymbol{\lambda}_\tau^{\gamma,k+1} = \boldsymbol{0} \qquad \text{for all } \gamma \in \mathcal{I}_n^{k+1},$$

$$\lambda_\nu^{\gamma,k+1} = c_\nu \tilde{D}_\nu^{\gamma,k,n+1} + c_\nu \frac{([D_\nu^{\gamma,k,n+1}]_+)^{\alpha-1} - \tilde{D}_\nu^{\gamma,k,n+1}}{D_\nu^{\gamma,k,n+1} - D_\nu^{\gamma,k,n}} (D_\nu^{\gamma,k+1,n+1} - D_\nu^{\gamma,k,n+1})$$
$$\text{for all } \gamma \in \mathcal{A}_\nu^{k+1},$$

$$\boldsymbol{\lambda}_\tau^{\gamma,k+1} = c_\tau \dot{\boldsymbol{q}}_\tau^{\gamma,k+1} \qquad \text{for all } \gamma \in \mathcal{A}_\tau^{k+1} \cap \mathcal{A}_\nu^{k+1},$$

$$\boldsymbol{\lambda}_\tau^{\gamma,k+1} = \mu\lambda_\nu^{\gamma,k+1} \frac{\dot{\boldsymbol{q}}_\tau^{\gamma,k}}{\|\dot{\boldsymbol{q}}_\tau^{\gamma,k}\|} + \Big(\boldsymbol{\lambda}_\tau^{\gamma,k} \frac{(\dot{\boldsymbol{q}}_\tau^{\gamma,k})^T}{\|\dot{\boldsymbol{q}}_\tau^{\gamma,k}\|^2} - \mu\lambda_\nu^{\gamma,k} \frac{\boldsymbol{I}_2}{\|\dot{\boldsymbol{q}}_\tau^{\gamma,k}\|}\Big)(\dot{\boldsymbol{q}}_\tau^{\gamma,k+1} - \dot{\boldsymbol{q}}_\tau^{\gamma,k})$$
$$\text{for all} \quad \gamma \in \mathcal{I}_\tau^{k+1} \cap \mathcal{A}_\nu^{k+1}.$$

(v) If $\|(\boldsymbol{q}^{\gamma,k+1}, \boldsymbol{\lambda}^{\gamma,k+1}) - (\boldsymbol{q}^{\gamma,k}, \boldsymbol{\lambda}^{\gamma,k})\| \leq \epsilon$, $\mathcal{A}_\nu^{k+1} = \mathcal{A}_\nu^k$ and $\mathcal{A}_\tau^{k+1} = \mathcal{A}_\tau^k$ stop, else goto (ii).

We also need to note that it is essential to underline that one of the specific characteristics of the PDAS methods is to directly and strictly impose the conditions of contact calculated by the iterative process of the semi-smooth Newton approach. From the algorithm, we can see that only one Active set iteration is performed for each contact at each global iteration of the NLGS loop, unlike other methods.

### 4.2.4 Linear solver

At each iteration, we formulate a complete linear system, which is typically non-symmetric and must be solved efficiently. To achieve this, we utilize optimized solvers to minimize the computational cost.

We now focus on solving the system

$$Ax = b, \tag{4.36}$$

where $A \in \mathbb{R}^{n\times n}$ is an invertible (though not necessarily symmetric) matrix, and $b \in \mathbb{R}^n$ is the corresponding right-hand side vector. Various solver techniques are employed to enhance the computational efficiency in solving this system.

#### 4.2.4.1 LAPACK and Sparse Matrix Strategies

The LAPACK (Linear Algebra PACKage) library consolidates most linear algebra algorithms. It encompasses all the functions for solving linear systems. LAPACK operations maximize the use of Basic Linear Algebra Subprograms (BLAS), especially employing a block-partitioning approach for LU factorization to leverage BLAS3 operations. For our original code, we use `Dgesv` subroutine which is a function in the LAPACK for solving linear algebra problems (4.36), particularly for dense systems.

Problems involving contact with friction can lead to poorly conditioned linear systems, where small changes in input parameters can result in large variations in the solution. In the presence of friction and contact problems, the system matrices can be non-symmetric. This leads to the computational cost being large.

We have noted that linear systems characterized by sparsity often arise from hollow configurations, where a significant portion of the matrix elements are zero. An effective approach involves reformulating the tangential matrix into a sparse matrix format and employing specialized solvers for the solution of (4.36) to enhance the computational efficiency of CPU time of the system.

#### 4.2.4.2 Assembly sparse - Preconditioned Conjugate Gradient

In many numerical simulations, the discretization of the problem leads to handling a very large matrix, most of whose coefficients are zero. In this case, we consider a sparse matrix, which considerably reduces the complexity of system resolutions by not storing a significant portion of the zero coefficients.

The main idea of the CSR (Compressed Sparse Row) sparse storage is to consider only the non-vanishing terms of the matrix. Then, two additional tables are necessary to point to the row name of each element and the first element of each line (see Gibbs et al. [82] for example).

When matrix A is symmetric positive definite (frictionless $\mu = 0.0$), to handle the system (4.36), we can apply a preconditioner in the `Pcg` subroutine or `Jpcg` of the SPARSKIT library for symmetric matrices, follow these steps:

- Convert the matrix to CSR format: Transform your dense symmetric matrix to the CSR (Compressed Sparse Row) format.
- Choose and configure a preconditioner: Select and set up a preconditioner, such as a Jacobi or ILU (Incomplete LU) preconditioner.
- Use the `Pcg` subroutine with the preconditioner or Jacobi `Jpcg` subroutine: Pass the CSR matrix, preconditioner, initial guess vector, solution vector, and other necessary parameters to the `Pcg`/`Jpcg` subroutine.

#### 4.2.4.3 Direct solvers (LU and supernodal LU for sparse matrices)

Another version, we use SPARSKIT created by Youcef Saad [150] which is a pivotal software library in computational linear algebra. It specializes in efficient numerical computations involving sparse matrices, offering optimized routines for operations like matrix-vector multiplication, iterative solvers, and eigenvalue computations to enhance

the efficiency of computations involving large-scale sparse matrices. More precisely, in SPARSKIT, DNSCSR is a crucial routine in computation that converts matrices from a Dense format to a Compressed Row Sparse (CSR). While CSRDNS converts a CSR matrix back to a dense matrix, DNSCSR performs the inverse operation, converting a densely stored matrix into a row-oriented compactly sparse matrix. The primary purpose of this conversion is to facilitate efficient computational operations. It is essential to note that DNSCSR does not perform a check to determine if an element in the matrix is small. Instead, it operates on the assumption that matrix A(i,j) is considered zero if and only if it is exactly equal to zero. This approach simplifies the conversion process and ensures accuracy in the transformation without relying on threshold checks for small values.

When matrix A is non-symmetric, especially in case of friction ($0 < \mu \leq 1$), it is necessary to apply a permutation of columns or rows to ensure the numerical stability of the LU factorization. Generally, this strategy is implemented during the calculation. The pivot selection criterion is based on both the minimal degree algorithm, to minimize fill-in, and the partial pivot, for numerical stability. In the `Superlu` code, the strategy is applied a priori, before initiating the computations, enabling to analysis factorization similar to Cholesky. In SPARSKIT, `ILUT` is an Incomplete LU factorization (ILU) precondition. This routine carries out incomplete LU factorization with a dual truncation mechanism. Sorting is done for both L and U. Then, `LUSOL0` performs a forward followed by a backward solve for the LU matrix as produced by `ILUT` (more details in [150]).

#### 4.2.4.4 Iterative solvers: GMRES with ILU Preconditioner

In the case of a non-symmetric matrix $A$, the Krylov method associated with this scenario is the GMRES solver (Generalized Minimal Residual). It is characterized by

$$\begin{aligned} \|r_n\|_2 &= \min_{x \in x_0 + \mathcal{K}_n(A, r_0)} \|b - Ax\|_2 \\ \text{or} \qquad & x_n \in x_0 + \mathcal{K}_n(A, r_0), \\ & r_n \perp A\mathcal{K}_n(A, r_0). \end{aligned}$$

The convergence is monotonic, and the solution is attained within a maximum of n iterations (more details in [149] and references therein).

In general, Krylov methods converge too slowly and the system must be preconditioned. The idea is to find a matrix C such that CA is better conditioned than A and their product is not too expensive. Iterative methods are often applied to sparse matrices, a large part of which coefficients are zero. One of the drawback of direct methods applied to sparse matrices lies in the storage costs incurred during Gauss or Cholesky factorization. One idea to limit the filling is to use incomplete factorizations (ILU). Consequently, the matrix A can be expressed as $A = LU + R$, with the preconditioning defined by $C = U^{-1}L^{-1}$. There are various strategies to limit padding, but incomplete factorization may not exist. However, it exists for any strategy if A is an M-matrix [148]. The simplest case is the ILU(0) method where no filling is allowed.

Another solver is GMRES with ILU Preconditioner, namely `pgmres`, leveraging the synergy of ILUT and GMRES. The ILUT preconditioner uses a dual strategy for

dropping elements instead of the usual level of-fill-in approach. PGMRES uses the L and U matrices generated from the subroutine `Ilut` to precondition the GMRES algorithm.

## 4.3 Numerical simulation without friction

In this section, we will present several simulations based on the mechanical setting outlined above, specifically implementing the Newmark-INC-PDAS method, using programming language Fortran 90. The corresponding code for this method is provided in Appendix A.

### 4.3.1 Some remarks concerning the INC method

Let us consider the INC: $\Lambda = \dfrac{\alpha}{2r}[\tilde{d}_\nu^{n+1}]_+\nu^{n+1}$ $(\alpha > 2)$ with

$$\tilde{d}_\nu^{n+1} = (-1)^\alpha \frac{([D_\nu^{n+1}]_+)^\alpha - ([D_\nu^n]_+)^\alpha}{\alpha(D_\nu^{n+1} - D_\nu^n)}.$$

To simplify, we set:

$$c_\nu = \frac{1}{2r} \text{ and } \tilde{D}_\nu^{n+1} = (-1)^\alpha \frac{([D_\nu^{n+1}]_+)^\alpha - ([D_\nu^n]_+)^\alpha}{D_\nu^{n+1} - D_\nu^n}.$$

Thus, the improved normal compliance becomes

$$\boxed{\Lambda = c_\nu[\tilde{D}_\nu^{n+1}]_+\nu^{n+1}.}$$

Then the derivative of $\tilde{D}_\nu^{n+1}$ with respect to $D_\nu^{n+1}$, namely $\nabla\tilde{D}_\nu^{n+1}$, reads

$$\nabla\tilde{D}_\nu^{n+1} = \frac{(-1)^\alpha\alpha\big([D_\nu^{k,n+1}]_+\big)^{\alpha-1} - \tilde{D}_\nu^{n+1}}{D_\nu^{k,n+1} - D_\nu^n}. \tag{4.37}$$

**Remark 4.3.** We have some noteworthy comments as follows.

- The mathematical analysis (exact) of tildegap can be presented as

$$\begin{aligned}
\tilde{D}_\nu^{n+1} &= (-1)^\alpha \frac{([D_\nu^{n+1}]_+)^\alpha - ([D_\nu^n]_+)^\alpha}{D_\nu^{n+1} - D_\nu^n} \\
&= (-1)^\alpha \frac{(D_\nu^{n+1} - D_\nu^n)\big((D_\nu^{n+1})^{\alpha-1} + (D_\nu^{n+1})^{\alpha-2}(D_\nu^n) + ... + (D_\nu^{n+1})(D_\nu^n)^{\alpha-2} + (D_\nu^n)^{\alpha-1}\big)}{D_\nu^{n+1} - D_\nu^n} \\
&= (-1)^\alpha(D_\nu^{n+1})^{\alpha-1} + (D_\nu^{n+1})^{\alpha-2}(D_\nu^n) + ... + (D_\nu^{n+1})(D_\nu^n)^{\alpha-2} + (D_\nu^n)^{\alpha-1} \\
&= (-1)^\alpha \sum_{i=0}^{\alpha-1}(D_\nu^{n+1})^{\alpha-2-i}(D_\nu^n)^i.
\end{aligned}$$

- If $D_\nu^{k,n+1} - D_\nu^n \to 0$, means that $D_\nu^{k,n+1} = D_\nu^n + \varepsilon$ where $\varepsilon \to 0$, it leads to (passing the limit)

$$\tilde{D}_\nu^{n+1} = (-1)^\alpha \alpha \big(D_\nu^{k,n+1}\big)^{\alpha-1}$$

and then its derivative

$$\nabla \tilde{D}_\nu^{n+1} = (-1)^\alpha \alpha(\alpha-1)\big(D_\nu^{k,n+1}\big)^{\alpha-2}$$

where $\varepsilon \to 0$.

- If $D_\nu^{k,n+1} \neq D_\nu^n$, we calculate at by (4.37).

### 4.3.2 Test cases

We provide some simulations in 2D as follow.

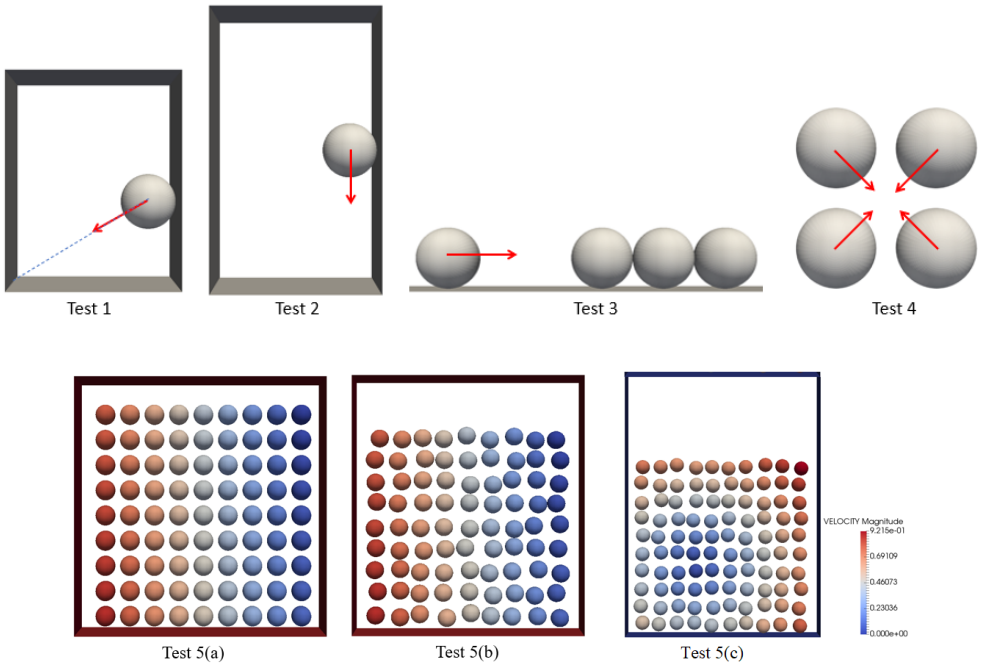



**Figure 4.12:** Physical setting of some simualtions in 2D.

1. A ball moving with velocity $v = (-0.8, -0.4)$ from position $(1.0, 0.6)$ in 3s.
2. A ball falling along the vector $y = -1.0$ from $(1.0, 1.0)$ in 3s.
3. Simulation of four particles as a Newton model in $1,5$s.
4. Simulation of four particles that occurs the contact simultaneously in $1,1$s.
5. Several particles in a box.
   (a) 81 particles in a box with fixed velocities and positions;
   (b) 81 particles in a box with random velocities and positions;

(c) 100 particles in a box.

We use the notation INC$^2$ (or INC$^3$, INC$^4$) instead of INC with $\alpha = 2$ (or $\alpha = 3, 4$). Also, remember that in the simulations, R corresponds to $1/c_\nu$.

#### 4.3.2.1 Test 1: A moving ball

First, we consider the movement of one particle in 2D with different time-stepping and different penalization parameters (see Tab. 4.9 and Fig. 4.13).

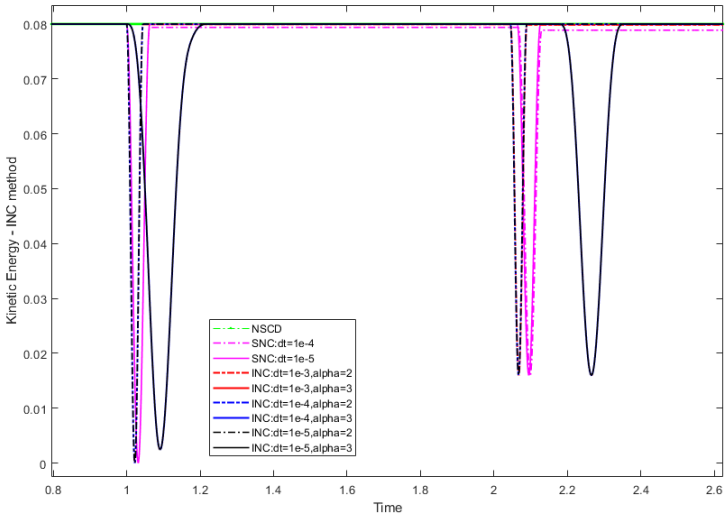



**Figure 4.13:** Kinetic energy of the system - Simulation of one particle in 2D with different time steps.

**Remark 4.4.** Some considerations on time step selection.

(i) We can observe the kinetic energy of the system in Fig. 4.13. With the same penalisation parameters $R$, the result is better when $\Delta t$ is smaller and $\alpha$ is larger.

(ii) NSCD-NLGS method gives up a good resolution even with a large duration of movement $\Delta t$, whereas with DEM, it is necessary to select a smaller time step to ensure accurate collision timing.

(iii) With SNC, we should choose a very small time step (Tab. 4.9). If the time step is large, the overlap will be skipped and the solution is no longer correct.

(iv) With INC, we can choose a time step bigger than SNCs (from $10^{-3}$) and it also guarantee the time to collide between particles. In addition, the smaller the time interval, the more accurate the solution value (velocity and kinetic energy).

#### 4.3.2.2 Test 2: Falling ball

We study this example to verify the case of one particle moving along one boundary all the time and check the *contact function* in our code.

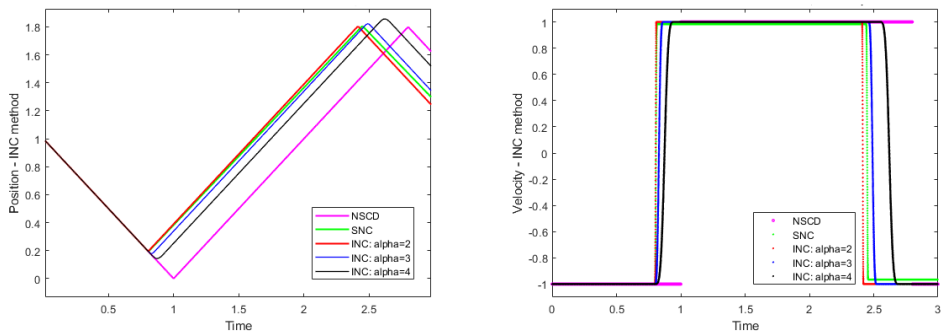
**Remark 4.5.** Some remarkable comments.

| $\Delta t$ | $R$ | $\Delta\mathcal{E}_{c\ relative}$ | | | | Relative penetration | | | |
|---|---|---|---|---|---|---|---|---|---|
| $10^{-3}$ | | NSCD: $2.15{\cdot}10^{-7}$ | | | | 0.000 | | | |
| | | DEM | SNC | INC $^2$ | INC $^3$ | DEM | SNC | INC $^2$ | INC $^3$ |
| | $10^{-3}$ | | | $1.820{\cdot}10^{-3}$ | $1.407{\cdot}10^{-6}$ | | | $2.828{\cdot}10^{-2}$ | $1.259{\cdot}10^{-1}$ |
| | $10^{-4}$ | | | $1.499{\cdot}10^{-2}$ | $5.968{\cdot}10^{-7}$ | | | $8.943{\cdot}10^{-3}$ | $5.848{\cdot}10^{-2}$ |
| | $10^{-5}$ | | | | $6.430{\cdot}10^{-5}$ | | | | $2.714{\cdot}10^{-1}$ |
| | $10^{-6}$ | | | | $1.715{\cdot}10^{-3}$ | | | | $1.260{\cdot}10^{-2}$ |
| | $10^{-7}$ | | | | $9.260{\cdot}10^{-3}$ | | | | $5.834{\cdot}10^{-3}$ |
| $10^{-4}$ | | NSCD: $1.79{\cdot}10^{-8}$ | | | | 0.000 | | | |
| | | DEM | SNC | INC $^2$ | INC $^3$ | DEM | SNC | INC $^2$ | INC $^3$ |
| | $10^{-3}$ | $2.021{\cdot}10^{-2}$ | $1.977{\cdot}10^{-2}$ | $2.730{\cdot}10^{-5}$ | $1.399{\cdot}10^{-9}$ | $2.821{\cdot}10^{-2}$ | $2.852{\cdot}10^{-2}$ | $2.828{\cdot}10^{-2}$ | $1.259{\cdot}10^{-1}$ |
| | $10^{-4}$ | $9.613{\cdot}10^{-2}$ | $6.097{\cdot}10^{-2}$ | $3.347{\cdot}10^{-4}$ | $1.175{\cdot}10^{-10}$ | $9.346{\cdot}10^{-3}$ | $8.866{\cdot}10^{-3}$ | $8.944{\cdot}10^{-3}$ | $5.848{\cdot}10^{-2}$ |
| | $10^{-5}$ | | | $5.274{\cdot}10^{-3}$ | $1.136{\cdot}10^{-7}$ | | | $2.828{\cdot}10^{-3}$ | $2.714{\cdot}10^{-2}$ |
| | $10^{-6}$ | | | $4.657{\cdot}10^{-3}$ | $1.409{\cdot}10^{-6}$ | | | $8.943{\cdot}10^{-4}$ | $1.259{\cdot}10^{-2}$ |
| | $10^{-7}$ | | | | $6.261{\cdot}10^{-7}$ | | | | $5.848{\cdot}10^{-3}$ |
| $10^{-5}$ | | NSCD: $2.61{\cdot}10^{-9}$ | | | | 0.000 | | | |
| | | DEM | SNC | INC $^2$ | INC $^3$ | DEM | SNC | INC $^2$ | INC $^3$ |
| | $10^{-3}$ | $2.001{\cdot}10^{-3}$ | $1.997{\cdot}10^{-3}$ | $6.773{\cdot}10^{-7}$ | $2.410{\cdot}10^{-12}$ | $2.831{\cdot}10^{-2}$ | $2.828{\cdot}10^{-2}$ | $2.828{\cdot}10^{-2}$ | $1.259{\cdot}10^{-1}$ |
| | $10^{-4}$ | $6.343{\cdot}10^{-3}$ | $6.301{\cdot}10^{-3}$ | $5.435{\cdot}10^{-6}$ | $7.153{\cdot}10^{-13}$ | $8.968{\cdot}10^{-3}$ | $8.937{\cdot}10^{-3}$ | $8.944{\cdot}10^{-3}$ | $5.848{\cdot}10^{-2}$ |
| | $10^{-5}$ | $2.935{\cdot}10^{-2}$ | $1.978{\cdot}10^{-2}$ | $4.037{\cdot}10^{-5}$ | $3.422{\cdot}10^{-11}$ | $2.868{\cdot}10^{-3}$ | $2.821{\cdot}10^{-3}$ | $2.828{\cdot}10^{-3}$ | $2.714{\cdot}10^{-2}$ |
| | $10^{-6}$ | $9.613{\cdot}10^{-2}$ | $6.103{\cdot}10^{-2}$ | $2.293{\cdot}10^{-4}$ | $1.409{\cdot}10^{-9}$ | $9.346{\cdot}10^{-4}$ | $8.866{\cdot}10^{-4}$ | $8.944{\cdot}10^{-4}$ | $1.259{\cdot}10^{-2}$ |
| | $10^{-7}$ | | | $1.967{\cdot}10^{-3}$ | $3.438{\cdot}10^{-10}$ | | | $2.828{\cdot}10^{-4}$ | $5.848{\cdot}10^{-3}$ |

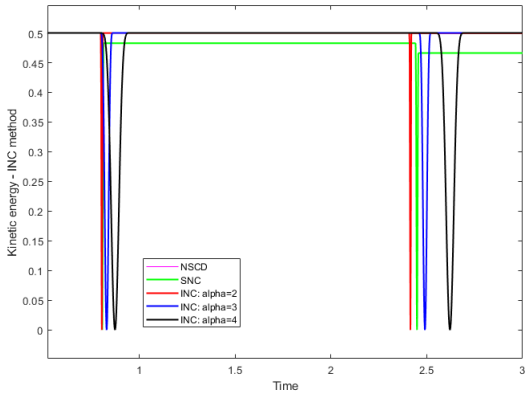
**Tab 4.9:** Simulation one particle in 2D with different time steps $\Delta t$ and some different $R$.

| Penalisation | Method | $\Delta\mathcal{E}_{c\ relative}$ | Penetration |
|---|---|---|---|
| | NSCD | $7.45{\cdot}10^{-9}$ | $\approx 0.00$ |
| $R = 10^{-4}$ | DEM | $3.19{\cdot}10^{-2}$ | $2.53{\cdot}10^{-2}$ |
| | SNC | $3.09{\cdot}10^{-2}$ | $2.49{\cdot}10^{-2}$ |
| | INC $^{2}$ | $2.96{\cdot}10^{-5}$ | $2.50{\cdot}10^{-2}$ |
| | INC $^{3}$ | $4.16{\cdot}10^{-9}$ | $1.16{\cdot}10^{-1}$ |
| | INC $^{4}$ | $1.81{\cdot}10^{-12}$ | $2.49{\cdot}10^{-1}$ |
| $R = 10^{-5}$ | DEM | $1.05{\cdot}10^{-1}$ | $8.21{\cdot}10^{-3}$ |
| | SNC | $9.44{\cdot}10^{-2}$ | $7.81{\cdot}10^{-3}$ |
| | INC $^{2}$ | $8.67{\cdot}10^{-4}$ | $7.90{\cdot}10^{-3}$ |
| | INC $^{3}$ | $1.79{\cdot}10^{-8}$ | $5.38{\cdot}10^{-2}$ |
| | INC $^{4}$ | $1.07{\cdot}10^{-11}$ | $1.40{\cdot}10^{-1}$ |
| $R = 10^{-6}$ | DEM | | |
| | SNC | | |
| | INC $^{2}$ | $5.68{\cdot}10^{-3}$ | $2.49{\cdot}10^{-3}$ |
| | INC $^{3}$ | $9.11{\cdot}10^{-7}$ | $2.49{\cdot}10^{-2}$ |
| | INC $^{4}$ | $1.09{\cdot}10^{-11}$ | $7.90{\cdot}10^{-2}$ |

**Tab 4.10:** Simulation of a falling ball with $v_y = -1.0$ and $\Delta t = 10^{-4}$.




(a) Position (left) and velocity (right)




(b) The system's energy

**Figure 4.14:** Simulation of a falling ball with $v_y = -1.0$, $c_\nu = 10^5, \Delta t = 10^{-4}$.

(i) With a very small $R$ penalty, both the DEM and SNC methods may exhibit suboptimal performance. In such cases, it becomes necessary to finetune value of $R$ in relation to $\Delta t$ in order to get a good energy conservation.

(ii) NSCD-NLGS method gives us a good resolution including velocity and conservation of system's energy (Fig. 4.14). However, with this method, it is lack of the behavior of pre- and pos-velocity of the movement and INC shows better the characterization of velocity and almost conversed energy of the system.

(iii) With INC, it ensures mostly conservation of Kinetic energy of the system when $\alpha$ becomes larger (see Fig. 4.14 - Kinetic energy).

#### 4.3.2.3 Test 3: Newton's cradle of four particles

Next, let us consider the Newton model of particles including 4 particles (see Fig. 4.15). The analysis solution after the shock is $v_1 = v_2 = v_3 = 0.0$ and $v_4 = 1.0$ with the total energy of system is $\mathcal{E}_c = 1.0$. The resolution by using NSCD-NLGS, DEM, SNC and INC method is shown in Tab. 4.11.

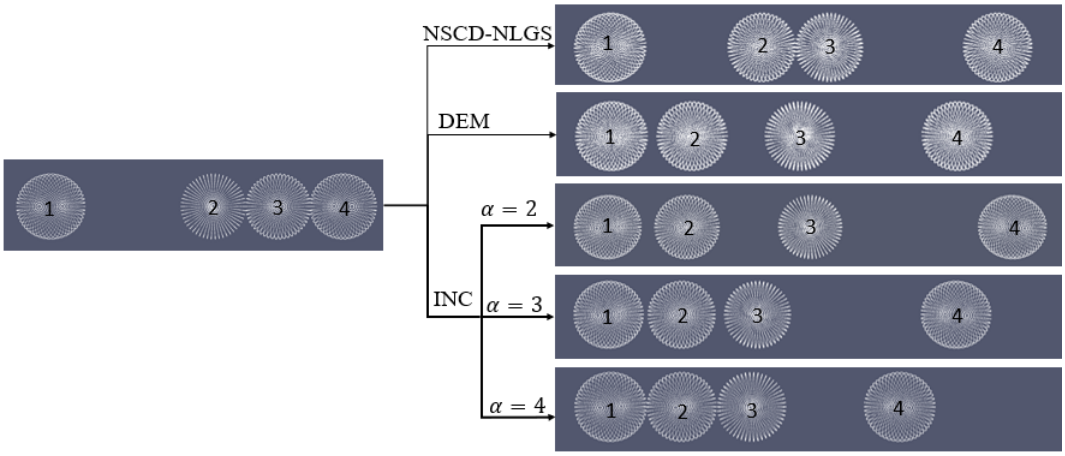



**Figure 4.15:** Position of particles - Newton's cradle model after $1.5s$ corresponding with NSCD-NLGS, DEM, SNC, INC: $\alpha = 2, 3, 4$.

**Remark 4.6.** Based on our numerical experiences, we notice that:

(i) Similar to the previous test, we have to choose a suitable R for DEM and SNC to get numerical results. With this model, it is evident that neither of these methods attains a high level of accuracy in representing particle behavior and conservation of energy.

(ii) NSCD-NLGS effectively preserves the energy of the system, even with a small time step. However, the motion of particles in Newton's model sometimes appears to deviate from the expected behavior, extending beyond the solutions predicted by the analysis. In contrast, INC improves this aspect of behavior (see Fig. 4.15).

(iii) With INC, the best approximate solution corresponds to the case of larger $\alpha$.

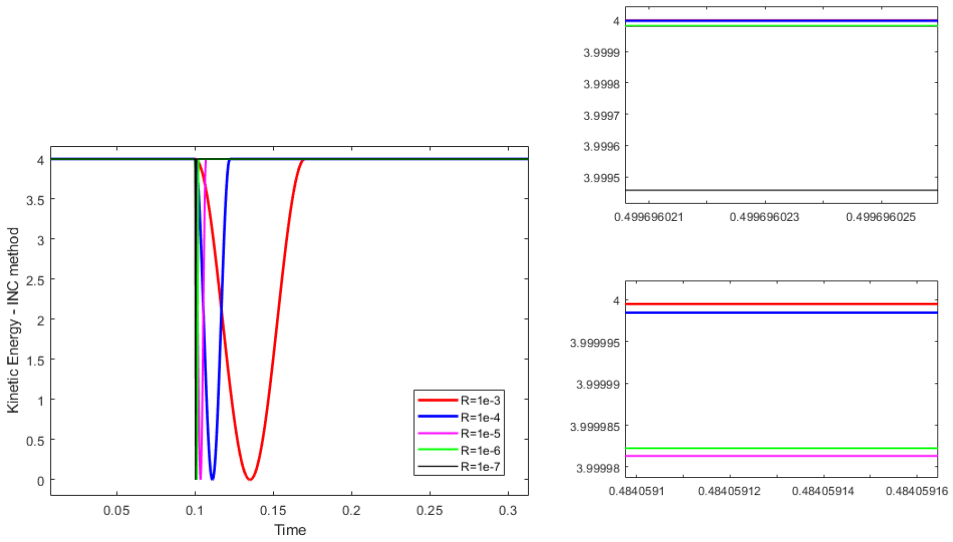
| $\Delta t$ | $R$ | | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $\Delta\mathcal{E}_{c\ relative}$ | Penetration |
|---|---|---|---|---|---|---|---|---|
| $10^{-3}$ | | NSCD | -0.333 | 0.222 | 0.222 | 0.888 | $1.591{\cdot}10^{-7}$ | $1.667{\cdot}10^{-3}$ |
| | $10^{-3}$ | DEM | -0.140 | $-7.238{\cdot}10^{-2}$ | 0.229 | 0.984 | $4.524{\cdot}10^{-2}$ | $8.138{\cdot}10^{-2}$ |
| | | SNC | -0.125 | $-5.787{\cdot}10^{-2}$ | 0.215 | 0.938 | $4.161{\cdot}10^{-2}$ | $7.960{\cdot}10^{-2}$ |
| | | INC [2] | -0.132 | $-6.499{\cdot}10^{-2}$ | 0.237 | 0.960 | $1.514{\cdot}10^{-4}$ | $8.048{\cdot}10^{-2}$ |
| | | INC [3] | $-3.789{\cdot}10^{-2}$ | $-1.009{\cdot}10^{-2}$ | $5.001{\cdot}10^{-2}$ | 0.997 | $2.809{\cdot}10^{-9}$ | 0.250 |
| | | INC [4] | $-1.039{\cdot}10^{-2}$ | $-8.913{\cdot}10^{-4}$ | $1.181{\cdot}10^{-2}$ | 0.999 | $7.972{\cdot}10^{-4}$ | 0.445 |
| | $10^{-4}$ | DEM | | | | | | |
| | | SNC | | | | | | |
| | | INC [2] | -0.132 | $-6.482{\cdot}10^{-2}$ | 0.238 | 0.959 | $1.592{\cdot}10^{-3}$ | $2.545{\cdot}10^{-2}$ |
| | | INC [3] | $-3.792{\cdot}10^{-2}$ | $-1.010{\cdot}10^{-2}$ | $5.005{\cdot}10^{-2}$ | 0.997 | $2.322{\cdot}10^{-6}$ | 0.116 |
| | | INC [4] | $-1.040{\cdot}10^{-2}$ | $-8.918{\cdot}10^{-4}$ | $1.141{\cdot}10^{-2}$ | 0.999 | $1.696{\cdot}10^{-9}$ | 0.250 |
| | $10^{-5}$ | DEM | | | | | | |
| | | SNC | | | | | | |
| | | INC [2] | -0.133 | $-6.201{\cdot}10^{-2}$ | 0.247 | 0.948 | $1.796{\cdot}10^{-2}$ | $8.053{\cdot}10^{-3}$ |
| | | INC [3] | $-3.807{\cdot}10^{-2}$ | $-1.015{\cdot}10^{-2}$ | $5.025{\cdot}10^{-2}$ | 0.997 | $1.256{\cdot}10^{-7}$ | $5.389{\cdot}10^{-2}$ |
| | | INC [4] | $-1.041{\cdot}10^{-2}$ | $-8.935{\cdot}10^{-4}$ | $1.142{\cdot}10^{-2}$ | 0.999 | $1.012{\cdot}10^{-8}$ | 0.141 |
| | $10^{-6}$ | DEM | | | | | | |
| | | SNC | | | | | | |
| | | INC [2] | | | | | | |
| | | INC [3] | $-3.870{\cdot}10^{-2}$ | $-1.037{\cdot}10^{-2}$ | $5.127{\cdot}10^{-2}$ | 0.997 | $1.553{\cdot}10^{-4}$ | $2.502{\cdot}10^{-2}$ |
| | | INC [4] | $-1.046{\cdot}10^{-2}$ | $-8.987{\cdot}10^{-4}$ | $1.147{\cdot}10^{-2}$ | 0.999 | $3.918{\cdot}10^{-9}$ | $7.904{\cdot}10^{-2}$ |
| $10^{-4}$ | | NSCD | -0.333 | 0.222 | 0.222 | 0.888 | $1.641{\cdot}10^{-8}$ | $1.667{\cdot}10^{-4}$ |
| | $10^{-3}$ | DEM | -0.133 | $-6.570{\cdot}10^{-2}$ | 0.236 | 0.963 | $4.358{\cdot}10^{-3}$ | $8.057{\cdot}10^{-2}$ |
| | | SNC | -0.132 | $-6.425{\cdot}10^{-2}$ | 0.238 | 0.950 | $4.322{\cdot}10^{-3}$ | $8.039{\cdot}10^{-2}$ |
| | | INC [2] | -0.132 | $-6.497{\cdot}10^{-2}$ | 0.236 | 0.960 | $1.475{\cdot}10^{-6}$ | $8.048{\cdot}10^{-2}$ |
| | | INC [3] | $-3.788{\cdot}10^{-2}$ | $-1.009{\cdot}10^{-2}$ | $5.000{\cdot}10^{-2}$ | 0.997 | $1.996{\cdot}10^{-10}$ | 0.250 |
| | | INC [4] | $-1.039{\cdot}10^{-2}$ | $-8.911{\cdot}10^{-4}$ | $1.140{\cdot}10^{-2}$ | 0.999 | $-1.358{\cdot}10^{-12}$ | 0.445 |
| | $10^{-4}$ | DEM | -0.135 | $-6.731{\cdot}10^{-2}$ | 0.234 | 0.968 | $1.391{\cdot}10^{-2}$ | $2.554{\cdot}10^{-2}$ |
| | | SNC | -0.130 | $-6.269{\cdot}10^{-2}$ | 0.239 | 0.953 | $1.354{\cdot}10^{-2}$ | $2.536{\cdot}10^{-2}$ |
| | | INC [2] | -0.132 | $-6.499{\cdot}10^{-2}$ | 0.236 | 0.960 | $5.272{\cdot}10^{-6}$ | $2.545{\cdot}10^{-2}$ |
| | | INC [3] | $-3.788{\cdot}10^{-2}$ | $-1.009{\cdot}10^{-2}$ | $5.000{\cdot}10^{-2}$ | 0.997 | $3.904{\cdot}10^{-9}$ | 0.116 |
| | | INC [4] | $-1.039{\cdot}10^{-2}$ | $-8.911{\cdot}10^{-4}$ | $1.140{\cdot}10^{-2}$ | 0.999 | $-1.644{\cdot}10^{-12}$ | 0.250 |
| | $10^{-5}$ | DEM | -0.140 | $-7.500{\cdot}10^{-2}$ | 0.228 | 0.984 | $4.525{\cdot}10^{-2}$ | $8.135{\cdot}10^{-3}$ |
| | | SNC | -0.125 | $-5.787{\cdot}10^{-2}$ | 0.245 | 0.938 | $4.161{\cdot}10^{-2}$ | $7.960{\cdot}10^{-3}$ |
| | | INC [2] | -0.131 | $-6.513{\cdot}10^{-2}$ | 0.236 | 0.960 | $1.496{\cdot}10^{-4}$ | $8.045{\cdot}10^{-3}$ |
| | | INC [3] | $-3.789{\cdot}10^{-2}$ | $-1.009{\cdot}10^{-2}$ | $5.000{\cdot}10^{-2}$ | 0.997 | $2.186{\cdot}10^{-8}$ | $5.390{\cdot}10^{-2}$ |
| | | INC [4] | $-1.039{\cdot}10^{-2}$ | $-8.911{\cdot}10^{-4}$ | $1.140{\cdot}10^{-2}$ | 0.999 | $4.973{\cdot}10^{-13}$ | 0.141 |
| | $10^{-6}$ | DEM | | | | | | |
| | | SNC | | | | | | |
| | | INC [2] | -0.127 | $-6.417{\cdot}10^{-2}$ | 0.230 | 0.961 | $1.624{\cdot}10^{-3}$ | $2.537{\cdot}10^{-3}$ |
| | | INC [3] | $-3.789{\cdot}10^{-2}$ | $-1.009{\cdot}10^{-2}$ | $5.001{\cdot}10^{-2}$ | 0.997 | $2.879{\cdot}10^{-9}$ | $2.502{\cdot}10^{-2}$ |
| | | INC [4] | $-1.039{\cdot}10^{-2}$ | $-8.911{\cdot}10^{-2}$ | $1.140{\cdot}10^{-2}$ | 0.999 | $6.188{\cdot}10^{-12}$ | $7.906{\cdot}10^{-3}$ |

**Tab 4.11:** Simulation of Newton's cradle

(iv) With the same initial conditions, we can observe that the resolution by using INC is relatively better than SNC one, especially the kinetic energy of the system.

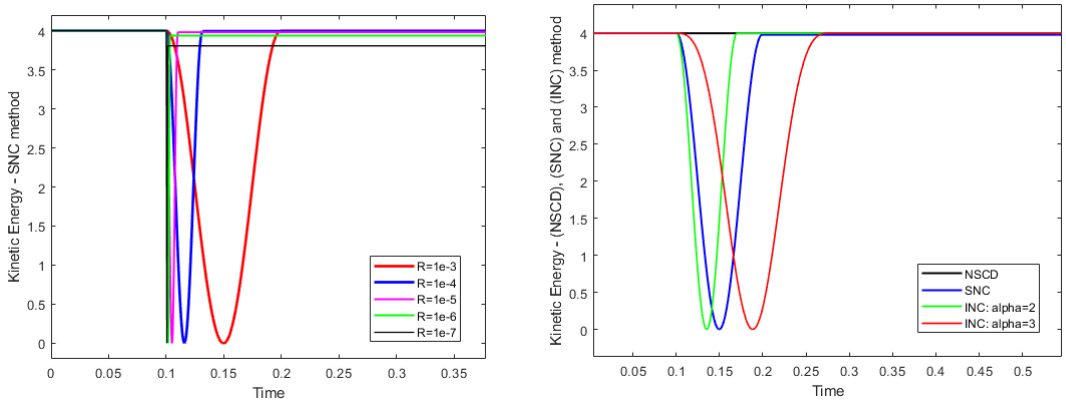
#### 4.3.2.4 Test 4: Four balls moving to a fixed point

The simulation involves the analysis of four particles utilizing various computational methods.




(a) INC$^2$ with different $R$ ( $\Delta t = 10^{-4}$)

(b) zoom out




(c) SNC with different $R$ ($\Delta t = 10^{-5}$)

(d) Comparison kinetic energy of the system ($\Delta t = 10^{-4}, R = 10^{-4}$)

**Figure 4.16:** Simulation of four particles moving to a fixed point in 2D.

**Remark 4.7.**

(i) With SNC, when the parameter R assumes larger values, the system's energy tends to be ensured (Fig. 4.16-(c)).

| $\Delta t$ | $R$ | Penetration | $\Delta\mathcal{E}_{c\ re.}$ | Penetration | $\Delta\mathcal{E}_{c\ re.}$ |
|---|---|---|---|---|---|
| $10^{-3}$ | NSCD | $1.863{\cdot}10^{-11}$ | $1.490{\cdot}10^{-8}$ | | |
| $10^{-4}$ | | $2.098{\cdot}10^{-13}$ | $1.863{\cdot}10^{-9}$ | | |
| $10^{-5}$ | | 0.000 | $1.164{\cdot}10^{-10}$ | | |
| | | DEM | | SNC | |
| $10^{-3}$ | $10^{-3}$ | $5.689{\cdot}10^{-2}$ | $7.317{\cdot}10^{-2}$ | $0.549{\cdot}10^{-1}$ | $0.675{\cdot}10^{-1}$ |
| | $10^{-4}$ | | | | |
| | $10^{-5}$ | | | | |
| | $10^{-6}$ | | | | |
| | $10^{-7}$ | | | | |
| $10^{-4}$ | $10^{-3}$ | $5.600{\cdot}10^{-2}$ | $7.055{\cdot}10^{-3}$ | $0.558{\cdot}10^{-1}$ | $0.699{\cdot}10^{-2}$ |
| | $10^{-4}$ | $1.778{\cdot}10^{-2}$ | $2.249{\cdot}10^{-2}$ | $0.219{\cdot}10^{-1}$ | $0.176{\cdot}10^{-1}$ |
| | $10^{-5}$ | | | | |
| | $10^{-6}$ | | | | |
| | $10^{-7}$ | | | | |
| $10^{-5}$ | $10^{-3}$ | $5.591{\cdot}10^{-2}$ | $7.028{\cdot}10^{-4}$ | $0.559{\cdot}10^{-1}$ | $0.702{\cdot}10^{-3}$ |
| | $10^{-4}$ | $1.769{\cdot}10^{-2}$ | $2.224{\cdot}10^{-3}$ | $0.177{\cdot}10^{-1}$ | $0.222{\cdot}10^{-3}$ |
| | $10^{-5}$ | $5.600{\cdot}10^{-3}$ | $7.055{\cdot}10^{-3}$ | $0.671{\cdot}10^{-2}$ | $0.558{\cdot}10^{-2}$ |
| | $10^{-6}$ | $1.778{\cdot}10^{-3}$ | $2.249{\cdot}10^{-2}$ | $0.176{\cdot}10^{-2}$ | $0.219{\cdot}10^{-1}$ |
| | $10^{-7}$ | $5.689{\cdot}10^{-4}$ | $7.317{\cdot}10^{-2}$ | $0.549{\cdot}10^{-3}$ | $0.675{\cdot}10^{-1}$ |

| $\Delta t$ | $R$ | Penetration | $\Delta\mathcal{E}_{c\ re.}$ | Penetration | $\Delta\mathcal{E}_{c\ re.}$ |
|---|---|---|---|---|---|
| | | $\text{INC}^2$ | | $\text{INC}^3$ | |
| $10^{-3}$ | $10^{-3}$ | $0.559{\cdot}10^{-1}$ | $0.135{\cdot}10^{-3}$ | 0.157 | $0.121{\cdot}10^{-5}$ |
| | $10^{-4}$ | $0.177{\cdot}10^{-1}$ | $0.126{\cdot}10^{-2}$ | $0.731{\cdot}10^{-1}$ | $0.131{\cdot}10^{-9}$ |
| | $10^{-5}$ | $0.549{\cdot}10^{-2}$ | $0.381{\cdot}10^{-2}$ | $0.339{\cdot}10^{-1}$ | $0.320{\cdot}10^{-6}$ |
| | $10^{-6}$ | | | $0.157{\cdot}10^{-1}$ | $0.158{\cdot}10^{-2}$ |
| | $10^{-7}$ | | | $0.729{\cdot}10^{-2}$ | $0.162{\cdot}10^{-2}$ |
| $10^{-4}$ | $10^{-3}$ | $0.559{\cdot}10^{-1}$ | $0.466{\cdot}10^{-5}$ | 0.157 | $0.105{\cdot}10^{-8}$ |
| | $10^{-4}$ | $0.177{\cdot}10^{-1}$ | $0.438{\cdot}10^{-5}$ | $0.731{\cdot}10^{-1}$ | $0.747{\cdot}10^{-12}$ |
| | $10^{-5}$ | $0.559{\cdot}10^{-2}$ | $0.135{\cdot}10^{-3}$ | $0.339{\cdot}10^{-1}$ | $0.926{\cdot}10^{-7}$ |
| | $10^{-6}$ | $0.177{\cdot}10^{-2}$ | $0.126{\cdot}10^{-2}$ | $0.157{\cdot}10^{-1}$ | $0.121{\cdot}10^{-5}$ |
| | $10^{-7}$ | $0.549{\cdot}10^{-3}$ | $0.382{\cdot}10^{-2}$ | $0.731{\cdot}10^{-2}$ | $0.114{\cdot}10^{-8}$ |
| $10^{-5}$ | $10^{-3}$ | $0.559{\cdot}10^{-1}$ | $0.532{\cdot}10^{-6}$ | 0.157 | $0.320{\cdot}10^{-12}$ |
| | $10^{-4}$ | $0.177{\cdot}10^{-1}$ | $0.391{\cdot}10^{-6}$ | $0.731{\cdot}10^{-1}$ | $0.311{\cdot}10^{-13}$ |
| | $10^{-5}$ | $0.559{\cdot}10^{-2}$ | $0.467{\cdot}10^{-5}$ | $0.339{\cdot}10^{-1}$ | $0.455{\cdot}10^{-12}$ |
| | $10^{-6}$ | $0.177{\cdot}10^{-2}$ | $0.444{\cdot}10^{-5}$ | $0.157{\cdot}10^{-1}$ | $0.105{\cdot}10^{-8}$ |
| | $10^{-7}$ | $0.549{\cdot}10^{-3}$ | $0.136{\cdot}10^{-3}$ | $0.731{\cdot}10^{-2}$ | $0.993{\cdot}10^{-11}$ |

**Tab 4.12:** Resolution of simulation for four balls moving to a fixed point.

(ii) With INC, for a given $\alpha$, the best approximate solution (resp. the energy of the system balances) usually corresponds to a small parameter $R$ (i.e. $R = 10^{-3}, R = 10^{-4}$) and decreasing when $R$ is smaller (see Fig. 4.16).

(iii) When the value $\alpha$ is gradually increased, the error of kinetic energy compared with the initial energy seems to get better. With INC, for each $\alpha$, the penatration of particles is gradually increased which directly affects the calculation of the reaction force after the collision (observe Tab. 4.12 with $\alpha = 2, 3, 4$ in case of $\Delta t = 10^{-5}, R = 10^{-4}$).

#### 4.3.2.5 Test 5: several particles in a box

We provide some tests as follow.

(a) It is used to test the ability of the particles to move vertically and the successive collisions that occur between layers of particles stacking on each other.

(b) This test is for 81 particles in a box with random input (velocities and positions), the results are shown in Tab. 4.14 and Fig. 4.18.

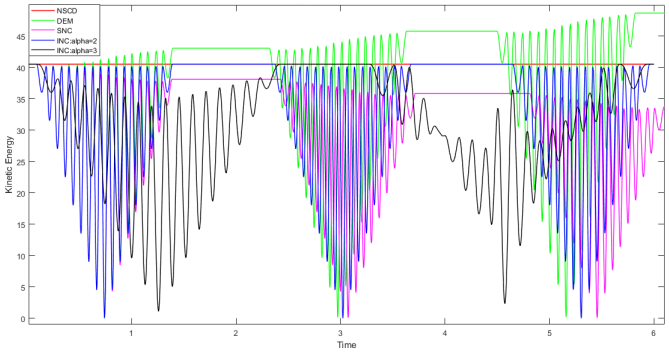
(c) Another example we conducted involved 100 particles moving and colliding with each other at a fixed point in the region.

| R | Method | $\mathcal{E}_c$/Initial | $\Delta\mathcal{E}_{c\ relative}$ | Penetration |
|---|---|---|---|---|
| | NSCD | 40.499/40.5 | $2.28 \cdot 10^{-7}$ | $6.25 \cdot 10^{-4}$ |
| $10^{-3}$ | DEM | 48.645/40.5 | $2.01 \cdot 10^{-1}$ | $1.53 \cdot 10^{-1}$ |
| | SNC | 33.724/40.5 | $1.71 \cdot 10^{-1}$ | $1.39 \cdot 10^{-1}$ |
| | INC [2] | 40.490/40.5 | $2.48 \cdot 10^{-4}$ | $1.40 \cdot 10^{-1}$ |
| | INC [3] | 40.499/40.5 | $1.13 \cdot 10^{-8}$ | $4.03 \cdot 10^{-1}$ |
| | INC [4] | 40.500/40.5 | $1.84 \cdot 10^{-13}$ | $6.69 \cdot 10^{-1}$ |
| $10^{-4}$ | DEM | | | |
| | SNC | | | |
| | INC [2] | 40.379/40.5 | $2.99 \cdot 10^{-3}$ | $4.42 \cdot 10^{-2}$ |
| | INC [3] | 40.499/40.5 | $2.61 \cdot 10^{-7}$ | $1.83 \cdot 10^{-1}$ |
| | INC [4] | 40.499/40.5 | $9.92 \cdot 10^{-12}$ | $3.73 \cdot 10^{-1}$ |
| $10^{-5}$ | DEM | | | |
| | SNC | | | |
| | INC [2] | 39.121/40.5 | $3.40 \cdot 10^{-2}$ | $1.40 \cdot 10^{-2}$ |
| | INC [3] | 40.499/40.5 | $3.34 \cdot 10^{-6}$ | $8.48 \cdot 10^{-2}$ |
| | INC [4] | 40.499/40.5 | $4.17 \cdot 10^{-10}$ | $2.09 \cdot 10^{-1}$ |

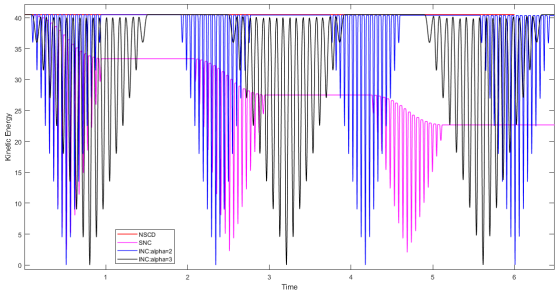
**Tab 4.13:** Simulation of 81 particles with fixed initial conditions.

**Remark 4.8.**

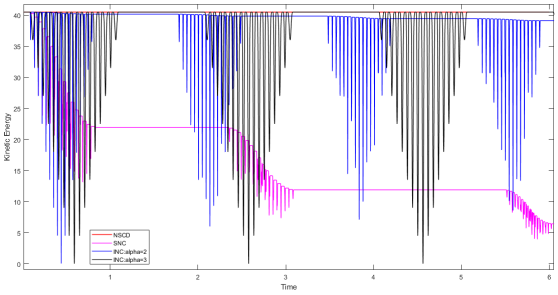
(i) In test (a), the time to complete a single period of collisions between particles is different since penetrations corresponding to each $\alpha$ value are different (see Fig. 4.17).




(a) $\Delta t = 10^{-4}, c_\nu = 10^3$




(b) $\Delta t = 10^{-4}, c_\nu = 10^4$




(c) $\Delta t = 10^{-4}, c_\nu = 10^5$

**Figure 4.17:** Simulation of 81 particles with fixed initial conditions.

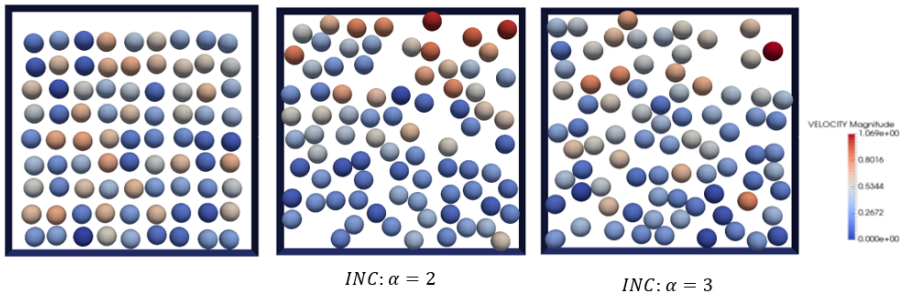
| | Energy | $\Delta\mathcal{E}_{c\ relative}$ | Penetration |
|---|---|---|---|
| DEM | 16.7219349 | $1.4866 \cdot 10^{+1}$ | $1.4798 \cdot 10^{-2}$ |
| NSCD-NLGS | 6.7248957 | $4.8171 \cdot 10^{-8}$ | $3.2928 \cdot 10^{-4}$ |
| SNC | 5.4813453 | $1.8492 \cdot 10^{-1}$ | $2.5967 \cdot 10^{-4}$ |
| INC [2] | 6.6738305 | $7.5935 \cdot 10^{-3}$ | $8.8043 \cdot 10^{-3}$ |
| INC [3] | 6.7248927 | $4.8092 \cdot 10^{-7}$ | $6.4869 \cdot 10^{-2}$ |

**Tab 4.14:** Resolution of simulation of 81 particles in a box with random initial positions and velocities, $\Delta t = 10^{-4}, c_\nu = 10^4$ - NSCD-NLGS, DEM, SNC, INC: $\alpha = 2, 3$.

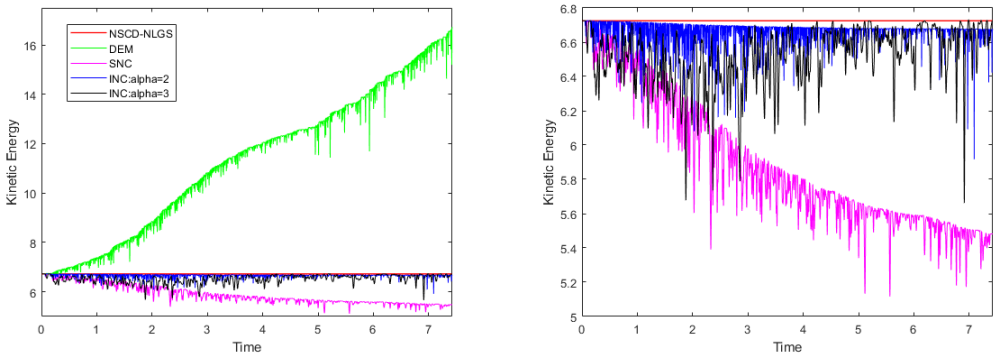
(ii) In test (b) with random initial conditions, it is clear that with NSCD-NLGS, we optimal the CPU time while DEM, SNC, and INC take more time due to the difference of the procedure at the local contact (observe Tab. 4.14).

(iii) In experiments (b) and (c), the DEM and SNC method typically demonstrates good performance with $R = 10^{-3}, 10^{-4}$, consistent with findings from previous tests. Hence, careful consideration should be given to the selection of this parameter.

(iv) In two tests, it is observed that with a larger $\alpha$ value, as employed in the INC method, the preservation of the system's kinetic energy is achieved with greater accuracy, as Tab. 4.13, 4.14, 4.15.




(a) Visualization




(b) Kinetic energy of the system

**Figure 4.18:** Simulation of 81 particles with random initial positions and velocities, $\Delta t = 10^{-4}, c_\nu = 10^4$ - using different methods: NSCD-NLGS, DEM, SNC, INC: $\alpha = 2, 3$.

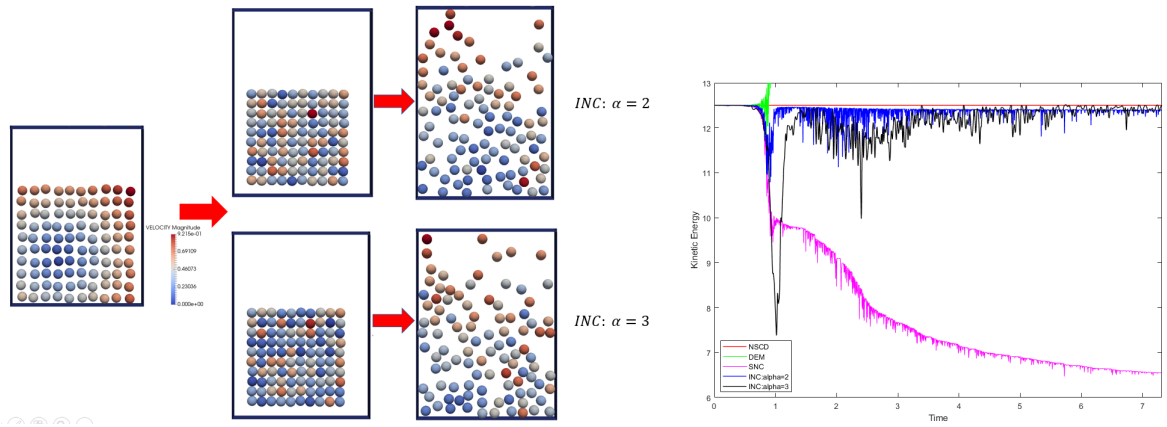
| R | Method | NSCD | DEM | SNC | INC [2] | INC [3] |
|---|---|---|---|---|---|---|
| $10^{-4}$ | $\Delta\mathcal{E}_{c\ relative}$ | $5.105 \cdot 10^{-8}$ | 0.236 | 0.186 | $3.717 \cdot 10^{-4}$ | $1.506 \cdot 10^{-3}$ |
| | Penetration | $3.942 \cdot 10^{-2}$ | 3.390 | $3.865 \cdot 10^{-2}$ | $3.750 \cdot 10^{-2}$ | 0.129 |
| $10^{-5}$ | $\Delta\mathcal{E}_{c\ relative}$ | | | 0.476 | $8.327 \cdot 10^{-3}$ | $1.752 \cdot 10^{-7}$ |
| | Penetration | | | $9.587 \cdot 10^{-3}$ | $1.239 \cdot 10^{-2}$ | $6.278 \cdot 10^{-2}$ |
| $10^{-6}$ | $\Delta\mathcal{E}_{c\ relative}$ | | | 0.849 | $8.983 \cdot 10^{-2}$ | $3.378 \cdot 10^{-7}$ |
| | Penetration | | | $1.554 \cdot 10^{-3}$ | $3.255 \cdot 10^{-3}$ | $3.135 \cdot 10^{-2}$ |

**Tab 4.15:** Simulation of 100 particles with random initial conditions.

## 4.4 Restitution coefficient for impact law

The restitution coefficient, often denoted by the symbol "$e$", is a parameter used in mechanics to describe the elasticity or "bounciness" of a collision between two objects. It represents the ratio of the relative velocity of separation after the collision to the relative velocity of approach before the collision. Values of e range from 0 to 1, where $e = 0$ denotes a perfectly inelastic collision, $e = 1$ represents a perfectly elastic collision, and $0 < e < 1$ indicates a partially elastic collision with dissipation of kinetic energy (Impact Laws, see Section 3.2.2).

In this work, focused on addressing displacement contact problems, we introduce an extension of Newmark-INC-PDAS with a restitution coefficient $e$. The normal reaction




(a) Visualization-INC: $\alpha = 2, 3$ (b) Kinetic engergy

**Figure 4.19:** Simulation of 100 particles in a box with random initial conditions ($\Delta t = 10^{-4}, c_\nu = 10^5$).

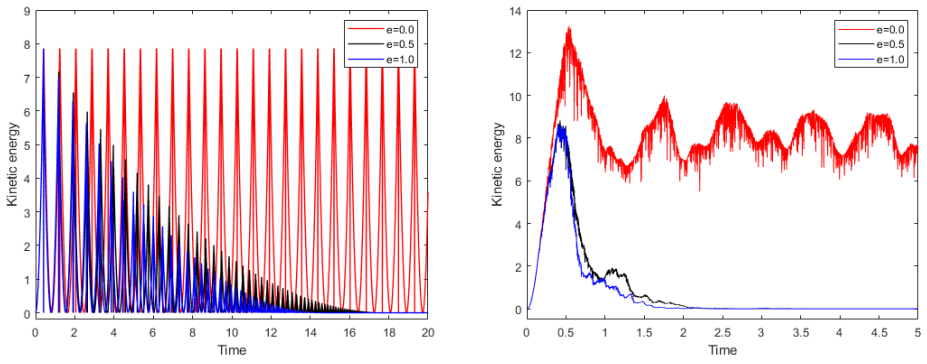
force is formulated as follows

$$\lambda_\nu^{n+1} = c_\nu \left( \frac{1}{1+e}\, \tilde{D}_\nu^{n+1} + \frac{e}{1+e} D_\nu^{n+1} \right)$$

where $e \in [0, 1]$.

Now, let us consider some examples.

- A ball under gravity with $c_\nu = 10^6, v_0 = (0, 0)$, $g = 9.80655$ during $T = 20$s and $\Delta t = 10^{-4}$. The system's energy is shown in Fig. 4.20-b corresponds to $e = 0.0/0.5/1.0$.
- Simulation of 25 balls under gravity with $c_\nu = 10^6$, random initial position, $v_0^i = (0, 0)$ and $g = 9.80655$ during $T = 5$s, $\Delta t = 10^{-4}$, enr $< 10^{-6}$ (see Fig. 4.20-b).




(a) 1 particle (b) 25 particles

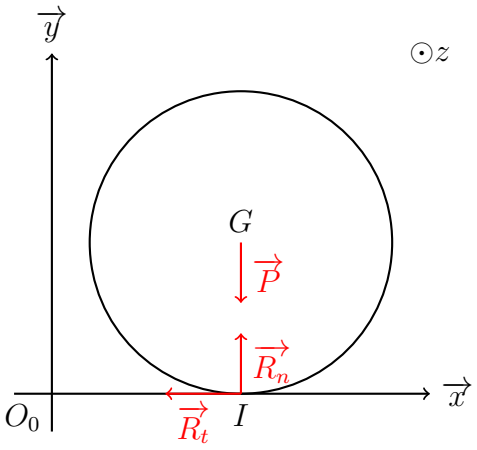
**Figure 4.20:** Simulation of 1 and 25 particles with different restitution coefficients.

**Remark 4.9.** The system's energy gradually diminishes as the restitution coefficient $e$ value increases, but it's worth noting that this reduction occurs over an extended period. It is important to emphasize that this behavior is distinct from the restitution coefficient's impact law concerning velocity, leading to an absence of specific conclusions regarding the cases of $e = 0$ or $e = 1$ in this context.

## 4.5 Friction conditions

This section is dedicated to investigating the integration of friction conditions into the Newmark-INC-PDAS model. The first part 4.5.1 provides a description of a particular simple friction model. Subsequently, we focus on the primary objective of studying the mechanical problem under friction conditions 4.5.2, followed by the execution of simulations 4.5.3 based on the newly formulated theory.

### 4.5.1 Analytical solution




Let us consider a sliding/rolling ball on a horizontal plane. The balance of the forces is

$$\vec{P} = -mg\vec{y} \quad \Rightarrow \mathrm{R}_n = mg\vec{y} \quad \Rightarrow \quad R_t = -\mu mg \cdot \mathrm{sign}(v_0)\vec{x} \text{ if slip.}$$

The dynamic of the system is $m\vec{\Gamma}(G) = m\ddot{x}$, equivalently

$$\begin{cases} m\ddot{x} = -\mu mg \cdot \mathrm{sign}(v_0)\vec{x} \\ x(0) = 0, \ \dot{x}(0) = v_0. \end{cases}$$

It follows that

$$x(t) = -\frac{1}{2}\mu g \cdot \mathrm{sign}(v_0)t^2 + v_0 t.$$

In addition, the Moment of inertia with respect to the $(G, \vec{z})$ axis, namely $I$, is

$$I\ddot{\theta}\vec{z} = r\vec{y} \wedge (-\mu g \cdot \mathrm{sign}(v_0)\vec{x}) = -\mu rmg \cdot \mathrm{sign}(v_0)\vec{z}.$$

Thus we have

$$\begin{cases} \ddot{\theta}(t) = -\dfrac{\mu rmg}{I}\mathrm{sign}(v_0) \\ \theta(0) = \dot{\theta}(0) = 0 \end{cases}$$

Solving the above second order differential equation, we get that

$$\theta(t) = -\frac{\mu r m g}{2I}\text{sign}(v_0)t^2.$$

The particle undergoes a transition from sliding to rolling when its relative velocity reaches zero. To determine the sliding speed necessary for this transition, we calculate the sliding speed as

$$\vec{V}(I) = \vec{V}(C) + \dot{\theta}\vec{z} \wedge \vec{CI} = \big( -\mu g \text{sign}(v_0)t + v_0 - \frac{\mu r^2 m g}{I}\text{sign}(v_0)t\big)\vec{x},$$

then the transition occurs when $\vec{V}(I) = 0$, equivalently at

$$\bar{t} = \frac{|v_0|}{\mu g\left(1 + \dfrac{r^2 m}{I}\right)}.$$

During the process, the horizontal velocity of the system decreases from its initial value $v = v_0$ at $t = 0$ to $v(\bar{t})$ at time $\bar{t}$, where $\mu$ represents a relevant factor. Subsequently, the horizontal velocity remains constant. Simultaneously, the rotational velocity decreases from 0 at $t = 0$ to $t(\bar{t})$ at time $\bar{t}$, and then stabilizes at a constant value. The value of $v(\bar{t})$ and $\theta(\bar{t})$ are given by

$$\dot{\theta}(\bar{t}) = -\mu m g\frac{r}{I}\text{sign}(v_0)\bar{t} \quad \text{and} \quad \vec{V}(C) = \vec{V}(I) - \dot{\theta}(\bar{t})\vec{z} \wedge \vec{CI}.$$

### 4.5.2 Newmark-INC method with friction condition

We consider the mechanic problem under the friction conditions given by

$$\begin{cases} \text{If } \|\boldsymbol{\lambda}_\tau^{n+1}\| < \mu\,\lambda_\nu^{n+1} \text{ then } \boldsymbol{\lambda}_\tau^{n+1} & = c_\tau \dot{\boldsymbol{d}}_\tau^{n+1}, \\ \text{if } \|\boldsymbol{\lambda}_\tau^{n+1}\| \geq \mu\,\lambda_\nu^{n+1} \text{ then } \boldsymbol{\lambda}_\tau^{n+1} & = \mu\,\lambda_\nu^{n+1}\dfrac{\dot{\boldsymbol{d}}_\tau^{n+1}}{\|\dot{\boldsymbol{d}}_\tau^{n+1}\|} \end{cases}$$

where $c_\tau$ is a constant, $\mu$ is a coefficient friction constant, and the tangential relative velocity $\dot{\boldsymbol{d}}_\tau^{n+1}$ between particles $i$ and $j$ is defined ( for a 2D problem) by

$$\dot{\boldsymbol{d}}_\tau^{n+1} = [\dot{\boldsymbol{q}}_i(1:2) - \dot{\boldsymbol{q}}_j(1:2)] \cdot \boldsymbol{\tau} + r_i\dot{\boldsymbol{q}}_i(3) - r_j\dot{\boldsymbol{q}}_j(3)$$

with $r_i$, $r_j$ respectively are the radius of particles $i$ and $j$.

Friction can be broadly categorized into two types: static friction and kinetic (or dynamic) friction. In the case of static friction, there is no relative motion between the two surfaces in contact, and the frictional force prevents any motion from occurring. This is also known as the "stick" case. In the case of kinetic (or dynamic) friction, there is relative motion between the two surfaces, and the frictional force opposes this motion. This is also known as the "slip" case. Therefore, friction encompasses both slip and stick cases.

To accurately model the dynamics of a Newmark system involving friction, it is essential to determine the tangential matrix of motion at the contact point. First, let

us remind that when a contact occurs, $\lambda_\nu^{n+1} = c_\nu \tilde{D}_\nu^{n+1}$. Hence, the two first conditions on the frictional contact are

$$\boldsymbol{\lambda} = c_\nu \tilde{D}_\nu \boldsymbol{\nu} + \begin{cases} c_\tau \big(\boldsymbol{\tau}^T \dot{\boldsymbol{q}} + r_i \dot{\boldsymbol{q}}(3)\big)\boldsymbol{\tau} & \text{if } \ c_\tau \|\dot{\boldsymbol{d}}_\tau\| \leq \mu c_\nu \tilde{D}_\nu, \\ \varepsilon \mu \lambda_\nu \boldsymbol{\tau} & \text{if } \ c_\tau \|\dot{\boldsymbol{d}}_\tau\| > \mu c_\nu \tilde{D}_\nu \end{cases}$$

where $\varepsilon = \pm 1 = \dot{\boldsymbol{d}}_\tau^{n+1} / \|\dot{\boldsymbol{d}}_\tau^{n+1}\|$. For the third condition, we need to transport the momentum of the contact force written at the contact point $I$ to the gravity center $G$ of the particle. This leads to $\boldsymbol{\lambda} = \vec{GI} \wedge (\overline{\lambda}_\tau \boldsymbol{\tau}) = r_i \overline{\lambda}_\tau \boldsymbol{\nu} \wedge \boldsymbol{\tau} = r_i \overline{\lambda}_\tau \mathbf{e}_3$ where $\overline{\lambda}_\tau = \boldsymbol{\lambda}_\tau \cdot \boldsymbol{\tau}$. Hence, it follows (can be simplified using $\varepsilon$ and set $\overline{d}_\tau = \dot{\boldsymbol{d}}_\tau \cdot \boldsymbol{\tau}$) that

$$\boldsymbol{\lambda} = \begin{cases} \mathbf{0} & \text{if } \ \tilde{D}_\nu \leq 0, \\ c_\nu \tilde{D}_\nu \boldsymbol{\nu} + c_\tau \overline{d}_\tau (\boldsymbol{\tau} + r_i \mathbf{e}_3) & \text{if } \ \tilde{D}_\nu > 0 \text{ and } c_\tau \|\dot{\boldsymbol{d}}_\tau\| \leq \mu c_\nu \tilde{D}_\nu, \\ c_\nu \tilde{D}_\nu \boldsymbol{\nu} + \varepsilon \mu c_\nu \tilde{D}_\nu (\boldsymbol{\tau} + r_i \mathbf{e}_3) & \text{if } \ \tilde{D}_\nu > 0 \text{ and } c_\tau \|\dot{\boldsymbol{d}}_\tau\| > \mu c_\nu \tilde{D}_\nu. \end{cases} \tag{4.38}$$

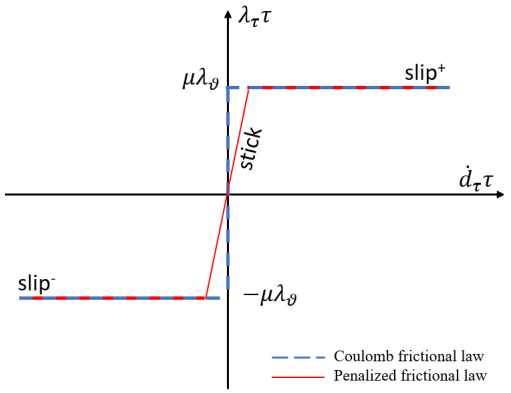



**Figure 4.21:** The penalized frictional law.

The tangential matrix provides a valuable tool for calculating the partial derivatives of the stiffness matrix with respect to the variables involved in the Newmark system. Now, combining with $\nabla_q = \frac{2}{\Delta t} \nabla_{\dot{q}}$ the Jacobian of $\boldsymbol{\lambda}$, namely $J_C = \nabla_q \boldsymbol{\lambda}$, can be calculated in the following specific cases:

- Gap case ($\tilde{D}_\nu \leq 0$).

$$J_C^{q_i} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- Contact + stick case ($\tilde{D}_\nu > 0$ and $c_\tau \|\dot{\boldsymbol{d}}_\tau\| \leq \mu c_\nu \tilde{D}_\nu$).

$$J_C^{q_i} = \begin{bmatrix} c_\nu \nabla \tilde{D}_\nu^{n+1} \boldsymbol{\nu}\boldsymbol{\nu}^T + \dfrac{2}{\Delta t} c_\tau \boldsymbol{\tau}\boldsymbol{\tau}^T & \dfrac{2}{\Delta t} c_\tau r_i \boldsymbol{\tau} \\ \dfrac{2}{\Delta t} c_\tau r_i \boldsymbol{\tau}^T & \dfrac{2}{\Delta t} c_\tau r_i^2 \end{bmatrix}.$$

- Contact + slip case ($\tilde{D}_\nu > 0$ and $c_\tau \|\dot{\boldsymbol{d}}_\tau\| > \mu c_\nu \tilde{D}_\nu$).

$$J_C^{q_i} = \begin{bmatrix} c_\nu \nabla \tilde{D}_\nu^{n+1} \boldsymbol{\nu}\boldsymbol{\nu}^T + \varepsilon \mu c_\nu \nabla \tilde{D}_\nu^{n+1} \boldsymbol{\tau}\boldsymbol{\nu}^T & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}$$

  where $\varepsilon = 1$ with slip$^+$ case and $\varepsilon = -1$ with slip$^-$ one.

Thus, the Jacobian of the reaction force between $q_i$ and $q_j$ can be represented in the form of

$$J_C^{q_i,q_j} = \begin{bmatrix} J_C^{q_i} & -J_C^{q_j} \\ -J_C^{q_i} & J_C^{q_j} \end{bmatrix}.$$

### 4.5.3 Numerical experiments

#### 4.5.3.1 Contact of a ball and the wall

With our model in case of friction, we firstly provide an example of the horizontal movement of a ball along the x-axis, considering the effects of gravity and friction with input: $r = 0.2001, v^0 = (1.0), q^0 = (0.2, 0), \Delta t = 10^{-4}, T = 1.1$ under gravity $g = 9.80655$. A series of results obtained through various experiments by systematically varying the parameters are given as follows.

**Analytical solution** We first provide an analytical result with different coefficients **mu** in Fig. 4.22-(a)) (based on Section 4.5.1).

**Resolution** • **Test 1**: Different coefficients of friction $\mu$ (observe Fig. 4.22: (b) NSCD-NLGS; (c) SNC; (d) INC: $\alpha = 2$; (e) INC: $\alpha = 3$).

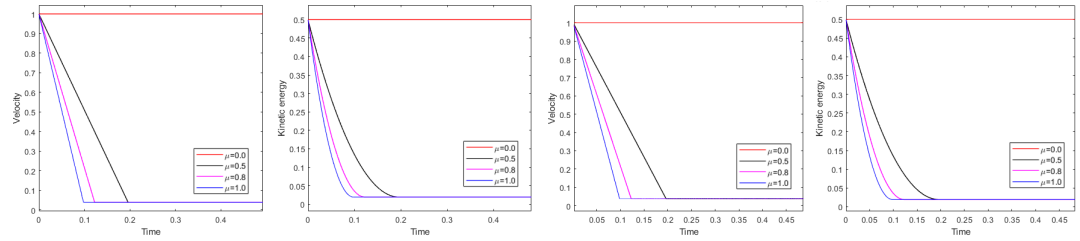
**Remark 4.10.** (i) In all methods, the system's energy decreases more rapidly as the coefficient of friction is larger.

(ii) As we established the radius ($r = 0.2001$) to exceed the initial ball position ($q^0(1) = 0.2$), penetration occurs during the initial time step for both SNC and INC methods, observe Fig. 4.22-(c,d,e), different from NSCD-NLGS one (Fig. 4.22-(a)) as well as Fig. 4.22-(a). Consequently, the kinetic energy of the system exceeds its original value. Moreover, following the conclusion of the frictional process (after 0.2s), the object continues to oscillate in its position due to the combined effects of gravity and the impulse associated with penetration (pertaining to the object's displacement).

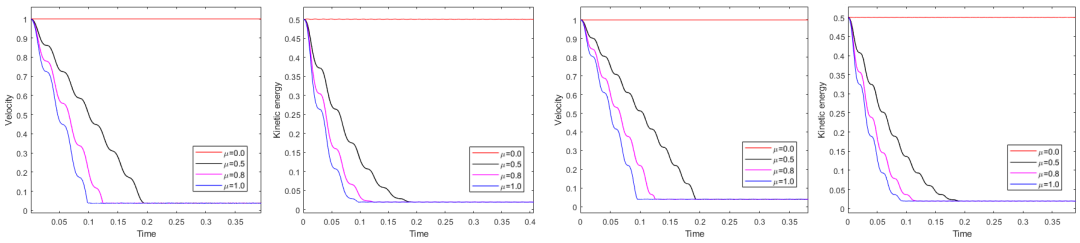
- **Test 2**: Different penalization coefficients $c_\nu$ (Fig. 4.23).

**Remark 4.11.** After the sliding-rolling process, we observe that the penalization, especially with INC: $\alpha = 3$, under the gravity, significantly impact the amplitude of particle oscillations due to penetration.

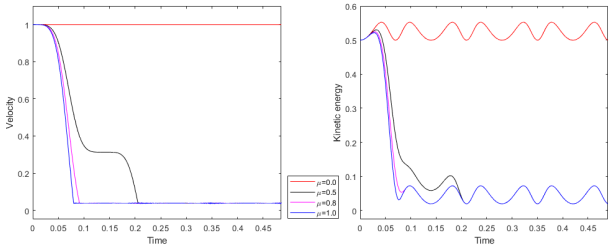
- **Test 3**: Different tangential coefficients $c_\tau$ (Fig. 4.24).




(a) Analytical solution (b) NSCD-NLGS ($c_\nu = c_\tau = 10^4$ )




(c) SNC ($c_\nu = 10^5, c_\tau = 10^4$ ) (d) INC: $\alpha = 2$ ($c_\nu = 10^5, c_\tau = 10^4$)




(e) INC: $\alpha = 3$ ($c_\nu = 10^5, c_\tau = 10^4$)

**Figure 4.22:** Different friction coefficients $\mu$.

**Remark 4.12.** Theoretically, we can choose the initial values for the tangential contact coefficients. A larger value for this coefficient tends to bring the frictional model with penalization (Fig. 4.21) closer to the Coulomb friction model. Consequently, opting for an initial value of $c_\tau$ that approximates $c_\nu$ furnishes us with results of sufficient precision, observing Fig. 4.24.

- **Test 4**: Different masses (Fig. 4.25).

**Remark 4.13.** In dynamic friction, the mass of the object affects the object's speed under the influence of friction, specifically a heavier object is characterized by an expedited frictional process, resulting in a shortened duration of frictional action. Additionally, it also relates to the extent of object penetration to the contacting surface.

**Comparision** Finally, a comparison between some different methods ($c_\nu = c_\tau = 10^6$, $\mu = 0.5$) with result is shown in Fig. 4.26.

**Remark 4.14.** In the context of this frictional conditions, SNC and INC: $\alpha = 2$ exhibit greater effectiveness, as they inherently incorporate energy dissipation during the process. For INC: $\alpha = 3$, after frictional interaction, the body remains to oscillate

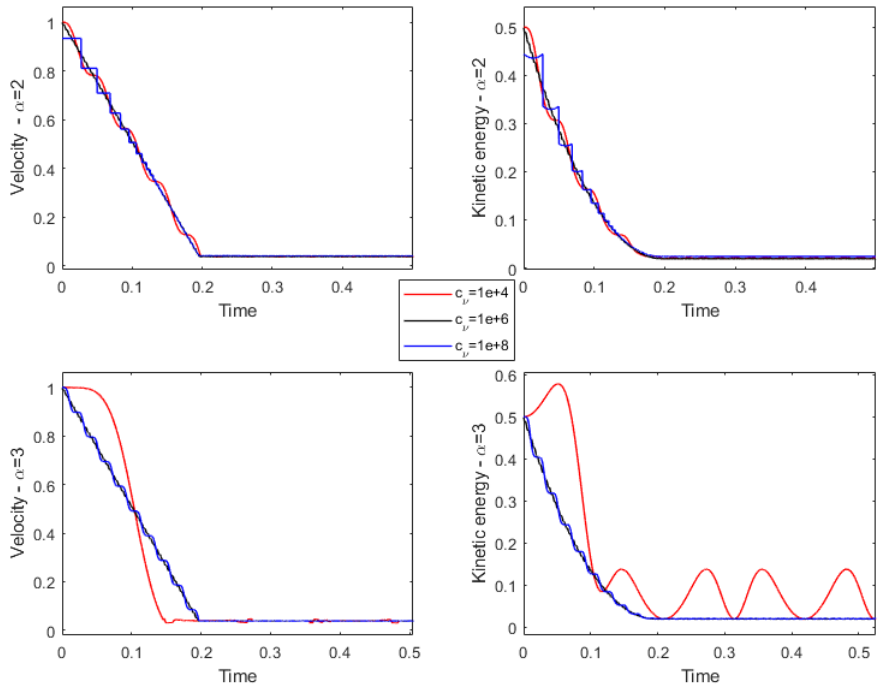



**Figure 4.23:** Different penalization coefficients $c_\nu$ ($c_\tau = 10^6, \mu = 0.5$).

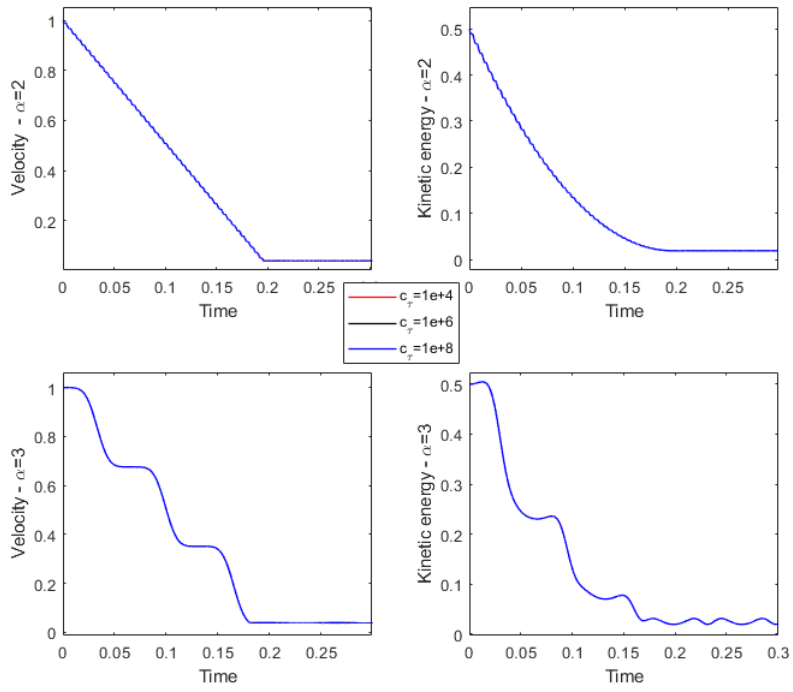



**Figure 4.24:** Different tangential coefficients $c_\tau$ ($c_\nu = 10^6, \mu = 0.5$).

under gravity. In such cases, to maximize energy dissipation in the object, adjustments to the parameters $mu$ or $R$ can be considered (Fig. 4.26 and also 4.28).

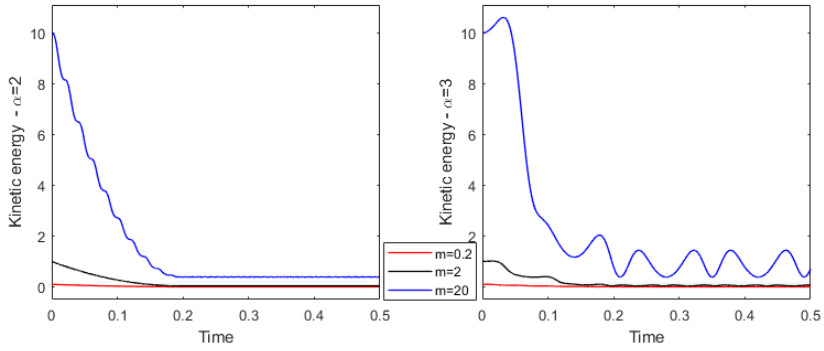



**Figure 4.25:** Different masses of particles $m$ ($c_\nu = 10^6, c_\tau = 10^6, \mu = 0.5$).

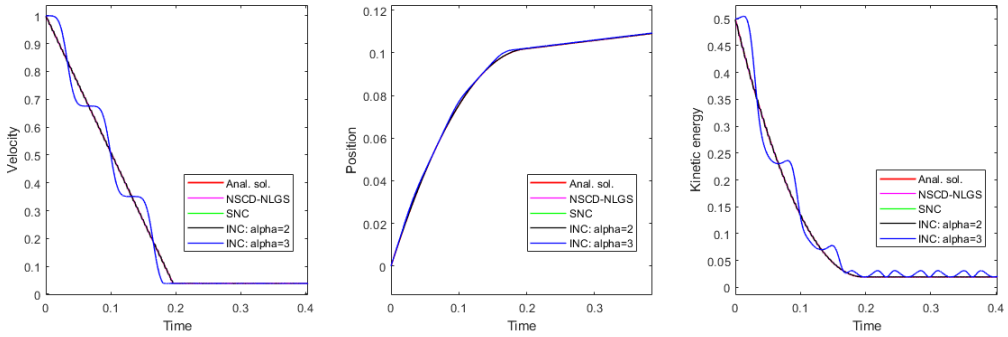



**Figure 4.26:** Bouncing ball ($c_\nu = c_\tau = 10^6, \mu = 0.5$) - Comparison: Analytical solution, NSCD-NLGS, SNC, INC: $\alpha = 2, 3$).

### 4.5.3.2 Contact between particles

The aim of this test is to combine friction conditions to the impulse at contact between particles. The result is shown in Fig. 4.27 with different friction coefficients by using INC method and another is the comparison between NSCD-NLGS and INC: $\alpha = 2, 3$ during $T = 0.5$s with $\Delta t = 10^{-4}$, see Fig. 4.28.

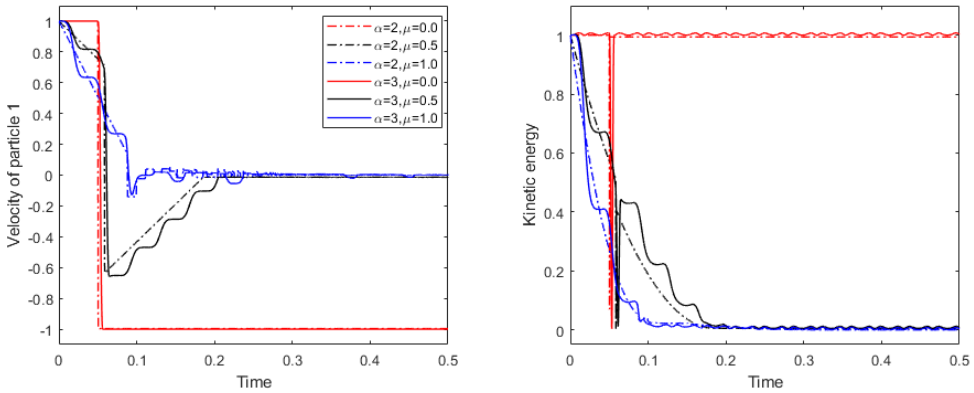



**Figure 4.27:** Simulation of two particles with different friction coefficients ($c_\nu = c_\tau = 10^7$) by using INC

**Remark 4.15.** Similar to above, the comments are satisfied in the case of friction between two bodies.

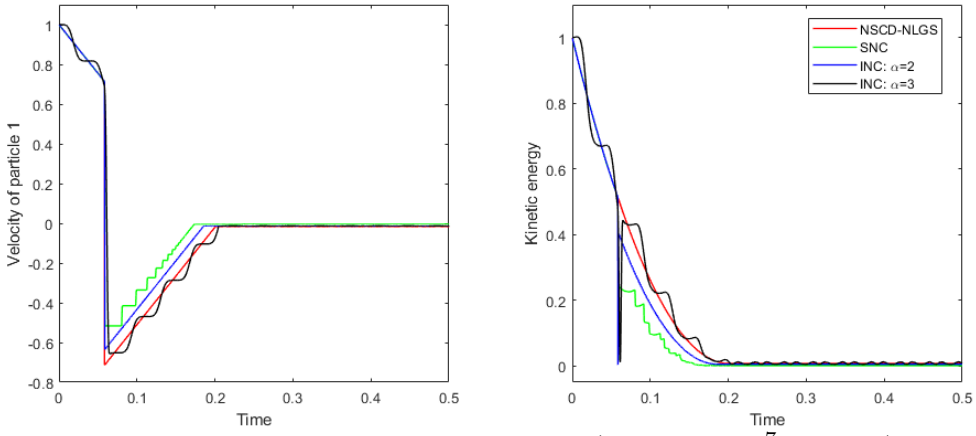
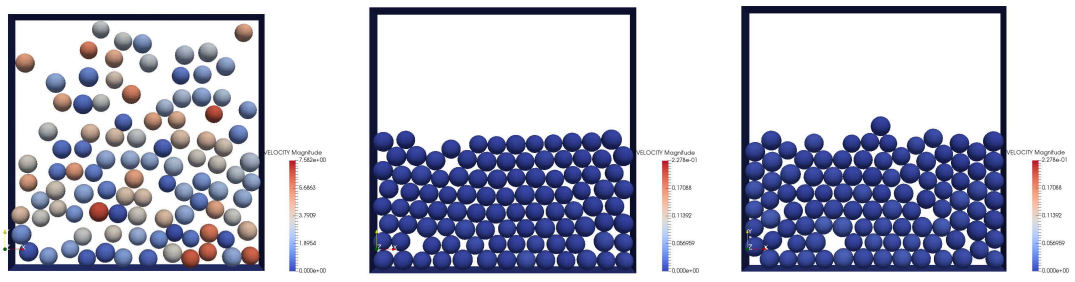


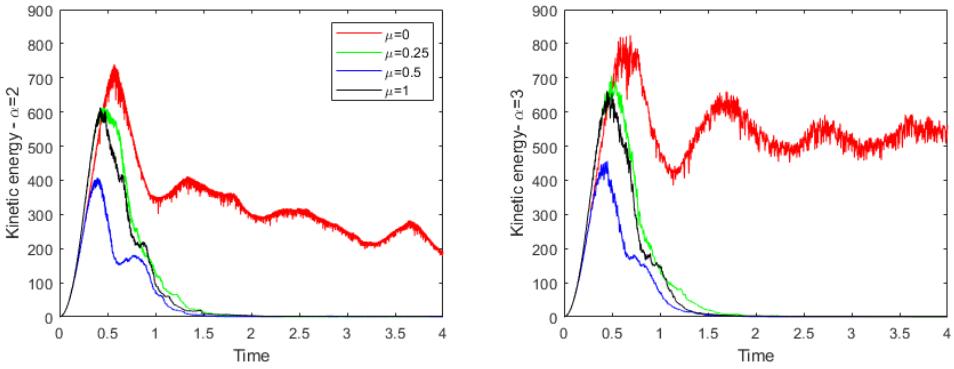

**Figure 4.28:** Collision of two particles with friction ($c_\nu = c_\tau = 10^7, \mu = 0.5$) - Comparison different methods: NSCD-NLGS, SNC, INC: $\alpha = 2, 3$.

#### 4.5.3.3 Simulation of 100 particles in a box

Finally, it is about 100 particles with random positions and radiuses that are moving under gravity and friction conditions ($c_\nu = 10^6, c_\tau = 10^4$ using INC during $4s$).



(a) $\mu = 0.0$ (b) $\mu = 0.25$ (c) $\mu = 0.5$




(d) The system's energy

**Figure 4.29:** Simulation of 100 particles in a box with different frictional coefficients ($c_\nu = 10^6, c_\tau = 10^{15}$).

**Remark 4.16.** With a higher coefficient of friction, the energy dissipation in the granular system occurs more rapidly, more precisely, with $\mu = 0.25/0.5/1.0$, the system

reaches a stable state, observe Fig. 4.29-(d).

## 4.6 CPU time and Nonlinear number iterations

In this section, we present the specific computation times for the examples under both frictionless and friction conditions. The numerical results below summarize the performance of INC methods combined with various linear solvers including `Dgesv`, `Superlu`, `Gmres` with ILT preconditioner, `Pcg`, `Jpcg` and a direct sparse version using `Pcg`, `Jpcg` for the frictionless case (referred to as `Sparse+Pcg`, `Sparse+Jpcg` in the table) and `Sparse+Superlu`, `Sparse+Gmres`. This comprehensive analysis allows us to assess the impact of these solvers on CPU time. In the following, we also use some notation as follow:

- **Pene.** is the maximum penetration at contact;
- **N.L.** is the total number of nonlinear iterations;
- **N.C.** is the total number of contacts;
- **Max.nlit** is the maximum number of nonlinear iterations in a loop;
- **enr = norm(dq)** is the condition for convergence of each time iteration.

All calculations in this study were conducted on a `Dell Inc. Latitude 5400` computer, utilizing the Fortran 90 programming language along with the `LAPACK` and `Sparskit` library packages.

### 4.6.1 Frictionless

#### 4.6.1.1 Newton's cradle

We begin our analysis with Newton's cradle (see again 4.3.2.3) using the INC with $\alpha = 3$ in Tab. 4.16.

#### 4.6.1.2 81 particles in a box without friction

The tables 4.17, 4.18, and figure 4.30 present the results concerning 81 particles in frictionless case, as described in test 5 in 4.3.2.5.

**Remark 4.17.** *Some comments for frictionless case:*

- `Dgesv` exhibits moderate performance with relatively low CPU time but slightly higher energy error compared to other methods.
- `Superlu`, `Gmres` achieves comparable accuracy with `Dgesv` but at a significantly reduced CPU time.
- `Sparse` combined with either `Superlu` or `Gmres` show promising results, offering a balance between accuracy and efficiency, especially noticeable in scenarios with higher complexity ($\alpha = 3$).

| $R$ | Method | $\Delta\mathcal{E}_{c\ relative}$ | Pene. | N.C. | N.L. | CPU |
|---|---|---|---|---|---|---|
| $10^3$ | `Dgesv` | $1.996 \cdot 10^{-10}$ | 0.25 | 25,933 | 30,000 | 0.188 |
| | `Superlu` | $1.372 \cdot 10^{-8}$ | 0.25 | 25,933 | 30,000 | 0.160 |
| | `Gmres` | $5.060 \cdot 10^{-8}$ | 0.25 | 25,933 | 30,000 | 0.164 |
| | `Sparse+Superlu` | $1.994 \cdot 10^{-10}$ | 0.25 | 25,934 | 30,000 | 0.160 |
| | `Sparse+Gmres` | $1.010 \cdot 10^{-7}$ | 0.25 | 25,934 | 30,000 | 0.160 |
| $10^4$ | `Dgesv` | $3.803 \cdot 10^{-9}$ | 0.116 | 18,465 | 30,000 | 0.184 |
| | `Superlu` | $8.071 \cdot 10^{-8}$ | 0.116 | 18,465 | 30,000 | 0.152 |
| | `Gmres` | $5.124 \cdot 10^{-7}$ | 0.116 | 18,465 | 30,000 | 0.160 |
| | `Sparse+Superlu` | $3.799 \cdot 10^{-9}$ | 0.116 | 18,466 | 30,000 | 0.156 |
| | `Sparse+Gmres` | $1.020 \cdot 10^{-6}$ | 0.116 | 18,466 | 30,000 | 0.160 |
| $10^5$ | `Dgesv` | $2.170 \cdot 10^{-8}$ | $5.390 \cdot 10^{-2}$ | 15,000 | 30,000 | 0.188 |
| | `Superlu` | $2.599 \cdot 10^{-7}$ | $5.390 \cdot 10^{-2}$ | 15,000 | 30,000 | 0.156 |
| | `Gmres` | $6.770 \cdot 10^{-7}$ | $5.390 \cdot 10^{-2}$ | 15,000 | 30,000 | 0.164 |
| | `Sparse+Superlu` | $2.168 \cdot 10^{-8}$ | $5.390 \cdot 10^{-2}$ | 15,001 | 30,000 | 0.152 |
| | `Sparse+Gmres` | $1.043 \cdot 10^{-5}$ | $5.390 \cdot 10^{-2}$ | 15,001 | 30,000 | 0.156 |
| $10^6$ | `Dgesv` | $3.340 \cdot 10^{-8}$ | $2.502 \cdot 10^{-2}$ | 13,393 | 30,000 | 0.188 |
| | `Superlu` | $8.513 \cdot 10^{-6}$ | $2.502 \cdot 10^{-2}$ | 13,393 | 30,000 | 0.156 |
| | `Gmres` | $1.393 \cdot 10^{-6}$ | $2.502 \cdot 10^{-2}$ | 13,393 | 30,000 | 0.156 |
| | `Sparse+Superlu` | $3.348 \cdot 10^{-8}$ | $2.502 \cdot 10^{-2}$ | 13,394 | 30,000 | 0.156 |
| | `Sparse+Gmres` | $1.571 \cdot 10^{-5}$ | $2.502 \cdot 10^{-2}$ | 13,394 | 30,000 | 0.156 |

**Tab 4.16:** Simulation of Newton's cradle in 1.5s using different solvers for linear system.

| $R$ | Method | $\mathcal{E}_c$ | $\Delta\mathcal{E}_{c\ relative}$ | Pene. | N.C. | N.L. | CPU |
|---|---|---|---|---|---|---|---|
| $10^5$ | `Dgesv` | 40.499 | $2.226{\cdot}10^{-6}$ | $8.483{\cdot}10^{-2}$ | 344,502 | 70,000 | 163.33 |
| | `Superlu` | 40.502 | $4.079{\cdot}10^{-5}$ | $8.483{\cdot}10^{-2}$ | 344,511 | 70,000 | 26.78 |
| | `Gmres` | 40.499 | $2.289{\cdot}10^{-6}$ | $8.483{\cdot}10^{-2}$ | 344,502 | 70,000 | 22.76 |
| | `Sparse+Superlu` | 40.499 | $2.151{\cdot}10^{-6}$ | $8.483{\cdot}10^{-2}$ | 362,781 | 70,000 | 14.44 |
| | `Sparse+Gmres` | 40.499 | $2.788{\cdot}10^{-6}$ | $8.483{\cdot}10^{-2}$ | 326,781 | 70,000 | 15.26 |
| $10^6$ | `Dgesv` | 40.499 | $2.322{\cdot}10^{-5}$ | $3.937{\cdot}10^{-2}$ | 159,921 | 73,363 | 165.756 |
| | `Superlu` | 40.497 | $8.170{\cdot}10^{-5}$ | $3.939{\cdot}10^{-2}$ | 159,885 | 73,364 | 22.176 |
| | `Gmres` | 40.499 | $2.326{\cdot}10^{-5}$ | $3.937{\cdot}10^{-2}$ | 159,921 | 73,363 | 23.276 |
| | `Sparse+Superlu` | 40.499 | $1.849{\cdot}10^{-5}$ | $3.937{\cdot}10^{-2}$ | 169,921 | 70,000 | 14,89 |
| | `Sparse+Gmres` | 40.499 | $2.440{\cdot}10^{-5}$ | $3.937{\cdot}10^{-2}$ | 159,894 | 70,000 | 14.84 |

**Tab 4.17:** CPU-81 particles in a box with fixed initial conditions using different solvers for linear system (INC: $\alpha = 3$, enr $\leq 10^{-4}$ during T = 3.5s).

## 4.6.2 Friction

### 4.6.2.1 A ball moving along the x-axis

First, we present the computational time aspect of a simple example: A ball moving along the x-axis through the table 4.19.

This example is straightforward; hence, there is minimal fluctuation or variation

| | Solver | $\mathcal{E}_c$ | $\Delta\mathcal{E}_{c\ relative}$ | Pene. | N.L. | CPU |
|---|---|---|---|---|---|---|
| $\alpha = 2$ | `Dgesv` | 6.55018 | $2.60{\cdot}10^{-2}$ | $5.152{\cdot}10^{-3}$ | 37,092 | 202.072 |
| | `Superlu` | 6.57157 | $2.28{\cdot}10^{-2}$ | $5.038{\cdot}10^{-3}$ | 37,514 | 28.992 |
| | `Gmres` | 6.63137 | $1.39{\cdot}10^{-2}$ | $5.120{\cdot}10^{-3}$ | 37,775 | 28.584 |
| | `Sparse+Superlu` | 6.63130 | $1.39{\cdot}10^{-2}$ | $5.684{\cdot}10^{-3}$ | 37,575 | 15.816 |
| | `Sparse+Gmres` | 6.60192 | $1.83{\cdot}10^{-2}$ | $5.207{\cdot}10^{-3}$ | 36,999 | 15.784 |
| $\alpha = 3$ | `Dgesv` | 6.72489 | $9.29{\cdot}10^{-7}$ | $4.294{\cdot}10^{-2}$ | 390,308 | 208.304 |
| | `Superlu` | 6.69253 | $4.81{\cdot}10^{-3}$ | $4.929{\cdot}10^{-2}$ | 387,318 | 30.712 |
| | `Gmres` | 6.67396 | $7.57{\cdot}10^{-3}$ | $4.295{\cdot}10^{-2}$ | 386,931 | 29.812 |
| | `Sparse+Superlu` | 6.72489 | $1.47{\cdot}10^{-6}$ | $4.294{\cdot}10^{-2}$ | 390,306 | 16.052 |
| | `Sparse+Gmres` | 6.68559 | $5.85{\cdot}10^{-3}$ | $5.135{\cdot}10^{-2}$ | 376,254 | 16.464 |

**Tab 4.18:** 81 particles in a box with random initial conditions using different solvers for linear system (INC: $\alpha = 3$, $c_\nu = 10^5$, enr $\leq 10^{-4}$ during T = 4.4984s).

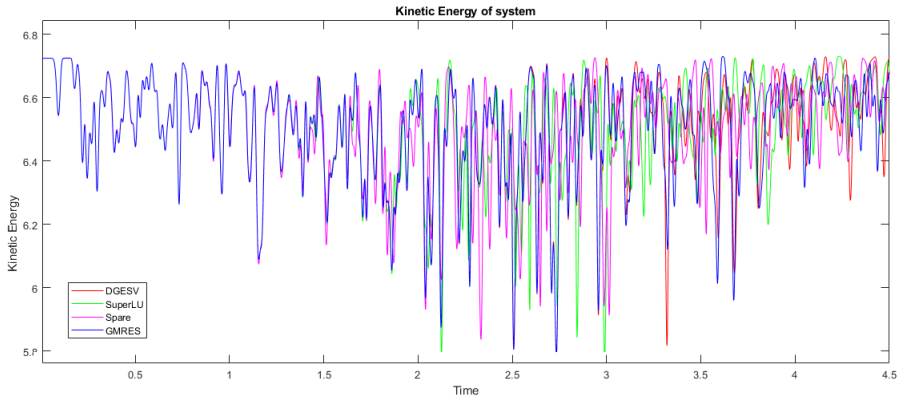



**Figure 4.30:** System's energy of 81 particles in a box with random initial conditions using different solvers for linear system (INC: $\alpha = 3$, enr $\leq 10^{-4}$, $c_\nu = 10^5$ during T = 4.4984s).

in the number of contacts, with only a slight increase as $\alpha$ increases, as well as in the number of nonlinear iterations. Specifically, the total number of nonlinear iterations across all solvers is 22,000.

#### 4.6.2.2 Four Particles from Square Corners to Center Point

Next, we investigate the behavior of a system consisting of four particles moving from square corners towards a central point. By varying parameters such as the contact and friction coefficients ($c_\nu, c_\tau$, epss) and friction coefficient ($\mu$) without gravity, we explore how these factors influence the efficiency and computational performance, measured in CPU time, of two different solvers: NSCD, INC with either Lapack or `Spare+Gmres` with $\Delta t = 10^{-4}$. The table below presents the results of our simulations, providing insights into the interplay between these parameters and the solver's performance.

We observe in the table 4.20 that the conservation of energy in the system in the case of frictionless. Conversely, this example is quite sensitive to parameter selection in case of friction. Here, we can adjust the value of epss to be small (epss $= 10^6, 10^7$),

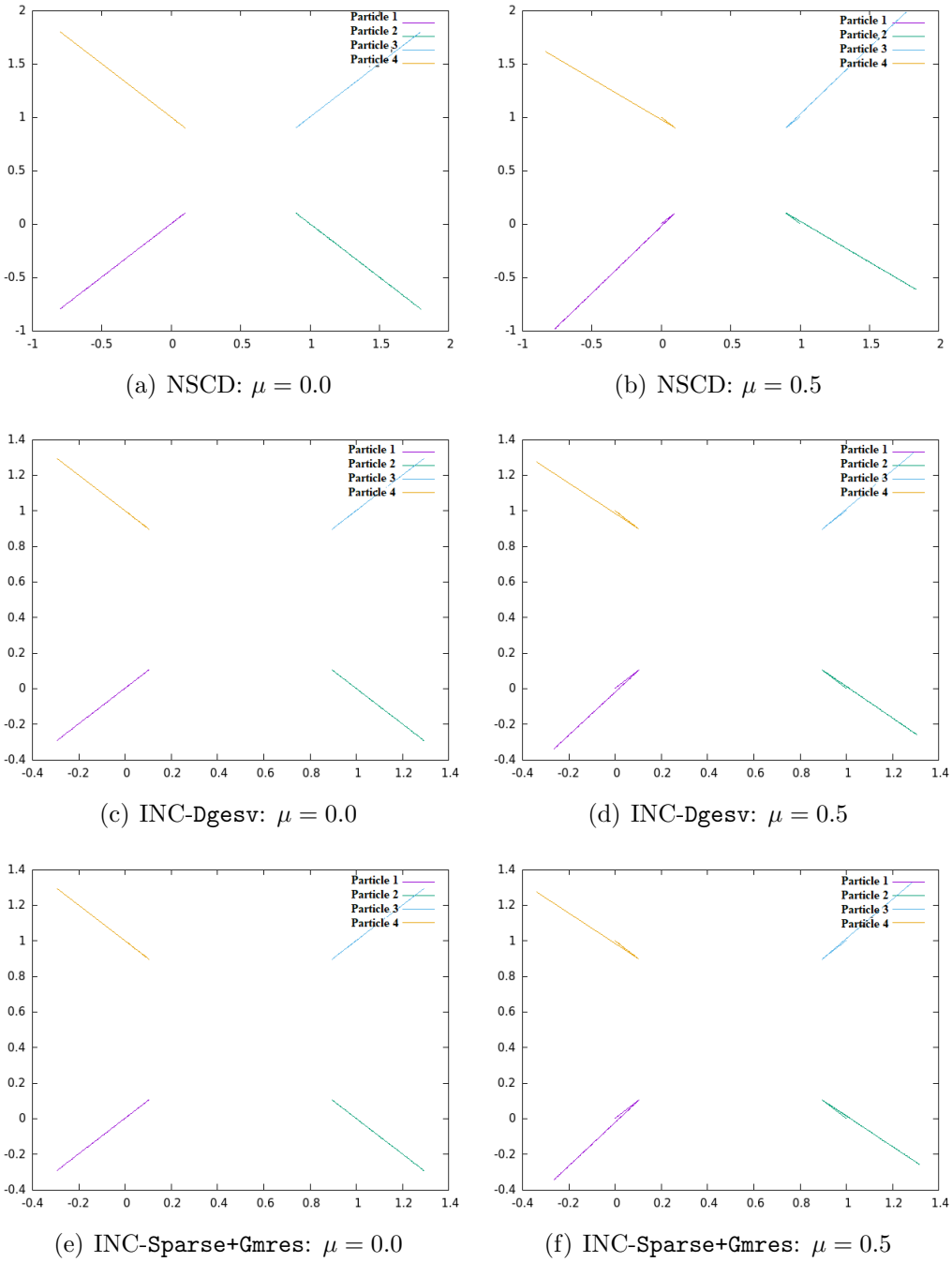
| Frictionless | $\mu = 0.0$ | CPU | N.C. |
|---|---|---|---|
| | NSCD | 0.048 | 11,002 |
| $\alpha = 2$ | `Dgesv` | 0.112 | 10,570 |
| | `Superlu` | 0.999 | 10,570 |
| | `Gmres` | 0.104 | 10,570 |
| | `Sparse+Superlu` | 0.056 | 10,570 |
| | `Sparse+Gmres` | 0.052 | 10,570 |
| | `Pcg` | 0.076 | 10,570 |
| | `Jpcg` | 0.080 | 10,570 |
| | `Sparse+Pcg` | 0.068 | 10,570 |
| | `Sparse+Jpcg` | 0.072 | 10,570 |
| $\alpha = 3$ | `Dgesv` | 0.208 | 11,223 |
| | `Superlu` | 0.152 | 11,223 |
| | `Gmres` | 0.164 | 11,223 |
| | `Sparse+Superlu` | 0.056 | 11,223 |
| | `Sparse+Gmres` | 0.056 | 11,223 |
| | `Pcg` | 0.084 | 11,223 |
| | `Jpcg` | 0.080 | 11,223 |
| | `Sparse+Pcg` | 0.076 | 11,223 |
| | `Sparse+Jpcg` | 0.72 | 11,223 |

| Friction | $\mu$ | 0.25 | | 0.5 | | 1.0 | |
|---|---|---|---|---|---|---|---|
| | | CPU | N.C. | CPU | N.C. | CPU | N.C. |
| | NSCD | 0.052 | 11,000 | 0.052 | 11,000 | 0.056 | 11,000 |
| $\alpha = 2$ | `Dgesv` | 0.120 | 10,547 | 0.128 | 10,547 | 0.124 | 10,547 |
| | `Superlu` | 0.104 | 10,547 | 0.104 | 10,547 | 0.108 | 10,547 |
| | `Gmres` | 0.108 | 10,547 | 0.108 | 10,547 | 0.108 | 10,547 |
| | `Sparse+Superlu` | 0.056 | 10,547 | 0.052 | 10,547 | 0.056 | 10,547 |
| | `Sparse+Gmres` | 0.060 | 10,547 | 0.056 | 10,547 | 0.060 | 10,547 |
| $\alpha = 3$ | `Dgesv` | 0.204 | 11,000 | 0.192 | 11,000 | 0.200 | 11,000 |
| | `Superlu` | 0.152 | 11,000 | 0.148 | 11,000 | 0.152 | 11,000 |
| | `Gmres` | 0.172 | 11,000 | 0.168 | 11,000 | 0.164 | 11,000 |
| | `Sparse+Superlu` | 0.060 | 11,000 | 0.056 | 11,000 | 0.056 | 11,000 |
| | `Sparse+Gmres` | 0.060 | 11,000 | 0.056 | 11,000 | 0.060 | 11,000 |

**Tab 4.19:** A ball moving along the x-axis using different solvers for linear system ($c_\nu = 10^6,\ c_\tau = 10^4$ during T=1.1s).

| $c_\nu, c_\tau$ | epss | $\mu = 0.0$ | | $\mu = 0.25$ | | $\mu = 0.5$ | | $\mu = 1.0$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{E}_c$ | CPU | $\mathcal{E}_c$ | CPU | $\mathcal{E}_c$ | CPU | $\mathcal{E}_c$ | CPU |
| NSCD | | 4.50000 | 0.048 | 4.25450 | 0.048 | 4.3036 | 0.052 | 4.3629 | 0.052 |
| `Dgesv` | | | | | | | | | |
| $10^5, 10^7$ | $10^{-5}$ | 4.50000 | 0.080 | 4.46410 | 0.084 | 4.48966 | 0.084 | 4.51028 | 0.080 |
| | $10^{-6}$ | 4.50000 | 0.080 | 4.42403 | 0.084 | 4.42922 | 0.088 | 4.42471 | 0.088 |
| | $10^{-7}$ | 4.49999 | 0.076 | 4.42698 | 0.080 | 4.42748 | 0.080 | 4.42478 | 0.086 |
| $10^7, 10^7$ | $10^{-5}$ | 4.50103 | 0.080 | 5.90099 | 0.084 | 5.97627 | 0.084 | 5.30436 | 0.086 |
| | $10^{-6}$ | 4.50014 | 0.076 | 4.42150 | 0.084 | 4.41822 | 0.088 | 4.42648 | 0.088 |
| | $10^{-7}$ | 4.50000 | 0.080 | 4.41915 | 0.080 | 4.41386 | 0.086 | 4.41196 | 0.088 |
| $10^7, 10^5$ | $10^{-5}$ | 4.50091 | 0.080 | 5.90114 | 0.080 | 5.97984 | 0.084 | 4.63217 | 0.084 |
| | $10^{-6}$ | 4.50001 | 0.084 | 4.41510 | 0.080 | 4.41445 | 0.086 | 4.50575 | 0.088 |
| | $10^{-7}$ | 4.50000 | 0.080 | 4.41234 | 0.084 | 4.44369 | 0.080 | 4.47444 | 0.088 |
| `Sparse+Gmres` | | | | | | | | | |
| $10^5, 10^7$ | $10^{-5}$ | 4.50000 | 0.048 | 4.46406 | 0.056 | 4.48965 | 0.056 | 4.51032 | 0.056 |
| | $10^{-6}$ | 4.50000 | 0.068 | 4.42403 | 0.052 | 4.42922 | 0.056 | 4.42471 | 0.060 |
| | $10^{-7}$ | 4.50000 | 0.068 | 4.42698 | 0.056 | 4.42748 | 0.052 | 4.42478 | 0.068 |
| $10^7, 10^7$ | $10^{-5}$ | 4.50107 | 0.048 | 5.90099 | 0.056 | 5.97625 | 0.048 | 5.30436 | 0.056 |
| | $10^{-6}$ | 4.50018 | 0.048 | 4.42150 | 0.048 | 4.41803 | 0.056 | 4.42644 | 0.052 |
| | $10^{-7}$ | 4.50004 | 0.052 | 4.41915 | 0.064 | 4.41388 | 0.060 | 4.40188 | 0.060 |
| $10^7, 10^5$ | $10^{-5}$ | 4.50107 | 0.052 | 5.90114 | 0.052 | 5.97985 | 0.052 | 4.55111 | 0.068 |
| | $10^{-6}$ | 4,50018 | 0.068 | 4.41789 | 0.068 | 4.43939 | 0.068 | 4.46983 | 0.056 |
| | $10^{-7}$ | 4.50004 | 0.052 | No convergence | | | | | |

**Tab 4.20:** Simulation of 4 particles - NSCD, INC with `Dgesv`/`Sparse+Gmres`




(a) NSCD: $\mu = 0.0$ (b) NSCD: $\mu = 0.5$

(c) INC-`Dgesv`: $\mu = 0.0$ (d) INC-`Dgesv`: $\mu = 0.5$

(e) INC-`Sparse+Gmres`: $\mu = 0.0$ (f) INC-`Sparse+Gmres`: $\mu = 0.5$

**Figure 4.31:** Particle motion of four particles - NSCD, INC with Dgsev/`Sparse+Gmres`

corresponding to a large value of $c_\tau$ ($c_\tau = 10^7$) to achieve high accuracy. This indicates that care must be taken with parameter combinations that may lead to non-convergence issues.

The figure Fig. 4.31 shows the position of each particle corresponding to $\mu = 0$ and $\mu = 0.5$. It can be observed that the direction of motion of the particles after collision is consistent in both cases.

#### 4.6.2.3 529 particles moving in a box under gravity

This test presents the computational performance of a simulation involving 529 particles confined within a box (see Tab. 4.21 for frictionless and Tab. 4.22 for friction). The input parameters are as follows: $c_\nu = 10^6$, $c_\tau = 10^{15}$, enr $< 10^{-4}$, $\Delta t = 10^{-4}$ over a duration of 7.5 seconds. The primary focus is on evaluating the efficiency of different parameter settings and solver configurations to understand their performance in large-scale simulations.

| | CPU | Pene. | N.C. | N.L. | Max.nlit |
|---|---|---|---|---|---|
| NSCD | 50 | $6.68{\cdot}10^{-3}$ | 202863 | 1720586 | 90 |
| $\alpha = 2$ | | | | | |
| `Dgesv` | 105676 | $3.27{\cdot}10^{-2}$ | 3131332 | 154593 | 3 |
| `Superlu` | 4891 | $2.79{\cdot}10^{-2}$ | 3041011 | 154426 | 3 |
| `Gmres` | 4566 | $3.27{\cdot}10^{-2}$ | 3150433 | 154378 | 3 |
| `Sparse+Superlu` | 1066 | $2.82{\cdot}10^{-2}$ | 3073025 | 154561 | 3 |
| `Sparse+Gmres` | 1031 | $3.76{\cdot}10^{-2}$ | 3072327 | 154884 | 3 |
| `Pcg` | 3754 | $2.97{\cdot}10^{-2}$ | 3021952 | 154958 | 3 |
| `Jpcg` | 3978 | $3.56{\cdot}10^{-2}$ | 3041722 | 154446 | 3 |
| `Sparse+Pcg` | 1458 | $3.12{\cdot}10^{-2}$ | 3070717 | 154760 | 3 |
| `Sparse+Jpcg` | 881 | $3.11{\cdot}10^{-2}$ | 3129424 | 154317 | 3 |
| $\alpha = 3$ | | | | | |
| `Dgesv` | 137764 | $1.47{\cdot}10^{-1}$ | 11976601 | 171650 | 3 |
| `Superlu` | 5098 | $1.44{\cdot}10^{-1}$ | 11965401 | 171731 | 3 |
| `Gmres` | 5154 | $1.52{\cdot}10^{-1}$ | 12075593 | 173718 | 3 |
| `Sparse+Superlu` | 1300 | $1.48{\cdot}10^{-1}$ | 12079145 | 172546 | 3 |
| `Sparse+Gmres` | 1264 | $1.43{\cdot}10^{-1}$ | 11999571 | 171511 | 3 |
| `Pcg` | 4153 | $1.47{\cdot}10^{-1}$ | 11980447 | 171342 | 3 |
| `Jpcg` | 4171 | $1.45{\cdot}10^{-1}$ | 12031884 | 172717 | 3 |
| `Sparse+Pcg` | 1667 | $1.45{\cdot}10^{-1}$ | 12021243 | 173692 | 3 |
| `Sparse+Jpcg` | 876 | $1.47{\cdot}10^{-1}$ | 11911307 | 172228 | 3 |

**Tab 4.21:** Simulation of 529 particles in a box during 7.5s using different solvers for linear system - frictionless.

**Remark 4.18.** *Based on the numerical experiments, the following notable observations are as follows:*

1. **Computational time (CPU):**
   - NSCD method has the lowest computation time compared to other methods, with a CPU time when $\mu = 0.0$ (frictionless case).
   - INC-`Dgesv` has the highest CPU time since the problem involving to contact with friction often leads to the poorly conditional linear system and the tangential matrices are basically non-symmetric.

| $\mu$ | | NSCD | Dgesv | Superlu | Gmres | Sparse+Superlu | Sparse+Gmres |
|---|---|---|---|---|---|---|---|
| | | | $\alpha = 2$ | | | | |
| 0.25 | CPU | 1032 | 107062 | 4654 | 4475 | 1147 | 1008 |
| | Pene. | $3.13{\cdot}10^{-3}$ | $1.36{\cdot}10^{-2}$ | $1.61{\cdot}10^{-2}$ | $1.27{\cdot}10^{-2}$ | $1.47{\cdot}10^{-2}$ | $1.46{\cdot}10^{-2}$ |
| | N.C. | 27629423 | 54992692 | 54879707 | 56293411 | 54004181 | 53790852 |
| | N.L. | 35268922 | 150002 | 150000 | 150000 | 150000 | 150002 |
| | Max.nlit | 2407 | 3 | 2 | 2 | 2 | 3 |
| 0.5 | CPU | 1033 | | 5160 | 4469 | 1192 | 1032 |
| | Pene. | $3.02{\cdot}10^{-3}$ | | $1.27{\cdot}10^{-2}$ | $1.25{\cdot}10^{-2}$ | $1.46{\cdot}10^{-2}$ | $1.20{\cdot}10^{-2}$ |
| | N.C. | 23362775 | | 46945641 | 52071251 | 45232578 | 45297339 |
| | N.L. | 44729043 | | 150156 | 150088 | 150301 | 150085 |
| | Max.nlit | 3402 | | 4 | 4 | 4 | 4 |
| 1.0 | CPU | 1175 | | 5070 | 4585 | 1193 | 1030 |
| | Pene. | $3.55{\cdot}10^{-3}$ | | $1.27{\cdot}10^{-2}$ | $1.53{\cdot}10^{-2}$ | 1.23-2 | $1.19{\cdot}10^{-2}$ |
| | N.C. | 17754274 | | 45998717 | 46747000 | 43673608 | 43307010 |
| | N.L. | 47322241 | | 154861 | 155284 | 155237 | 155421 |
| | Max.nlit | 105688 | | 5 | 5 | 5 | 5 |
| | | | $\alpha = 3$ | | | | |
| 0.25 | CPU | | 146888 | 4751 | 4502 | 1161 | 1003 |
| | Pene. | | $8.62{\cdot}10^{-2}$ | 0.10 | $9.81{\cdot}10^{-2}$ | $9.22{\cdot}10^{-2}$ | $9.74{\cdot}10^{-2}$ |
| | N.C. | | 69815300 | 47873079 | 48985794 | 48766559 | 48495958 |
| | N.L. | | 150000 | 150000 | 150000 | 150002 | 150000 |
| | Max.nlit | | 3 | 2 | 2 | 2 | 2 |
| 0.5 | CPU | | 147576 | 4496 | 4511 | 985 | 996 |
| | Pene. | | $8.87{\cdot}10^{-2}$ | 0.13 | 0.15 | 0.11 | 0.13 |
| | N.C. | | 66446916 | 21414022 | 21172789 | 21728441 | 21913137 |
| | N.L. | | 223796 | 151058 | 151172 | 151179 | 151050 |
| | Max.nlit | | 4 | 4 | 4 | 4 | 4 |
| 1.0 | CPU | | | 5838 | 5820 | 1270 | 1256 |
| | Pene. | | | 0.16 | 0.15 | 0.16 | 0.17 |
| | N.C. | | | 16988046 | 16934803 | 16415800 | 16835946 |
| | N.L. | | | 198779 | 198395 | 198128 | 198812 |
| | Max.nlit | | | 6 | 5 | 5 | 6 |

**Tab 4.22:** Simulation for 529 particles in a box during 7.5s using different solvers for linear system - friction.

- The combined solvers (`Sparse+Superlu` and `Sparse+Gmres`) not only have lower CPU times compared to individual solvers like `Superlu` or `Gmres` but also approximate the CPU time of the NSCD method for the friction case, making them efficient and competitive alternatives.
- When $\alpha$ increases from 2 to 3, the CPU time for all methods rises. INC methods in the form of a `Sparse` matrix (such as `Sparse+Superlu`, `Sparse+Gmres`, `Sparse+Pcg`, `Sparse+Jpcg`) have shorter CPU times compared to their original matrix form, with INC combined with `Sparse+Jpcg` solver being the fastest in both cases.

2. **Number of Contacts (N.C.):**

   - NSCD method has the fewest contacts when $\mu = 0.0$, but the number of contacts increases significantly as $\mu$ increases. In contrast, the contact counts for INC methods with all solvers are generally higher than those for the NSCD method since INC is the displacement contact problem taking into account each contact interaction more comprehensively. This detailed consideration of individual contacts results in a higher overall contact count for INC methods, reflecting a more precise and granular approach to modeling the interactions between particles.

3. **Number of Nonlinear iterations (N.L.):**

   - For INC methods, the maximum number of nonlinear iterations per step is typically 5 or 6, indicating greater stability compared to the NSCD method with NLGS solvers. As a result, the NSCD method tends to have the highest number of nonlinear iterations, particularly when $\mu = 1.0$.

#### 4.6.2.4 Sedimentation of several particles

We extend our model to a larger system with 1225 particles. This subsection explores the computational challenges and performance metrics associated with simulating a more complex and densely populated particle system. The input parameters are as follows: $c_\nu = 10^8$, $c_\tau = 10^{15}$, enr $< 10^{-5}$, $\Delta t = 10^{-4}$ over a duration of 7.5 seconds.

A visualization of the simulation of 1,225 particles in a box using the INC method with the `Sparse+Gmres` version for the linear system is provided in Figure 4.32 and Figure A, illustrating the interaction forces between the particles in the system.

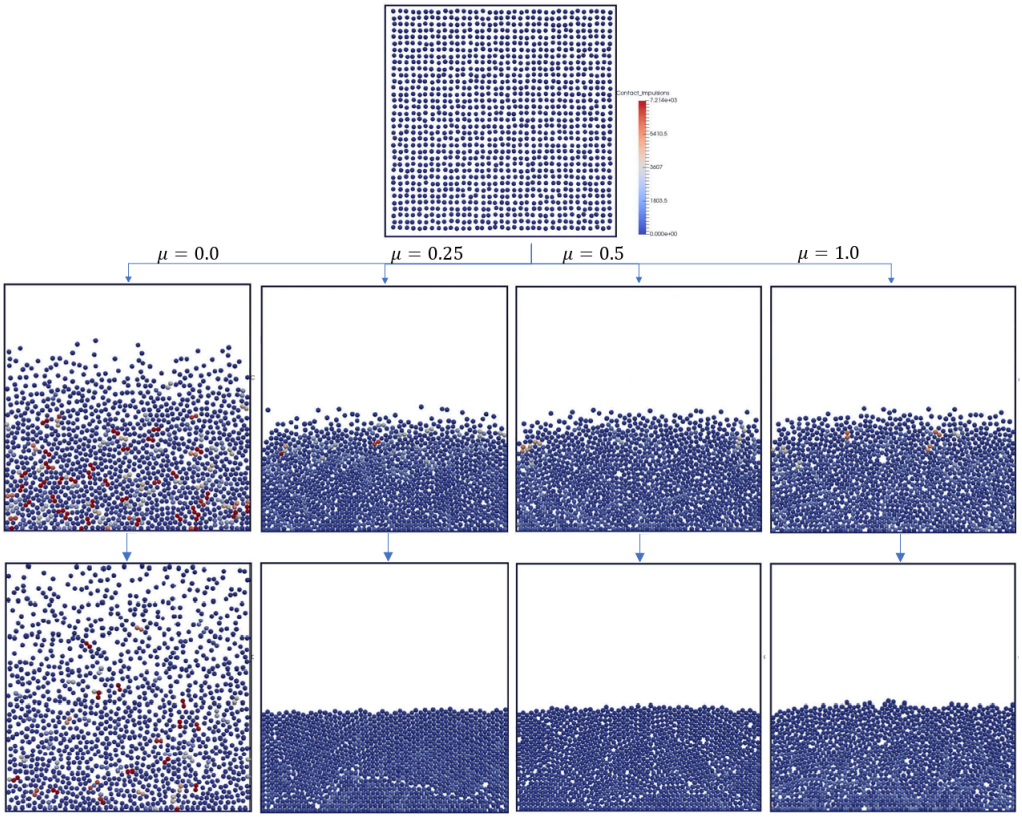
**Remark 4.19.** In this simulation, we configure the maximum number of nonlinear iterations (Max.nlit) to be 100 for the INC method and $2 \cdot 10^4$ for NSCD, ensuring optimal compatibility with each method's specific requirements. However, with NSCD, in the presence of friction with $\mu = 1.0$, we can observe from Table 4.24 that it reaches the maximum number of iterations, indicating a lack of convergence in the algorithm (even with Max.nlit= $2 \cdot 10^6$). In contrast, INC particularly with the `Sparse+Gmres` solver, performs better and achieves convergence but INC-Dgesv (friction case).

## 4.7 Perspectives

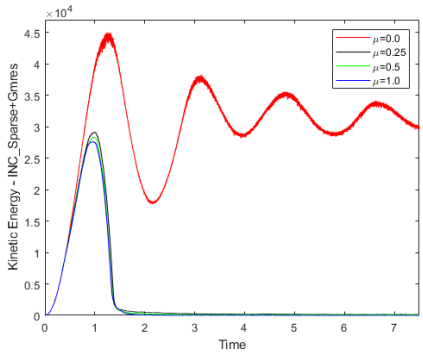
In this section, we explore several perspectives for advancing the methodologies and applications discussed earlier. These perspectives aim to enhance algorithmic and computational efficiency, explore new application domains, and rigorously establish theoretical foundations.

1. *Algorithmic and Computational Parallelism*

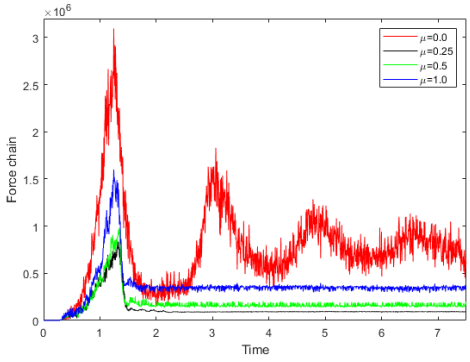
   - Algorithmic Parallelism: One primary objective of this work is to explore algorithmic parallelism, specifically through the Domain Decomposition Method



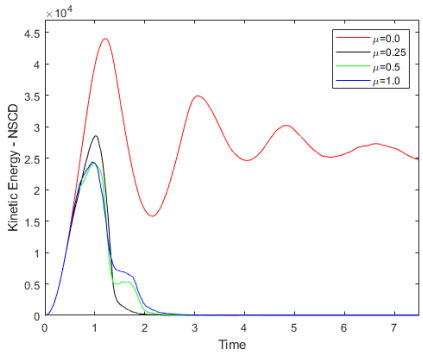

(a) Visualization-INC: $\alpha = 3$




(b) Kinetic Energy-INC: $\alpha = 3$




(c) Force chain-INC: $\alpha = 3$




(d) Kinetic Energy-NSCD

**Figure 4.32:** Visualization, system's energy and force chain of simulation of 1225 particles in a box - INC_`Sparse+Gmres`.

| $\alpha = 3$ | CPU | $\mathcal{E}_c$ | Pene. | N.L. | Max.nlit |
|---|---|---|---|---|---|
| NSCD | 187 | 25311.8 | $1.14 \cdot 10^{-2}$ | 2221418 | 106 |
| `Dgesv` | $2.7 \cdot 10^{6}$ | 29581.3 | $5.03 \cdot 10^{-2}$ | 291379 | 4 |
| `Superlu` | 67945 | 30449.7 | $5.02 \cdot 10^{-2}$ | 291232 | 4 |
| `Gmres` | 61962 | 30024.0 | $5.63 \cdot 10^{-2}$ | 291272 | 4 |
| `Sparse+Superlu` | 7141 | 31410.3 | $4.60 \cdot 10^{-2}$ | 219972 | 3 |
| `Sparse+Gmres` | 7265 | 30234.0 | $5.26 \cdot 10^{-2}$ | 291257 | 4 |
| `Pcg` | 28941 | 31004.6 | $4.99 \cdot 10^{-2}$ | 293099 | 4 |
| `Jpcg` | 27943 | 31370.5 | $4.72 \cdot 10^{-2}$ | 291276 | 4 |
| `Sparse+Pcg` | 10435 | 30814.2 | $4.81 \cdot 10^{-2}$ | 291235 | 4 |
| `Sparse+Jpcg` | 6328 | 31597.5 | $4.93 \cdot 10^{-2}$ | 292193 | 4 |

**Tab 4.23:** Simulation for 1225 particles in a box during 7.5s with different linear solvers - frictionless.

| $\mu$ | | NSCD | `Superlu` | `Gmres` | `Sparse+Superlu` | `Spare+Gmres` |
|---|---|---|---|---|---|---|
| 0.25 | CPU | 5556 | 84194 | 86110 | 12570 | 10595 |
| | $\mathcal{E}_c$ | 0.33 | 0.72 | 0.72 | 1.81 | 0.61 |
| | Pene. | $3.85 \cdot 10^{-3}$ | $2.70 \cdot 10^{-2}$ | $2.70 \cdot 10^{-2}$ | $2.57 \cdot 10^{-2}$ | $2.54 \cdot 10^{-2}$ |
| | N.L. | 92448520 | 366400 | 366400 | 367947 | 367623 |
| | Max.nlit | 11708 | 9 | 9 | 9 | 8 |
| 0.5 | CPU | 6150 | 96423 | 90724 | 14102 | 12699 |
| | $\mathcal{E}_c$ | 1.23 | 2.28 | 3.13 | 0.60 | 1.03 |
| | Pene. | $3.37 \cdot 10^{-3}$ | $2.45 \cdot 10^{-2}$ | $2.48 \cdot 10^{-2}$ | $2.39 \cdot 10^{-2}$ | $2.35 \cdot 10^{-2}$ |
| | N.L. | 110217192 | 419192 | 417389 | 418593 | 421344 |
| | Max.nlit | 10738 | 11 | 12 | 12 | 11 |
| 1.0 | CPU | 5862 | 100686 | 101037 | 15599 | 14776 |
| | $\mathcal{E}_c$ | 1.77 | 4.08 | 23.75 | 3.54 | 3.84 |
| | Pene. | $4.28{\cdot}10^{-3}$ | $2.38 \cdot 10^{-2}$ | $2.33 \cdot 10^{-2}$ | $2.36 \cdot 10^{-2}$ | $2.37 \cdot 10^{-2}$ |
| | N.L | 105161949 | 499773 | 496547 | 503009 | 501563 |
| | Max.nlit | 20001 | 15 | 14 | 15 | 15 |

**Tab 4.24:** Simulation for 1225 particles in a box during 7.5s with different linear solvers - friction.

(DDM) with an additive Schwarz method integrating coarse modes. This approach would allow the problem to be divided into smaller, independent subdomains, facilitating the parallel resolution of linear system. Particular attention will be paid to optimizing algorithms to enhance convergence and computational efficiency.

- Computational Parallelism: Simultaneously, computational parallelism will be explored to optimize matrix-vector product operations and other intensive calculations. By leveraging modern parallel computing architectures (such as GPUs and computing clusters), it would be possible to significantly accelerate simulations, making the Newmark-PDAS-INC approach more applicable to large-scale problems.

The combination of algorithmic and computational parallelism is essential to make numerical methods more efficient and applicable to large problems. By detailing the specific advantages of domain decomposition methods and modern parallel architectures, a clear direction for future work is outlined.

2. *Potential Applications*

   - Fluid-Structure Interaction: A second perspective involves applying the developed methodology to fluid-structure interaction problems, such as fluidized beds and crowd movements. These applications will validate the relevance and robustness of the Newmark-PDAS-INC approach in various contexts. For example, in fluidized beds, the method could help model the complex interactions between solid particles and the surrounding fluid.
   - Validation and Extension: By validating the method in these new contexts, not only will its versatility be demonstrated, but potential improvements and extensions to better capture the specific dynamics of each application will also be identified.

   Identifying concrete applications like fluidized beds and crowd movements grounds the theoretical work in practical and relevant contexts. This reinforces the importance of the developed method and opens opportunities for interdisciplinary collaborations.

3. *Existence Theory for the Second Order Moreau Sweeping Process*

   - Existence of the Solution: Finally, an essential theoretical perspective is to rigorously establish the existence of solutions for the second-order Moreau sweeping process, used to model contact dynamics with Moreau-Yosida regularization. This theoretical study will ensure the mathematical robustness of the model and reinforce confidence in the numerical results obtained.
   - Moreau-Yosida Regularization: By deepening the understanding of Moreau-Yosida regularization, it would be possible to better characterize the conditions under which the model remains valid and stable, thereby paving the way for new applications and methodological improvements.

   The theoretical validation of the existence of solutions is crucial to ensure the scientific rigor of the model. By focusing on Moreau-Yosida regularization, a key technical aspect that can have profound implications for the stability and accuracy of simulations is highlighted.

# Part II

# Weak and strong convexity

# Chapter 5

# Farthest distance function to strongly convex sets in Hilbert spaces

Throughout this chapter, $X$ stands for a (real) Hilbert space not reduced to $0_X$ endowed with the inner product $\langle \cdot, \cdot \rangle$ and its associated norm $\sqrt{\|\cdot\|}$. The distance function to a convex set is known to admit diverse descriptions involving duality relations from convex analysis. Namely, given a closed convex set $C$ in $X$, the distance function $d_C$ has been described in the following ways (see, e.g., [102]) :

(i) in terms of the support function $\sigma(\cdot, C)$: for an appropriate $x^\star \in X$

$$d_C(x) = \langle x^\star, x \rangle - \sigma(x^\star, C);$$

(ii) by the duality property:

$$d_C(x) = \max_{\|\bar{x}\|=1} \inf_{y \in C} \langle \bar{x}, x - y \rangle = \inf_{y \in C} \max_{\|\bar{x}\|=1} \langle \bar{x}, x - y \rangle ;$$

(iii) in terms of supporting hyperplanes: the distance of a point to a convex set is the maximum of distances to appropriate hyperplanes separating the set and the point.

Extensions of (i)-(ii) and (iii) to prox-regular sets have been recently considered in [84].

The chapter is organized as follows. In Section 5.1, we recall some definitions and properties of strong convex and prox-regular sets. Section 5.2 is devoted to the semiconcavity property of the farthest distance function associated to a strongly convex set. In Section 5.3 of the present work, we develop the analogs of the above properties (i)-(ii) and (iii) in Section 1.4.1 for a strongly convex set $C$.

## 5.1 Strongly convex and prox-regular sets

This paragraph is devoted to the needed preliminaries on strong convexity and prox-regularity of sets in Hilbert spaces. We start with the definition of a strongly

convex set. For more details on such a class of sets, we refer, e.g., to the book by E. S. Polovinkin and M. V. Balashov [119], to the survey by G. E. Ivanov and V. V. Goncharov ([105]) and the references therein.

**Definition 5.1.** Let $C$ be a nonempty subset in $X$. One says that $C$ is $R$-strongly convex for some real $R > 0$ whenever there is a nonempty set $L \subset X$ such that

$$C = \bigcap_{x \in L} B[x, R].$$

It is clear that every $R$-strongly convex set is closed with diameter less or equal than $2R$. Strongly convex sets can be characterized through the farthest distance function. Before giving next Theorem 5.2 (whose arguments can be found, e.g., in [90, 105, 126, 133] (see also the references therein)), let us set

$$\mathcal{E}_\rho(C) := \{x \in X : \mathrm{dfar}_C(x) > \rho\}$$

for any nonempty subset $C \subset X$ and any real $\rho > 0$.

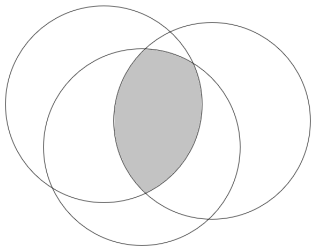


**Figure 5.1:** Illustration the strongly convex set.

**Theorem 5.2.** *Let $C$ be a nonempty closed convex bounded subset not reduced to a singleton in $X$ and let $R > 0$ be a real. The following assertions are equivalent:*
*(a) the set $C$ is $R$-strongly convex;*
*(b) for all $x, x' \in C$ and for all $v \in N(C; x)$, one has*

$$\langle v, x' - x\rangle \le -\frac{\|v\|}{2R}\,\|x' - x\|^2;$$

*(c) the mapping $\mathrm{far}_C$ is well defined on $\mathcal{E}_R(C)$ and for every real $s > R$, for all $x, x' \in \mathcal{E}_s(C)$,*

$$\|\mathrm{far}_C(x) - \mathrm{far}_C(x')\| \le \frac{1}{(s/R) - 1}\|x - x'\|;$$

*(d) the function $\mathrm{dfar}^2_C(\cdot)$ is differentiable on $\mathcal{E}_R(C)$ with a locally Lipschitz derivative and*

$$\nabla \mathrm{dfar}^2_C(x) = 2\big(x - \mathrm{far}_C(x)\big) \quad \textit{for all}\, x \in \mathcal{E}_R(C);$$

*(e) for any $u \in \mathcal{E}_R(C)$ such that $\mathrm{far}_C(u) =: x$ is well defined, one has*

$$x = \mathrm{far}_C\big(x - \frac{t}{\mathrm{dfar}_C(u)}(x - u)\big) \quad \textit{for all}\, t \in ]R, +\infty[.$$

**Remark 5.3.** *Let $C$ be an $R$-strongly convex subset of $X$ for some real $R > 0$. The assertions (c) and (e) in Theorem 5.2 above guarantee that for any $u \in \mathcal{E}_R(C)$ the vector $x := \mathrm{far}_C(u)$ is well defined along with*

$$x \in \mathrm{Far}_C\big(x - \frac{R}{\mathrm{dfar}_C(u)}(x-u)\big).$$

We now develop the concept of prox-regularity (see [118]). Historical comments and applications can be found, e.g., in the survey [100].

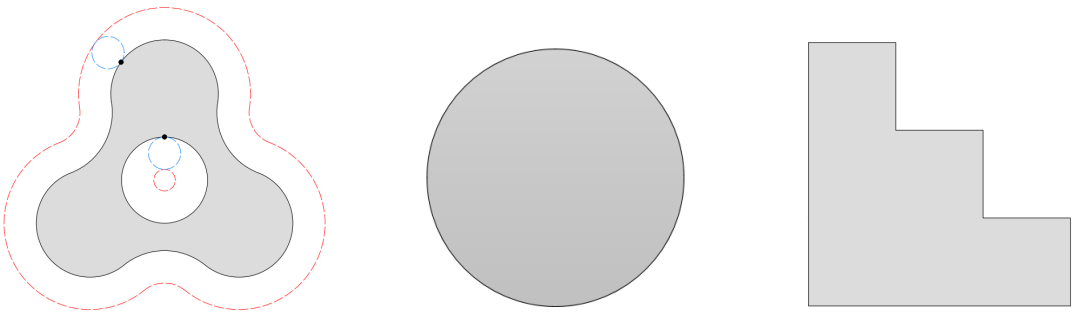
**Definition 5.4.** Let $S$ be a nonempty closed subset of $X$ and $r \in ]0,+\infty]$. One says that $S$ is *$r$-prox-regular* (or uniformly prox-regular with constant $r$) whenever, for all $x \in S$, for all $v \in N^P(S;x) \cap \mathbb{B}$ and for every real $t \in ]0,r]$, one has $x \in \mathrm{Proj}_S(x+tv)$.

A crucial class of uniformly prox-regular sets is given by the complements of open balls. More precisely, given any $x \in X$ and any real $r > 0$, it is known (and not difficult to check) that the set $S := X \setminus B(x,r)$ is $r$-prox-regular. It will be convenient for us to use as definition of $r$-prox-regularity of sets ([118]) the following property.

**Definition 5.5.** A nonempty closed set $S$ of a Hilbert space $\mathcal{H}$ is said to be *$r$-prox-regular* (or prox-regular with constant/thickness $r$) for some $r \in ]0,+\infty]$ if

$$\langle v, x' - x\rangle \leq \frac{\|v\|}{2r}\|x'-x\|^2, \tag{5.1}$$

for all $x, x' \in S$, for all $v \in N^C(S;x)$ (or $N^F(S;x)$ (or $N^L(S;x)$).



(a) Prox-regular set (b) Nonempty closed convex ($\infty$-prox-regular) (c) Lack of prox-regularity

**Figure 5.2:** Illustration the prox-regular and non prox-regular set.

The next theorem provides some useful characterizations and properties of uniform prox-regular sets. For its proof, we refer, e.g., to [100, 118]. Given a subset $S \subset X$ and an extended real $r > 0$, we recall the notation in the introduction

$$U_r(S) := \{x \in X : d_S(x) < r\}.$$

**Theorem 5.6.** *Let $S$ be a nonempty closed subset of $X$ and let $r \in ]0,+\infty]$. The following assertions are equivalent:*

*(a) the set $S$ is $r$-prox-regular;*
*(b) for all $x, x' \in S$, for all $v \in N^P(S;x)$, one has*

$$\langle v, x' - x\rangle \leq \frac{1}{2r}\,\|v\|\,\|x - x'\|^2\,;$$

*(c) the mapping $\mathrm{proj}_S(\cdot)$ is well defined on $U_r(S)$, and for every real $s \in ]0, r[$, for all $x, x' \in U_s(S)$,*

$$\|\mathrm{proj}_S(x) - \mathrm{proj}_S(x')\| \leq \frac{1}{1-(s/r)}\,\|x - x'\|\,;$$

*(d) the function $d_S^2(\cdot)$ is differentiable on $U_r(S)$ with a locally Lipschitz derivative and*

$$\nabla d_S^2(x) = 2\big(x - \mathrm{proj}_S(x)\big) \quad \textit{for all}\, x \in U_r(S);$$

*(e) for any $u \in U_r(S) \setminus S$ such that $\mathrm{proj}_S(u) =: x$ is well defined, one has*

$$x = \mathrm{proj}_S\big(x + \frac{t}{d_S(u)}(u - x)\big) \quad \textit{for all}\, t \in [0, r[.$$

We end this part with two results on prox-regular sets which will be at the heart of the development below. The first one is concerned with complements of prox-regular sets.

**Theorem 5.7** ([115])**.** *Let $S$ be an $r$-prox-regular subset of $X$ with $r \in ]0, +\infty[$. Then, for all $s \in ]0, r[$, the set $X \setminus S$ is the union of a family of closed balls of $X$ of radius $s$.*

The second result by G. E. Ivanov [108, Theorem 1.12.3] gives a sufficient condition ensuring the prox-regularity for the Minkowski sum.

**Theorem 5.8** ([108])**.** *Let $C$ be an $R$-strongly convex subset of $X$ with $R \in ]0, +\infty]$ and let $S$ be an $r$-prox-regular subset of $X$ with $r \in ]0, +\infty]$ such that $0 < R < r$. Then, the set $C + S$ is $(r - R)$-prox-regular, in particular closed.*

## 5.2 Semiconcavity of the farthest distance function

This section is devoted to the study of the semiconcavity property of the farthest distance function. Given a real $\sigma \geq 0$, a function $f : U \to \mathbb{R} \cup \{+\infty\}$ defined on a (not necessarily open) nonempty convex subset $U$ of $X$ is said to be linearly *$\sigma$-semiconvex* (or semiconvex with constant $\sigma$) on $U$ provided that for every $t \in ]0, 1[$ and every $x, y \in U$

$$f\big(tx + (1-t)y\big) \leq tf(x) + (1-t)f(y) + \frac{\sigma}{2}t(1-t)\,\|x - y\|^2\,.$$

It is well known (and not difficult to check) that the function $f$ is $\sigma$-semiconvex on $U$ for some $\sigma \geq 0$ if and only if the function $f(\cdot) + \sigma/2\,\|\cdot\|^2$ is convex on $U$. The function $f$ is said to be *locally semiconvex* if $f$ is semiconvex on a neighborhood of each point of $U$. If $-f$ is $\sigma$-semiconvex on $U$ for some real $\sigma \geq 0$, then $f$ is said to be *$\sigma$-semiconcave* on $U$. Similarly, $f$ is said to be *locally semiconcave* if $f$ is semiconcave on a neighborhood of each point in the set $U$.

Proposition 18 in [100] proved that a closed set $S$ of $X$ is $r$-prox-regular if and only if, for any positive real $s < r$, the square distance function $d_S^2$ is $s/(r-s)$-semiconvex on any convex set included in $U_s(S)$. Regarding the distance function $d_S$ itself, it has been shown by M. V. Balashov [88] that the distance $d_S$ associated to an $r$-prox-regular set $S$ is $(r-s)^{-1}$-semiconvex on any convex set included in the open $s$-enlargement $U_s(S)$. Another different proof of this property has been recently established by F. Nacry and L. Thibault ([115]) through Theorem 5.7 and Lemma 5.9. The arguments for Lemma 5.9 are contained in [95, Proposition 2.2.2] and [115, Proposition 7].

**Lemma 5.9** ([95]). *Let $S$ be a nonempty subset of $X$. The following hold.*
*(a) The squared distance function $d_S^2$ is linearly semiconcave with coefficient 2.*
*(b) For any nonempty convex subset $U$ of $X$ and for any real $\delta > 0$ such that $U \cap \big(S + B(0,\delta)\big) = \emptyset$, the distance function $d_S$ is $\delta^{-1}$-semiconcave on $U$. So, $d_S$ is locally linearly semiconcave on $X \setminus S$.*
*(c) If $S$ is the union of a collection of closed balls with a common radius $r > 0$, then on each nonempty convex set $U$ included in $\mathrm{cl}(X \setminus S)$, the distance function $d_S$ is $r^{-1}$-semiconcave.*

Our aim in this section is to show that such a lemma can also be successfully used to establish the semiconcavity of the farthest distance function associated to a strongly convex set. As said in the very introduction, our approach and proofs are quite different from those involved in the work [89] by M. V. Balashov and M. O. Golubev; notice that some results in certain Banach spaces are also contained in [89]. We start with the following result which has its own interest and where for a real $r > 0$ and a function $\varphi : X \to \mathbb{R} \cup \{+\infty\}$, we use notation $\{\varphi \leq r\} := \{x \in X : \varphi(x) \leq r\}$ along with analogous one for $\{\varphi < r\}$, etc.

**Proposition 5.10.** *Let $C$ be a nonempty bounded subset of $X$ and let $R > 0$ be a positive real. The following hold.*
*(a) The equalities*

$$\{\mathrm{dfar}_C \leq R\} = \bigcap_{c \in C} B[c,R] \quad \textit{and} \quad \{\mathrm{dfar}_C < R\} = \mathrm{int}\left(\bigcap_{c \in C} B[c,R]\right),$$

*are satisfied, so in particular the sublevel set $\{\mathrm{dfar}_C \leq R\}$ is $R$-strongly convex.*
*(b) One has*

$$\{\mathrm{dfar}_C > R\} = (X \setminus R\mathbb{B}) + C.$$

*(c) One has*

$$\mathrm{cl}\,\{\mathrm{dfar}_C > R\} = \{\mathrm{dfar}_C \geq R\},$$

*or equivalently,*

$$\mathrm{int}\{\mathrm{dfar}_C \leq R\} = \{\mathrm{dfar}_C < R\}.$$

*If in addition the set $C$ is $R$-strongly convex, then the following assertions also hold true:*
*(d) For every real $s \geq R$, one has with $F_s := \{\mathrm{dfar}_C \leq s\}$*

$$\mathrm{dfar}_C(u) = s + d_{F_s}(u) \quad \textit{for all } u \in X \setminus F_s = \mathcal{E}_s(C).$$

*(e) For every real $s > R$, one has*

$$\{\mathrm{dfar}_C > s\} = \mathrm{int}\,\{\mathrm{dfar}_C \geq s\} \quad \textit{and} \quad \{\mathrm{dfar}_C \leq s\} = \mathrm{cl}\,\{\mathrm{dfar}_C < s\}\,.$$

*(f) For every real $s > R$, one has*

$$\{\mathrm{dfar}_C \geq s\} = (X \setminus s\mathbb{U}) + C,$$

*in particular the set $\{\mathrm{dfar}_C \geq s\}$ is $(s-R)$-prox-regular.*

*Proof.* (a) The first equality being obvious, we only show the second one. Consider first the inclusion of the left-hand side into the right-hand one. Set $K := \bigcap_{c\in C} B[c,R]$. Let any $x \in X$ with $\mathrm{dfar}_C(x) < R$. Choose some real $\eta > 0$ such that $\mathrm{dfar}_C(x)+\eta < R$. For all $c \in C$ and all $b \in \mathbb{B}$, it is readily seen that

$$\|c - (x+\eta b)\| \leq \|c-x\| + \eta\|b\| \leq \mathrm{dfar}_C(x) + \eta < R,$$

hence $x + \eta\mathbb{B} \subset K$. This means that $\{\mathrm{dfar}_C < R\} \subset \mathrm{int}\,K$.

Conversely, let $z \in \mathrm{int}\,K$. Choose a real $\delta \in ]0, R[$ such that $z + \delta\mathbb{B} \subset B[c,R]$ for every $c \in C$. Then for each $c \in C$ we have

$$z + \delta\mathbb{B} \subset c + (R-\delta)\mathbb{B} + \delta\mathbb{B},$$

hence by the Rådström cancellation lemma (see, e.g., [124, Property (1.22)]) $z \in c + (R-\delta)\mathbb{B}$, that is, $\|z-c\| \leq R-\delta$. It ensues that $\mathrm{dfar}_C(z) \leq R-\delta < R$, so $\mathrm{int}\,K \subset \{\mathrm{dfar}_C < R\}$ This finishes the proof of (a).

(b) The inclusion $\supset$ is obvious while the converse one $\subset$ follows from the fact that for any $u \in X$ with $\mathrm{dfar}_C(u) > R$, we can find some $c \in C$ with $\|u-c\| > R$, hence $u = (u-c) + c \in (X \setminus R\mathbb{B}) + C$.

(c) The second claimed equality directly follows from the first one by taking complements. Let us then focus on the first desired equality. The inclusion $\subset$ being obvious by continuity of $\mathrm{dfar}_C$, we only show the converse one $\supset$. We may suppose that $C$ is not a singleton. Fix any $x \in X$ with $\mathrm{dfar}_C(x) = R$. Take any real $\varepsilon > 0$. Using the definition of $\mathrm{dfar}_C(x)$, we can find a sequence $(c_n)_{n\geq 1}$ in $C$ such that

$$R_n := \|c_n - x\| \to \mathrm{dfar}_C(x) = R.$$

Since $C$ is not a singleton, we have $R_n > 0$ for $n$ large enough, and this allows us to choose some integer $N \geq 1$ such that

$$a := \frac{R}{R_N} - 1 < \frac{\varepsilon}{R_N} =: b.$$

Pick any $t \in ]a, b[$ and observe that

$$\mathrm{dfar}_C\big(-tc_N + (1+t)x\big) \geq (1+t)\|x - c_N\| = (1+t)R_N > R$$

along with

$$\|-tc_N + (1+t)x - x\| = t\|x - c_N\| = tR_N < \varepsilon.$$

Consequently, we have $B(x,\varepsilon) \cap \{\mathrm{dfar}_C > R\} \neq \emptyset$ and this guarantees the inclusion $x \in \mathrm{cl}\{\mathrm{dfar}_C > R\}$. This confirms the first equality in (c).

Assume now that $C$ is $R$-strongly convex for a given real $R > 0$.

(d) Consider any real $s \geq R$. Fix any $u \in X$ with $\mathrm{dfar}_C(u) > s$. Thanks to Theorem 5.2(c), we know that $z := \mathrm{far}_C(u)$ is well defined. Setting $v := z - s\frac{z-u}{\|z-u\|}$, we also have (see Theorem 5.2(e)) that $\mathrm{far}_C(v)$ is well defined along with $z = \mathrm{far}_C(v)$. Hence, we see that

$$\mathrm{dfar}_C(v) = \|\mathrm{far}_C(v) - v\| = \|z - v\| = s,$$

in particular $v \in F_s := \{\mathrm{dfar}_C \leq s\}$. This and a direct computation give

$$d_{F_s}(u) \leq \|u - v\| = \|(u - z)\big(1 - \frac{s}{\|z-u\|}\big)\| = \mathrm{dfar}_C(u) - s. \tag{5.2}$$

On the other hand, we observe that

$$\mathrm{dfar}_C(u) \leq \mathrm{dfar}_C(x) + \|u - x\| \leq s + \|u - x\| \quad \text{for all } x \in F_s,$$

and this obviously entails

$$\mathrm{dfar}_C(u) \leq s + d_{F_s}(u). \tag{5.3}$$

It suffices then to put together (5.2) and (5.3) to obtain the desired equality in (d).

(e) Fix any real $s > R$. Observe that the second claimed equality directly follows from the first one by taking complements. Further, note that $\mathcal{C} := \mathrm{cl}\,\{\mathrm{dfar}_C < s\}$ is obviously included (keeping in mind the continuity of $\mathrm{dfar}_C$) in $\{\mathrm{dfar}_C \leq s\}$. Then, it remains to justify the inclusion $\{\mathrm{dfar}_C \leq s\} \subset \mathcal{C}$. Fix any $u \in X$ with $\mathrm{dfar}_C(u) \leq s$. We may suppose that $\mathrm{dfar}_C(u) = s$ (otherwise there is nothing to prove). Since $C$ is $R$-strongly convex and $s > R$, we know (see Theorem 5.2(c)) that $\mathrm{far}_C(u)$ is well defined. Take any sequence $(s_n)_{n\geq 1}$ of $]R, s[$ with $s_n \to s$. The latter convergence and the equality $\mathrm{dfar}_C(u) = s$ obviously give

$$v_n := \mathrm{far}_C(u) - \frac{s_n}{\mathrm{dfar}_C(u)}\big(\mathrm{far}_C(u) - u\big) \to u.$$

On the other hand, Theorem 5.2(e) guarantees that for every integer $n \geq 1$, the vector $\mathrm{far}_C(v_n)$ is well defined along with $\mathrm{far}_C(u) = \mathrm{far}_C(v_n)$. Hence, we can write for every integer $n \geq 1$

$$\mathrm{dfar}_C(v_n) = \|\mathrm{far}_C(v_n) - v_n\| = \|\mathrm{far}_C(u) - v_n\| = s_n < s,$$

from which we derive that $u \in \mathrm{cl}\,\{\mathrm{dfar}_C < s\}$, and hence the desired inclusion $\{\mathrm{dfar}_C \leq s\} \subset \mathcal{C}$.

(f) Fix any real $s > R$. According to (b) above, we have

$$\{\mathrm{dfar}_C > s\} = (X \setminus s\mathbb{B}) + C,$$

hence by (c)

$$\{\mathrm{dfar}_C \geq s\} = \mathrm{cl}\,\{\mathrm{dfar}_C > s\} = \mathrm{cl}\big((X \setminus s\mathbb{B}) + C\big). \tag{5.4}$$

We claim that

$$\mathrm{cl}\big((X \setminus s\mathbb{B}) + C\big) = (X \setminus s\mathbb{U}) + C.$$

First, we note that the inclusion $\supset$ holds by the obvious fact that $(X \setminus s\mathbb{U}) + C \subset \{\mathrm{dfar}_C \geq s\}$ and by (5.4). Let us justify the converse inclusion. The set $S := \big(X \setminus$

$s\mathbb{U}\big) + C$ is the Minkowski's sum of an $s$-prox-regular set and an $R$-strongly convex with constant $R < s$. According to Theorem 5.8, we know that the set $S$ is closed, so

$$\mathrm{cl}\big((X \setminus s\mathbb{B}) + C\big) \subset \mathrm{cl}\, S = S.$$

The proof is complete. □

We are now in a position to provide our proof of the following characterizations of $R$-strongly convex sets.

**Theorem 5.11.** *Let $C$ be a nonempty closed bounded subset of $X$ and let $R > 0$ be a positive real. The following assertions are equivalent:*
*(a) the set $C$ is $R$-strongly convex;*
*(b) for any real $s > R$, the function $-\mathrm{dfar}_C + \frac{1}{2(s-R)}\|\cdot\|^2$ is convex on any nonempty convex subset $V$ of $\mathcal{E}_s(C)$, that is, $-\mathrm{dfar}_C$ is linearly semiconvex on $V$ with $(s-R)^{-1}$ as coefficient;*
*(c) there exists a real $s > R$ such that the function $-\mathrm{dfar}_C + \frac{1}{2(s-R)}\|\cdot\|^2$ is convex on any nonempty convex subset $V$ of $\mathcal{E}_s(C)$;*
*(d) the function $-\mathrm{dfar}_C$ is locally linearly semiconvex on $\mathcal{E}_R(C)$, that is, linearly semiconvex near each point in $\mathcal{E}_R(C)$.*

*Proof.* $(a) \Rightarrow (b)$, Assume that $C$ is $R$-strongly convex and fix any real $s > R$. Let $V$ be any nonempty convex subset of $\mathcal{E}_s(C)$. Let $t \in ]R, s[$. According to (f) in Proposition 5.10, the set $D := \{\mathrm{dfar}_C \geq s\}$ is $(s-R)$-prox-regular, hence (see Theorem 5.7) the set $S := X \setminus D = \{\mathrm{dfar}_C < s\}$ is the union of a collection of closed balls of common radius $s - t$. Thanks to (e) in Proposition 5.10, we know that $\mathrm{cl}\, S = \{\mathrm{dfar}_C \leq s\}$, so

$$d(\cdot, S) = d(\cdot, \mathrm{cl}\, S) = d\big(\cdot, \{\mathrm{dfar}_C \leq s\}\big).$$

On the other hand, Lemma 5.9 says that $d(\cdot, S)$ is $(s-t)^{-1}$-linearly semiconcave on every nonempty convex set $U \subset \mathrm{cl}(X \setminus S) = D$, in particular on the set $V$. By virtue of Proposition 5.10(d) again, we can write

$$\mathrm{dfar}_C(u) = s + d\big(u, \{\mathrm{dfar}_C \leq s\}\big) \quad \text{for all } u \in \mathcal{E}_s(C).$$

It follows that $\mathrm{dfar}_C(\cdot)$ is $(s-t)^{-1}$-linearly semiconcave on $V$, that is, $\mathrm{dfar}_C(\cdot) - \frac{1}{2(s-t)}\|\cdot\|^2$ is concave on $V$. Since $t$ has been arbitrarily choosen in $]R, s[$, the desired property is then justified.
$(b) \Rightarrow (c)$ and $(c) \Rightarrow (d)$, Obvious.
$(d) \Rightarrow (a)$, Assume that $-\mathrm{dfar}_C$ is locally linearly semiconvex on any convex subset of $\mathcal{E}_R(C)$. Fix any $x \in \mathcal{E}_R(C)$. We can then find two reals $\rho, \delta > 0$ such that the function $f := -\mathrm{dfar}_C(\cdot) + \rho\|\cdot\|^2$ is convex on $B(x, \delta) \subset \mathcal{E}_R(C)$. It directly follows from the $C^{1,1}$-property of $\|\cdot\|^2$ (see (2.9)) that

$$\partial_P f(x) = \partial_P(-\mathrm{dfar}_C)(x) + \rho\nabla\|\cdot\|^2(x).$$

Therefore, we have $\partial_P(-\mathrm{dfar}_C)(x) \neq \emptyset$, so $-\mathrm{dfar}_C$ is Fréchet differentiable at $x$. We conclude that $C$ is $R$-strongly convex according to Theorem 5.2(d). The proof is complete. □

## 5.3 Farthest distance and separating balls

Given a nonempty closed convex set $C$ in the Hilbert space $X$ and an exterior point of $C$, say $x \in X \setminus C$, it is well-known (and not difficult to check) that we have with $x_\star := d_C(x)^{-1}\big(x - \mathrm{proj}_C(x)\big)$ the following separation property for some real $\alpha$

$$C \subset \{\langle x_\star, \cdot\rangle < \alpha\} \quad \text{and} \quad \langle x_\star, x\rangle > \alpha. \tag{5.5}$$

Replacing the above half-space $\{\langle x_\star, \cdot\rangle < \alpha\}$ by a set in the more general form $\left\{\langle x_\star, x\rangle - \frac{\|x\|^2}{2r} < \alpha\right\}$ (which is nothing but the complement of a closed ball (see Proposition 5.15)) allows to extend such a separation property, with a suitable vector $x_\star$, to the context of $r$-prox-regular sets.

**Theorem 5.12** ([83])**.** *Let $S$ be an $r$-prox-regular subset of $X$ with $r \in ]0, +\infty]$, $x \in X$ with $\delta := d_S(x) \in ]0, r[$. Then, one has with $x_\star := (\frac{1}{r} - \frac{1}{\delta})\mathrm{proj}_S(x) + \frac{1}{\delta}x$ the following separation property for some $\alpha \in \mathbb{R}$*

$$S \subset \left\{\langle x_\star, \cdot\rangle - \frac{\|\cdot\|^2}{2r} < \alpha\right\} \quad \textit{and} \quad \langle x_\star, x\rangle - \frac{\|x\|^2}{2r} > \alpha. \tag{5.6}$$

Our first aim in this section is to see how the above separation property can be reinforced in the case of strongly convex sets.

**Lemma 5.13.** *Let $C$ be an $R$-strongly convex subset of $X$ for some real $R > 0$. Let also $(x, y) \in \mathrm{gph}\,\mathrm{Far}_C$ with $\mathrm{dfar}_C(x) > 0$ and let $\rho \in ]0, \mathrm{dfar}_C(x)[$. Set $x^\star := \mathrm{dfar}_C(x)^{-1}(x - y)$. Then, one has for every $c \in C$*

$$\langle x^\star, x\rangle - \frac{1}{\rho}\mathrm{dfar}_C^2(x) < \langle x^\star, y\rangle \le \langle x^\star, c\rangle - \frac{1}{2R}\|c - y\|^2,$$

*in particular there exists $\alpha \in \mathbb{R}$ such that for every $c \in C$*

$$\langle x^\star, x\rangle - \frac{1}{\rho}\mathrm{dfar}_C^2(x) < \alpha < \langle x^\star, c\rangle - \frac{1}{2R}\|c - y\|^2.$$

*Proof.* Fix any $c \in C$. Using the inclusion $-x^\star \in N(C; y) \cap \mathbb{S}$ (see (2.7)) and Theorem 5.2(b), we obtain

$$\langle -x^\star, c - y\rangle \le -\frac{1}{2R}\|c - y\|^2,$$

or equivalently,

$$\langle x^\star, y\rangle \le \langle x^\star, c\rangle - \frac{1}{2R}\|c - y\|^2. \tag{5.7}$$

On the other hand, we easily see (thanks to the definition of $x^\star$ and the inclusion $\rho \in ]0, \mathrm{dfar}_C(x)[$) that

$$\langle x^\star, x\rangle - \frac{1}{\rho}\mathrm{dfar}_C^2(x) < \langle x^\star, y\rangle. \tag{5.8}$$

It then suffices to combine (6.11) and (5.8) to finish the proof. □

In (5.5) and (5.6) we have a separation property with half-spaces and complements of balls for convex sets and prox-regular sets, respectively. Here, we are in a position to refine these results for strongly convex sets in establishing for such sets a separation property with balls. Given any $x^\star \in X$, any real $R > 0$ and any $R$-strongly convex set $C$ in $X$, it will be convenient for us to set

$$q_{x^\star,R}(x) := \langle x^\star, x\rangle - \frac{\|x\|^2}{2R} \quad \text{for all } x \in X \tag{5.9}$$

and

$$\Upsilon_{C,R}(x) := \big(\frac{1}{R} - \frac{1}{\mathrm{dfar}_C(x)}\big)\mathrm{far}_C(x) + \frac{1}{\mathrm{dfar}_C(x)}x \quad \text{for all } x \in \mathcal{E}_R(C). \tag{5.10}$$

**Theorem 5.14.** *Let $C$ be an $R$-strongly convex set in $X$ for some real $R > 0$ and let $x \in X$ with $\delta := \mathrm{dfar}_C(x) > R$. Then one has with $x^\star = \Upsilon_{C,R}(x) := \big(R^{-1} - \delta^{-1}\big)\mathrm{far}_C(x) + \delta^{-1}x$*

$$C \subset \left\{\langle x^\star, \cdot\rangle - \frac{\|\cdot\|^2}{2R} \geq \inf_{c\in C} q_{x^\star,R}(c)\right\} \quad \textit{and} \quad q_{x^\star,R}(x) \leq \inf_{c\in C} q_{x^\star,R}(c). \tag{5.11}$$

*If in addition $\delta > 2R$ (so $x \notin C$), then one has the following strict separation property for some $\alpha \in \mathbb{R}$,*

$$C \subset \left\{\langle x^\star, \cdot\rangle - \frac{\|\cdot\|^2}{2R} > \alpha\right\} \quad \textit{and} \quad q_{x^\star,R}(x) < \alpha \leq \inf_{c\in C} q_{x^\star,R}(c).$$

*Proof.* Set $y := \mathrm{far}_C(x)$, $\delta := \mathrm{dfar}_C(x)$ and $u^\star := \delta^{-1}(x - y)$ (so $x^\star = u^\star + R^{-1}y$). Assume first that $\delta > R$. The inclusion in (5.11) is obvious since $\kappa := \inf_{c\in C} q_{x^\star,R}(c) < +\infty$. Noticing that $u^\star \in -N(C; y) \cap \mathbb{S}$ and using Theorem 5.2(b), we get

$$\langle u^\star, c - y\rangle \geq \frac{1}{2R}\|c - y\|^2 \quad \text{for all } c \in C,$$

or equivalently,

$$\Big\langle u^\star + \frac{y}{R}, c\Big\rangle - \frac{\|c\|^2}{2R} \geq \Big\langle u^\star + \frac{y}{R}, x\Big\rangle - \frac{\|x\|^2}{2R} = q_{x^\star,R}(x) \quad \text{for all } c \in C.$$

We then arrive to the inequality $q_{x^\star,R}(x) \leq \kappa$.

Now, assume that $\delta > 2R$. Applying Lemma 5.13 with $\rho := 2R$ gives some real $\beta$ such that for all $c \in C$

$$\langle u^\star, x\rangle - \frac{\delta^2}{2R} < \beta < \langle u^\star, c\rangle - \frac{1}{2R}\|c - y\|^2. \tag{5.12}$$

Fix any $c \in C$. Through elementary computations, we observe that

$$\langle u^\star, c\rangle - \frac{1}{2R}\|c - y\|^2 = \langle x^\star, c\rangle - \frac{1}{2R}(\|y\|^2 + \|c\|^2) \tag{5.13}$$

and

$$\langle u^\star, x\rangle - \frac{\delta^2}{2R} = \langle x^\star, x\rangle - \frac{1}{2R}(\|x\|^2 + \|y\|^2). \tag{5.14}$$

Putting together (5.12), (5.13) and (5.14) yields

$$\langle x^\star, x\rangle - \frac{\|x\|^2}{2R} < \beta + \frac{\|y\|^2}{2R} < \langle x^\star, c\rangle - \frac{\|c\|^2}{2R}.$$

It remains to set $\alpha := \beta + \frac{\|y\|^2}{2R}$ to finish the proof. □

Setting

$$\Phi_{C,R}(x^\star) := \inf_{c\in C} q_{x^\star,R}(c) \quad \text{and} \quad L_{x^\star,R,\alpha} := \{q_{x^\star,R} \geq \alpha\} \tag{5.15}$$

and noticing the elementary equalities

$$\|Rx^\star - x\|^2 = R^2\|x^\star\|^2 - 2R\,\langle x^\star, x\rangle + \|x\|^2$$

along with

$$R^2\|x^\star\|^2 - 2R\Phi_{C,R}(x^\star) = R^2\|x^\star\|^2 + \sup_{c\in C}\big(-2R\,\langle x^\star, c\rangle + \|c\|^2\big) = \sup_{c\in C}\|Rx^\star - c\|^2,$$

it is not difficult to check that an element $x^\star \in L_{x^\star,R,\alpha}$ if and only if $R\,\|x^\star - x\|^2 \leq R^2\|x^\star\|^2 - 2R\alpha$. It leads to the statement below

$$x^\star \in L_{x^\star,R,\alpha} \Leftrightarrow \begin{cases} x^\star \in B[Rx^\star, \sqrt{R^2\|x^\star\|^2 - 2R\alpha}] & \text{if } R^2\|x^\star\|^2 - 2R\alpha \geq 0\\ x^\star = \emptyset & \text{otherwise.}\end{cases}$$

In addition, if we take $\alpha = \Phi_{C,R}(x^\star)$ for some closed subset $C$, it follows that

$$\begin{aligned} R^2\|x^\star\|^2 - 2R\alpha &= R\|x^\star\|^2 - 2R\sup_{x\in C}\left(\langle x^\star, x\rangle + \frac{1}{2R}\|x\|^2\right)\\ &= \sup_{x\in C}\big(R\|x^\star\|^2 - 2R\,\langle x^\star, x\rangle + \|x\|^2\big)\\ &= \sup_{x\in C}\|Rx^\star - x\|^2 = \mathrm{dfar}_C(Rx^\star)^2 \end{aligned}$$

and thus, one has $L_{x^\star,R,\Phi_{C,R}(x^\star)} = B[Rx^\star, \mathrm{dfar}_C(Rx^\star)]$.

Now, let us consider when $C$ be an $R$-strongly convex subset of $\mathcal{H}$ for some real $R > 0$. Take $v := \dfrac{\mathrm{far}_C(x) - x}{\mathrm{dfar}_C(x)} \in N(C, \mathrm{far}_C(x))$, we observe that

$$\mathrm{far}_C(x) - Rv = \mathrm{far}_C(x)\left(1 - \frac{R}{\mathrm{dfar}_C(x)}\right) + \frac{R}{\mathrm{dfar}_C(x)}x =: R\Upsilon_{C,R}(x)$$

where

$$\Upsilon_{C,R}(x) = \left(\frac{1}{R} - \frac{1}{\mathrm{dfar}_C(x)}\right)\mathrm{far}_C(x) + \frac{x}{\mathrm{dfar}_C(x)}.$$

Going then, we directly establish the following important description of the upper level set $L_{x^\star,R,\alpha}$.

**Proposition 5.15.** *Let $x^\star \in X$, $\alpha \in \mathbb{R}$ and $R \in ]0, +\infty[$. Let also $C$ be a nonempty subset of $X$. The following hold with $\rho := R^2\|x^\star\|^2 - 2R\alpha$.*
*(a) One has*

$$L_{x^\star,R,\alpha} = \begin{cases} B[Rx^\star, \sqrt{\rho}] & \text{if } \rho \geq 0, \\ \emptyset & \text{otherwise.} \end{cases}$$

*(b) If $\alpha = \Phi_{C,R}(x^\star)$, then one has $\rho = \mathrm{dfar}_C^2(Rx^\star)$, in particular*

$$L_{x^\star,R,\Phi_{C,R}(x^\star)} = B[Rx^\star, \mathrm{dfar}_C(Rx^\star)].$$

Theorem 5.14 naturally leads to investigate some properties of $\Upsilon_{C,R}(x)$. The following lemma will be needed.

**Lemma 5.16.** *Let $x^\star \in X, \alpha \in \mathbb{R}, R \in ]0, +\infty[$. Assume that $C \subset L_{x^\star,R,\alpha}$ and $x \notin L_{x^\star,R,\alpha}$. Then, one has $\beta := \Phi_{C,R}(x^\star) \leq \alpha$ and*

$$C \subset L_{x^\star,R,\beta} \subset L_{x^\star,R,\alpha} \quad \text{and} \quad x \notin L_{x^\star,R,\beta},$$

*where $\Phi_{C,R}$ and $L_{x^\star,R,\alpha}$ are as defined in* (5.15).
*In particular, one has*

$$\mathrm{dfar}(x, C) \leq \mathrm{dfar}(x, L_{x^\star,R,\beta}) \leq \mathrm{dfar}(x, L_{x^\star,R,\alpha}).$$

*Proof.* Due to the definition of $\beta$, it is evident that $C \subset L_{x^\star,R,\beta}$. On the other hand, we observe that

$$\langle x^\star, c\rangle - \frac{\|c\|^2}{2R} \geq \alpha \quad \text{for all } c \in C,$$

hence $\beta \geq \alpha$. This obviously implies that $L_{x^\star,R,\beta} \subset L_{x^\star,R,\alpha}$ along with (keeping in mind that $x \notin L_{x^\star,R,\alpha}$) $x \notin L_{x^\star,R,\beta}$. The proof is complete. □

Now, we can state and prove:

**Proposition 5.17.** *Let $C$ be an $R$-strongly subset of $X$ with $R \in ]0, +\infty[$ and let $x \in X$ with $\delta := \mathrm{dfar}_C(x) > R$. Let $\Upsilon_{C,R}(x)$ be as defined in* (5.10). *The following hold with $t := \delta^{-1}R \in ]0, 1[$.*
*(a) The inclusion*

$$\mathrm{far}_C(x) \in \mathrm{Far}_C\big(R\Upsilon_{C,R}(x)\big)$$

*is satisfied along with the equalities*

$$R\Upsilon_{C,R}(x) = tx + (1-t)\mathrm{far}_C(x) \quad \text{and} \quad \|R\Upsilon_{C,R}(x) - x\| = \delta - R. \tag{5.16}$$

*(b) One has*

$$\{R\Upsilon_{C,R}(u) : u \in \mathcal{E}_R(C)\} = \{\mathrm{dfar}_C = R\} \cap \mathrm{Dom}\,\mathrm{Far}_C =: \Lambda_R(C) \tag{5.17}$$

*along with*

$$\mathrm{cl}_{\|.\|}\big(\Lambda_R(C)\big) = \{\mathrm{dfar}_C = R\}.$$

*(c) One has with $x^\star := \Upsilon_{C,R}(x)$ and $L_{x^\star,R,\Phi_{C,R}(x^\star)}$ as defined in* (5.15)

$$R^2\|x^\star\|^2 - 2R\Phi_{C,R}(x^\star) = R^2 \quad \text{and} \quad L_{x^\star,R,\Phi_{C,R}(x^\star)} = B[Rx^\star, R].$$

*(d) One has*

$$\mathcal{E}_R(C) = \{\tau R\Upsilon_{C,R}(u) + (1-\tau)\mathrm{far}_C(u) : u \in \mathcal{E}_R(C), \tau \in ]1, +\infty[\}\,.$$

*(e) For all* $\tau \in ]1, +\infty[$*, one has*

$$\Upsilon_{C,R}\big(\tau R\Upsilon_{C,R}(x) + (1-\tau)\mathrm{far}_C(x)\big) = \Upsilon_{C,R}(x).$$

*(f) One has*

$$C = \bigcap_{y \in \Lambda_R(C)} B[y, R].$$

*Proof.* (a) The inclusion is a direct consequence of Theorem 5.2(e) (see also Remark 5.3) while the equalities directly follow from the definition of $\Upsilon_{C,R}(x)$.
(b) Fix any $u \in \mathcal{E}_R(C)$. Thanks to Theorem 5.2(c) and (a) above we know that $\mathrm{far}_C(u)$ is well defined along with $\mathrm{far}_C(u) \in \mathrm{Far}_C\big(R\Upsilon_{C,R}(u)\big)$. The latter inclusion and the definition of $v := R\Upsilon_{C,R}(u)$ easily give

$$\mathrm{dfar}_C(v) = \|\mathrm{far}_C(u) - v\| = R,$$

which justifies the inclusion $\subset$ concerning the first equality in (b). Let us show the converse inclusion. Fix any $z \in \mathrm{Dom}\,\mathrm{Far}_C$ such that $\mathrm{dfar}_C(z) = R$. Pick any $v \in \mathrm{Far}_C(z)$ and any real $\lambda > 1$. It is not difficult to check that $u := \lambda z + (1-\lambda)v \in \mathcal{E}_R(C)$ along with $\mathrm{far}_C(u) = v$ (see (2.8)). Set $\sigma := \mathrm{dfar}_C(u)$ and note that (having in mind $\|z - v\| = R$)

$$\sigma := \mathrm{dfar}_C(u) = \|u - v\| = \lambda\|z - v\| = \lambda R,$$

hence $\lambda = \sigma/R$. This and the obvious equality $u - z = (\lambda - 1)(z - v)$ gives

$$\|u - z\| = \frac{\sigma - R}{R}\|z - v\| = \sigma - R$$

and

$$\langle v - z, z - u\rangle = (\lambda - 1)\|v - z\|^2 = (\lambda - 1)R^2 = (\sigma - R)R.$$

We then arrive to

$$\begin{aligned}\left\|(1 - \frac{R}{\sigma})v + \frac{R}{\sigma}u - z\right\|^2 &= \left\|(1 - \frac{R}{\sigma})(v - z) - \frac{R}{\sigma}(z - u)\right\|^2 \\ &= \big(1 - \frac{R}{\sigma}\big)^2\|v - z\|^2 + \frac{R^2}{\sigma^2}\|z - u\|^2 \\ &\quad - 2\frac{R}{\sigma}(1 - \frac{R}{\sigma})\langle v - z, z - u\rangle \\ &= (\sigma - R)^2\frac{R^2}{\sigma^2} + \frac{R^2}{\sigma^2}(\sigma - R)^2 - 2\frac{R^2}{\sigma^2}(\sigma - R)^2 = 0,\end{aligned}$$

that is, $z = (1 - \frac{R}{\sigma})v + \frac{R}{\sigma}u$. This means that $z = R\Upsilon_{C,R}(u)$ with $u \in \mathcal{E}_R(C)$, which confirms the desired inclusion $\supset$ for the first equality in (b). The first equality in (b) is then proved.

Regarding the second equality in (b), take any $y \in X$ such that $\mathrm{dfar}_C(y) = R$. According to Proposition 5.10, we can choose a sequence $(u_n)_{n\geq 1}$ in $\{\mathrm{dfar}_C > R\}$ with

$u_n \to y$. Keeping in mind that $C$ is bounded and $\mathrm{dfar}_C(\cdot)$ is continuous, we easily see that $v_n := R\Upsilon_{C,R}(u_n) \to y$. It remains to use (5.17) above to get that $\mathrm{dfar}_C(v_n) = R$ for every integer $n \geq 1$. This justifies the second equality in (b).
(c) According to Proposition 5.15, we have $R^2\|x^\star\|^2 - 2R\Phi_{C,R}(x^\star) = \mathrm{dfar}_C^2(Rx^\star)$ and

$$L_{x^\star,R,\Phi_{C,R}(x^\star)} = B[Rx^\star, \mathrm{dfar}_C(Rx^\star)].$$

It then suffices to use (a) above to get $\mathrm{dfar}_C(Rx^\star) = \|Rx^\star - \mathrm{far}_C(x)\| = R$.
(d) Since any $u \in \mathcal{E}_R(C)$ can be written by the first equality in (5.16) as

$$u = \frac{1}{t}R\Upsilon_{C,R}(u) + \left(1 - \frac{1}{t}\right)\mathrm{far}_C(u) \quad \text{with } \frac{1}{t} > 1,$$

the inclusion $\subset$ directly follows.

Conversely, fix any $u \in \mathcal{E}_R(C)$ and any real $\tau > 1$. Set $\omega := \tau R\Upsilon_{C,R}(u) + (1-\tau)\mathrm{far}_C(u)$. Using the inclusion $\mathrm{far}_C(u) \in C$ and the definition of $\Upsilon_{C,R}$, we easily see that

$$\mathrm{dfar}_C(\omega) \geq \|\omega - \mathrm{far}_C(u)\| = \tau\|R\Upsilon_{C,R}(u) - \mathrm{far}_C(u)\| = \tau R > R.$$

The desired equality is then established.
(e) Fix any real $\tau > 1$. According to (d) above we know that $y := \tau R\Upsilon_{C,R}(x) + (1-\tau)\mathrm{far}_C(x) \in \mathcal{E}_R(C)$. Further, the inclusion $\mathrm{far}_C(x) \in \mathrm{Far}_C\big(R\Upsilon_{C,R}(x)\big)$ in (a) above combined with the property (2.8) gives that

$$\begin{aligned}\mathrm{far}_C(x) &= \mathrm{far}_C\Big(R\Upsilon_{C,R}(x) + (\tau-1)\big(R\Upsilon_{C,R}(x) - \mathrm{far}_C(x)\big)\Big)\\ &= \mathrm{far}_C\big(\tau R\Upsilon_{C,R}(x) + (1-\tau)\mathrm{far}_C(x)\big) = \mathrm{far}_C(y). \end{aligned} \tag{5.18}$$

Therefore, elementary computations yield with $\delta' := \mathrm{dfar}_C(y)$ (recalling that $\delta := \mathrm{dfar}_C(x)$)

$$\begin{aligned} w := \Upsilon_{C,R}(x) - \Upsilon_{C,R}(y) &= \big(\frac{1}{\delta'} - \frac{1}{\delta}\big)\mathrm{far}_C(x) + \frac{x}{\delta} - \frac{y}{\delta'}\\ &= \big(\frac{1}{\delta'} - \frac{1}{\delta}\big)\big(\mathrm{far}_C(x) - x\big) + \frac{1}{\delta'}(x-y). \end{aligned} \tag{5.19}$$

Using again the definition of $\Upsilon_{C,R}(x)$, we can easily check that

$$y - x = \frac{\delta - \tau R}{\delta}\big(\mathrm{far}_C(x) - x\big) \tag{5.20}$$

and

$$y - \mathrm{far}_C(x) = y - x + x - \mathrm{far}_C(x) = \frac{\tau R}{\delta}\big(x - \mathrm{far}_C(x)\big). \tag{5.21}$$

If $\delta - \tau R \geq 0$ (resp. $\delta - \tau R < 0$), we have by (5.20) and (5.21) and by the equality $\mathrm{far}_C(x) = \mathrm{far}_C(y)$ in (5.18) that

$$\|y - x\| = \delta - \tau R = \delta - \|y - \mathrm{far}_C(x)\| = \delta - \delta'$$

(resp.

$$\|y - x\| = \tau R - \delta = \|y - \mathrm{far}_C(x)\| - \delta = \delta' - \delta).$$

In both cases $\tau R = \delta'$, so (5.19) easily gives $\omega = 0$, which means that (e) holds true.
(f) Set $C_0 := \bigcap_{x^\star \in \Lambda_R(C)} B[x^\star, R]$. According to Theorem 5.14 and to (c) and (b) above, we can write

$$C \subset \bigcap_{x^\star \in \Upsilon_{C,R}(\mathcal{E}_R(C))} L_{x^\star,R,\Phi_{C,R}(x^\star)} = \bigcap_{x^\star \in \Upsilon_{C,R}(\mathcal{E}_R(C))} B[Rx^\star, R] = C_0.$$

We are going to show that $C_0 \subset C$. By contradiction, assume that there is $z \in C_0 \setminus C$. Set $p := \mathrm{proj}_C(z)$ and $d := d_C(z) > 0$. Note first (see (2.6)) that $v := \frac{z-p}{d} \in N(C; p)$. Fix any real $R' > R$ and put $q := p - R'v$. Thanks to Theorem 5.2(b), we see that for every $c \in C$

$$\langle p - q, c - p\rangle = R' \langle v, c - p\rangle \leq -\frac{R'}{2R}\|c - p\|^2 \leq -\frac{1}{2}\|c - p\|^2$$

and this entails (see (2.5)) that $p \in \mathrm{Far}_C(q)$. In fact, we have $p = \mathrm{far}_C(q)$ since $\mathrm{dfar}_C(q) = \|p - q\| = R' > R$. Therefore, we can apply Theorem 5.2(e) (see also Remark 5.3) to get

$$p = \mathrm{far}_C(q) \in \mathrm{Far}_C\big(p - R\frac{p-q}{\mathrm{dfar}_C(q)}\big) = \mathrm{Far}_C\big(p - \frac{R}{R'}(p-q)\big).$$

Set $w := p - \frac{R}{R'}(p - q)$ and note that $\mathrm{dfar}_C(w) = \|p - w\| = \frac{R}{R'}\|p - q\| = R$. Then, it follows from the definition of $C_0$ that $z \in B[w, R]$. On the other hand, we easily observe that

$$z - w = z - p + \frac{R}{R'}(p - q) = z - p + Rv = \big(1 + \frac{R}{d}\big)(z - p),$$

hence

$$d + R = \big(1 + \frac{R}{d}\big)\|z - p\| \leq \|z - w\| \leq R.$$

Consequently, we get $d = 0$ and this is the desired contradiction. The equality $C = C_0$ is then established and the proof of the proposition is complete. □

Our second aim in the present section is to provide some analytic formulation for the farthest distance function to a strongly convex set. Doing so, we complement the following well known formula (see, e.g., [102, Theorem 6.23])

$$d_C(x) = \langle x_\star, x\rangle - \sigma(x_\star, C), \tag{5.22}$$

where $x$ is an exterior point of a convex set $C$ and $\sigma(x_\star, C)$ denotes the support function of the set $C$ at $x_\star := d_C(x)^{-1}(x - \mathrm{proj}_C(x))$. We point out that such an equality has been extended to the context of a prox-regular set in [115]. Keeping notation of Theorem 5.12, the equality analogous to (5.22) for an $r$-prox-regular set $S$ is given by

$$d_S(x)\left(1 - \frac{d_S(x)}{2r}\right) = q_{x_\star,r}(x) - \phi_{S,r}(x_\star),$$

where $\phi_{S,r}(x_\star) := \sup_{u \in S} q_{x_\star,r}(u)$.

**Theorem 5.18.** *Let $C$ be an $R$-strongly convex subset of $X$ for some real $R > 0$ and let $x \in X$ with $\operatorname{dfar}_C(x) > R$. Then, there exists one and only one $x^\star \in X$ with $\|x^\star - R^{-1}x\| = R^{-1}\operatorname{dfar}_C(x) - 1$ (namely, $x^\star := \Upsilon_{C,R}(x)$ as in (5.10)) such that*

$$\operatorname{dfar}_C(x)\left(1 - \frac{\operatorname{dfar}_C(x)}{2R}\right) = q_{x^\star,R}(x) - \Phi_{C,R}(x^\star),$$

*where $q_{x^\star,R}$ and $\Phi_{C,R}$ are as defined in (5.9) and (5.15), respectively.*

*Proof.* Set $v := \operatorname{far}_C(x)$ and $\delta := \operatorname{dfar}_C(x)$. The proof is divided into two parts.

**Existence.** Putting together the inclusion $v - x \in N(C; v)$ (see (2.5)) and Theorem 5.2, we see that for every $c \in C$

$$\langle v - x, c - v\rangle \le -\frac{\delta}{2R}\|c - v\|^2,$$

or equivalently,

$$\frac{\delta}{2R}\|c - v\|^2 + \langle v - x, c - x\rangle \le \|v - x\|^2 = \delta^2.$$

This inequality and the inclusion $v \in C$ easily give

$$\delta^2 \le \sup_{c\in C}\Big(\frac{\delta}{2R}\|c - v\|^2 + \langle v - x, c - x\rangle\Big) \le \delta^2,$$

that is,

$$\delta = \sup_{c\in C}\Big(\frac{\|c - v\|^2}{2R} + \delta^{-1}\langle v - x, c - x\rangle\Big).$$

Keeping in mind the definition of $x^\star := \Upsilon_{C,R}(x^\star)$ in (5.10), it is then not difficult to check that

$$\begin{aligned}
\delta - \frac{\delta^2}{2R} &= \sup_{c\in C}\left(\frac{1}{2R}(\|c - v\|^2 - \|x - v\|^2) + \delta^{-1}\langle v - x, c - x\rangle\right)\\
&= \left\langle \delta^{-1}(v - x) - R^{-1}v, -x\right\rangle - \frac{\|x\|^2}{2R} + \sup_{c\in C}\left(\left\langle \delta^{-1}(v - x) - R^{-1}v, c\right\rangle + \frac{\|c\|^2}{2R}\right)\\
&= q_{x^\star,R}(x) - \inf_{c\in C}\left(\langle x^\star, c\rangle - \frac{\|c\|^2}{2R}\right)\\
&= q_{x^\star,R}(x) - \Phi_{C,R}(x^\star).
\end{aligned}$$

**Uniqueness.** Let $x_1^\star, x_2^\star \in X$ be such that for each $i \in \{1, 2\}$

$$\left\|x_i^\star - \frac{x}{R}\right\| = \frac{\delta}{R} - 1 \quad \text{and} \quad \delta\big(1 - (2R)^{-1}\delta\big) = q_{x_i^\star,R}(x) - \Phi_{C,r}(x_i^\star).$$

It is readily seen with $u^\star := 2^{-1}(x_1^\star + x_2^\star)$ that

$$\left\|u^\star - R^{-1}x\right\| \le 2^{-1}\left\|x_1^\star - R^{-1}x\right\| + 2^{-1}\left\|x_2^\star - R^{-1}x\right\| = \frac{\delta}{R} - 1. \tag{5.23}$$

Setting $q_i := q_{x_i^\star,R}$ for each $i \in \{1, 2\}$, it is also straightforward to check that

$$q_{u^\star,R}(x) = 2^{-1}[q_1(x) + q_2(x)].$$

We deduce from what precedes

$$\begin{aligned} q_{u^\star,R}(x) - \Phi_{C,R}(u^\star) &= 2^{-1}\big[q_1(x) + q_2(x) - \inf_{c\in C}\big(q_1(c) + q_2(c)\big)\big] \\ &\le 2^{-1}\big[q_1(x) + q_2(x) - \inf_{c\in C} q_1(c) - \inf_{c\in C} q_2(c)\big] \\ &= 2^{-1}\big(q_1(x) - \Phi_{C,R}(x_1^\star) + q_2(x) - \Phi_{C,R}(x_2^\star)\big) \\ &= \delta\big(1 - \frac{\delta}{2R}\big). \end{aligned} \tag{5.24}$$

Combining (5.24), the definition of $q_{u^\star,R}$ and Proposition 5.15(b), we obtain

$$\begin{aligned} \delta(2R-\delta) =& 2R\big(q_{u^\star,R}(x) - \Phi_{C,R}(u^\star)\big) \\ =& 2Rq_{u^\star,R}(x) - R^2\|u^\star\|^2 + R^2\|u^\star\|^2 - 2R\Phi_{C,R}(u^\star) \\ =& -\|Ru^\star - x\|^2 + \mathrm{dfar}_C^2(Ru^\star). \end{aligned} \tag{5.25}$$

Using (5.25), the 1-Lipschitz property of $\mathrm{dfar}_C$, the inequality $\delta - \|Ru^\star - x\| \ge 0$ and (5.23), we get

$$\begin{aligned} \delta(2R-\delta) \ge& -\|Ru^\star - x\|^2 + \big(\delta - \|Ru^\star - x\|\big)^2 \\ =& \delta^2 - 2\delta\|Ru^\star - x\| \\ \ge& \delta^2 - 2\delta(\delta - R) = \delta(2R-\delta). \end{aligned} \tag{5.26}$$

Through (5.26), we then see $\|Ru^\star - x\| = \delta - R$, or equivalently,

$$\|u^\star - \frac{x}{R}\| = \frac{\delta}{R} - 1 =: \kappa.$$

We conclude that $x_1^\star, x_2^\star, u^\star$ lie on the sphere $R^{-1}x + \kappa\mathbb{S}_X$ with $u^\star = (x_1^\star + x_2^\star)/2$. It remains to invoke the strict convexity of the Hilbert norm $\|\cdot\|$ to obtain $x_1^\star = x_2^\star$. The proof is complete. □

Next we provide, in addition to the preceding theorem, a proposition describing (for a strongly convex set $C$) the farthest distance $\mathrm{dfar}_C(x)$ in terms of the unique function $q_{x^\star,R}$. The expression also reverses appropriate supremum and minimum.

**Proposition 5.19.** *Let $C$ be an $R$-strongly convex subset of $X$ for some real $R > 0$, $x \in X$ with $\delta := \mathrm{dfar}_C(x) > R$ and $\mathcal{L} := \{y^\star \in X^\star : \|y^\star - R^{-1}x\| = R^{-1}\delta - 1\}$. Then, one has with $q_{y^\star,R}$ as defined in* (5.9)

$$\delta\big(1 - \frac{\delta}{2R}\big) = \min_{y^\star\in\mathcal{L}} \sup_{c\in C}\big(q_{y^\star,R}(x) - q_{y^\star,R}(c)\big) = \sup_{c\in C}\min_{y^\star\in\mathcal{L}}\big(q_{y^\star,R}(x) - q_{y^\star,R}(c)\big).$$

*Proof.* For every $y^\star \in X$, set

$$\mu(y^\star) := \sup_{c\in C}\left(\frac{\|c\|^2 - \|x\|^2}{2R} + \langle y^\star, x - c\rangle\right) = \sup_{c\in C}\big(q_{y^\star,R}(x) - q_{y^\star,R}(c)\big).$$

Thanks to Theorem 5.18, we have with $x^\star := \Upsilon_{C,R}(x)$

$$\mu(x^\star) = \sup_{c\in C}\big(q_{x^\star,R}(x) - q_{x^\star,R}(c)\big) = q_{x^\star,R}(x) - \Phi_{C,R}(x^\star) = \delta\big(1 - \frac{\delta}{2R}\big). \tag{5.27}$$

We obviously have for every $y^\star \in \mathcal{L}$

$$\begin{aligned}\sup_{c\in C}\left(\frac{\|c\|^2-\|x\|^2}{2R}+\langle y^\star, x-c\rangle\right) &= \sup_{c\in C}\left(\Big\langle y^\star-\frac{x}{R}, x-c\Big\rangle+\frac{\|x-c\|^2}{2R}\right)\\ &\geq \sup_{c\in C}\left(-\|y^\star-\frac{x}{R}\|\|x-c\|+\frac{\|x-c\|^2}{2R}\right)\\ &= \sup_{c\in C}\left(\big(1-\frac{\delta}{R}\big)\|x-c\|+\frac{\|x-c\|^2}{2R}\right)=:\kappa,\end{aligned}$$

so $\inf_{y^\star\in\mathcal{L}}\mu(y^\star)\geq\kappa$. On the other hand, let $(c_n)_{n\geq 1}$ be a sequence in $C$ with $\|c_n-x\|\to\delta$. It is readily seen that

$$\big(1-\frac{\delta}{R}\big)\|x-c_n\|+\frac{\|x-c_n\|^2}{2R}\leq\kappa\quad\text{for all } n\geq 1,$$

hence by (5.27)

$$\mu(x^\star)=\delta\big(1-\frac{\delta}{2R}\big)=\big(1-\frac{\delta}{R}\big)\delta+\frac{\delta^2}{2R}\leq\kappa\leq\inf_{y^\star\in\mathcal{L}}\mu(y^\star).$$

This and the inclusion $x^\star\in\mathcal{L}$ ensure that

$$\mu(x^\star)=\delta\big(1-\frac{\delta}{2R}\big)=\kappa=\min_{y^\star\in\mathcal{L}}\mu(y^\star).\tag{5.28}$$

The left desired equality of the proposition is then established. Regarding the right equality, let us first write

$$\begin{aligned}\inf_{y^\star\in\mathcal{L}}\left(\frac{\|c\|^2-\|x\|^2}{2R}+\langle y^\star, x-c\rangle\right) &= \frac{\|c\|^2-\|x\|^2}{2R}+\inf_{y^\star\in\mathcal{L}}\langle y^\star, x-c\rangle\\ &= \frac{\|c\|^2-\|x\|^2}{2R}+\Big\langle\frac{x}{R}, x-c\Big\rangle+\inf_{y^\star\in\mathcal{L}}\Big\langle y^\star-\frac{x}{R}, x-c\Big\rangle,\end{aligned}$$

which yields

$$\begin{aligned}\inf_{y^\star\in\mathcal{L}}\left(\frac{\|c\|^2-\|x\|^2}{2R}+\langle y^\star, x-c\rangle\right) &= \frac{\|x-c\|^2}{2R}+\inf_{\|y^\star\|=R^{-1}\delta-1}\langle y^\star, x-c\rangle\\ &= \frac{\|x-c\|^2}{2R}-\big(\frac{\delta}{R}-1\big)\sup_{\|y^\star\|=1}\langle y^\star, c-x\rangle\\ &= \frac{\|x-c\|^2}{2R}+\big(1-\frac{\delta}{R}\big)\|x-c\|.\end{aligned}$$

Therefore, we have for every $c\in C$,

$$\begin{aligned}\inf_{y^\star\in\mathcal{L}}\left(\frac{\|c\|^2-\|x\|^2}{2R}+\langle y^\star, x-c\rangle\right) &= \min_{y^\star\in\mathcal{L}}\left(\frac{\|c\|^2-\|x\|^2}{2R}+\langle y^\star, x-c\rangle\right)\\ &= \frac{\|x-c\|^2}{2R}+\big(1-\frac{\delta}{R}\big)\|x-c\|,\end{aligned}$$

where the above minimum being justified by reflexivity of $X$ since the first term is the infimum of a continuous affine functional over the sphere $\mathcal{L}$. We deduce that

$$\sup_{c\in C}\min_{y^\star\in\mathcal{L}}\left(\frac{\|c\|^2-\|x\|^2}{2R}+\langle y^\star,x-c\rangle\right)=\kappa,$$

and this finishes the proof according to (5.28). □

**Remark 5.20.** *Let $C$ be an $R$-strongly convex subset of $X$ for some real $R>0$. Consider any real $\rho>R$ and any nonempty convex set $V$ with $V\subset\mathcal{E}_\rho(C)$. According to Theorem 5.11, we know that the function $-\mathrm{dfar}_C$ is $\sigma:=(\rho-R)^{-1}-$semiconvex on set $V$, or equivalently, $-\mathrm{dfar}_C+\psi_V$ is $\sigma$-semiconvex on $X$. Proceeding as in [83, Remark 3.2], we can establish that*

$$\begin{aligned}
-\mathrm{dfar}_C(x)&=\sup_{x^\star\in X}\big(q_{x^\star,\rho-R}(x)-(-\mathrm{dfar}_C+\psi_V)^{\star,\rho-R}(x^\star)\big)\\
&=\sup_{x^\star\in X}\left[q_{x^\star,\rho-R}(x)-\sup_{y\in X}\big(\langle x^\star,y\rangle-\frac{\|y\|^2}{2(\rho-R)}+(\mathrm{dfar}_C-\psi_V)(y)\big)\right]\\
&=\sup_{x^\star\in X}\left[q_{x^\star,\rho-R}(x)-\sup_{y\in V}\big(\langle x^\star,y\rangle-\frac{\|y\|^2}{2(\rho-R)}+\mathrm{dfar}_C(y)\big)\right]\\
&=\sup_{x^\star\in X}\left[q_{x^\star,\rho-R}(x)-\sup_{y\in V}\big(\langle x^\star,y\rangle-\frac{\|y\|^2}{2(\rho-R)}+(\mathrm{dfar}_C-\psi_V)(y)\big)\right]\\
&=\sup_{x^\star\in X}\left[q_{x^\star,\rho-R}(x)+\inf_{y\in V}\big(-\langle x^\star,y\rangle+\frac{\|y\|^2}{2(\rho-R)}-\mathrm{dfar}_C(y)\big)\right]\\
&=\sup_{x^\star\in X}\inf_{y\in V}\left(\langle x^\star,x-y\rangle-\frac{\|x\|^2}{2(\rho-R)}+\frac{\|y\|^2}{2(\rho-R)}-\mathrm{dfar}_C(y)\right)\\
&=\sup_{x^\star\in X}\inf_{y\in V}\big(q_{x^\star,\rho-R}(x)-q_{x^\star,\rho-R}(y)-\mathrm{dfar}_C(y)\big).
\end{aligned}$$

*In sum up, one has*

$$-\mathrm{dfar}_C(x)=\sup_{x^\star\in X}\inf_{y\in V}\big(q_{x^\star,\rho-R}(x)-q_{x^\star,\rho-R}(y)-\mathrm{dfar}_C(y)\big)\quad\textit{for all}\,x\in C.$$

■

Another important result that can be seen as a geometrical characterization of the farthest distance from a strongly convex set is presented in Theorem 5.21. It is the analogous for strongly convex sets of a similar result for convex sets [102, Chapter 6] recently extended to prox-regular sets in [84, Theorem 7]. It tells us that the farthest distance of a given point $x$ to a strongly convex set is the minimum of the farthest distance from $x$ to suitable closed balls separating the set and the point $x$.

**Theorem 5.21.** *Let $C$ be an $R$-strongly subset of $X$ for some $R\in]0,+\infty[$ and let $x\in X$ with $\delta:=\mathrm{dfar}(x,C)>2R$. Then, one has with $L_{y^\star,R,\alpha}$ as defined in* (5.15)

$$\delta=\min\left\{\mathrm{dfar}(x,L_{y^\star,R,\alpha}):(y^\star,\alpha)\in X\times\mathbb{R},C\subset L_{y^\star,R,\alpha},x\notin L_{y^\star,R,\alpha}\right\}. \tag{5.29}$$

*The minimum is attained at* $(x^\star, \beta)$ *with* $x^\star := \Upsilon_{C,R}(x)$ *and* $\beta := \Phi_{C,R}(x^\star)$ *(see* (5.10) *and* (5.15)*).*

*Further, for all* $y^\star \in X$ *with* $\|y^\star - R^{-1}x\| = R^{-1}\delta - 1$ *and all* $\alpha \in \mathbb{R}$*, one has the following implication*

$$\left.\begin{array}{l} \delta = d(x, L_{y^\star,r,\alpha}), \\ C \subset L_{y^\star,r,\alpha}, x \notin L_{y^\star,r,\alpha} \end{array}\right\} \Rightarrow (y^\star, \alpha) = \big(x^\star, \Phi_{C,R}(x^\star)\big).$$

*Proof.* Set $v := \mathrm{far}_C(x)$. Thanks to Lemma 5.16, we have

$$\inf\{\mathrm{dfar}(x, L_{y^\star,R,\alpha}) : (y^\star, \alpha) \in X \times \mathbb{R}, C \subset L_{y^\star,R,\alpha}, x \notin L_{y^\star,R,\alpha}\} \geq \delta. \tag{5.30}$$

On the other hand, Theorem 5.14 gives some real $\kappa$ such that $C \subset L_{x^\star,R,\kappa}$ and $x \notin L_{x^\star,R,\kappa}$. Lemma 5.16 again says that $C \subset L_{x^\star,R,\beta} =: L$ and $x \notin L_{x^\star,R,\beta}$ with $\beta := \Phi_{C,R}(x^\star)$. According to Proposition 5.17(c), we have $L = B[Rx^\star, R]$, hence (see Proposition 5.17(a))

$$\mathrm{dfar}(x, L) = R + \|Rx^\star - x\| = R + (\delta - R) = \delta. \tag{5.31}$$

The desired equality (5.29) directly follows from (5.30) and (5.31).

Fix any $y^\star \in X$ with $\|y^\star - R^{-1}x\| = R^{-1}\delta - 1$. Let $t \in \mathbb{R}$ be such that $\mathrm{dfar}(x, C) = \mathrm{dfar}(x, L_{y^\star,R,t})$ along with

$$C \subset L_{y^\star,R,t} \quad \text{and} \quad x \notin L_{y^\star,R,t} =: L_t. \tag{5.32}$$

According to Lemma 5.16, we have with $\theta := \Phi_{C,R}(y^\star)$

$$C \subset L_{y^\star,R,\theta} \subset L_t \quad \text{and} \quad x \notin L_{y^\star,R,\theta} =: L_\theta.$$

Hence, we get

$$\delta \leq \mathrm{dfar}(x, L_\theta) \leq \mathrm{dfar}(x, L_t) = \delta.$$

From Proposition 5.15, we then see that $\rho := R^2\|y^\star\|^2 - 2R\theta = \mathrm{dfar}_C^2(Ry^\star)$ along with

$$\delta = \mathrm{dfar}(x, L_\theta) = \mathrm{dfar}\big(x, B[Ry^\star, \sqrt{\rho}]\big) = \sqrt{\rho} + \|Ry^\star - x\|.$$

This and the equality $\|Ry^\star - x\| = \delta - R$ furnish

$$\rho = R^2\|y^\star\|^2 - 2R\theta = \big(\delta - \|Ry^\star - x\|\big)^2 = R^2,$$

or equivalently,

$$\theta = \frac{R}{2}\big(\,\|y^\star\|^2 - 1\big). \tag{5.33}$$

From definitions of $q_{y^\star,R}(x)$ and $\Phi_{C,R}(y^\star)$ and from (5.33), we obtain that

$$\begin{aligned} q_{y^\star,R}(x) - \Phi_{C,R}(y^\star) &= q_{y^\star,R}(x) - \theta \\ &= \langle y^\star, x\rangle - \frac{\|x\|^2}{2R} - \frac{R}{2}\big(\|y^\star\|^2 - 1\big) \\ &= -\frac{R}{2}\|y^\star - \frac{x}{R}\|^2 + \frac{R}{2} \\ &= -\frac{R}{2}(R^{-1}\delta - 1)^2 + \frac{R}{2} = \delta\big(1 - \frac{\delta}{2R}\big). \end{aligned}$$

The equality above and Theorem 5.18 yield $y^\star = x^\star = \Upsilon_{C,R}(x)$. It remains to show that $t = \Phi_{C,R}(y^\star)$. Note that the inclusion in (5.32) and the definition of $\Phi_{C,R}$ easily give $\Phi_{C,R}(y^\star) \geq t$. If $t < \Phi_{C,R}(y^\star)$, then Proposition 5.15 and Proposition 5.17(c) would entail

$$\begin{aligned}
\delta = \operatorname{dfar}(x, L_{x^\star,R,t}) &= \sqrt{R^2\|x^\star\|^2 - 2Rt} + \|Rx^\star - x\| \\
&> \sqrt{R^2\|x^\star\|^2 - 2R\Phi_{C,R}(x^\star)} + \|Rx^\star - x\| \\
&= R + \|Rx^\star - x\| \\
&= \operatorname{dfar}(x, L) = \delta.
\end{aligned}$$

which obviously leads to a contradiction. We conclude that $t = \Phi_{C,R}(y^\star)$. The proof is then complete. □

# Chapter 6

# Metric subregularity and $\omega(\cdot)$-normal regularity properties

This chapter is organized as follows. We recall the concept of normally $\omega(\cdot)$-regularity and metric subregularity respectivley in Section 6.1 and 6.2. Then, in section 6.3, we provide general sufficient conditions ensuring the preservation of normal $\omega(\cdot)$-regularity for generalized equations. Sections 6.4 and 6.5 focus on normally regular versions of Robin-Ursescu theorem. Preservation results in the line of Section 6.3 are also provided.

In the whole chapter, $(X, \|\cdot\|)$ is stand for a (real) normed space. The (topological) dual $X^\star$ of $X$ is endowed with its natural norm $\|\cdot\|_\star$ defined by

$$\|x^\star\|_\star := \sup_{x\in\mathbb{B}_X} \langle x^\star, x\rangle \quad \text{for all } x^\star \in X^\star,$$

where $\langle x^\star, x\rangle := x^\star(x)$.

## 6.1 Normally $\omega(\cdot)$-regular sets

This section is devoted to the class of normally $\omega(\cdot)$-regular sets introduced in [128]. Let us start by giving the definition of such sets:

**Definition 6.1.** Let $S$ be a subset of a normed space $(X, \|\cdot\|)$ and let $\omega : \mathbb{R}_+ \to \mathbb{R}$ be a function with $\omega(0) = 0$. Given a concept of normal cone $\mathcal{N}$ in $X$, one says that $S$ is *$\mathcal{N}$-normally $\omega(\cdot)$-regular* relative to an open set $V \subset X$ (with respect to the norm $\|\cdot\|$) whenever

$$\langle x^\star, x' - x\rangle \le \|x^\star\|_\star \omega(\|x' - x\|),$$

for all $x, x' \in S \cap V$ and for all $x^\star \in \mathcal{N}(S; x)$. When $V$ is the whole space $X$, we will just say that $S$ is $\mathcal{N}$-normally $\omega(\cdot)$-regular. It will be also convenient to say that $S$ is $C$-normally (resp. $F$-normally) $\omega(\cdot)$-regular whenever $\mathcal{N}$ is the normal cone $N^C$ (resp. $N^F$).

The class of normally $\omega(\cdot)$-regular sets contains the class of $(\sigma, \delta)$-subsmooth sets ([146]) which (roughly speaking) expresses a variational behavior of order one.

**Definition 6.2.** Let $S$ be a subset of a normed space $(X, \|\cdot\|)$ and let $\overline{x} \in S$. One says that $S$ is $(\sigma, \delta)$-*subsmooth* at $\overline{x}$ for some reals $\sigma, \delta > 0$ provided that

$$\langle x^\star, x' - x\rangle \leq \sigma \|x^\star\|_\star \|x' - x\|,$$

for all $x, x' \in S \cap B(\overline{x}, \delta)$ and for all $x^\star \in N^C(S; x)$.

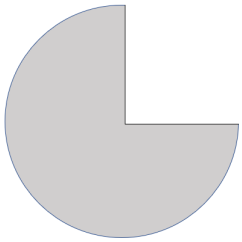


**Figure 6.1:** Illustration of the subsmooth set.

Clearly, if a subset $S$ of a normed space $X$ is $(\sigma, \delta)$-subsmooth at $\overline{x} \in S$, then it is $C$-normally $\omega(\cdot)$-regular relative to $B(\overline{x}, \delta)$ with $\omega(t) := \sigma t$ for every real $t \geq 0$. The above definition of $(\sigma, \delta)$-subsmooth property is quite related to the original definition of subsmooth sets by D. Aussel, A. Daniilidis and L. Thibault [129] where the authors declare that the set $S$ is subsmooth at $\overline{x}$ whenever for every $\varepsilon > 0$ we can find some $\delta > 0$ such that

$$\langle x^\star, x' - x\rangle \leq \varepsilon \|x^\star\|_\star \|x' - x\|,$$

for all $x, x' \in S \cap B(\overline{x}, \delta)$ and all $x^\star \in N^C(S; x)$. It is readily seen that the set $S$ is subsmooth at $\overline{x} \in S$ if and only if for all $\sigma > 0$ there is a real $\delta_\sigma > 0$ such that $S$ is $(\sigma, \delta_\sigma)$-subsmooth at $\overline{x} \in S$. Given a real $\sigma > 0$, there are $(\sigma, \delta)$-subsmooth sets $S$ in $\mathbb{R}^2$ for every $\delta > 0$ which fails to be Fréchet-Clarke regular at $\overline{x} \in S$ (i.e., $N^F(S; \overline{x}) = N^C(S; \overline{x})$). Such sets are not hemi-subsmooth at $\overline{x}$ (hence one-sided subsmooth at $\overline{x}$/subsmooth at $\overline{x}$ (see, e.g., [143, Chapter 8])). An example of such a set $S$ has been given by X.Y. Zheng and Q.H. He in their 2014 paper [146]: namely, for a real $\sigma > 0$ the set $S = \operatorname{epi} f$ where

$$f(x) := \begin{cases} 0 & \text{if } x \leq 0, \\ -\sigma x & \text{otherwise.} \end{cases}$$

is $(\sigma, \delta)$-subsmooth at $\overline{x} := (0, 0)$ for every real $\delta > 0$ for the 1-norm $\|\cdot\|_1$ in $\mathbb{R}^2$ and fails to be Fréchet-Clarke regular at $(0, 0)$. Indeed, set $S := \operatorname{epi} f$, then S is

1. for points $(x, y) \in S$ with $x \leq 0$, $y \geq 0$;
2. for points $(x, y) \in S$ with $x > 0$, $y \geq -\sigma x$.

Let $\delta > 0$, we are going to show that S is $(\sigma, \delta)$-subsmooth. Let $(x, y) \in S$ with $\|(x, y)\|_1 \leq \delta$ and let $(x^\star, y^\star) \in N^C(S; (0, 0))$ with $\|(x^\star, y^\star)\|_\infty \leq 1$. According to what precedes, there are $\lambda \geq 0$ and $t \geq 0$ such that $(x^\star, y^\star) = -(t\sigma, \lambda + t)$. Let us distinguish two cases:

- Case 1: $\|(x^\star, y^\star)\|_\infty = t\sigma$. We have $t \leq \lambda + t \leq t\sigma \leq 1$ and

$$x^\star x + y^\star y \leq t\sigma|x| + (\lambda + t)|y| \leq \sigma(|x| + |y|).$$

- Case 2: $\|(x^\star, y^\star)\|_\infty = \lambda + t$. It $y \geq 0$, then

$$x^\star x + y^\star y \leq t\sigma|x| \leq \sigma(|x| + |y|).$$

If $y < 0$, then $x > 0$ (having in mind that$(x, y) \in S$) and $-\sigma x \leq y < 0$. It follows that $|y| \leq \sigma|x|$, hence (since $x^\star \leq 0$)

$$x^\star x + y^\star y \leq (\lambda + t)|y| \leq |y| \leq \sigma|x|.$$

We conclude that S is $(\sigma, \delta)$-subsmooth at (0,0). In addition, we can check that

$$N^F(S; (0,0)) \neq N^C(S; (0,0)) = 0 \times \mathbb{R}_- + \mathbb{R}_-(\sigma, 1),$$

hence S is not hemi-subsmooth at $(0, 0)$. More precisely, the Fréchet normal cone $N^F(S; (0, 0))$ at a point $(0, 0) \in S$ is defined as the set of vectors $x$ such that

$$\limsup_{(x,y)\to(0,0),(x,y)\in S} \frac{\langle v, (x, y) - (0, 0)\rangle}{\|(x, y) - (0, 0)\|} \leq 0.$$

Thus, the vector $v = (v_1, v_2)$ must satisfy the limiting condition:

$$\limsup_{(x,y)\to(0,0),(x,y)\in S} \frac{\langle (v_1, v_2), (x, y)\rangle}{\|(x, y)\|} \leq 0.$$

Let's consider each region separately:

1. For $x \leq 0$ and $y \geq 0$: $\langle (v_1, v_2), (x, y)\rangle = v_1 x + v_2 y \leq 0$. For $v_1 \geq 0$, this is satisfied if $v_2 \leq 0$.

2. For $x > 0$ and $y \geq -\sigma x$: $\langle (v_1, v_2), (x, y)\rangle = v_1 x + v_2 y = x(v_1 - \sigma v_2) \leq 0$. Since $x > 0$, this implies $v_1 \leq \sigma v_2$.

Combining these conditions, for $v$ to be in the Fréchet normal cone:

$$\begin{cases} v_1 \geq 0, v_2 \leq 0, \\ v_1 \leq \sigma v_2. \end{cases}$$

The third condition $v_1 \leq \sigma v_2$ with $v_2 \leq 0$ means $v_1 \leq 0$, but this contradicts $v_1 \geq 0$ unless $v_1 = 0$. Thus:

$$N^F(S; (0,0)) = \{(0, v_2) \mid v_2 \leq 0\}.$$

On the other hand, the Clarke normal cone $N^C(S; (0, 0))$ at a point $(0, 0) \in S$ is defined as the set of vectors $v$ such that

$$v \in \operatorname{co}\left\{\lim_{i\to\infty} \nabla g(x_i, y_i) \mid (x_i, y_i) \to (0, 0), (x_i, y_i) \in S\right\},$$

where co denotes the convex hull.

1. For $x \leq 0$, $y \geq 0$, the subdifferential is vertical since the function is constant.

2. For $x > 0$, $y \geq -\sigma x$, the subdifferential comes from the line with slope $-\sigma$.

The Clarke normal cone at $(0,0)$ is thus the convex hull of the union of these subdifferentials:

$$N^C(S;(0,0)) = \{0\} \times \mathbb{R}_- + \mathbb{R}_-(\sigma, 1).$$

Clearly,

$$N^F(S;(0,0)) \subset N^C(S;(0,0)),$$

but they are not equal since $N^C(S;(0,0))$ includes additional directions involving $(\sigma, 1)$ that are not present in $N^F(S;(0,0))$. Thus,

$$N^F(S;(0,0)) \neq N^C(S;(0,0)) = \{0\} \times \mathbb{R}_- + \mathbb{R}_-(\sigma, 1).$$

The class of normally $\omega(\cdot)$-regular sets also contains the class of uniform prox-regular sets in Hilbert spaces. Prox-regularity has been well recognized as a fundamental tool in variational analysis which allows to go beyond convexity property in many topics of modern analysis (see the survey [100] and the references therein).

An $r$-prox-regular closed set $S$ of a Hilbert space $\mathcal{H}$ is clearly $C$-normally $\omega(\cdot)$-regular relative to the whole space $\mathcal{H}$ with $\omega(t) := \frac{1}{2r}t^2$ for every real $t \geq 0$. It is also known (and not difficult to check) that the (nonempty closed) set $S \subset \mathcal{H}$ is $r$-prox-regular if and only if (5.1) holds for all $x, x' \in S$ with $\|x' - x\| < 2r$ and for all $v \in N^F(S;x) \cap \mathbb{B}_{\mathcal{H}}$. The $r$-prox-regularity of the (nonempty closed) set $S$ in $\mathcal{H}$ is often defined by means of the characterization property requiring that for any $x \in S$ and any nonzero $v \in N^F(S;x)$ with $\|v\| \leq 1$, one has $x \in \mathrm{Proj}_S(x + tv)$ for every non-negative real $t \leq r$. The closed subset $S$ in $\mathcal{H}$ is also known to be $r$-prox-regular if and only if $\mathrm{Proj}_S(x)$ is a singleton for every $x \in U_r(S) := \{d_S < r\}$ and the induced (single-valued) mapping is continuous on $U_r(S)$.

Given a real $r > 0$, we can check (see, e.g., [127, Theorem 4.1]) that the epigraph of an $r^{-1}$-semiconvex continuous function $f : \mathcal{H} \to \mathbb{R}$ (that is, $f + \frac{1}{2r}\|\cdot\|^2$ is convex continuous) is $r$-prox-regular. Recall also (see [143, Proposition 15.35]) that a mapping $F : \mathcal{H} \to \mathcal{H}$ (resp. a function $F : \mathcal{H} \to \mathbb{R}$) which is differentiable with $L$-Lipschitz derivative has its graph (resp. epigraph) $1/L$-prox-regular.

It is known (see, e.g., the survey [153]) that a strongly convex set is nothing but the intersection of a family of closed balls with common radius $R$ (hence convex and bounded). Of course, a strongly convex set is $\omega(\cdot)$-normally regular with $\omega(t) := -\frac{t^2}{2R}$ for all $t \in [0, +\infty[$.

The class of $\mathcal{N}$-normally $\omega(\cdot)$-regular sets is also quite related to other previous concepts of nonsmooth sets: $C^{1,\varphi}$-regularity for functions and sets ([159]), super-regularity ([160]), Clarke regularity ([161]). Before closing this subsection devoted to nonsmooth sets, let us give a result ensuring the normal $\omega(\cdot)$-regularity for the graph of a multimapping $M$ given as a sum of a mapping and a set. For the proof, we refer the reader to [128, Theorem 4.4].

**Proposition 6.3.** *Let $f : X \to Y$ be a mapping between two normed spaces $X$ and $Y$ and let $S$ be a subset of $Y$ which is $C$-normally $\omega(\cdot)$-regular relative to the whole space*

*$Y$ for some nondecreasing function $\omega(\cdot):\mathbb{R}_+\to\mathbb{R}_+$ with $\omega(0)=0$. Assume that:*
*(i) there exists a real $K\geq 0$ such that*

$$\|f(x)-f(x')\|\leq K\|x-x'\| \quad \textit{for all } x,x'\in X;$$

*(ii) the mapping $f$ is differentiable on $X$ and there exists a real $L\geq 0$ such that*

$$\|Df(x)-Df(x')\|\leq L\,\|x-x'\| \quad \textit{for all } x,x'\in X.$$

*Let $\|\cdot\|_{X\times Y}$ a norm on $X\times Y$ associated to the product topology (of the norm topologies of $X$ and $Y$) and such that*

$$\max(\|x-x'\|,\|y-y'\|)\leq\|(x,y)-(x',y')\|_{X\times Y} \quad \textit{for all } x,x'\in X, \textit{all } y,y'\in Y.$$

*Then, the graph of the multimapping $M(\cdot):=f(\cdot)+S$ is $C$-normally $\rho(\cdot)$-regular where $\rho:\mathbb{R}_+\to\mathbb{R}_+$ is defined by*

$$\rho(t):=\omega\big(\max(1,K)t\big)+\frac{Lt^2}{2} \quad \textit{for all } t\in\mathbb{R}_+.$$

## 6.2 Metric subregularity

This section is devoted to the necessary background on metric subregularity theory needed in the paper. For more details on this topic, we refer to [107, 120, 124] and the references therein.

Let $M:X\rightrightarrows Y$ be a multimapping from a normed space $X$ to another normed space $Y$ and let $(\overline{x},\overline{y})\in\operatorname{gph}M$. One says that $M$ is *metrically subregular* at $\overline{x}$ for $\overline{y}$ whenever there exist a real $\gamma\geq 0$ and a neighborhood $U$ of $\overline{x}$ such that

$$d\big(x,M^{-1}(\overline{y})\big)\leq\gamma d\big(\overline{y},M(x)\big) \quad \text{for all } x\in U. \tag{6.1}$$

The *modulus of metric subregularity* $\operatorname{subreg}[M](\overline{x}\mid\overline{y})$ of $M$ at $\overline{x}$ for $\overline{y}$ is defined as the infimum of all $\gamma\in[0,+\infty[$ for which there is a neighborhood $U$ of $\overline{x}$ such that the inequality (6.1) is fulfilled.

The multimapping $M^{-1}:Y\rightrightarrows X$ denotes the inverse of the multimapping $M$ defined by $M^{-1}(y):=\{x\in X:y\in M(x)\}$. And the core of $M(X)$ is defined by

$$\operatorname{core}M(X):=\{y\in M(X):\forall y'\in Y,\exists r>0,y+ry'\in M(X)\}.$$

One important property of $\operatorname{core}M(X)$ is that it contains all points in the interior of $M(X)$), meaning every point in $\operatorname{int}M(X)$ is also in $\operatorname{core}M(X)$. In fact, the latter inclusion $\bar{y}\in\operatorname{core}M(X)$ guarantees the existence of a positive constant $c$ such that

$$d\big(x,M^{-1}(y)\big)\leq(c-\|y-\overline{y}\|)^{-1}(1+\|x-\overline{x}\|)d\big(y,M(x)\big) \quad \text{for all } x\in X, y\in B(\overline{y},c).$$

The following proposition is due to X.Y. Zheng and K.F. Ng ([140]). It provides important quantitative properties on a multimapping $M$ which fails to fulfill the metric subregularity inequality (6.1) for some point $x$, that is,

$$\gamma d\big(\overline{y},M(x)\big)<d\big(x,M^{-1}(\overline{y})\big).$$

**Proposition 6.4.** *Let $M : X \rightrightarrows Y$ be a multimapping with closed graph between two Banach spaces $X$ and $Y$ and let $\overline{y} \in Y$. Assume that there exist $x \in X$ and two reals $\gamma, r \in ]0, +\infty[$ such that*

$$\gamma d\big(\overline{y}, M(x)\big) < r < d\big(x, M^{-1}(\overline{y})\big).$$

*Then, for all reals $\eta, \varepsilon \in ]0, +\infty[$, there exist $(u, v) \in \operatorname{gph} M$ satisfying:*
*(i)* $\|u - x\| < r$ *and* $0 < \|v - \overline{y}\| < \min\{\frac{r}{\gamma}, d(\overline{y}, M(u)) + \varepsilon\}$;
*(ii)* $\|v - \overline{y}\| \leq \|b - \overline{y}\| + \frac{1}{\gamma}\big(\|a - u\| + \eta\|b - v\|\big)$ *for all* $(a, b) \in \operatorname{gph} M$;
*(iii)* $(0, 0) \in \{0\} \times \partial\|\cdot\|(v - \overline{y}) + \frac{1}{\gamma}\big(\mathbb{B}_{X^\star} \times \eta\mathbb{B}_{Y^\star}\big) + N^C\big(\operatorname{gph} M; (u, v)\big)$.

With the above result at hands, Zheng and Ng provide in [140] an important estimate for the modulus of subregularity:

**Proposition 6.5.** *Let $M : X \rightrightarrows Y$ be a multimapping with closed graph between two Banach spaces and let $(\overline{x}, \overline{y}) \in \operatorname{gph} M$. Then, one has*

$$\operatorname{subreg}[M](\overline{x}|\overline{y}) \leq \inf_{\varepsilon>0} \sup \left\{ \sup_{b \in \mathbb{B}_{X^\star}} \|D_C M(x, y)^{-1}(b^\star)\| : \begin{cases} x \in B(\overline{x}, \varepsilon) \setminus M^{-1}(\overline{y}) \\ y \in M(x) \cap B(\overline{y}, \varepsilon) \end{cases} \right\}.$$

**Remark 6.6.** *We point out that Proposition 6.4 (resp. Proposition 6.5) also holds for the Mordukhovich limiting normal cone $N^L\big(\operatorname{gph} M; (u, v)\big)$ (resp. the coderivative $D_L M(x, y)^{-1}(b^\star)$) in the context of Asplund spaces $X$ and $Y$.* ■

## 6.3 Preservation of normal $\omega(\cdot)$-regularity of sets under metric subregularity

We first give sufficient conditions ensuring the normal $\omega(\cdot)$-regularity (so in particular the prox-regularity) for sets of the form

$$\{x \in X : \overline{y} \in M(x)\} =: M^{-1}(\overline{y}), \tag{6.2}$$

where $M : X \rightrightarrows Y$ is a multimapping and $\overline{y} \in Y$. It should be noted that a large number of sets can be rewritten as in (6.2). For instance, this is the case of constraint sets, namely sets of the form

$$\{f_1 \leq 0, \ldots, f_m \leq 0, f_{m+1} = 0, \ldots, f_{m+n} = 0\} \tag{6.3}$$

which is nothing but $M^{-1}(\overline{y})$ with $\overline{y} = 0$ and

$$M(x) := -\big(f_1(x), \ldots, f_m(x), f_{m+1}(x), \ldots, f_{m+n}(x)\big) + ]-\infty, 0]^m \times \{0_{\mathbb{R}^n}\}. \tag{6.4}$$

Another case which deserves to be stated lies in the intersection of finitely many sets. Indeed, for any sets $S_1, \ldots, S_n \subset X$ we note that

$$\bigcap_{k=1}^{n} S_k = \{x \in X : 0 \in M(x)\} \quad \text{with } M(x) := -(x, \ldots, x) + \prod_{k=1}^{n} S_k. \tag{6.5}$$

More generally, for a single-valued mapping $f : X \to Y$ and $B \subset Y$, the inverse image $f^{-1}(B)$ reduces to $M^{-1}(\overline{y})$ with $\overline{y} = 0$ and

$$M(x) := -f(x) + B. \tag{6.6}$$

Observe that in the above cases (6.4), (6.6) and (6.5), the involved multimapping $M(\cdot)$ is nothing but the translation of a fixed set. Such multimappings will be at the heart of Section 6.5.

Let us mention that the uniform prox-regularity of level and sublevel sets has been studied in [125, 127, 128, 152]. In [125], J.-P. Vial established the uniform prox-regularity (called weak convexity therein) of a sublevel set $S = \{f \leq 0\}$ in $\mathbb{R}^n$ of a weakly convex function $f$ satisfying

$$\inf_{\zeta \in \partial f(x), x \in \operatorname{bdry} S} \|\zeta\| > 0.$$

We also mention the work [152] where the author gives sufficient conditions ensuring the prox-regularity of the set $S' = \{f_1 \leq 0, \ldots, f_m \leq 0\}$ with $f_i \in C^2(\mathbb{R}^n)$ and for some positive constants $\alpha, \beta, M$

$$\alpha \leq |\nabla f_i(x)| \leq \beta \quad \text{and} \quad |D^2 f_i(x)| \leq M.$$

In [127], the authors establish the prox-regularity of the set $S'$ for possibly nonsmooth functions $f_i$ defined on a Hilbert space $\mathcal{H}$ with their $C$-subdifferentials enjoying an hypomonotone property and under a generalized Slater's condition, namely the existence of a real $\delta > 0$ such that for every $x \in \operatorname{bdry} S'$ there is a unit vector $v_x$ such that for all $k = 1, \ldots, m$ and for all $\zeta \in \partial_C f_k(x)$,

$$\langle \zeta, v_x \rangle \leq -\delta.$$

The prox-regularity of the level $L := \{F = 0\}$ of a smooth function $F$ is also developed in [127], under an openness condition, say for some real $\delta > 0$

$$\delta \mathbb{B} \subset DF(x)(\mathbb{B}) \quad \text{for all } x \in \operatorname{bdry} L.$$

Regarding the prox-regularity of a constraint set with finitely many inequality and equality constraints (see (6.3)), let us say that is has been studied in [127] and [128] through two approaches, namely two different openness conditions either on the derivatives or on a perturbation of the involved graphs.

Sublevel and level sets can be seen as a particular case of inverse image. The prox-regularity of the inverse image $f^{-1}(B)$ has been first studied in the survey [100] under a theoretical condition

$$N^F(f^{-1}(B); x) \cap \mathbb{B} \subset Df(x)^\star(N(B; f(x) \cap \gamma \mathbb{B})$$

and also investigated in the papers [127] and [128]. Subsmoothness of inverse images have been studied in [137]. Besides the prox-regularity of inverse images, let us point out that the direct image case has been also examinated in [100, 128, 133].

Let us also mention that intersection of two uniformly prox-regular sets, say $S_1$ and $S_2$ may fail to be prox-regular, even in $\mathbb{R}^2$ (see, e.g., [100]). To the best of our knowledge,

prox-regularity of the intersection $S_1 \cap S_2$ holds under anyone of the following conditions:
- the strong convexity of either $S_1$ or $S_2$;
- an openness condition on involved tangent cones, namely the existence of a real $s > 0$ such that for all $\overline{x} \in \mathrm{bdry}(S_1 \cap S_2)$,

$$s\mathbb{B} \subset T(S_1; x_1) \cap \mathbb{B} - T(S_2; x_2) \cap \mathbb{B}, \quad x_i \in S_i, \text{ near } \overline{x},$$

- an openness condition on $S_1, S_2$, more precisely, the existence of $\alpha, \beta, s > 0$ such that for all $\overline{x} \in \mathrm{bdry}(S_1 \cap S_2)$,

$$\beta\mathbb{B}_{\mathcal{H}^6} \subset -\Delta_{B[(\overline{x},\overline{x},\overline{x}),\alpha]^2} + \Delta_{\mathcal{H}^3} \times \mathcal{H} \times S_1 \times S_2, \tag{6.7}$$

where $\Delta_{E^m} := \{(x, \ldots, x) : x \in E\} \subset E^m$.

For a detailed overview on the preservation of nonsmooth sets under set operations, we refer the reader to the recent book by L. Thibault [143] (see the comments at the end of Chapters 8, 15 and 16).

We start our study of preservation of $\omega(\cdot)$-regularity in this section by giving an important estimate for the normal cone $N(M^{-1}(\overline{y}); \overline{x})$ under a metric subregularity assumption on the multimapping $M$. Given a multimapping $M : X \rightrightarrows Y$ between two normed spaces $X$ and $Y$, we define $\Delta_M : X \times Y \to \mathbb{R} \cup \{+\infty\}$ by setting

$$\Delta_M(x, y) := d\big(y, M(x)\big) \quad \text{for all } (x, y) \in X \times Y.$$

It is known that a certain Lipschitz behavior of $\Delta_M(\cdot, y)$ is equivalent to the Aubin-Lipschitz property of $M$ (see, e.g., [124, Proposition 7.7]). Subgradients of the function $\Delta_M$ have been studied in the literature by L. Thibault [123], B.S. Mordukhovich and N.M. Nam ([162, 163]) and M. Bounkhel [155, Chapter 4]. It is also known (see, e.g., [124, Proposition 4.162]) that

$$\partial_F \Delta_M(\overline{x}, \overline{y}) = N^F\big(\mathrm{gph}\, M; (\overline{x}, \overline{y})\big) \cap (X^\star \times \mathbb{B}_{Y^\star})$$

and

$$N^F\big(\mathrm{gph}\, M; (\overline{x}, \overline{y})\big) = \mathbb{R}_+ \partial_F \Delta_M(\overline{x}, \overline{y}).$$

Similar equalities hold true for the Mordukhovich-limiting normal cone and subdifferentials whenever $X$ and $Y$ are Banach spaces and $\mathrm{gph}\, M$ is closed.

**Lemma 6.7.** *Let $M : X \rightrightarrows Y$ be a multimapping between two normed spaces $X$ and $Y$ and let $(\overline{x}, \overline{y}) \in \mathrm{gph}\, M$. The following hold:*
*(a) One has*

$$\partial_F d\big(\overline{y}, M(\cdot)\big)(\overline{x}) \times \{0\} \subset \{0\} \times \mathbb{B}_{Y^\star} + \partial_C \Delta_M(\overline{x}, \overline{y})$$

*and*

$$\partial_F d\big(\overline{y}, M(\cdot)\big)(\overline{x}) \subset D_C M(\overline{x}, \overline{y})(\mathbb{B}_{Y^\star}).$$

*(b) Assume that $X$ and $Y$ are Asplund spaces. Then, one has*

$$\partial_F d\big(\overline{y}, M(\cdot)\big)(\overline{x}) \subset D_L M(\overline{x}, \overline{y})(\mathbb{B}_{Y^\star}).$$

*If in addition $\Delta_M$ is lower semicontinuous near $(\overline{x}, \overline{y})$, then one has*

$$\partial_F d\big(\overline{y}, M(\cdot)\big)(\overline{x}) \times \{0\} \subset \{0\} \times \mathbb{B}_{Y^\star} + \partial_L \Delta_M(\overline{x}, \overline{y}). \tag{6.8}$$

(*c*) *Assume that there exists a real* $\gamma \geq 0$ *and a real* $\delta > 0$ *such that*

$$d\big(x, M^{-1}(\overline{y})\big) \leq \gamma d\big(\overline{y}, M(x)\big) \quad \textit{for all } x \in B(\overline{x}, \delta).$$

*Then, one has*

$$N^F\big(M^{-1}(\overline{y}); \overline{x}\big) \cap \mathbb{B}_{X^\star} = \partial_F d\big(\cdot; M^{-1}(\overline{y})\big)(\overline{x}) \subset \gamma \partial_F d\big(\overline{y}, M(\cdot)\big)(\overline{x}). \tag{6.9}$$

*Proof.* (*a*) We denote $\|\cdot\|_1$ the 1-norm on $X \times Y$. Let $x^\star \in \partial_F d\big(\overline{y}, M(\cdot)\big)(\overline{x})$. Define $\varphi, \theta : X \times Y \to \mathbb{R}$ by setting for all $(x, y) \in X \times Y$

$$\varphi(x, y) := \Delta_M(x, y) + \|y - \overline{y}\| \quad \text{and} \quad \theta(x, y) := \psi_{\operatorname{gph} M}(x, y) + \|y - \overline{y}\|,$$

It is readily seen that

$$\varphi \leq \theta \quad \text{and} \quad \varphi(\overline{x}, \overline{y}) = \theta(\overline{x}, \overline{y}) = 0. \tag{6.10}$$

Fix any real $\varepsilon > 0$. Through the definition of $F$-subgradients, we can find some real $\eta > 0$ such that

$$\begin{aligned} \langle (x^\star, 0), (x', y') - (\overline{x}, \overline{y}) \rangle &\leq d_{M(x')}(\overline{y}) - d_{M(\overline{x})}(\overline{y}) + \varepsilon \|x' - \overline{x}\| \\ &\leq \Delta_M(x', y') + \|y' - \overline{y}\| + \varepsilon \|(x', y') - (\overline{x}, \overline{y})\|_1 \\ &\leq \varphi(x', y') - \varphi(\overline{x}, \overline{y}) + \varepsilon \|(x', y') - (\overline{x}, \overline{y})\|_1, \end{aligned} \tag{6.11}$$

for all $x' \in B(\overline{x}, \eta)$ and all $y' \in Y$. It follows that (see Theorem 2.16)

$$(x^\star, 0) \in \partial_F \varphi(\overline{x}, \overline{y}) \subset \partial_C \varphi(\overline{x}, \overline{y}) \subset \{0\} \times \mathbb{B}_{Y^\star} + \partial_C \Delta_M(\overline{x}, \overline{y}).$$

The first inclusion in (a) is then established. Regarding the second inclusion, it suffices to note that (thanks to (6.11) and (6.10))

$$\langle (x^\star, 0), (x', y') - (\overline{x}, \overline{y}) \rangle \leq \theta(x', y') - \theta(\overline{x}, \overline{y}) + \varepsilon \|(x', y') - (\overline{x}, \overline{y})\|_1,$$

for all $x' \in B(\overline{x}, \eta)$ and all $y' \in Y$ to get (as above)

$$(x^\star, 0) \in \partial_F \theta(\overline{x}, \overline{y}) \subset \partial_C \theta(\overline{x}, \overline{y}) \subset \{0\} \times \mathbb{B}_{Y^\star} + N^C\big(\operatorname{gph} M; (\overline{x}, \overline{y})\big).$$

(*b*) The proof is similar to (*a*) (see Theorem 2.21 for the inclusion (6.8)).
(*c*) Let $x^\star \in N^F\big(M^{-1}(\overline{y}); \overline{x}\big) \cap \mathbb{B}_{X^\star} = \partial_F d_{M^{-1}(\overline{y})}(\overline{x})$ (see (2.4)). Fix any real $\varepsilon > 0$. By definition of $F$-subgradients, there is a real $\eta > 0$ with $\eta < \delta$ such that

$$\langle x^\star, x' - \overline{x} \rangle \leq d_{M^{-1}(\overline{y})}(x') - d_{M^{-1}(\overline{y})}(\overline{x}) + \gamma \varepsilon \|\overline{x} - x'\|,$$

for every $x' \in B(\overline{x}, \eta)$. It follows from this

$$\big\langle \gamma^{-1} x^\star, x' - \overline{x} \big\rangle \leq d_{M(x')}(\overline{y}) - d_{M(\overline{x})}(\overline{y}) + \varepsilon \|\overline{x} - x'\| \quad \text{for all } x' \in B(\overline{x}, \eta)$$

and this translates the inclusion $x^\star \in \gamma \partial_F d\big(\overline{y}, M(\cdot)\big)(\overline{x})$. □

We easily derive from (c) of the above proposition the following result:

**Proposition 6.8.** *Let $M : X \rightrightarrows Y$ be a multimapping between two normed spaces and let $\overline{y} \in Y$ with $M^{-1}(\overline{y}) \neq \emptyset$. Assume that:*
*(i) There exists a real $\gamma \geq 0$ such that for each $\overline{x} \in M^{-1}(\overline{y})$ there is a real $\delta > 0$ satisfying*

$$d\big(x, M^{-1}(\overline{y})\big) \leq \gamma d\big(\overline{y}, M(x)\big) \quad \text{for all } x \in B(\overline{x}, \delta).$$

*(ii) There exists a function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$ and a real $c \geq 0$ such that for all $x, x' \in M^{-1}(\overline{y})$ and for all $x^\star \in \partial_F d\big(\overline{y}, M(\cdot)\big)(x)$,*

$$\langle x^\star, x' - x\rangle \leq \big(\|x^\star\| + c\big)\omega(\|x' - x\|).$$

*Then, the set $M^{-1}(\overline{y})$ is $F$-normally $\rho(\cdot)$-regular with $\rho(\cdot) := (1 + \gamma c)\omega(\cdot)$.*

*Proof.* If suffices to observe that for any $x, x' \in M^{-1}(\overline{y})$ and any $x^\star \in N^F\big(M^{-1}(\overline{y}); x\big) \cap \mathbb{B}_{X^\star}$, we have (see Lemma 6.7) $\gamma^{-1}x^\star \in \partial_F d\big(\overline{y}, M(\cdot)\big)(x)$ which gives through $(ii)$

$$\langle x^\star, x' - x\rangle \leq \gamma(\gamma^{-1}\|x^\star\| + c)\omega(\|x' - x\|) \leq (1 + \gamma c)\omega(\|x' - x\|).$$

The proof is complete. □

**Remark 6.9.** *For the above $F$-normal $\omega(\cdot)$-regularity of $M^{-1}(\overline{y})$, we claim that $(ii)$ in Proposition 6.8 can be replaced by the assumption of $C$-normal $\omega(\cdot)$-regularity of* $\operatorname{gph} M$ *at $(\overline{x}, \overline{y}) \in \operatorname{gph} M$. We endow $X \times Y$ with the norm $\|\cdot\|_2$ whose dual norm $\|\cdot\|_\star$ is known (and easily seen) to satisfy*

$$\|(x^\star, y^\star)\|_\star \leq \|x^\star\| + \|y^\star\| \quad \text{for all } (x^\star, y^\star) \in X^\star \times Y^\star. \tag{6.12}$$

*Fix any $x, x' \in M^{-1}(\overline{y})$ and any $x^\star \in N^F\big(M^{-1}(\overline{y}); x\big) \cap \mathbb{B}_{X^\star}$. According to Lemma 6.7(c)-(a), we can find $y^\star \in \mathbb{B}_{Y^\star}$ such that*

$$(x^\star, \gamma y^\star) \in N^C(\operatorname{gph} M; (x, y))$$

*and this allows us to use the $C$-normal $\omega(\cdot)$-regularity of* $\operatorname{gph} M$ *at $(\overline{x}, \overline{y}) \in \operatorname{gph} M$ to get*

$$\langle (x^\star, \gamma y^\star), (x', \overline{y}) - (x, \overline{y})\rangle \leq \|(x^\star, \gamma y^\star)\|_\star \omega(\|(x', \overline{y}) - (x, \overline{y})\|_2).$$

*This and the inequality* (6.12) *easily give*

$$\langle x^\star, x' - x\rangle \leq (\|x^\star\| + \gamma\|y^\star\|)\omega(\|(x' - x, 0)\|) \leq (1 + \gamma)\omega(\|x' - x\|)$$

*which translates the $F$-normal $(1 + \gamma)\omega(\cdot)$-regularity of the set $M^{-1}(\overline{y})$ at the point $\overline{x}$.*

*We point out that such a remark obviously holds for the $L$-normal $\omega(\cdot)$-regularity of* $\operatorname{gph} M$ *at $(\overline{x}, \overline{y})$ whenever $X$ and $Y$ are Asplund spaces.* ■

**Remark 6.10.** *Putting together (c) and the second inclusion of (a) in Lemma 6.7, we easily see that the $F$-normal $\omega(\cdot)$-regularity property for the set $M^{-1}(\overline{y})$ at $\overline{x} \in M^{-1}(\overline{y})$ also holds true under the following inequality*

$$\langle x^\star, x' - x\rangle \leq (\|x^\star\| + c)\omega(\|x' - x\|),$$

*for all $x, x' \in M^{-1}(\overline{y})$ and all $x^\star \in D_C M(x, \overline{y})(\mathbb{B}_{Y^\star})$.* ■

We pass now to the normal $\omega(\cdot)$-regularity for a (solution set of) generalized equation, say $S := \{x \in X : f(x) \in F(x)\}$ with $f$ (resp. $F$) single-valued (resp. set-valued). Setting $M(x) := -f(x) + F(x)$ for every $x \in X$, we observe that $S = M^{-1}(0)$ so the above proposition gives sufficient conditions ensuring the desired normal $\omega(\cdot)$-regularity of the set $S$. For the above multimapping $M$, note that $(i)$ and $(ii)$ of Proposition 6.8 can be rewritten as:
$(i')$: There exists a real $\gamma \geq 0$ such that for each $\overline{x} \in S$ there is a real $\delta > 0$ satisfying

$$d(x, S) \leq \gamma d\big(f(x), F(x)\big) \quad \text{for all } x \in B(\overline{x}, \delta). \tag{6.13}$$

$(ii')$: There exists a function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$ and a real $c \geq 0$ such that for all $x, x' \in S$ and all $x^\star \in \partial_F d\big(f(\cdot), F(\cdot)\big)(x)$

$$\langle x^\star, x' - x\rangle \leq \big(\|x^\star\| + c\big)\omega(\|x' - x\|).$$

According to Remark 6.9, we know that $(ii')$ above could be replaced by the $C$-normal $\omega(\cdot)$-regularity of $\operatorname{gph} M$ (and by the $L$-regularity in the context of Asplund spaces). As shown by the next result, such a regularity can only be required for $\operatorname{gph} F$.

**Proposition 6.11.** *Let $X, Y$ be normed spaces, $F : X \rightrightarrows Y$ be a multimapping and let $f : X \to Y$ be a mapping. Assume that $S := \{x \in X : f(x) \in F(x)\} \neq \emptyset$ along with:*
*(i) There exists a real $\gamma \geq 0$ such that for each $\overline{x} \in S$, there exists a real $\delta > 0$ satisfying*

$$d(x, S) \leq \gamma d\big((x, f(x)), \operatorname{gph} F\big) \quad \textit{for all } x \in B(\overline{x}, \delta).$$

*(ii) There exists a nondecreasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$ such that the set $\operatorname{gph} F$ is $C$-normally $\omega(\cdot)$-regular with respect to the product norm $\|\cdot\|_2$ on $X \times Y$.*
*(iii) The mapping $f$ is $K$-Lipschitz continuous on $X$ for some real $K \geq 0$ and differentiable on $X$ with $L$-Lipschitz continuous derivative $Df$ for some real $L \geq 0$.*

*Then, the set $S$ is $F$-normally $\widehat{\omega}(\cdot)$-regular with $\widehat{\omega} : \mathbb{R}_+ \to \mathbb{R}_+$ defined by*

$$\widehat{\omega}(t) := \gamma\big(\omega(\kappa t) + \frac{L}{2}t^2\big) \quad \textit{for all } t \in \mathbb{R}_+,$$

*where $\kappa := (1 + K^2)^{1/2}$.*

*Proof.* We denote $\|\cdot\|_2$ the 2-norm on $X \times Y$ and $\mathbb{B}_\star$ the unit ball of $(X \times Y)^\star$. First, let us define the function $\theta : X \times Y \to \mathbb{R}$ by setting

$$\theta(x, y) := d\big((x, y), \operatorname{gph} F\big) \quad \text{for all } (x, y) \in X \times Y.$$

We also need to consider the mapping $\varphi : X \to X \times Y$ defined by

$$\varphi(x) := \big(x, f(x)\big) \quad \text{for all } x \in X$$

which is obviously $\kappa$-Lipschitz continuous and differentiable at each point $x \in X$ with

$$D\varphi(x)(h) = \big(h, Df(x)(h)\big) \quad \text{for all } h \in X.$$

Let $x, x' \in S$ and $u^\star \in N^F(S; x) \cap \mathbb{B}_{X^\star}$. Putting together the equalities $d_S(x) = 0 = (\theta \circ \varphi)(x)$, the inequality given by $(i)$ and the fact that the Fréchet subdifferential is always included in the Clarke's one, we get

$$u^\star \in \partial_F d_S(x) \subset \gamma \partial_F(\theta \circ \varphi)(x) \subset \gamma \partial_C(\theta \circ \varphi)(x). \tag{6.14}$$

Using Proposition 2.17 and the inclusion (2.3.1) we obtain

$$\partial_C(\theta\circ\varphi)(x) = D\varphi(x)^\star\big(\partial_C\theta(\varphi(x))\big) \subset D\varphi(x)^\star\Big(N^C\big(\operatorname{gph} F;(x,f(x))\big)\cap\mathbb{B}_\star\Big). \quad (6.15)$$

The inclusions (6.15) and (6.14) obviously give some $(x^\star,y^\star)\in N^C\big(\operatorname{gph} F;(x,f(x))\big)\cap \mathbb{B}_{X^\star\times Y^\star}$ such that $u^\star=\gamma D\varphi(x)^\star(x^\star,y^\star)$. We then have

$$\big\langle\gamma^{-1}u^\star,h\big\rangle = D\varphi(x)^\star(x^\star,y^\star)(h) = \langle(x^\star,y^\star),(h,Df(x)(h))\rangle \quad \text{for all } h\in X.$$

Now, we observe that

$$\begin{aligned}\big\langle\gamma^{-1}u^\star,x'-x\big\rangle &= \langle(x^\star,y^\star),(x'-x,Df(x)(x'-x)\rangle\\ &= \langle(x^\star,y^\star),(x',f(x'))-(x,f(x))\rangle + \langle y^\star,-f(x')+f(x)+Df(x)(x'-x)\rangle\,.\end{aligned}$$

Combining $(ii)$, the inclusions $(x',f(x')),(x,f(x))\in\operatorname{gph}F$ and $(x^\star,y^\star)\in\mathbb{B}_\star$ and $(iii)$, we get

$$\begin{aligned}\langle(x^\star,y^\star),(x',f(x'))-(x,f(x))\rangle &\le \omega\big(\|(x',f(x'))-(x,f(x))\|_2\big)\\ &\le \omega\big(\kappa\|x'-x\|\big).\end{aligned}$$

On the other hand, we have

$$\langle y^\star,-f(x')+f(x)+Df(x)(x'-x)\rangle = \left\langle y^\star, Df(x)(x'-x)-\int_0^1 Df(x+t(x'-x))(x'-x)dt\right\rangle,$$

hence (noticing that $\|y^\star\|\le 1$)

$$\begin{aligned}\langle y^\star,-f(x')+f(x)+Df(x)(x'-x)\rangle &\le \int_0^1 \|Df(x)-Df(x+t(x'-x))\|\|x'-x\|dt\\ &\le L\|x'-x\|^2\int_0^1 tdt = \frac{L}{2}\|x'-x\|^2.\end{aligned}$$

We conclude that

$$\langle u^\star,x'-x\rangle \le \gamma\omega\big(\kappa\|x'-x\|\big)+\frac{\gamma L}{2}\|x'-x\|^2$$

and this translates the fact that the set $S$ is $F$-normally $\widehat{\omega}(\cdot)$-regular. □

**Remark 6.12.** *In view of* (6.15)*, it is clear that we can replace (ii) by (ii"):*
*(ii") There exists a nondecreasing function $\omega:\mathbb{R}_+\to\mathbb{R}_+$ with $\omega(0)=0$ such that*

$$\langle(x^\star,y^\star),(x',f(x'))-(x,f(x))\rangle \le \omega(\|(x',f(x'))-(x,f(x))\|),$$

*for all $x,x'\in S$ and all $(x^\star,y^\star)\in\partial_C d(\cdot,\operatorname{gph}F)(x,f(x))$.* ■

If $F$ is constant in Proposition 6.11 (say $F\equiv B\subset Y$) we easily see that $(i)$ (of Proposition 6.11) becomes

$$d(x,S)\le\gamma d\big((x,f(x)),X\times B\big) \quad \text{for all } x\in B(\overline{x},\delta).$$

This inequality obviously holds for the distance $d$ on $X \times Y$ associated to the norm $\|\cdot\|_2$ whenever

$$d\big(x, f^{-1}(B)\big) \le \gamma d\big(f(x), B\big) \quad \text{for all } x \in B(\overline{x}, \delta),$$

which is exactly the estimate in (6.13). Therefore, a direct application of Proposition 6.11 gives the normal $\widehat{\omega}(\cdot)$-regularity of $S = \{x \in X : f(x) \in B\} = f^{-1}(B)$. In fact, a direct and similar proof for the constant case $F \equiv B$ allows to slightly improve the above modulus $\widehat{\omega}$ of normal regularity:

**Proposition 6.13.** *Let $X, Y$ be normed spaces, $B \subset Y$ and let $f : X \to Y$ be a mapping. Assume that $S := f^{-1}(B) \neq \emptyset$ along with:*
*(i) There exists a real $\gamma \ge 0$ such that for each $\overline{x} \in S$, there exists a real $\delta > 0$ satisfying*

$$d\big(x, f^{-1}(B)\big) \le \gamma d\big(f(x), B\big) \quad \textit{for all } x \in B(\overline{x}, \delta).$$

*(ii) There exists a nondecreasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$ such that the set $B$ is $C$-normally $\omega(\cdot)$-regular.*
*(iii) The mapping $f$ is $K$-Lipschitz continuous on $X$ for some real $K \ge 0$ and differentiable on $X$ with $L$-Lipschitz continuous derivative $Df$ for some real $L \ge 0$.*

*Then, the set $S$ is $F$-normally $\widehat{\omega}(\cdot)$-regular with $\widehat{\omega} : \mathbb{R}_+ \to \mathbb{R}_+$ defined by*

$$\widehat{\omega}(t) := \gamma\big(\omega\big(Kt\big) + \frac{L}{2}t^2\big) \quad \textit{for all } t \in \mathbb{R}_+.$$

*Proof.* Let $x, x' \in S$ and $u^\star \in N^F(S; x) \cap \mathbb{B}_{X^\star}$. According to Lemma 6.7, we have

$$u^\star \in \gamma \partial_F d\big(0, -f(\cdot) + B\big)(\overline{x}) = \gamma \partial_F d(f(\cdot), B)(\overline{x}).$$

Applying Proposition 2.17 and using the inclusion (2.3.1), we can write $u^\star = \gamma Df(x)^\star(y^\star) = \gamma\big(y^\star \circ Df(x)\big)$ for some $y^\star \in N^C\big(B; f(x)\big) \cap \mathbb{B}_{Y^\star}$. It then suffices to observe (as in the proof of Proposition 6.11) that

$$\langle y^\star, f(x') - f(x)\rangle \le \omega\big(\|f(x') - f(x)\|\big) \le \omega\big(K\|x' - x\|\big).$$

and

$$\begin{aligned}\langle y^\star, -f(x') + f(x) + Df(x)(x' - x)\rangle &\le \int_0^1 \|Df(x) - Df(x + t(x' - x))\|\|x' - x\| dt \\ &\le L\|x' - x\|^2 \int_0^1 t dt = \frac{L}{2}\|x' - x\|^2\end{aligned}$$

to get the desired $F$-normal $\widehat{\omega}(\cdot)$-regularity of $S$. □

## 6.4 Metric subregularity for multimappings with normally $\omega(\cdot)$-regular graph

The main aim of the present section is to replace the metric subregularity assumption in Proposition 6.8 and Proposition 6.11 by some openness type conditions. This leads to develop the following result ensuring the metric subregularity of a multimapping with normally $\omega(\cdot)$-regular graph. We follow for a large part the proof of [128, Theorem 2] (see the introduction for the precise statement). As usual, we set $1/\rho := 0$ whenever $\rho := +\infty$.

**Theorem 6.14.** *Let $X, Y$ be Banach spaces, $M : X \rightrightarrows Y$ be a multimapping with closed graph, $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ be an upper semicontinuous nondecreasing function with $\omega(0) = 0$. Let also $\overline{y} \in Y$ with $M^{-1}(\overline{y}) \neq \emptyset$. Assume that:*
*(i) there exist $\alpha, \beta \in ]0, +\infty[$ and $\rho \in ]0, +\infty]$ such that*

$$\beta > \frac{\alpha}{\rho} + \Big(1 + \frac{1}{\rho}\Big)\omega\Big(\big(\alpha^2 + \beta^2\big)^{\frac{1}{2}}\Big) \quad \text{and} \quad B(\overline{y}, \beta) \subset M(B[\overline{x}, \alpha]) \quad \text{for all } \overline{x} \in M^{-1}(\overline{y}).$$

*(ii) the set $\operatorname{gph} M$ is $C$-normally $\omega(\cdot)$-regular (in $X \times Y$ endowed with the 2-norm $\|\cdot\|_2$) relative to the open set $V := \bigcup_{x \in M^{-1}(\overline{y})} B\big((x, \overline{y}), \sqrt{\alpha^2 + \beta^2}\big)$;*

*Then, there exists a real $\gamma \in [0, \rho[$ such that for every $\overline{x} \in M^{-1}(\overline{y})$, there exists a real $\delta > 0$ satisfying*

$$d\big(x, M^{-1}(\overline{y})\big) \leq \gamma d\big(\overline{y}, M(x)\big) \quad \text{for all } x \in B(\overline{x}, \delta).$$

*Proof.* By contradiction, assume that for each $\gamma \in [0, \rho[$, there is $\overline{x} \in M^{-1}(\overline{y})$ such that for every real $\delta > 0$, there is $x \in B(\overline{x}, \delta)$ satisfying

$$d\big(x, M^{-1}(\overline{y})\big) > \gamma d(\overline{y}, M(x)).$$

Fix for a moment any $\rho' \in ]0, \rho[$. Let $(\varepsilon_n)_{n \geq 1}$ be a sequence of $]0, \rho'[$ with $\varepsilon_n \to 0$. Choose some integer $N \geq 1$ such that $\theta_n := \frac{1}{n(\rho' - \varepsilon_n)} < \frac{\beta}{2}$ and $\frac{4}{n^2} + \theta_n^2 < \alpha^2 + \beta^2$ for every integer $n \geq N$. Fix any integer $n \geq N$. There are $\overline{x}_n \in M^{-1}(\overline{y})$ and $x_n \in X$ with $\|x_n - \overline{x}_n\| < \frac{1}{n}$ such that

$$d\big(x_n, M^{-1}(\overline{y})\big) > (\rho' - \varepsilon_n) d\big(\overline{y}, M(x_n)\big).$$

According to Proposition 6.4, there is $(u_n, v_n) \in \operatorname{gph} M$ such that

$$\|u_n - x_n\| < d\big(x_n, M^{-1}(\overline{y})\big) \quad \text{and} \quad 0 < \|v_n - \overline{y}\| < \frac{d\big(x_n, M^{-1}(\overline{y})\big)}{\rho' - \varepsilon_n} \tag{6.16}$$

along with

$$(0, 0) \in \{0\} \times \partial \|\cdot\|(v_n - \overline{y}) + \frac{1}{\rho' - \varepsilon_n}(\mathbb{B}_{X^\star} \times \varepsilon_n \mathbb{B}_{Y^\star}) + N^C\big(\operatorname{gph} M; (u_n, v_n)\big).$$

Since $v_n - \overline{y} \neq 0$, there is $z_n^\star \in \mathbb{S}_{Y^\star}$ such that $\langle z_n^\star, v_n - \overline{y} \rangle = \|v_n - \overline{y}\|$ and $(x_n^\star, y_n^\star) \in N^C\big(\operatorname{gph} M; (u_n, v_n)\big)$ satisfying

$$(x_n^\star, y_n^\star) \in (0, -z_n^\star) + \frac{1}{\rho' - \varepsilon_n}(\mathbb{B}_{X^\star} \times \varepsilon_n \mathbb{B}_{Y^\star}).$$

Therefore, we can write $y_n^\star = -z_n^\star + \frac{\varepsilon_n}{\rho' - \varepsilon_n} b_n^\star$ for some $b_n^\star \in \mathbb{B}_{Y^\star}$ along with

$$\|x_n^\star\| \leq \frac{1}{\rho' - \varepsilon_n} \quad \text{and} \quad \|y_n^\star\| \leq 1 + \frac{\varepsilon_n}{\rho' - \varepsilon_n}. \tag{6.17}$$

We also note that

$$d\big(x_n, M^{-1}(\overline{y})\big) \leq \|x_n - \overline{x}_n\| \leq \frac{1}{n}$$

which obviously gives (see (6.16))

$$\|v_n-\overline{y}\|<\frac{d\big(x_n,M^{-1}(\overline{y})\big)}{\rho'-\varepsilon_n}\le\frac{1}{n(\rho'-\varepsilon_n)}=\theta_n.$$

and (see again (6.16))

$$\|u_n-\overline{x}_n\|\le\|u_n-x_n\|+\|x_n-\overline{x}_n\|<d\big(x_n,M^{-1}(\overline{y})\big)+\|x_n-\overline{x}_n\|\le\frac{2}{n}. \tag{6.18}$$

Now, set $\zeta_n:=(\beta-\theta_n)\frac{v_n-\overline{y}}{\|v_n-\overline{y}\|}$ and observe that (keeping in mind the choice of $N$)

$$\|v_n-\zeta_n-\overline{y}\|=\big|\|v_n-\overline{y}\|-\beta+\theta_n\big|=\beta-\|v_n-\overline{y}\|-\theta_n<\beta,$$

that is, $v_n-\zeta_n\in B(\overline{y},\beta)$. By assumption, we can find some $w_n\in B[\overline{x}_n,\alpha]$ such that $v_n-\zeta_n\in M(w_n)$. Through (6.18), we see that

$$\|w_n-u_n\|\le\|w_n-\overline{x}_n\|+\|\overline{x}_n-u_n\|<\alpha+\frac{2}{n}. \tag{6.19}$$

On the other hand, we have

$$\|(w_n,v_n-\zeta_n)-(\overline{x}_n,\overline{y})\|_2^2=\|w_n-\overline{x}_n\|^2+\|v_n-\zeta_n-\overline{y}\|^2<\alpha^2+\beta^2$$

and

$$\|(u_n,v_n)-(\overline{x}_n,\overline{y})\|_2^2=\|u_n-\overline{x}_n\|^2+\|v_n-\overline{y}\|^2\le\frac{4}{n^2}+\theta_n^2<\alpha^2+\beta^2.$$

Thus, we have the inclusions

$$(w_n,v_n-\zeta_n),(u_n,v_n)\in V.$$

Denote $\|\cdot\|_\star$ the dual norm of the product norm $\|\cdot\|_2$ on $X\times Y$. Putting together the fact that $\operatorname{gph}M$ is $C$-normally $\omega(\cdot)$-regular, (6.17), (6.19) and the definition of $\zeta_n$, we get

$$\begin{aligned}\langle(x_n^\star,y_n^\star),(w_n,v_n-\zeta_n)-(u_n,v_n)\rangle&\le\|(x_n^\star,y_n^\star)\|_\star\omega\big(\|(w_n-u_n,-\zeta_n)\|\big)\\&\le\big(\|x_n^\star\|+\|y_n^\star\|\big)\omega\Big(\big(\|w_n-u_n\|^2+\|\zeta_n\|^2\big)^{\frac12}\Big)\\&\le\big(1+\frac{1+\varepsilon_n}{\rho'-\varepsilon_n}\big)\omega\Big(\big((\alpha+\frac{2}{n})^2+(\beta-\theta_n)^2\big)^{\frac12}\Big).\end{aligned}$$

By (6.19) and (6.17), we also have

$$\langle x_n^\star,w_n-u_n\rangle\ge-\|x_n^\star\|\|w_n-u_n\|\ge-\frac{1}{\rho'-\varepsilon_n}(\alpha+\frac{2}{n}).$$

Combining the definition of $\zeta_n$ with the equality $\langle z_n^\star,v_n-\overline{y}\rangle=\|v_n-\overline{y}\|$, we obtain

$$\langle y_n^\star,\zeta_n\rangle=\left\langle-z_n^\star+\frac{\varepsilon_n}{\rho'-\varepsilon_n}b_n^\star,\zeta_n\right\rangle\le\theta_n-\beta+\frac{\varepsilon_n}{\rho'-\varepsilon_n}(\beta-\theta_n).$$

Putting what precedes together, we arrive to (having in mind that $\omega(\cdot)$ is upper semicontinuous)

$$-\frac{1}{\rho'-\varepsilon_n}(\alpha+\frac{2}{n}) \leq \theta_n-\beta+\frac{\varepsilon_n}{\rho'-\varepsilon_n}(\beta-\theta_n)+\big(1+\frac{1+\varepsilon_n}{\rho'-\varepsilon_n}\big)\omega\Big(\big((\alpha+\frac{2}{n})^2+(\beta-\theta_n)^2\big)^{\frac{1}{2}}\Big).$$

Letting $n\to\infty$ and $\rho'\uparrow\rho$ obviously yields

$$\beta \leq \frac{\alpha}{\rho}+\big(1+\frac{1}{\rho}\big)\omega\Big(\big(\alpha^2+\beta^2\big)^{\frac{1}{2}}\Big)$$

which is the desired contradiction. The proof is then complete. □

**Remark 6.15.** *If $\rho=+\infty$, the inequality in (ii) of Theorem 6.14 is reduced to*

$$\beta > \omega\Big((\alpha^2+\beta^2)^{\frac{1}{2}}\Big). \tag{6.20}$$

*Assume there is a real $r>0$ such that $\omega(t):=\frac{t^2}{2r}$ for every $t\geq 0$ (which is the case if $X$ and $Y$ are Hilbert spaces and* $\operatorname{gph} M$ *is an $r$-prox-regular set of $X\times Y$). We then observe that the latter inequality* (6.20) *can be written as*

$$\beta > \frac{\alpha^2+\beta^2}{2r}$$

*which is obviously satisfied for every $\alpha,\beta>0$ with $\alpha<r$ and $|\beta-r|<\sqrt{r^2-\alpha^2}$.* ■

We derive from the latter theorem sufficient conditions ensuring the normal $\omega(\cdot)$-regularity of an inverse image $M^{-1}(\overline{y})$. Doing so, we complement Proposition 6.8.

**Proposition 6.16.** *Let $X,Y$ be two Banach spaces, $M:X\rightrightarrows Y$ be a multimapping whose graph is closed and let $\omega:\mathbb{R}_+\to\mathbb{R}_+$ be an upper semicontinuous nondecreasing function with $\omega(0)=0$. Let $\overline{y}\in Y$ with $S:=M^{-1}(\overline{y})\neq\emptyset$. Assume that (i) and (ii) of Theorem 6.14 hold with $\rho<+\infty$. Then, the set $S$ is $F$-normally $\theta(\cdot)$-regular with $\theta=(1+\rho)\omega$.*

*Proof.* Let $x,x'\in S$ and $x^\star\in N^F(S;x)\cap\mathbb{B}_{X^\star}$. Combining Theorem 6.14 and Lemma 6.7(a)-(c) we can find $b^\star\in\mathbb{B}_{Y^\star}$ such that $x^\star\in\rho D_C M(x,\overline{y})(b^\star)$, that is,

$$(\rho^{-1}x^\star,-b^\star)\in N^C\big(\operatorname{gph} M;(x,\overline{y})\big).$$

Let $\|\cdot\|_2$ be the product 2-norm of the norms of $X$ and $Y$ and let $\|\cdot\|_\star$ its associated dual norm. We can then write

$$\begin{aligned}\langle x^\star,x'-x\rangle=\langle(x^\star,-\rho b^\star),(x',\overline{y})-(x,\overline{y})\rangle &\leq \|(x^\star,-\rho b^\star)\|_\star\omega\big(\|(x',\overline{y})-(x,\overline{y})\|_2\big)\\ &\leq \big(\|x^\star\|+\rho\|b^\star\|\big)\omega\big(\|x'-x\|\big)\\ &\leq (1+\rho)\omega\big(\|x'-x\|\big)\end{aligned}$$

which translates the desired $F$-normal $\theta(\cdot)$-regularity of the set $S$. □

**Remark 6.17.** *If $\rho = +\infty$ in the latter proposition, then we can conclude with similar arguments that there exists a real $\lambda > 0$ such that the set $S$ is $F$-normally $\theta(\cdot)$-regular with $\theta(\cdot) := \lambda\omega(\cdot)$.* ■

**Remark 6.18.** *It is readily seen that we can develop Theorem 6.14 and Proposition 6.16 under the L-normal regularity of* $\operatorname{gph} M$ *and the Asplund property of the Banach spaces $X$ and $Y$.* ■

We now focus on the normal $\omega(\cdot)$-regularity of the solution set of a generalized equation, say $f(x) \in F(x)$ for a single-valued mapping $f$ and a multimapping $F$. We point out that the normal $\omega(\cdot)$-regularity for a set of the form $\{x \in X : 0 \in F_1(x) + F_2(x)\}$ with $F_1, F_2$ two multimappings has been established in [128] under an openness condition in the product space $(X \times Y)^2$, namely

$$\beta\mathbb{U}_{(X\times Y)^2} \subset -\{((x,y),(x,y)) : (x,y) \in (\overline{x},\overline{y}) + \alpha\mathbb{B}_{X\times Y}\} + \operatorname{gph} F_1 \times \operatorname{gph}(-F_2),$$

for two constants $\alpha, \beta > 0$ satisfying (1.4) for some real $\rho > 0$.

**Proposition 6.19.** *Let $X, Y$ be Banach spaces, $F : X \rightrightarrows Y$ be a multimapping with closed graph and let $f : X \to Y$ be a mapping. Assume that $S := \{x \in X : f(x) \in F(x)\} \neq \emptyset$ along with:*
*(i) The mapping $f$ is $K$-Lipschitz continuous on $X$ for some real $K \geq 0$ and differentiable on $X$ with $L$-Lipschitz continuous derivative $Df$ for some real $L \geq 0$.*
*(ii) there exists an upper semicontinuous nondecreasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$, two reals $\alpha, \beta \in ]0, +\infty[$ and an extended real $\rho \in ]0, +\infty]$ satisfying with $\kappa := (1 + K^2)^{\frac{1}{2}}$*

$$\beta > \frac{\alpha}{\rho} + \Big(1 + \frac{1}{\rho}\Big)\Big(\omega\big(\max(1,\kappa)(\alpha^2+\beta^2)^{\frac{1}{2}}\big) + \frac{L}{2}(\alpha^2+\beta^2)\Big)$$

*such that the set* $\operatorname{gph} F$ *is $C$-normally $\omega(\cdot)$-regular and*

$$\beta\mathbb{U}_{(X\times Y)} \subset -\operatorname{gph} f \cap \big(B[\overline{x},\alpha] \times Y\big) + \operatorname{gph} F \quad \textit{for all } \overline{x} \in S. \tag{6.21}$$

*Then, there exists a real $\gamma \in [0, \rho[$ such that for each $\overline{x} \in S$, there exists a real $\delta > 0$ satisfying*

$$d(x, S) \leq \gamma d\big((x, f(x)), \operatorname{gph} F\big) \quad \textit{for all } x \in B(\overline{x}, \delta). \tag{6.22}$$

*Further, the set $S$ is $F$-normally $\widehat{\omega}(\cdot)$-regular with $\widehat{\omega} : \mathbb{R}_+ \to \mathbb{R}_+$ defined by*

$$\widehat{\omega}(t) := \gamma\big(\omega\big(\kappa t\big) + \frac{L}{2}t^2\big) \quad \textit{for all } t \in \mathbb{R}_+.$$

*Proof.* First, we define the multimapping $M : X \rightrightarrows X \times Y$ by setting

$$M(x) := -(x, f(x)) + \operatorname{gph} F \quad \text{for all } x \in X.$$

It is readily seen that the multimapping $M$ has its graph closed. It is also evident that $M^{-1}(0,0) = S$ and

$$d\big((0,0), M(x)\big) = d\big((x, f(x)), \operatorname{gph} F\big) \quad \text{for all } x \in X.$$

On the other hand, the mapping $\varphi : X \to X \times Y$ defined by

$$\varphi(x) := (x, f(x)) \quad \text{for all } x \in X$$

is obviously $\kappa$-Lipschitz continuous on $X$ endowed with the 2-norm $\|\cdot\|_2$ and differentiable at each point $x \in X$ with

$$D\varphi(x)(h) := \big(h, Df(x)(h)\big) \quad \text{for all } h \in X.$$

Further, we observe that the derivative $D\varphi$ is $L$-Lipschitz continuous on $X$ since we have for every $x, x' \in X$

$$\begin{aligned}\sup_{h\in\mathbb{B}_X} \|D\varphi(x)(h) - D\varphi(x')(h)\|_2 &= \sup_{h\in\mathbb{B}_X} \|\big(0, (Df(x) - Df(x'))(h)\big)\|_2 \\ &\leq \|Df(x) - Df(x')\| \leq L\|x - x'\|.\end{aligned}$$

According to Proposition 6.3, we know that $M$ has a $C$-normally $\omega_0(\cdot)$ regular graph (with respect to the norm $\|\cdot\|_2$) where $\omega_0(t) := \omega(\kappa)t) + \frac{L}{2}t^2$ for all $t > 0$. The inequality (6.22) then follows from Theorem 6.14. It remains to apply Proposition 6.11 to conclude the proof. □

## 6.5 Metric subregularity for multimappings with normally ω-regular values

Theorem 6.14 requires the normal $\omega(\cdot)$-regularity for the graph $\operatorname{gph} M$ of the involved multimapping $M$. Unfortunately, there are numerous and various multimappings which fail to enjoy such a property for a given function $\omega(\cdot)$. This can be easily seen with the subdifferential of a nonsmooth function, for instance

$$\operatorname{gph}\big(\partial|\cdot|\big) = ]-\infty, 0[\times\{-1\} \cup \{0\} \times [-1, 1] \cup ]0, +\infty[\times\{1\},$$

which is obviously non-prox-regular (even not subsmooth) in $\mathbb{R}^2$.

Our aim in the present section is to provide some metric subregularity properties for multimappings normally $\omega(\cdot)$-regular-valued. Doing so, we need to develop an appropriate version of Proposition 6.4 where the normal cone $N^C\big(\operatorname{gph} M; (u, v)\big)$ is replaced by $N\big(M(x); v\big)$. Our arguments follow those in the work [145].

**Proposition 6.20.** *Let $M : X \rightrightarrows Y$ be a multimapping from a normed space $X$ into a Banach space $Y$ and let $\overline{y} \in Y$. Assume that there exists $x \in X$ with $M(x)$ nonempty and closed and two reals $\gamma, r \in ]0, +\infty[$ such that*

$$\gamma d\big(\overline{y}, M(x)\big) < r < d\big(x, M^{-1}(\overline{y})\big). \tag{6.23}$$

*Then, for all reals $\eta, \varepsilon \in ]0, +\infty[$, there exist $v \in M(x)$ satisfying:*
*(i)* $0 < \|v - \overline{y}\| < \min\{\frac{r}{\gamma}, d(\overline{y}, M(x)) + \varepsilon\}$;
*(ii)* $\|v - \overline{y}\| \leq \|y - \overline{y}\| + \frac{\eta}{\gamma}\|y - v\|$ *for all* $y \in M(x)$;
*(iii)* $0 \in \partial\|\cdot\|(v - \overline{y}) + \frac{\eta}{\gamma}\mathbb{B}_{Y^\star} + N^C\big(M(x); v\big)$.
*If in addition $Y$ is an Asplund space, then (iii) can be replaced by*
*(iii′)* $0 \in \partial\|\cdot\|(v - \overline{y}) + \frac{\eta}{\gamma}\mathbb{B}_{Y^\star} + N^L\big(M(x); v\big)$.

*Proof.* Let $\eta, \varepsilon \in ]0, +\infty[$. Choose some real $\gamma' > \gamma$ and some real $r' < r$ such that

$$\gamma' d\big(\overline{y}, M(x)\big) < r' < d\big(x, M^{-1}(\overline{y})\big).$$

According to the first inequality, we can find some $y_0 \in M(x)$ such that $\|y_0 - \overline{y}\| < \frac{r'}{\gamma'}$. Thanks to the second inequality in (6.23), we obviously have

$$B[x, r] \cap M^{-1}(\overline{y}) = \emptyset. \tag{6.24}$$

Let us set $\theta(\cdot) := \|\cdot - \overline{y}\| + \psi_{M(x)}(\cdot)$ which is lower semicontinuous and proper (keeping in mind that $M(x)$ is nonempty and closed). Note that

$$\theta(y_0) = \|y_0 - \overline{y}\| < \frac{r'}{\gamma'} \leq \inf_{y \in Y} \theta(y) + \frac{r'}{\gamma'}.$$

Fix any real $\eta' \in ]0, \min\{\gamma, \eta\}[$ with $\frac{2\eta'\|y_0 - \overline{y}\|}{\gamma' - \eta'} < \varepsilon$. Observe that $(Y, N)$ is a Banach space with $N(\cdot) := \eta'\|\cdot\|$. We are in a position to apply Ekeland variational principle. Doing so, we get some $v \in Y$ such that

$$N(y_0 - v) \leq r' < r,$$

$$\theta(v) \leq \theta(y_0) = \|y_0 - \overline{y}\| < \frac{r'}{\gamma'} < \frac{r}{\gamma} \tag{6.25}$$

and

$$\theta(v) \leq \theta(y) + \frac{1}{\gamma'} N(y - v) \quad \text{for all } y \in Y. \tag{6.26}$$

From (6.25), we see that $v \in M(x)$ (so $v \neq \overline{y}$ by (6.24)) and $\|v - \overline{y}\| \leq \|y_0 - \overline{y}\|$. Then, by (6.26) we have

$$\theta(v) = \|v - \overline{y}\| \leq \|y - \overline{y}\| + \frac{\eta'}{\gamma'}\|y - v\| \quad \text{for all } y \in M(x). \tag{6.27}$$

Now, let us define $f : Y \to \mathbb{R}$ by setting

$$f(y) := \|y - \overline{y}\| + \frac{\eta'}{\gamma'}\|y - v\| \quad \text{for all } y \in Y.$$

Through (6.27), we see that $v$ is a global minimizer of $f + \psi_{M(x)}$ and this implies that

$$0 \in \partial_C f(v) \subset \partial\|\cdot\|(v - \overline{y}) + \frac{\eta'}{\gamma'}\mathbb{B}_{Y^\star} + N^C\big(M(x); v\big) \tag{6.28}$$

along with

$$\begin{aligned} f(v) = \|v - \overline{y}\| &\leq \|y - \overline{y}\| + \frac{\eta'}{\gamma'}\|y - v\| \\ &\leq (1 + \frac{\eta'}{\gamma'})\|y - \overline{y}\| + \frac{\eta'}{\gamma'}\|\overline{y} - v\| \end{aligned}$$

for every $y \in M(x)$. Hence, we have

$$\|v-\overline{y}\| \leq \frac{1+\eta'/\gamma'}{1-\eta'/\gamma'}\|y-\overline{y}\| = \frac{\gamma'+\eta'}{\gamma'-\eta'}\|y-\overline{y}\| \quad \text{for all } y \in M(x).$$

We deduce from this

$$\|v-\overline{y}\|_Y \leq \frac{\gamma'+\eta'}{\gamma'-\eta'}d\big(\overline{y},M(x)\big) = d\big(\overline{y},M(x)\big) + \frac{2\eta'}{\gamma'-\eta'}d\big(\overline{y},M(x)\big). \tag{6.29}$$

Coming back to (6.25) and using the definition of $\eta'$ yield

$$d\big(\overline{y},M(x)\big) \leq \|v-\overline{y}\| \leq \|\overline{y}-y_0\| < \frac{(\gamma'-\eta')\varepsilon}{2\eta'}. \tag{6.30}$$

It remains to put together (6.29) and (6.30) to obtain

$$\|v-\overline{y}\| \leq d\big(\overline{y},M(x)\big) + \varepsilon.$$

The proof of $(i)-(ii)-(iii)$ is complete.

Regarding $(iii')$, if $Y$ is an Asplund space we can write (see (6.28))

$$0 \in \partial_L f(v) \subset \partial\|\cdot\|(v-\overline{y}) + \frac{\eta'}{\gamma'}\mathbb{B}_{Y^\star} + N^L\big(M(x);v\big).$$

This finishes the proof. □

**Remark 6.21.** *Assume that $M(\cdot) = -f(\cdot)+S$ for some mapping $f : X \to Y$ and some set $S \subset Y$. Given any $v \in Y$, we obviously have $N^C(M(x);v) = N^C(S;v+f(x))$, so $(iii)$ in Proposition 6.20 can be rewritten as*

$$0 \in \partial\|\cdot\|(v-\overline{y}) + \frac{\eta}{\gamma}\mathbb{B}_{Y^\star} + N^C(S;v+f(x))$$

*without any assumption on the mapping $f$.* ■

We are now in a position to establish the following result which complements Theorem 6.14.

**Theorem 6.22.** *Let $M : X \rightrightarrows Y$ be a multimapping from a normed space $X$ into a Banach (resp. Asplund) space $Y$ and let $\overline{y} \in Y$. Assume that $M(\cdot)$ is closed and $C$-normally (resp. $L$-normally) $\omega(\cdot)$-regular valued for some upper semicontinuous nondecreasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$. Assume also that there exist $\alpha,\beta,\kappa \in ]0,+\infty[$ with $\beta > \kappa\alpha + \omega(\beta+\kappa\alpha)$ such that:*
*$(i)$ for all $\overline{x} \in M^{-1}(\overline{y})$, there exists a real $\eta > 0$ satisfying*

$$\operatorname{exc}\big(M(x),M(x')\big) \leq \kappa\|x'-x\| \quad \textit{for all } x \in B[\overline{x},\alpha],\ \textit{all } x' \in B[\overline{x},\eta];$$

*$(ii)$ for all $\overline{x} \in M^{-1}(\overline{y})$, one has $B(\overline{y},\beta) \subset M(B[\overline{x},\alpha])$.*

*Then, there exists a real $\gamma \geq 0$ such that for every $\overline{x} \in M^{-1}(\overline{y})$, there exists a real $\delta > 0$ satisfying*

$$d\big(x,M^{-1}(\overline{y})\big) \leq \gamma d\big(\overline{y},M(x)\big) \quad \textit{for all } x \in B(\overline{x},\delta).$$

*Proof.* We only deal with the $C$-normal $\omega(\cdot)$-regularity for the values of $M(\cdot)$ in the Banach space $Y$. We proceed as in the proof of Theorem 6.14. By contradiction, assume that for each $\gamma \geq 0$, there is $\overline{x} \in M^{-1}(\overline{y})$ such that for every real $\delta > 0$, there is $x \in B(\overline{x}, \delta)$ satisfying

$$d\big(x, M^{-1}(\overline{y})\big) > \gamma d\big(\overline{y}, M(x)\big). \tag{6.31}$$

Fix any real $\gamma > 0$ and any real $\kappa' > \kappa$. According to assumption (i), for each $\overline{x} \in M^{-1}(\overline{y})$ we can find a real $\eta_{\overline{x}} > 0$ (see (2.1)) such that

$$M(x) \subset M(x') + \kappa' \|x - x'\| \mathbb{B}_Y \quad \text{for all } x \in B[\overline{x}, \alpha], \text{ all } x' \in B[\overline{x}, \eta_{\overline{x}}]; \tag{6.32}$$

Pick any sequence $(\varepsilon_n)_{n\geq 1}$ in $]0, \gamma[$ with $\varepsilon_n \to 0$. Choose some integer $N \geq 1$ such that $\theta_n := \frac{1}{n(\gamma - \varepsilon_n)} < \frac{\beta}{2}$ for every integer $n \geq N$. Fix for a moment any integer $n \geq N$. Thanks to (6.31), there are $\overline{x}_n \in M^{-1}(\overline{y})$ and $x_n \in X$ with $\|x_n - \overline{x}_n\| < \min\{\frac{1}{n}, \eta_{\overline{x}_n}\}$ such that

$$d\big(x_n, M^{-1}(\overline{y})\big) > (\gamma - \varepsilon_n) d\big(\overline{y}, M(x_n)\big).$$

According to Proposition 6.20, there is $v_n \in M(x_n)$ such that

$$0 < \|v_n - \overline{y}\| < \frac{d\big(x_n, M^{-1}(\overline{y})\big)}{\gamma - \varepsilon_n}$$

along with

$$0 \in \partial \|\cdot\|(v_n - \overline{y}) + \frac{\varepsilon_n}{\gamma - \varepsilon_n} \mathbb{B}_{Y^\star} + N^C\big(M(x_n); v_n\big).$$

Since $v_n - \overline{y} \neq 0$, the latter inclusion gives $y_n^\star \in N^C\big(M(x_n); v_n\big)$, $z_n^\star \in \mathbb{S}_{Y^\star}$ with $\langle z_n^\star, v_n - \overline{y}\rangle = \|v_n - \overline{y}\|$ and $b_n^\star \in \mathbb{B}_{Y^\star}$ such that

$$y_n^\star = -z_n^\star + \frac{\varepsilon_n}{\gamma - \varepsilon_n} b_n^\star.$$

Setting $\zeta_n := (\beta - \theta_n) \frac{v_n - \overline{y}}{\|v_n - \overline{y}\|}$ and noticing that $v_n - \zeta_n \in B(\overline{y}, \beta)$, we can find some $w_n \in B[\overline{x}_n, \alpha]$ such that $v_n - \zeta_n \in M(w_n)$. Thanks to (6.32), there is $\xi_n \in M(x_n)$ and some $b_n \in \mathbb{B}_Y$ such that

$$v_n - \zeta_n = \xi_n + \kappa' \|x_n - w_n\| b_n.$$

Using the fact that $M(x_n)$ is $C$-normally $\omega(\cdot)$-regular, we obtain

$$\begin{aligned}
\langle y_n^\star, -\zeta_n\rangle &= \langle y_n^\star, \xi_n - v_n + \kappa' \|x_n - w_n\| b_n\rangle \\
&= \langle y_n^\star, \xi_n - v_n\rangle + \kappa' \|x_n - w_n\| \langle y_n^\star, b_n\rangle \\
&\leq \|y_n^\star\| \Big(\omega(\|\xi_n - v_n\|) + \kappa' \|x_n - w_n\|\Big) \\
&\leq \Big(1 + \frac{\varepsilon_n}{\gamma - \varepsilon_n}\Big) \Big(\omega(\|\xi_n - v_n\|) + \kappa' \|x_n - w_n\|\Big).
\end{aligned}$$

On the other hand, we have

$$\|x_n - w_n\| \leq \|x_n - \overline{x}_n\| + \|\overline{x}_n - w_n\| < \frac{1}{n} + \alpha$$

and this entails

$$\|\xi_n - v_n\| = \| - \zeta_n - \kappa'\|x_n - w_n\|b_n\| \leq \ \beta - \theta_n + \kappa'(\frac{1}{n} + \alpha).$$

Combining the definition of $\zeta_n$ with the equality $\langle z_n^\star, v_n - \overline{y}\rangle = \|v_n - \overline{y}\|$, we obtain

$$\langle y_n^\star, \zeta_n\rangle = \left\langle -z_n^\star + \frac{\varepsilon_n}{\gamma - \varepsilon_n} b_n^\star, \zeta_n \right\rangle \leq \theta_n - \beta + \frac{\varepsilon_n}{\gamma - \varepsilon_n}(\beta - \theta_n).$$

Putting what precedes together, we arrive to

$$\beta - \theta_n - \frac{\varepsilon_n}{\gamma - \varepsilon_n}(\beta - \theta_n) \leq \big(1 + \frac{\varepsilon_n}{\gamma - \varepsilon_n}\big)\Big(\omega\big(\beta - \theta_n + \kappa'(\frac{1}{n} + \alpha)\big) + \kappa'(\frac{1}{n} + \alpha)\Big).$$

Keeping in mind that the function $\omega(\cdot)$ is upper semicontinuous and letting $n \to \infty$ and $\kappa' \downarrow \kappa$ give

$$\beta \leq \kappa\alpha + \omega(\beta + \kappa\alpha)$$

which is the desired contradiction. The proof is then complete. □

The Lipschitz behavior with respect to the Hausdorff-Pompeiu excess in (ii) of Theorem 6.22 obviously holds for the Lipschitz translation of a fixed set, say $M(x) = -f(x) + B$ for some Lipschitz mapping $f$.

**Corollary 6.23.** Let $f : X \to Y$ be a $\kappa$-Lipschitz continuous mapping between a normed space $X$ and a Banach (resp. Asplund) space $Y$ with $\kappa \geq 0$. Let also $B$ be a closed $C$-normally (resp. $L$-normally) $\omega(\cdot)$-regular set for some upper semicontinuous nondecreasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$. Assume that there exist two reals $\alpha, \beta > 0$ such that

$$\beta > \kappa\alpha + \omega(\beta + \kappa\alpha) \quad \text{and} \quad \beta\mathbb{U}_Y \subset -f(B[\overline{x}, \alpha]) + B \quad \text{for all } \overline{x} \in f^{-1}(B).$$

Then, there exists a real $\gamma \geq 0$ such that for every $\overline{x} \in f^{-1}(B)$, there exists a real $\delta > 0$ satisfying

$$d\big(x, f^{-1}(B)\big) \leq \gamma d\big(f(x), B\big) \quad \text{for all } x \in B(\overline{x}, \delta).$$

*Proof.* It suffices to apply Theorem 6.22 with the multimapping $M(\cdot) := -f(\cdot) + B$ and the point $\overline{y} := 0$. □

**Remark 6.24.** *We keep notation and assumptions of the latter corollary. If in addition $f$ is strictly Hadamard differentiable, then we can combine Proposition 6.5 and Proposition 2.18 to obtain*

$$\gamma \leq \inf_{\varepsilon > 0} \sup_{y^\star \in \Omega_\varepsilon} \|y^\star\|_{Y^\star},$$

*where for each real $\varepsilon > 0$, we denote $\Omega_\varepsilon$ the set of $y^\star \in Y^\star$ for which there are $x \in B(\overline{x}, \varepsilon)$ with $\overline{y} + f(x) \notin S$ and $y \in B(\overline{y}, \varepsilon)$ with $y + f(x) \in S$ such that $y^\star \in N(S; f(x) + y)$ and $\|Df(x)^\star(y^\star)\| \leq 1$.* ■

Coming back to Proposition 6.19 with $F \equiv B$, we see that the openness condition (6.21) can be written as

$$\beta\mathbb{U}_{(X\times Y)} \subset -\{(x, f(x)) : x \in B[\overline{x}, \alpha]\} + X \times B.$$

Using Corollary 6.23 allows us to drop the whole space $X$ in the latter formula. More precisely:

**Proposition 6.25.** *Let $X$ be a normed space, $Y$ be a Banach (resp. an Asplund) space, $B \subset Y$ and let $f : X \to Y$ be a mapping. Assume that $S := f^{-1}(B) \neq \emptyset$ along with:*
*(i) There exists a nondecreasing upper semicontinuous function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$ such that the set $B$ is $C$-normally (resp. $L$-normally) $\omega(\cdot)$-regular.*
*(ii) The mapping $f$ is $K$-Lipschitz continuous on $X$ for some real $K \geq 0$ and differentiable on $X$ with $L$-Lipschitz continuous derivative $Df$ for some real $L \geq 0$.*
*(iii) There exist two reals $\alpha, \beta > 0$ such that*

$$\beta > K\alpha + \omega(\beta + K\alpha) \quad \text{and} \quad \beta\mathbb{U}_Y \subset -f(B[\overline{x}, \alpha]) + B \quad \text{for all}\, \overline{x} \in f^{-1}(B).$$

*Then, there exists a real $\gamma \geq 0$ such that the set $S$ is $F$-normally $\widehat{\omega}(\cdot)$-regular with $\widehat{\omega} : \mathbb{R}_+ \to \mathbb{R}_+$ defined by*

$$\widehat{\omega}(t) := \gamma\big(\omega\big(Kt) + \frac{L}{2}t^2\big) \quad \text{for all}\, t \in \mathbb{R}_+.$$

*Proof.* It directly follows from Proposition 6.13 and Corollary 6.23. □

Proposition 6.25 allows to get sufficient conditions ensuring the normal regularity of an intersection set, say $S_1 \cap S_2$ with $S_1 \times S_2$ normally $\omega(\cdot)$-regular. Indeed, with $f(x) := (x, x)$ (which is obviously Lipschitz continuous with Lipschitz derivative) we see that

$$f^{-1}(S_1 \times S_2) = S_1 \cap S_2.$$

Hence, we see through Proposition 6.25 that the latter set $S_1 \cap S_2$ is normally regular under an openness condition of the form

$$\beta\mathbb{U}_{X^2} \subset -\{(x, x) : x \in B[\overline{x}, \alpha]\} + S_1 \times S_2 \quad \text{for all} \in \overline{x} \in S_1 \cap S_2.$$

This complements (6.7) which comes from [128, Proposition 7]. A similar remark holds for the normal regularity of the constraint set (6.3) denoted $S$. In such a case, a suitable openness condition is given by

$$\beta\mathbb{U}_{\mathbb{R}^{m+n}} \subset -F(B[\overline{x}, \alpha]) + ]-\infty, 0]^m \times \{0_{\mathbb{R}^n}\} \quad \text{for all}\, \overline{x} \in S,$$

where

$$F(x) := \big(f_1(x), \ldots, f_m(x), f_{m+1}(x), \ldots, f_{m+n}(x)\big).$$

## 6.6 Perspectives

1. In our paper[3], we established results for the farthest distance function within the context of Hilbert spaces. The next step is to extend these findings to the broader setting of Banach spaces.

2. Consider the state-dependent sweeping process described by:

$$\begin{cases} -\dot{u}(t) \in \mathrm{N}\big(C(t,u(t)); u(t)\big), \\ u(0) = u_0. \end{cases}$$

   In existing results concerning this problem, the compactness assumption on the moving set $C(t,x)$ plays a crucial role. We would like to understand the necessity of such an assumption. An initial approach is to investigate the scenario where the moving set is a strongly convex set, such as a moving ball, which is not compact in infinite-dimensional settings.

3. Consider the differential inclusion, known as the *steepest descent problem*, given by

$$\begin{cases} -\dot{u}(t) \in \partial_P f(u(t)) \quad \text{a.e. in } I, \\ u(T_0) = u_0. \end{cases}$$

   where $T_0 \in [0, +\infty[$ is the initial time and $f$ is a prox-regular (or primal lower regular) function. The existence of local and global solutions is established by Theorems 2.9 and 3.2 in [169]. We aim to analyze the asymptotic behavior of the solutions to the time-dependent differential inclusion:

$$\begin{cases} -\dot{u}(t) \in \partial_P f_t(u(t)) \quad \text{a.e. in } I, \\ u(T_0) = u_0. \end{cases}$$

   Here, $f_t$ denotes a time-dependent function, and the focus is on understanding how the time dependency. An open question is what happens in the classical convex case $f_t = \mathrm{dfar}_{C(t)}$?

4. In paper [4], we established various preservation results concerning the preservation of normal $\omega$-regular sets (including prox-regular and subsmooth sets) and the metric subregularity of multimappings with normal $\omega(\cdot)$-regularity, either of the graph or the values. Building upon this, we aim to extend these preservation results to the setting of strongly convex sets, which are the duals of prox-regular sets.

# Chapter 7

# Synthèse de la thèse en français

## 7.1 Le concept général

Mathematiquement, les problèmes d'évolution, tels que la dynamique de contact, sont, dans la plupart des cas, régis par des inclusions différentielles, ou plus précisément, des inclusions différentielles de mesure. L'une des inclusions différentielles les plus célèbres est connue sous le nom de **processus de balayage** (*Sweeping Process*), introduit par J.J. Moreau en 1971 lors du célèbre Séminaire d'Analyse Convexe à Montpellier [42]. Ce problème d'évolution continue d'être largement étudié aujourd'hui, tant d'un point de vue théorique que numérique (voir [8, 9, 12–22, 29, 30, 44, 47] et les références associées). Le processus de balayage, conditionné par des contraintes d'inégalité (impliquant des problèmes tels que les milieux granulaires ou l'élastodynamique avec contact), constitue un outil fondamental pour l'analyse mathématique et la conception de schémas numériques liés à ces types d'applications en dynamique non régulière. Dans ce contexte, nous nous intéressons au processus de balayage sous ses aspects théoriques (modèle mathématique, existence de solutions, ...) et numériques (développement de schémas, analyse énergétique du système, ...) ainsi qu'à son application, en particulier à la dynamique de contact. La condition de contact unilatéral en termes de vitesse est cruciale dans le développement des méthodes de dynamique de contact non lisse. Le manque de régularité signifie que les lois de contact ne sont pas seulement non différentiables au sens classique, mais aussi multivoques. Cette condition permet d'obtenir de très bonnes propriétés de conservation de l'énergie, assurant ainsi la stabilité physique et numérique. En d'autres termes, les contraintes unilatérales sont exprimées comme une relation entre la vitesse et la quantité de mouvement, décrivant ainsi de manière cohérente la puissance mécanique du système. Ce type de processus de balayage du second ordre est une extension directe de la loi d'impact de Newton (loi de Moreau), formulée pour la première fois dans un cadre numérique discret. L'idée principale est d'étendre les modèles mathématiques sous forme d'inclusions différentielles mesurables pour les *milieux granulaires* et les matériaux déformables à la thermodynamique, c'est-à-dire au transfert de chaleur dans les milieux granulaires et à la thermoélasticité pour les matériaux déformables. En utilisant ces modèles précédents et en appliquant à la fois le processus de balayage et la méthode semi-lisse de Newton-Active set, nous visons à implémenter un schéma numérique dédié à l'intégration temporelle des systèmes non lisses avec contacts.

D'un point de vue théorique, l'étude des propriétés de l'ensemble en mouvement dans un processus de balayage joue également un rôle crucial. Elle permet d'élargir et de développer la compréhension des problèmes, notamment l'existence de solutions dans des scénarios non convexes, comme la prox-régularité ou sa classe duale, la forte convexité. Ici, la régularité métrique sert d'outil essentiel pour étudier la faible convexité de certains sous-ensembles. Cela revêt une importance particulière, car elle fournit une base théorique solide pour aborder des problèmes d'optimisation et de mécanique.

### 7.1.1 Processus de Balayage

D'un point de vue mécanique, nous visons à analyser le mouvement d'un point ou d'une particule confinée dans un ensemble convexe fermé en mouvement, noté $C(t)$. En fonction de la dynamique de cette région, le point restera soit stationnaire s'il évite de heurter la frontière de l'ensemble, soit il entraîne le domaine tout en restant à l'intérieur de $C(t)$ pour tous les instants $t$. Pour garantir que le point reste à l'intérieur de la région, sa vitesse doit pointer vers l'intérieur, c'est-à-dire dans la direction opposée à la normale de la vitesse par rapport à la frontière de l'ensemble. Un outil puissant pour traiter ce problème mécanique est le processus de balayage.

J.J. Moreau [42] a introduit la formulation mathématique du *processus de balayage* (au *processus de rafle*) qui implique des inclusions différentielles incorporant le cône normal $N\big(C(t);\cdot\big)$ à l'ensemble $C(t)$. L'objectif est de trouver une trajectoire solution $u : [0,T] \to X$ du problème 1 qui satisfait la condition de Cauchy, où $T > 0$ et un ensemble mobile $C(\cdot)$ existe dans l'espace de Hilbert $X$. L'existence et l'unicité d'une solution au problème 1 sont établies sous l'hypothèse que $C(t)$ est un ensemble convexe fermé en mouvement et qu'il est Lipschitz continu en temps $t$. Ceci est réalisé grâce à l'algorithme de rattrapage (plus de détails dans [42]). Dans le domaine de la convexité et de l'optimisation mathématique, l'investigation des solutions aux problèmes de processus de balayage d'ordre un et deux a engendré un domaine de recherche vibrant et profond. Kunze et Monteiro-Marques ont étudié l'existence de solutions lorsque les ensembles ont une variation Lipschitz en utilisant un schéma de discrétisation semi-implicite [12]. En utilisant une discrétisation explicite, Haddad et al. [13] ont prouvé l'existence de solutions d'un processus de balayage perturbé dépendant de l'état. De plus, Bounkhel et Castaing [14] ont obtenu le processus de balayage dépendant de l'état dans des espaces de Banach uniformément lisses et uniformément convexes. Plus tard, les auteurs Abderrahim Jourani et Emilio Vilches [9] ont prouvé l'existence de solutions de processus de balayage dépendant de l'état en utilisant la technique de régularisation Moreau-Yosida dans les ensembles à variation continue bornée avec des ensembles non réguliers (sous-lisses et positivement alpha-éloignés), etc.

Dans le cas non convexe, il est bien connu que Chemetov et Monteiro-Marques [15, 16] ont établi l'existence de solutions de processus de balayage perturbé dépendant de l'état avec des ensembles uniformément prox-réguliers en utilisant le théorème du point fixe de Schauder ou un argument de point fixe pour le processus de balayage perturbé dépendant de l'état. En outre, Castaing et al. [17] ont démontré l'existence de solutions pour le processus de balayage dépendant de l'état dans des ensembles uniformément prox-réguliers. Cela a été réalisé en employant une forme étendue du

théorème de Schauder en conjonction avec une méthode de discrétisation. Azzam-Laouir et al. [18] et Haddad et al. [19] ont montré le processus de balayage perturbé multivalué dépendant de l'état pour le cas uniformément prox-régulier dans le cadre fini-dimensionnel. De plus, l'auteur Noel [20], Noel et Thibault [21] ont considéré l'existence de processus de balayage perturbé multivalué dépendant de l'état avec des ensembles équi-uniformément sous-lisses et uniformément prox-réguliers, etc. D'autre part, les méthodes de régularisation Moreau et Yosida [9] ont été employées spécifiquement pour les ensembles convexes ou uniformément prox-réguliers pour aborder le problème du processus de balayage dépendant de l'état.

Notamment, les processus de balayage trouvent des manifestations naturelles dans un large éventail de problèmes pratiques, y compris la dynamique des mouvements de foule, le comportement des milieux granulaires, la modélisation élastoplastique, la dynamique des contacts, l'évolution des tas de sable et les circuits électriques non réguliers. Ces applications réelles soulignent la pertinence et l'universalité des problèmes de processus de balayage dans les investigations scientifiques et les entreprises pratiques.

## 7.1.2 Résolution des problèmes des milieux granulaires

### 7.1.2.1 Problèmes de milieux granulaires

Les milieux granulaires, composés de particules de tailles diverses (du grain de sable à un bloc de pierre de plusieurs mètres), de formes plus ou moins complexes (sphères ou polyèdres angulaires), interagissent entre eux par contact. Ces matériaux se retrouvent dans de nombreux secteurs industriels confrontés à divers problèmes, tels que dans les solides : le béton ou l'effondrement d'un bâtiment monumental sous contrainte sismique [70], les plaques tectoniques ou le mélange de comprimés pharmaceutiques [71]; dans les fluides : avalanche et glissement de terrain [72, 74], écoulement dans les silos [75], le ballast sous chargement cyclique [76], production et transmission de chaleur [78, 79], granulats collants [80], écoulements granulaires [81], etc. Les chercheurs de divers domaines sont attirés par les systèmes granulaires, non seulement pour leur pertinence pratique mais aussi en raison de leur comportement complexe (voir figure 1.2).

De nos jours, la compréhension des milieux granulaires et de leur comportement est devenue une préoccupation majeure pour les processus industriels ou les sciences environnementales. Un obstacle significatif dans la recherche sur les milieux granulaires est l'absence d'un cadre complet pouvant décrire universellement leurs propriétés, notamment en ce qui concerne les interactions multi-corps et multi-contacts. Par exemple, prédire l'écoulement granulaire nécessite une compréhension approfondie du modèle spécifique étudié, y compris les hypothèses sur les propriétés mécaniques et les lois d'interaction régissant ces systèmes multi-corps. Cette nature complexe a conduit à une utilisation croissante des simulations numériques, en tirant parti des progrès de la puissance de calcul pour analyser rapidement des systèmes granulaires réalistes et très complexes, bien que garantir l'exactitude des résultats numériques obtenus reste un défi. Par conséquent, les chercheurs doivent faire certaines hypothèses sur les états physiques des matériaux granulaires pour étudier efficacement leur comportement.

En résumé, les milieux granulaires représentent un domaine d'étude fascinant avec des implications pratiques couvrant diverses industries et disciplines scientifiques. Les chercheurs visent à découvrir les principes fondamentaux régissant ces matériaux, ou-

vrant la voie à des processus industriels plus efficaces et à une compréhension plus profonde des complexités du monde naturel.

#### 7.1.2.2 Méthodes de résolution numérique

Devant ces défis, des outils adaptés à la modélisation de la dynamique des milieux granulaires ont émergé. La Méthode des Éléments Discrets (*Discrete Element Method* ou DEM), initialement appelée Méthode des Éléments Distincts, a été développée par Cundall en 1971 [5] pour simuler la dynamique granulaire en suivant les interactions entre particules individuelles. La dynamique granulaire est alors régie par la deuxième loi de Newton, combinée à un modèle de contact régulier. En particulier, les simulations DEM peuvent être coûteuses en termes de calcul, en raison du schéma temporel explicite nécessitant de petits pas de temps. Par la suite, ces méthodes ont connu une évolution substantielle et ont donné naissance à diverses variantes, telles que la méthode de Dynamique Moléculaire (*Molecular Dynamics* ou MD) [35–37]. Bien que le comportement des matériaux granulaires puisse théoriquement être décrit dans le cadre de la mécanique des milieux continus classique, cette approche reste aujourd'hui encore insuffisante pour traiter adéquatement ces systèmes. Jean-Jacques Moreau a apporté des contributions majeures à l'analyse convexe, aux fonctions à variation bornée, à la théorie des mesures différentielles et à la théorie des processus de balayage — des outils mathématiques essentiels pour aborder la dynamique non régulière. Il a transformé ces concepts théoriques fondamentaux en une méthode numérique implicite innovante pour les systèmes non réguliers, connue sous le nom de *Contact Dynamics* (CD) [38, 43, 45, 46]. Cette méthode a ensuite été étendue par M. Jean [48] pour prendre en compte les corps déformables, conduisant ainsi au développement des méthodes de *Non-Smooth Contact Dynamics* (NSCD).

NSCD [23, 24, 38, 48] se concentre sur la modélisation et la simulation des forces de contact et de frottement entre les corps ou les particules. L'idée principale de NSCD est de traiter la dynamique des forces de contact comme un problème d'optimisation, soumis à des contraintes telles que la conservation de la quantité de mouvement et de l'énergie. La méthode implique l'introduction d'un ensemble de forces de contact inconnues, représentées sous forme de multiplicateurs de Lagrange, puis l'utilisation de techniques d'optimisation pour résoudre les forces de contact qui satisfont aux contraintes. Ces dernières sont dérivées de la loi de contact de Signorini en termes de vitesses et d'impulsions et de la loi de frottement de Coulomb liant la force de frottement à la composante tangentielle de la vitesse. Ensuite, M. Jean et J. J. Moreau [38, 46–48] ont développé un solveur itératif NonLinéaire Gauss-Seidel (NLGS) pour décrire les conditions de contact entre corps rigides. Ce solveur considère séquentiellement chaque contact jusqu'à ce que la convergence soit atteinte. De plus, d'autres solveurs alternatifs au NLGS, tels que les solveurs de gradient projeté conjugué, ont été proposés dans [49, 50]. En particulier, le terme non linéaire de NSCD est coûteux en calcul par rapport à DEM, comme souligné dans [10], bien que NSCD permette des pas de temps plus grands que DEM, permettant ainsi la résolution de plusieurs contacts simultanés. Les techniques couramment utilisées dans la littérature pour traiter les non-linéarités découlant de ces conditions sont basées sur plusieurs types de méthodes.

- Méthode Semi-Augmented Lagrangian (SAL) [51, 52]

La méthode SAL est utilisée pour résoudre le problème d'optimisation découlant de la dynamique du système, régie par la loi de Signorini - un modèle simplifié supposant que les forces de contact sont toujours perpendiculaires aux surfaces de contact. Elle introduit des multiplicateurs de Lagrange représentant ces forces et utilise une formulation lagrangienne augmentée pour assurer la régularité des forces et le respect des conditions du problème.

- Méthode du Bipotentiel [53, 54]
  La méthode du Bipotentiel est une extension du modèle de friction de Coulomb classique pour traiter les systèmes avec plusieurs corps rigides interagissant par friction. Elle utilise des multiplicateurs de Lagrange pour représenter les forces de contact et applique la formulation du bipotentiel, assurant la conservation de l'énergie et la stabilité.

- Méthode du gradient conjugué projeté (PCG) [50]
  Cette méthode est une technique numérique pour résoudre les problèmes d'optimisation sous contraintes, notamment dans les contextes de contact et de médias granulaires. Elle combine l'efficacité de la méthode du gradient conjugué, qui minimise une fonction en suivant des directions de gradient conjugué pour éviter le zigzag du gradient de descente standard, avec des techniques de projection qui gèrent les contraintes en projetant le point actuel sur l'ensemble faisable à chaque itération. Cela garantit que les contraintes sont respectées tout au long du processus d'optimisation, offrant un cadre de solution robuste et efficace.

- Méthode Primal-Dual Active Set (PDAS) [55, 56, 58]
  Elle a récemment émergé comme approche prometteuse et pertinente pour résoudre les problèmes de contact frictionnel en élastodynamique déformable, en raison de son efficacité et de sa facilité de mise en œuvre. Le principe principal est que les conditions de contact et de friction jouent comme deux éléments de fonctions de complémentarité non linéaire, dont la solution est directement fournie par l'approche semi-lisse de Newton. Dans la littérature, il y a peu de références aux méthodes PDAS pour résoudre les problèmes de contact rigide à corps multiples. Les méthodes PDAS pour les systèmes à corps multiples rigides ont véritablement commencé, en considérant PDAS comme une méthode pour le traitement local des conditions de contact sans friction dans le cadre de la NSCD, avec les travaux de M. Barboteu et S. Dumont. Par la suite, PDAS a été développée et a montré un niveau remarquable d'efficacité par rapport aux approches alternatives.

Par ailleurs, la méthode des Éléments Granulaires (GEM) [32] représente une variante de la méthode DEM qui se distingue par l'utilisation de lois de régularisation, la classant ainsi parmi les méthodes 'lisses'. La différence principale entre GEM et DEM réside dans son schéma de résolution, qui s'apparente à une méthode de rigidité, initialement proposée par Kawai [31] et par la suite étendue par Kishino. Cette méthode a été appliquée à la simulation quasi-statique des matériaux granulaires [32], ainsi qu'à l'étude de leur dynamique [33]. GEM se positionne comme une approche avancée qui offre un compromis entre la précision des simulations et la gestion des non-linéarités complexes inhérentes aux matériaux granulaires, tout en assurant une régularité dans les résultats. Dans la thèse de Mathieu Renouf (2004) de l'Université de Montpellier

[34], la méthode GEM est abordée en détail, notamment en ce qui concerne les aspects d'intégration numérique. Bien que la méthode GEM se concentre principalement sur la modélisation des interactions entre éléments granulaires, l'utilisation de méthodes d'intégration numérique, telles que la méthode de Newmark, est cruciale pour résoudre les équations de mouvement des particules au fil du temps. La méthode de Newmark est un schéma d'intégration numérique couramment utilisé pour résoudre les équations différentielles ordinaires qui apparaissent dans la dynamique des structures et des milieux continus. Elle est particulièrement populaire dans la mécanique des structures pour intégrer les équations du mouvement, en raison de sa stabilité et de sa capacité à contrôler l'amortissement numérique. Dans le contexte de la GEM, la méthode de Newmark est utilisée pour intégrer les équations de mouvement des particules granulaires au cours du temps. Ce schéma d'intégration est basé sur une approche implicite ou semi-implicite, ce qui permet de traiter de manière stable des systèmes dynamiques complexes, en particulier lorsque des interactions de contact multiples et non linéaires sont présentes.

Le schéma d'intégration numérique décrit dans la thèse de Mathieu Renouf concerne les techniques utilisées pour avancer dans le temps les solutions des équations de mouvement des particules. Le choix du schéma d'intégration est crucial car il affecte la précision et la stabilité des simulations. Les méthodes implicites comme celle de Newmark sont souvent privilégiées dans des simulations complexes de milieux granulaires, car elles permettent de plus grands pas de temps tout en maintenant la stabilité, même en présence de fortes non-linéarités dues aux interactions de contact. La méthode de Newmark permet de contrôler la précision de l'intégration en ajustant des paramètres spécifiques, comme l'amortissement numérique, pour éviter les oscillations non physiques. Dans la thèse, Mathieu Renouf a discuté de l'adaptation de la méthode de Newmark pour les besoins spécifiques de la GEM, en expliquant comment cette méthode permet de modéliser de manière réaliste le comportement dynamique des particules granulaires tout en gérant efficacement les contraintes de contact.

### 7.1.3 Quelques aspects théoriques en analyse variationnelle

L'analyse variationnelle est un domaine spécialisé des mathématiques qui se concentre sur l'optimisation, la théorie du contrôle, l'analyse multi-valuée et les problèmes connexes. Elle englobe un large éventail de sujets, y compris l'analyse convexe, l'analyse non linéaire, l'analyse non lisse et la théorie des ensembles multi-valués. Ce domaine d'étude intègre fréquemment des concepts issus de la théorie de la mesure, de la géométrie différentielle et de l'analyse fonctionnelle. L'analyse variationnelle possède une longue et riche histoire que l'on peut trouver dans les ouvrages de J.P. Aubin et H. Frankowska [154], R.T. Rockafellar et R.J. B. Wets [120], B.S. Mordukhovich [112], J.P. Penot [116], A.D. Ioffe [107], L. Thibault [124], etc. On peut dire que l'analyse variationnelle est l'extension et aussi le développement du Calcul des Variations classique et de l'Analyse Convexe vers une théorie plus générale. D'une part, les problèmes d'optimisation apparaissent souvent dans les sciences appliquées. D'autre part, résoudre des problèmes basés sur l'optimisation est une méthode efficace en mathématiques. Cela fait de l'analyse variationnelle un domaine d'intérêt tant du point de vue théorique qu'appliqué.

Dans notre travail, nous nous concentrons principalement sur deux problèmes et concepts en Analyse Variationnelle:

- Convexité faible et forte : Elles concernent des ensembles où les minima locaux sont des minima globaux avec des exigences de courbure moins strictes que celles de la convexité forte, qui assure une courbure extérieure uniforme. Cela est important pour les milieux granulaires et la dynamique de contact, où les ensembles peuvent ne pas être uniformément convexes mais peuvent être analysés en utilisant la convexité faible. L'un des concepts principaux de la convexité faible est la régularité proximale. Elle permet de traiter des ensembles non lisses ou irréguliers dans un cadre d'optimisation convexe, en assurant la définition précise du cône normal et des applications de projection, essentiels pour les simulations numériques de milieux granulaires. Chemetov et Monteiro-Marques, ainsi que Castaing et al., ont établi des solutions pour les processus de balayage perturbés dans des ensembles uniformément prox-réguliers [15, 16]. Castaing et al. ont démontré l'existence de solutions pour le processus de balayage dépendant de l'état dans des ensembles uniformément prox-réguliers en utilisant une extension du théorème de Schauder combinée à une méthode de discrétisation [17].

- Régularité métrique : elle fait le lien entre la convexité faible et les structures d'ensembles plus générales, en mesurant le comportement de la fonction distance par rapport à un ensemble, utile en optimisation et en analyse variationnelle, comme pour comprendre le comportement des particules dans les systèmes granulaires. Azzam-Laouir et al. [18] ainsi que Haddad et al. [19] ont démontré des processus de balayage perturbés multivoques efficaces dans des ensembles uniformément prox-réguliers en dimensions finies.

Intégrer ces concepts améliore les modèles numériques pour des scénarios complexes, renforce les fondations théoriques et étend l'applicabilité des processus de balayage en mécanique et en optimisation. Noel [20] et Thibault [21] ont exploré des processus de balayage perturbés multivoques dépendants de l'état avec une prox-régularité uniforme et ont utilisé la régularisation de Moreau-Yosida pour des ensembles uniformément prox-réguliers [9]. Cette transition enrichit notre compréhension théorique et pratique de ces phénomènes, avec la régularité métrique et la fonction de distance maximale jouant des rôles cruciaux en optimisation et en mécanique.

#### 7.1.3.1 Convexité faible et forte

L'existence de la solution au processus de balayage 1 est établie sous l'hypothèse que $C(t)$ est un ensemble convexe fermé en mouvement. Une extension naturelle pour parvenir à l'existence d'une solution consiste à considérer l'ensemble $C(t)$ comme faiblement convexe ou, à l'inverse, fortement convexe. Par conséquent, l'étude des propriétés de l'ensemble en mouvement constitue également un sujet important de cette thèse.

La prox-régularité est reconnue comme un concept fondamental en analyse variationnelle, permettant d'aller au-delà de la convexité (voir Poliquin, Rockafellar dans [118], Thibault dans [124] et quelques articles connexes [127], [84], [125]). Elle donne lieu à de nombreuses applications intéressantes liées à divers domaines des mathématiques tels que l'optimisation, l'analyse multivoque, la géométrie différentielle et les EDP, etc.

(voir, par exemple, [30, 91, 93, 100, 106, 108] et les références qui y sont citées). Une application importante est le **Sweeping Process**, régi par des contraintes d'inégalité, comme celles que l'on trouve dans les médias granulaires et l'élastodynamique impliquant le contact. Ce processus est un outil puissant pour l'analyse mathématique et le développement de schémas numériques pour ces dynamiques non-régulières. En outre, ce type d'ensemble est également appliqué pour considérer certaines propriétés particulièrement importantes telles que la *régularité métrique (submétrique et semi-métrique)*, la calmness, la transversalité, la propriété d'Aubin, etc. Les lecteurs peuvent en savoir plus dans les monographies [120], [100], [84], ... Il est bien établi dans [115] que le complément d'un ensemble prox-régulier n'est rien d'autre que l'union d'une famille de boules fermées de rayon commun, communément appelées ensembles $R$-fortement convexes. Par conséquent, l'étude de la prox-régularité dirige naturellement notre attention vers l'exploration de la forte convexité, approfondissant ainsi notre compréhension de la courbure des fonctions et des paysages d'optimisation. Parmi les contributeurs notables à ce domaine figurent H. Frankowska, C. Olech [104], J.P. Vial [125], E.S. Polovinkin [119], M.V. Balashov [88], G. E. Ivanov [105, 108, 110], A. Weber et G. Reibig [126]. Pour l'utilisation de tels ensembles dans divers problèmes mathématiques appliqués, nous nous référons, par exemple, le Sweeping Process [42, 96], contrôle optimal [104], l'optimisation numérique [111, 117], les jeux différentiels [108, 110], etc.

Les fonctions de distance jouent un rôle fondamental en analyse mathématique, y compris en analyse convexe [92, 116, 120], analyse variationnelle [97, 107, 112, 116, 120, 124], inclusions différentielles [87, 98, 114, 123], contrôle optimal [95, 113], théorie de l'approximation [102, 122], analyse de forme et EDP [85, 86, 101], etc. Il est bien connu (et facilement observable) que la convexité d'un sous-ensemble fermé $C$ dans un espace normé est équivalente à la convexité de sa fonction de distance associée (standard/usuelle). Une question naturellement posée est de savoir quel est l'ensemble pour lequel la distance est semi-convexe? Dans [88], M.V. Balashov étend dans un espace de Hilbert général une telle caractérisation de base en établissant que la semi-convexité (c'est-à-dire la convexité jusqu'à une norme carrée) de la fonction de distance $d_C$ (sur tout ensemble convexe d'un agrandissement approprié de $C$) est équivalente à la prox-régularité uniforme de l'ensemble fermé $C$. Rappelons qu'un sous-ensemble fermé $C$ dans un espace de Hilbert $X$ est uniformément prox-régulier ([118]) (également connu sous le nom de positivement atteint, faiblement convexe, $\varphi-$convexe, $\mathcal{O}(2)$-convexe, lisse proximale, voir, par exemple, [94, 99, 103, 121, 125] et les références qui y figurent) avec une constante $r > 0$ à condition que l'application du point le plus proche $\mathrm{proj}_C$ soit bien définie sur un agrandissement approprié de $C$ et continue à l'intérieur de celui-ci.

Il est connu que la fonction de distance à un ensemble convexe admet diverses descriptions impliquant des relations de dualité issues de l'analyse convexe (voir, par exemple, [102]). Dans le cadre de la régularité métrique, en plus de la fonction de distance, nous étudions les points les plus éloignés et leurs fonctions de distance correspondantes, qui mesurent la distance d'un point extérieur au point le plus éloigné d'un sous-ensemble donné. Les premières recherches sur ces fonctions ont été introduites par Jessen ([156], 1940) et des recherches supplémentaires ont été effectuées par Motzkin, Strauss, et Valentine ([158], 1953) dans un espace métrique et par Fitzpatrick ([157],

1980) dans un espace de Banach. Considérons, par exemple, une boule : le point le plus éloigné de son centre englobe tous les points sur sa frontière. Dans un cadre convexe, bien que la projection du point le plus proche sur un ensemble convexe soit unique, la fonction de distance la plus éloignée, à savoir $\mathrm{dfar}_S(x) := \sup_{y\in S} \|x - y\|$, se comporte différemment. Cet opérateur présente une déviation par rapport à la notion conventionnelle du point le plus proche, soulignant que l'existence du point le plus éloigné n'est pas garantie avec les concepts classiques d'ensembles convexes ou faiblement compacts. Il est naturel d'explorer des ensembles alternatifs mieux adaptés pour déterminer le point le plus éloigné.

Dans de nombreuses situations, il est également intéressant de renforcer la propriété de convexité. Cela peut être réalisé avec des ensembles $R$-fortement convexes ([105, 108, 119, 125, 126]) qui sont essentiellement l'intersection de boules fermées de rayon commun $R > 0$. Dans le travail [89], il est établi que pour tout ensemble fermé borné dans $X$, la propriété de forte convexité est équivalente à la semi-concavité de la fonction de distance (convexe) la plus éloignée $\mathrm{dfar}_C(x) := \sup_{c\in C} \|x - c\|$. Par conséquent, une idée est d'examiner la propriété de semi-concavité de la fonction de distance la plus éloignée associée à un ensemble fortement convexe. En outre, nous développons les analogues de certaines propriétés des ensembles convexes [102] ou des ensembles prox-réguliers [84] pour un ensemble fortement convexe $C$ lorsque l'ensemble $C$ satisfait à la condition renforcée de forte convexité.

#### 7.1.3.2 Régularité métrique

D'autre part, la régularité métrique [107, 112, 124, 131, 132, 139, 142, 144] est un concept en analyse mathématique, particulièrement dans le domaine de l'optimisation et de l'analyse variationnelle. Elle est étroitement liée à la stabilité et à la robustesse des solutions aux problèmes d'optimisation et aux systèmes d'équations. Plus précisément, la régularité métrique aide à comprendre la sensibilité des solutions optimales aux perturbations des paramètres ou des contraintes, tout en fournissant des bornes d'erreur pour les solutions approximatives aux équations et aux inégalités. De plus, en analyse variationnelle, la régularité métrique peut être un outil précieux pour étudier la convexité faible d'un sous-ensemble particulier. La régularité métrique est étroitement liée à d'autres concepts tels que la continuité Lipschitz [124], la régularité de Clarke [124, 161], le théorème de Robinson-Ursescu [142, 144], etc.

Soit $M : X \rightrightarrows Y$ une multimapping entre deux espaces de Banach avec un graphe convexe fermé $\mathrm{gph}\, M$ et soit $(\overline{x}, \overline{y}) \in \mathrm{gph}\, M$. En 1975-1976, C. Ursescu ([144]) et S.M. Robinson ([142]) ont établi indépendamment que l'existence d'un réel $\gamma \geq 0$ tel que l'inégalité 1.1 est équivalente à l'inclusion $\overline{y} \in \mathrm{core}\, M(X)$. On dit naturellement que la multimapping $M$ est $\gamma$-régulière métriquement en $\overline{x}$ pour $\overline{y}$. La propriété de régularité métrique a une longue et profonde histoire qui remonte aux travaux pionniers de L.A. Lyusternik ([139]) et L.M. Graves ([132]) et a été développée depuis dans de nombreux articles et livres (voir, par exemple, [107, 112, 124, 131] et les références qui s'y trouvent). Une telle propriété est connue pour être équivalente soit à un certain comportement lipschitzien de la multimapping $M^{-1}$ soit à une condition de type ouverture (avec un taux linéaire), à savoir l'existence de constantes positives $\alpha, \beta > 0$ telles que $B[y, \alpha\beta t] \subset M(B[x, t\alpha])$ pour tout $t \in ]0, 1]$ et tous $(x, y) \in \mathrm{gph}\, M$ proches de $(\overline{x}, \overline{y})$. Outre le théorème de Robinson-Ursescu (qui peut être considéré comme

une extension du célèbre théorème de l'application ouverte de Banach-Schauder), la régularité métrique est fortement impliquée dans le calcul sous-différentiel, les estimations de co-dérivées et les conditions d'optimalité (voir, par exemple, [112, 116, 124] et les références qui s'y trouvent).

Au fil des ans, des théorèmes de type Robinson-Ursescu pour les multimappings avec un graphe possiblement non convexe ont été fournis. Une première façon naturelle de dépasser la convexité dans un tel contexte réside dans le concept de paraconvexité. Dans [138], H. Huang et R.X. Li ont établi que si l'inverse de la multimapping $M$, à savoir $M^{-1}$, est paraconvexe, alors $M$ est métriquement régulier en $\overline{x}$ pour $\overline{y}$ lorsque 1.3 est vérifiée pour certains réels $\alpha, \beta > 0$. En 2012, X.Y. Zheng et K.F. Ng ([145]) ont montré que la prox-régularité des ensembles ([118]) est également un concept adapté pour développer des versions non convexes du théorème de Robinson-Ursescu. Deux ans plus tard, X.Y. Zheng et Q.H. He ont fourni dans [146] un théorème de type Robinson-Ursescu pour les multimappings avec un certain comportement variationnel d'ordre un, à savoir avec un graphe $(\sigma, \delta)$-subsmooth ([129]). Comme montré dans [128], les résultats ci-dessus de [145, 146] peuvent être étendus à la classe des multimappings $M$ avec un graphe normalement $\omega(\cdot)$-régulier. Plus précisément, il est établi dans [128] qu'une multimapping $M$ avec un graphe normalement $\omega(\cdot)$-régulier satisfaisant la condition d'ouverture (1.3) pour certains réels $\alpha, \beta, \rho > 0$ telle que l'inégalité 1.4 est $\gamma$-métriquement régulière en $\overline{x}$ pour $\overline{y}$ pour un certain réel $\gamma \leq \rho$. Les auteurs de [128] tirent de leur étude divers résultats de préservation pour les ensembles $\omega(\cdot)$-normalement réguliers qui complètent les travaux précédents consacrés à la stabilité des propriétés de prox-régularité et de subsmoothness [100, 125, 127, 137, 152] (voir aussi le récent article de G.E. Ivanov [133]).

Le théorème de Robinson-Ursescu a inspiré diverses extensions qui dépassent les limites de la convexité, élargissant ainsi le domaine de la régularité métrique et de ses applications. Dans cette thèse, il est intéressant de fournir des versions régulièrement normales du théorème de Robinson-Ursescu ainsi que des conditions suffisantes générales assurant la préservation de la $\omega$-régularité normale sous la subrégularité métrique.

## 7.2 Objectifs de la thèse et plan du manuscrit

### 7.2.1 Objectifs

**Aspect numérique:**
Dans ce travail, nous nous intéressons à un processus de balayage discontinu (mais de variation bornée), de second ordre de Moreau, modélisant la dynamique de contact de particules rigides ([47] et les références connexes [8, 23, 28, 147]). Rappelons que les processus de balayage [42] sont des inclusions différentielles particulières régies par le cône normal d'un ensemble mobile (non convexe). La loi de contact est modélisée à travers la régularisation de Moreau-Yosida [9, 27] de la condition unilatérale. Il semble que la régularisation de Moreau-Yosida avec le paramètre $\alpha \geq 2$ soit un outil approprié pour développer un modèle de contact régulier (compliance normale) qui maintient l'énergie cinétique du système tout en assurant que le contact n'autorise pas aux particules de passer à travers l'autre (non-pénétration). Le **comportement normal amélioré discret** (INC) est considéré comme une méthode appropriée pour

garantir la conservation de l'énergie dans le cadre continu. Pour résoudre le problème de non-linéarité, une combinaison du schéma de Newmark avec la méthode de Newton semi-régulier et de l'algorithme PDAS, qui utilise les fonctions de complémentarité des différents modèles de contact [60], sera employée.

L'objectif principal de ce travail est proposer une méthode de régularisation implicite pour laquelle la conservation de l'énergie et la non-pénétration sont assez similaires à NSCD-NLGS ainsi qu'à un coût computationnel approprié. Plusieurs expériences numériques sont rapportées à des fins de vérification et de validation, ainsi que pour évaluer l'efficacité et les performances de la méthode *Newmark-PDAS-INC* par rapport à d'autres méthodes numériques (DEM, NSCD-NLGS).

De plus, dans le cadre de notre configuration, un deuxième objectif essentiel consiste à développer un code autonome utilisé de manière exhaustive pour la méthode INC, en utilisant le langage FORTRAN avec la bibliothèque LAPACK-`Dgesv`. Ensuite, nous appliquons plusieurs solveurs pour résoudre les systèmes linéarisés afin d'améliorer le coût computationnel, à savoir LU, factorisation complète-`SuperLU`, `GMRES` avec préconditionneur ILU, Gradient Conjugué Préconditionné (préconditionneur ILU-`Pcg` ou préconditionneur Jacobi-`Jpcg`) en utilisant la bibliothèque SPARSKIT. De plus, nous développons également une version directe de la matrice creuse combinant soit `SuperLU/GMRES` pour le frottement, soit `Pcg/Jpcg` pour le cas sans frottement.

**Aspect théorique:**
Outre l'aspect numérique, nous abordons la partie théorique de l'analyse variationnelle pour considérer certaines nouvelles propriétés qui sont beaucoup plus utiles en mathématiques. Plus précisément, nous nous concentrerons sur deux travaux principaux.

Dans le premier travail, nous examinons les concepts de **fonction de distance la plus éloignée des ensembles fortement convexes dans les espaces de Hilbert**. Le premier objectif est de fournir une preuve alternative d'une telle caractérisation des ensembles fortement convexes. Notre approche s'inscrit dans la lignée de [115] en se basant sur le fait que le complément d'un ensemble prox-régulier n'est rien d'autre que l'union d'une famille de boules fermées de rayon commun. Le deuxième objectif du présent travail est de fournir en détail l'analyse des caractéristiques correspondantes (i), (ii) et (iii) dans la partie 1.4.1 lorsque l'ensemble $C$ satisfait la condition renforcée de convexité forte.

Le deuxième travail concerne les **propriétés de sous-régularité métrique et de régularité normale** $\omega(\cdot)$. Dans le présent article, nous montrons d'abord que la propriété de sous-régularité métrique du multimapping $M$ (c'est-à-dire l'inégalité (1.1) avec $y = \overline{y}$ (voir Section 6.2)) est une hypothèse appropriée pour obtenir la normale $\omega(\cdot)$-régularité de l'image inverse $M^{-1}(\overline{y})$. Nous établissons également la régularité normale d'un ensemble d'équations généralisé, disons $S := \{x \in X : f(x) \in M(x)\}$ avec $f : X \to Y$ une application (à valeur unique) sous une inégalité de sous-régularité métrique, à savoir $d(x, S) \leq \gamma d\Big(\big(x, f(x)\big), \operatorname{gph} M\Big)$, $x$ proche de $\overline{x}$. Bien sûr, nous exigeons dans les deux cas une propriété normale $\omega(\cdot)$ sur le multimapping impliqué $M$ (soit sur le codérivé soit sur le graphe). Nous remplaçons ensuite naturellement (dans la lignée de [128]) cette hypothèse de sous-régularité métrique par l'inclusion (1.3) avec $\alpha, \beta > 0$ tel que $\beta > \omega\big(\sqrt{\alpha^2 + \beta^2}\big)$. Enfin, nous montrons qu'un multimapping

Lipschitz (par rapport à la distance de Hausdorff-Pompeiu) avec des valeurs de normale $\omega(\cdot)$-régularité près de $\overline{x}$ bénéficie d'une certaine propriété de sous-régularité métrique en $\overline{x}$ pour $\overline{y}$.

### 7.2.2 Organisation de la thèse

Tout d'abord, nous commençons par le chapitre 2 : Notation and preliminaries où nous aborderons la compréhension des éléments de l'Analyse Variationnelle tels que les fonctions convexes, certains résultats classiques en optimisation, les cônes normaux et les sous-différentiels (au sens de Clarke, Fréchet et Mordukhovich, proximaux), les opérateurs maximaux monotones et la régularisation $\alpha$-Moreau Yosida respectivement à travers les sections 2.1 à 2.5.

La thèse se compose de deux parties.

Partie I est **Modélisation du Processus de Balayage**.

Nous commençons par formuler un problème dynamique pour analyser le mouvement des particules à l'intérieur d'une région prédéfinie, en mettant l'accent spécifiquement sur la loi de contact utilisant des conditions spécifiques de compliance normale, désignée comme Improved Normal Compliance (INC). Basé sur le travail de Hauret et Le Tallec [57], nous proposons une approche spécifique dans le cadre discret qui assure la conservation de l'énergie du système, en l'alignant avec le cas continu. Dans le Chapitre 3, nous approfondissons le processus de balayage de second ordre de Moreau pour modéliser la dynamique du contact, en incorporant la régularisation Moreau-Yosida et ainsi la méthode INC. Cette approche vise à garantir la conservation de l'énergie du système tout en évitant la pénétration entre les particules.

Dans le Chapitre 4, nous adoptons une approche de discrétisation temporelle en utilisant le schéma implicite de Newmark pour le modèle INC, combiné avec la méthode Primal-Dual Active Set (PDAS). Cette méthode intégrée, connue sous le nom de Newmark-PDAS-INC, facilite la résolution efficace du problème. De plus, nous effectuons une analyse comparative des solutions obtenues par différentes méthodes, y compris INC, la méthode des éléments discrets (DEM) et la dynamique de contact non lisse avec Gauss-Seidel non linéaire (NSCD-NLGS), pour évaluer l'efficacité de notre algorithme. Une analyse comparative est également menée par rapport à des méthodes numériques alternatives, telles que NSCD-NLGS et DEM, spécifiquement dans des scénarios sans frottement. Enfin, nous éclaircissons le traitement numérique du coefficient de restitution et des conditions de contact impliquant le frottement, dans le but d'offrir une solution robuste pour résoudre les problèmes de contact multi-corps rigides. La plupart des résultats de ce chapitre, ainsi que la section sur la régularisation de Moreau-Yosida pour la dynamique de contact dans le Chapitre 3, sont présentés dans les articles suivants [1] et [2].

Partie II est **Éléments de l'analyse variationnelle**.

La fonction de distance la plus éloignée est un outil géométrique utilisé pour mesurer le point le plus éloigné dans un ensemble par rapport à un point de référence, avec des applications dans des domaines allant de la géométrie computationnelle aux problèmes de localisation d'installations. Dans le Chapitre 5, nous proposons de nouvelles caractérisations de la fonction de distance la plus éloignée. D'autre part, le théorème de

Robinson-Ursescu (qui peut être considéré comme une extension du célèbre théorème de l'application ouverte de Banach-Schauder) est fortement impliqué dans le calcul sous-différentiel, les estimations des codérivées et les conditions d'optimalité. Nous passons le Chapitre 6 à fournir des conditions suffisantes générales garantissant la préservation de la $\omega$-régularité normale pour des équations généralisées et nous nous concentrons également sur les versions régulièrement normales du théorème de Robinson-Ursescu.

Enfin, dans le Chapitre 7, nous résumons la thèse en français (comprenant une vue d'ensemble, les objectifs, la structure, les résultats obtenus et les perspectives). Ensuite, l'Annexe A présente le code pour la méthode Newmark-PDAS-INC avec une version directe de la matrice creuse combinée avec le solveur `Gmres` pour le système linéaire.

## 7.3 Résultat

### 7.3.1 Aspect numérique

1. Basé sur l'importance croissante de comprendre les milieux granulaires et leurs comportements dans diverses industries, la première partie de cette thèse vise à proposer une nouvelle méthode différente de la Méthode des Éléments Discrets (DEM) et des approches de Dynamiques de Contact Non-Lisse (NSCD) pour modéliser la dynamique des granulaires. Nous nous concentrons sur un processus de balayage discontinu de second ordre de Moreau pour modéliser la dynamique des contacts, en incorporant la régularisation de Moreau-Yosida avec le paramètre alpha pour développer un modèle de contact régulier. Nous proposons la méthode de Compliance Normale Améliorée (INC) pour assurer la conservation de l'énergie et utilisons une combinaison de la méthode de Newmark et de l'ensemble actif Primal-Dual (PDAS) pour traiter la non-linéarité. Cette étude vise à comparer l'efficacité de notre approche avec d'autres techniques de modélisation numérique, telles que DEM et NSCD via la méthode de Gauss-Seidel Non-Linéaire (NLGS), en se concentrant sur l'amélioration de la conservation de l'énergie et du coût computationnel, plus de détails dans [1].

2. Un deuxième objectif clé consiste à développer un code autonome largement utilisé pour la méthode INC, en utilisant le langage FORTRAN et la bibliothèque LAPACK - `Dgesv`. Nous appliquerons ensuite plusieurs solveurs pour résoudre des itérations non linéaires afin d'améliorer le coût computationnel. Ces solveurs incluent la factorisation LU, la factorisation complète avec `Superlu`, `Gmres` avec préconditionneur ILU, et le gradient conjugué préconditionné `Pcg`/`Jpcg` en utilisant la bibliothèque SPARSKIT, en préparation de la publication [2].

### 7.3.2 Aspect théorique

1. La publication avec F. Nacry et L. Thibault [3] explore la fonction de distance maximale aux ensembles fortement convexes dans les espaces de Hilbert. Nous fournissons une preuve alternative de la caractérisation des ensembles fortement convexes, en nous inspirant des travaux antérieurs. De plus, nous menons une

analyse détaillée des caractéristiques spécifiques lorsque l'ensemble satisfait à une condition renforcée de convexité forte.

2. La publication avec F. Nacry et J. Venel [4] est consacrée au problème des préservations qui examine le comportement de la possibilité de convexité forte et des opérations sur les ensembles. Le résultat principal consiste à donner certaines conditions suffisantes garantissant la préservation de la régularité normale $\omega$ pour des équations généralisées.

## 7.4 Perspectives

### 7.4.1 Aspect numérique

Plusieurs perspectives sont envisagées pour améliorer l'efficacité algorithmique et informatique, explorer de nouveaux domaines d'application et établir rigoureusement des fondements théoriques.

1. *Parallélisme Algorithmique et Informatique*

   - Parallélisme Algorithmique: Une première perspective de ce travail est d'explorer le parallélisme algorithmique, notamment à travers la Méthode de Décomposition de Domaine (DDM) avec une méthode de Schwarz additive intégrant des modes grossiers. Cette approche permettrait de diviser le problème en sous-domaines plus petits et indépendants, facilitant ainsi la résolution parallèle des systèmes linéaires. Une attention particulière sera portée sur l'optimisation des algorithmes pour améliorer la convergence et l'efficacité computationnelle.
   - Parallélisme Informatique: Parallèlement, le parallélisme informatique sera exploré pour optimiser les opérations de produits matrice-vecteur et autres calculs intensifs. En exploitant les architectures modernes de calcul parallèle (comme les GPU et les clusters de calcul), il serait possible d'accélérer significativement les simulations, rendant l'approche Newmark-PDAS-INC plus applicable à des problèmes de grande échelle.

   La combinaison du parallélisme algorithmique et informatique est essentielle pour rendre les méthodes numériques plus efficaces et applicables à des problèmes de grande taille. En détaillant les avantages spécifiques des méthodes de décomposition de domaine et des architectures parallèles modernes, on montre une direction claire pour les travaux futurs.

2. *Applications Potentielles*

   - Interaction Fluide-Structure: Une deuxième perspective consiste à appliquer la méthodologie développée à des problématiques d'interaction fluide-structure, telles que les lits fluidisés et les mouvements de foules. Ces applications permettront de valider la pertinence et la robustesse de l'approche Newmark-PDAS-INC dans des contextes variés. Par exemple, dans les lits

fluidisés, la méthode pourrait aider à modéliser les interactions complexes entre les particules solides et le fluide environnant.

- Validation et Extension: En validant la méthode dans ces nouveaux contextes, on pourra non seulement démontrer sa polyvalence, mais aussi identifier d'éventuelles améliorations et extensions pour mieux capturer les dynamiques spécifiques à chaque application.

En identifiant des applications concrètes comme les lits fluidisés et les mouvements de foules, on ancre le travail théorique dans des contextes pratiques et pertinents. Cela renforce l'importance de la méthode développée et ouvre des opportunités pour des collaborations interdisciplinaires

3. *Théorie de l'Existence pour le Processus de Balayage de Second Ordre de Moreau*

   - Existence de la Solution: Enfin, une perspective théorique essentielle est d'établir rigoureusement l'existence de solutions pour le processus de balayage de second ordre de Moreau, utilisé pour modéliser la dynamique des contacts avec la régularisation de Moreau-Yosida. Cette étude théorique permettra de garantir la robustesse mathématique du modèle et de renforcer la confiance dans les résultats numériques obtenus.
   - Régularisation de Moreau-Yosida: En approfondissant la compréhension de la régularisation de Moreau-Yosida, il serait possible de mieux caractériser les conditions sous lesquelles le modèle reste valide et stable, ouvrant ainsi la voie à de nouvelles applications et à des améliorations méthodologiques.

   La validation théorique de l'existence des solutions est cruciale pour assurer la rigueur scientifique du modèle. En se concentrant sur la régularisation de Moreau-Yosida, on met en lumière un aspect technique clé qui peut avoir des implications profondes pour la stabilité et la précision des simulations.

### 7.4.2 Aspect théorique

1. Dans notre article [3], nous avons établi des résultats pour la fonction de distance maximale dans le contexte des espaces de Hilbert. La prochaine étape consiste à étendre ces résultats au cadre plus général des espaces de Banach.

2. Considérez le processus de balayage dépendant de l'état décrit par :

$$\begin{cases} -\dot{u}(t) \in \mathrm{N}\big(C(t,u(t)); u(t)\big), \\ u(0) = u_0. \end{cases}$$

   Dans les résultats existants concernant ce problème, l'hypothèse de compacité sur l'ensemble mobile $C(t,x)$ joue un rôle crucial. Nous souhaitons comprendre la nécessité de cette hypothèse. Une approche initiale consiste à examiner le scénario où l'ensemble mobile est un ensemble fortement convexe, tel qu'une boule mobile, qui n'est pas compacte dans les dimensions infinies.

3. Considérez l'inclusion différentielle, connue sous le nom de *problème de la descente la plus raide*, donnée par :

$$\begin{cases} -\dot{u}(t) \in \partial_P f(u(t)) \quad \text{p.p. dans } I, \\ u(T_0) = u_0. \end{cases}$$

où $T_0 \in [0, +\infty[$ est le temps initial et $f$ est une fonction prox-régulière (ou primalement inférieure régulière). L'existence de solutions locales et globales est établie par les Théorèmes 2.9 et 3.2 dans [169]. Nous visons à analyser le comportement asymptotique des solutions à l'inclusion différentielle dépendante du temps :

$$\begin{cases} -\dot{u}(t) \in \partial_P f_t(u(t)) \quad \text{p.p. dans } I, \\ u(T_0) = u_0. \end{cases}$$

Ici, $f_t$ désigne une fonction dépendante du temps, et l'accent est mis sur la compréhension de l'impact de la dépendance temporelle. Une question ouverte est ce qui se passe dans le cas convexe classique où $f_t = \mathrm{dfar}_{C(t)}$ ?

4. Dans l'article [4], nous avons établi divers résultats de préservation concernant la préservation des ensembles normaux $\omega$-réguliers (y compris les ensembles prox-réguliers et subsmooth) ainsi que la sous-régularité métrique des multimappings avec la $\omega(\cdot)$-régularité normale, que ce soit pour le graphe ou les valeurs. En nous appuyant sur ces résultats, nous visons à étendre ces résultats de préservation au cadre des ensembles fortement convexes, qui sont les duaux des ensembles prox-réguliers.

# Appendix A

# Newmark-PDAS-INC code testing

```
PROGRAM INC
    ! IMPROVE NORMAL COMPLIANCE + Sparse version+Gmres solver
    ! Consider case: Newmark scheme with gam = 1.0 and   beta = 0.5
    Implicit None
    integer :: np,NB_WALL,NB_CORNER,ntdl,i_alpha!,INFO
    integer, parameter :: dp = selected_real_kind(15, 307)
    real(kind=dp), dimension(:), allocatable :: qd_new,q_new,dq,dq_a,
       scm,scm_old
    real(kind=dp), dimension(:), allocatable :: f_free,PENET
    real(kind=dp), dimension(:), allocatable :: qd_old,q_old
    real(kind=dp), dimension(:), allocatable :: s, s_old,scl,scl_old,
       FC
    real(kind=dp), dimension(:), allocatable :: m,Mass
    real(kind=dp), dimension(:), allocatable :: dg,r,rhs,IPIV,p,
       impulse, ec, ec_0
    real(kind=dp), dimension(:), allocatable :: MAT_VAL
    real(kind=dp), dimension(2)   :: normal,normal_cl,tangent,
       tangent_cl
    real(kind=dp), dimension(2,2) :: nor_mat, tan_mat, tan_nor_mat
    real(kind=dp), dimension(3,3) :: A_II,A_IJ,A_JI,A_JJ,AFRIC
    real(kind=dp), dimension(:,:), allocatable :: WALLS,A
    real(kind=dp), dimension(:,:), allocatable :: LIN,DIAG
    integer, dimension(:,:), allocatable :: ROW
    integer, dimension(:)  , allocatable :: NJ,COL_IND,ROW_PTR
    real(kind=dp):: T_end,dt,NL_EPS,para,alpha,at,gap,tilde_gap,
       gap_old,c_nu,c_tau,v_tan
    real(kind=dp):: gapcl,gapcl_old,max_pen_loc,max_pen,som,epsy,crit,
       max_r,pesanteur
    real(kind=dp):: init_ec,final_ec,enr,enrs,INV2_DT
    logical      :: CONTACT, CONT_CL
    real(kind=dp):: START,FINISH
    real(kind=dp):: tol,epss,Total_fc
    real(kind=dp):: mu      !Friction coefficient
    real(kind=dp):: en      !Restitution coefficient
    integer      :: II,NNZ,NI
    integer      :: l,it_time,i,j,K,nl_it,NL_IT_MAX,unit,imax,noc,
       TEST_CASE,noc_cl
    integer      :: nb_contact_wall, nb_contact_par, total_nl_it
    integer      :: nzmax,maxits,iout,im,kmax,max_nlit

```

```
34      real(kind=dp), dimension(:), allocatable :: alu,wu,wl,vvv
35      integer, dimension(:), allocatable   :: jlu, ju
36      integer, dimension(:), allocatable   :: jr,jwl,jwu
37      integer :: ierr,lfil
38
39      T_end = 7.5D00
40      dt    = 1.D-4
41      alpha = 3
42      para  = 2.0D+08
43      c_nu  = 0.5D00*para
44      c_tau = 1.0D+015
45      mu    = 0.5D00
46      en    = 0.0D00
47      NL_IT_MAX = 50
48      NL_EPS = 1.D-5
49      imax  = INT(T_end/dt)
50      INV2_DT = 2.D00/DT
51      PESANTEUR = 1.D00 ! 0.D00
52      !--
53      !call command_line(dt)
54      !call command_line(alpha)
55      !call command_line(para)
56      !call command_line(en)
57      call command_line(mu)
58
59      call SYSTEM("rm -f ./VISU/*.vtk")
60      unit = 120
61
62      !!!!!!--------------------------------------------------
63      !> THE EXAMPLES
64      !TEST_CASE = 99  !one_particle    (TEST_difference bw 2 files)
65      !TEST_CASE = 100 !one_particle_a  (along to x)
66      !TEST_CASE = 101 !one_particle_b  (along to y)
67      !TEST_CASE = 102 !one_bounce      (free falling)
68      !TEST_CASE = 200 !two_particles_a (x and y)
69      !TEST_CASE = 201 !two_particles_b (along to x)
70      !TEST_CASE = 300 !three_particles (along to x)
71      !TEST_CASE = 400 !four_particles  (x and y)
72      !TEST_CASE = 401 !four_newton     (four particles along x)
73      !TEST_CASE = 500 !81_particles_free (y)
74      !TEST_CASE = 501 !81_particles    (x and y)
75      TEST_CASE = 502  !1225 particles
76      !TEST_CASE = 503 !25
77      !TEST_CASE = 504 !100   !fric
78      !TEST_CASE = 529 !529 particles
79      !!!!!!--------------------------------------------------
80
81      if     (TEST_CASE==99) THEN
82         call INITIALIZE_one_xy(q_old,qd_old,r,m,np,ntdl,WALLS)
83       else if (TEST_CASE==100) THEN
84         call INITIALIZE_one_a(q_old,qd_old,r,m,np,ntdl,WALLS)
85       else if (TEST_CASE==101) THEN
86         call INITIALIZE_one_b(q_old,qd_old,r,m,np,ntdl,WALLS)
87       else if (TEST_CASE==102) THEN
88         call INITIALIZE_bounce(q_old,qd_old,r,m,np,ntdl,WALLS)
89       else if (TEST_CASE==200) THEN
```

```
90        call INITIALIZE_two_a(q_old,qd_old,r,m,np,ntdl,WALLS)
91      else if (TEST_CASE==201) THEN
92        call INITIALIZE_two_b(q_old,qd_old,r,m,np,ntdl,WALLS)
93      else if (TEST_CASE==300) THEN
94        call INITIALIZE_three(q_old,qd_old,r,m,np,ntdl,WALLS)
95      else if (TEST_CASE==400) THEN
96        call INITIALIZE_four(q_old,qd_old,r,m,np,ntdl,WALLS)
97      else if (TEST_CASE==401) THEN
98        call INITIALIZE_newton(q_old,qd_old,r,m,np,ntdl,WALLS)
99      else if(TEST_CASE==500) THEN
100       call INITIALIZE_81_free(q_old,qd_old,r,m,np,ntdl,WALLS)
101     else if(TEST_CASE==501) THEN
102       call INITIALIZE_81(q_old,qd_old,r,m,np,ntdl,WALLS)
103     else if(TEST_CASE==502) THEN
104       call INITIALIZE_1225(q_old,qd_old,r,m,np,ntdl,WALLS)
105     else if(TEST_CASE==503) THEN
106       call INITIALIZE_25(q_old,qd_old,r,m,np,ntdl,WALLS)
107     else if(TEST_CASE==504) THEN
108       call INITIALIZE_100(q_old,qd_old,r,m,np,ntdl,WALLS)
109     else if (TEST_CASE==529) THEN
110       call INITIALIZE_529(q_old,qd_old,r,m,np,ntdl,WALLS)
111     end if
112
113     allocate(f_free(ntdl),p(ntdl),dq(ntdl),dq_a(ntdl),scm(ntdl),
          scm_old(ntdl))
114     allocate(qd_new(ntdl),q_new(ntdl)) !c(ntdl,ntdl)
115     allocate(dg(ntdl),rhs(ntdl),impulse(ntdl),ec(np),ec_0(np))
116     allocate(Mass(ntdl))
117     allocate(A(ntdl,ntdl))
118     allocate(s(np*np),scl(np*np))
119     allocate(s_old(np*np),scl_old(np*np))
120     allocate(IPIV(ntdl))
121     !WE SUPPOSE HERE THAT EACH PARTICLE HAS LESS THAN 10 CONTACTS
122     allocate(DIAG(ntdl,3),LIN(ntdl,30),ROW(NP,30))
123     allocate(MAT_VAL(30*ntdl),COL_IND(30*ntdl),NJ(NP))
124     allocate(ROW_PTR(ntdl+1))
125
126     ! ILU ou GMRES
127     nzmax = ntdl*ntdl
128     kmax  = 10 !50 IS THE MAXIMUM ALLOWED (OR CHANGE GEMRES PROGRAM)
129     lfil  = 5  !TO HAVE AN INCOMPLETE FACTORISATION: A diminuer
130     tol   = 5.D-2 !IS TO IMPOSE THE NUMBER OF ELEMENTS IN MATRIX LU
131     epss  = 1.D-05
132     maxits= ntdl
133     iout  = -1
134     im    = 10 !50 IS THE MAXIMUM ALLOWED
135
136     allocate(alu(nzmax),ju(ntdl),jlu(nzmax))
137     allocate(wu(ntdl+1),wl(ntdl),jr(ntdl),jwl(ntdl),jwu(ntdl))
138     allocate(vvv(ntdl*(IM+1)))
139     allocate(PENET(NP),FC(np))
140
141     max_pen = 0.D00 !Maximal Penetration over all the time steps
142     max_pen_loc = 0.D00 !Maximal Penetration on each time step
143     PENET       = 0.D00
144     FC              = 0.D00
```

```
145       nb_contact_wall = 0
146       nb_contact_par  = 0
147       total_nl_it     = 0
148       max_r           = 0.0D0
149       max_nlit        = 0
150
151       do i = 1,np
152          max_r = max(r(i),max_r)
153       enddo
154
155       !INITIAL KINETIC ENERGY
156       ec_0 = 0.0D00
157       do i = 1,np
158          ec_0(i) = ec_0(i) + 0.5D00*DOT_PRODUCT(M(3*I-2:3*I),qd_old(3*i
             -2:3*I)**2)
159       enddo
160       !INITIAL KINERTIC ENERGY OF SYSTEM (ALL PARTICLES)
161       init_ec  = 0.D00
162       do I = 1,np
163          init_ec  = init_ec  + ec_0(I)
164       end do
165
166       open(unit=160, file='kinetic_energy.txt', status='unknown')
167       write(160, *) 0.D00, init_ec
168
169       S       = 0.0D00
170       S_OLD   = 0.0D00
171       scl     = 0.0D00
172       scl_old = 0.0D00
173       AT      = 0.0D00
174       scm_old = 1.0D+4
175
176       call VISU_2D (0,WALLS,Q_OLD,QD_OLD,R,PENET,FC,NP,NB_WALL)
177       call cpu_time(start)
178       Mass = (2.D00/(dt**2))*M
179
180       !TIME ITERATIONS
181       do it_time = 1,imax
182          Total_fc = 0.0D0
183      !External forces
184          do i = 1,np
185             f_free(3*i-2:3*i) = [0.,-9.80665,0.]*PESANTEUR
186          end do
187          f_free= M * F_FREE   !matmul(M,f_free)
188
189          PENET = 0.D00
190          dq_a  = 0.D00
191
192          !Initialisation Newmark
193          qd_new = qd_old
194          q_new  = q_old + dt*qd_old
195
196          !NONLINEAR ITERATIONS
197          do nl_it = 1,NL_IT_MAX
198             ! At each contact
199             p(1:ntdl)   = 0.D00
```

```
200             s            = 0.D00
201             scl          = 0.D00
202             at           = 0.D00
203             tilde_gap    = 0.D00
204             max_pen_loc  = 0.D00
205
206             ROW       = 0
207             LIN       = 0.D00
208             NNZ       = 1
209             ROW_PTR(1)    = NNZ
210             NJ        = 1
211             DIAG      = 0.D00
212             MAT_VAL   = 0.D00
213             COL_IND   = 0
214
215             do i = 1,np
216                NI = NJ(I) !!! NUMBER OF THE DIAGONAL ELEMENT OF LINE I
217                LIN(3*I-2:3*I,NI:NI+2) = DIAG(3*I-2:3*I,1:3)
218                DO K = 0,2
219                   II = 3*I-2+K
220                   LIN(II,NI+K) = LIN(II,NI+K) + MASS(II)
221                END DO
222
223                ROW(I,NI:NI+2) = (/3*I-2,3*I-1,3*I/)
224
225                !CHECK CONTACT BETWEEN PARTICLE AND BOX
226                do j = 1, NB_WALL
227                   !Contact with the boundaries of the box
228                   cont_cl = is_cond_limit(q_new(3*i-2:3*i-1),q_old(3*i
                      -2:3*i-1),r(i),gapcl, &
229                                  gapcl_old,normal_cl,tangent_cl,WALLS(J
                                     ,1:2),WALLS(J,3:4))
230                   if (cont_cl) then
231                      A_II          = 0.D00
232                      noc_cl          = (i-1)*np+1
233                      scl(noc_cl)     = gapcl
234                      scl_old(noc_cl)= gapcl_old
235
236                      if (Scl(NOC_cl) .GT. -1.D-12) Scl(NOC_cl)
                         = 0.D0
237                      if (scl_OLD(NOC_cl) .GT. -1.D-12) Scl_OLD(NOC_cl)
                         = 0.D0
238
239                      if (nl_it .ne. 1) then  !Calculating tilde D_v
240                         som = 0.D00 !EXACT VALUE OF TILDE_GAP
241                         DO i_alpha = 0, int(alpha)-1
242                            som = som + scl(noc_cl)**(alpha-1-i_alpha)*
                                  scl_old(noc_cl)**i_alpha
243                         END DO
244                         tilde_gap = ((-1.D00)**alpha)*som
245
246                         if (DABS(scl(noc_cl) - scl_old(noc_cl))<1.D-10)
                             then
247                         !IF SMALL, TILDE_GAP AND AT REPLACED BY THE
                             LIMIT
```

```
248                         at = alpha*(alpha-1)*((- scl_old(noc_cl))**(
                               alpha-2))
249                       else
250                         at = ((-alpha)*((-scl(noc_cl))**(alpha-1)) -
                             tilde_gap)/(scl(noc_cl)-scl_old(noc_cl)
                               )
251                       endif
252                     else
253                       tilde_gap = scl(noc_cl)
254                       at = 1.0D00
255                     endif
256
257                     p(3*i-2:3*i-1) = p(3*i-2:3*i-1) + c_nu*(tilde_gap*
                          normal_cl+en*gapcl*normal_cl)/(1+en)
258                     nor_mat = (at/(1+en) + en/(1+en))*matmul(reshape(
                          normal_cl,(/2,1/)),reshape(normal_cl,(/1,2/)))
259
260                     A_II(1:2,1:2) =  A_II(1:2,1:2) + C_NU*NOR_MAT
261                     !friction conditions
262                     v_tan = qd_new(3*i-2)*tangent_cl(1) + qd_new(3*i
                          -1)*tangent_cl(2)
263                     v_tan = v_tan + r(i)*qd_new(3*i)
264                     V_TAN = C_TAU * V_TAN
265
266                     crit = dabs(v_tan) + mu*c_nu*tilde_gap
267                     if (crit <= 0.D-10) then
268                       !stick case
269                       !print*, 'stick case'
270                       p(3*i-2:3*i-1) = p(3*i-2:3*i-1) + v_tan*
                            tangent_cl
271                       p(3*i) = p(3*i) + v_tan*r(i)
272
273                       tan_mat =  matmul(reshape(tangent_cl,(/2,1/)),
                            reshape(tangent_cl,(/1,2/)))*INV2_DT*c_tau
274                       A_II(1:2,1:2) = A_II(1:2,1:2) + TAN_MAT
275                       A_II(1:2,  3) = A_II(1:2,  3) + INV2_DT*C_TAU*
                            TANGENT_CL*R(I)
276                       A_II(  3,1:2) = A_II(  3,1:2) + INV2_DT*C_TAU*
                            TANGENT_CL*R(I)
277                       A_II(  3,  3) = A_II(  3,  3) + INV2_DT*C_TAU*R
                            (I)*R(I)
278                     else
279                       !slip case
280                       EPSY = DSIGN(mu*c_nu,V_TAN)  ! SIGN OF THE SLIP
281                       !print*, 'slip case'
282                       p(3*i-2:3*i-1) = p(3*i-2:3*i-1) + epsy*
                            tilde_gap*tangent_cl
283                       p(3*i) = p(3*i) + epsy*tilde_gap*r(i)
284
285                       tan_nor_mat = at * matmul(reshape(tangent_cl
                            ,(/2,1/)),reshape(normal_cl,(/1,2/)))
286                       A_II(1:2,1:2) = A_II(1:2,1:2) + EPSY*
                            TAN_NOR_MAT
287                     endif
288                   !END OF "IF FRICTION"
289
```

```
290                 LIN(3*I-2:3*I,NI:NI+2) = LIN(3*I-2:3*I,NI:NI+2) + A_II
291                   end if !END OF "IF CONTACT"
292                 end do   ! END OF THE LOOP ON THE WALLS
293                 NJ(I) = NJ(I) + 3
294
295                 !CHECK CONTACT BETWEEN PARTICLES
296                 do j =  i+1,np
297                    CONTACT = is_contact(q_new(3*i-2:3*i-1),q_old(3*i
                          -2:3*i-1),q_new(3*j-2:3*j-1), &
298                                         q_old(3*j-2:3*j-1),r(i),r(j),
                                             gap,gap_old,normal,tangent
                                             )
299                    !to the rhs
300                    if (CONTACT) then
301                      A_II = 0.D00
302                      A_IJ = 0.D00
303                      A_JI = 0.D00
304                      A_JJ = 0.D00
305
306                      !to have the possibility of contact between each
                            couple (i,j)
307                      noc         = (i-1)*np+j
308                      s(noc)      = gap
309                      s_old(noc) = gap_old
310
311                      if (S(NOC) .GT. -1.D-12) S(NOC) = 0.D0
312                      if (S_OLD(NOC) .GT. -1.D-12) S_OLD(NOC) = 0.D0
313
314                      max_pen_loc = max(max_pen_loc,-gap)
315                      PENET(I) = PENET(I) - GAP
316                      PENET(J) = PENET(J) - GAP
317
318                      if (nl_it .ne. 1) then  !Calculating tilde D_v
319                         som = 0.D00
320                         !EXACT VALUE OF TILDE_GAP
321                         DO i_alpha=0, int(alpha)-1
322                            som = som + s(noc)**(alpha-1-i_alpha)*s_old
                                  (noc)**i_alpha
323                         END DO
324                         tilde_gap = ((-1.D00)**alpha)*som
325
326                         if (DABS(s(noc) - s_old(noc))<1.D-10) then
327                         !IF SMALL, TILDE_GAP AND AT REPLACED BY THE
                               LIMIT
328                            at = alpha*(alpha-1)*((- s_old(noc))**(
                                  alpha-2))
329                         else
330                            at = ( (-alpha)*((-s(noc))**(alpha-1)) -
                                  tilde_gap )/( s(noc)-s_old(noc))
331                         endif
332                      else
333                         tilde_gap = s(noc)
334                         at = 1.0D00
335                      endif
336
```

```
337                     nor_mat = (at/(1.D00+en) + en/(1.D00+en))*matmul(
                           reshape(normal,(/2,1/)),reshape(normal
                           ,(/1,2/)))*c_nu
338
339                     p(3*i-2:3*i-1) = p(3*i-2:3*i-1) + c_nu*(tilde_gap
                           +en*gap)*normal/(1.D00+en)
340                     p(3*j-2:3*j-1) = p(3*j-2:3*j-1) - c_nu*(tilde_gap
                           +en*gap)*normal/(1.D00+en)
341
342                     A_II(1:2,1:2) = A_II(1:2,1:2) + NOR_MAT
343                     A_IJ(1:2,1:2) = A_IJ(1:2,1:2) - NOR_MAT
344                     A_JI(1:2,1:2) = A_JI(1:2,1:2) - NOR_MAT
345                     A_JJ(1:2,1:2) = A_JJ(1:2,1:2) + NOR_MAT
346
347                  !!!!FRICTION CONDITIONS
348                     !v_tan = (qd_new(3*i-2)-qd_new(3*j-2))*tangent(1)
                           +(qd_new(3*i-1)-qd_new(3*j-1))*tangent(2)
349                     V_TAN = DOT_PRODUCT((qd_new(3*i-2:3*I-1)-qd_new
                           (3*j-2:3*J-1)),TANGENT(1:2))
350                     v_tan = v_tan + r(i)*qd_new(3*i) + r(j)*qd_new(3*
                           j)
351                     V_TAN = C_TAU*V_TAN
352
353                     crit = dabs(v_tan) + mu*c_nu*tilde_gap
354                     if (crit <= 0.D-10) then
355                        AFRIC = 0.D00
356                        !stick case
357                        tan_mat = c_tau*matmul(reshape(tangent,(/2,1/)
                              ),reshape(tangent,(/1,2/)))
358                        p(3*i-2:3*i) = p(3*i-2:3*i) + v_tan*(/tangent
                              (1:2),+r(i)/)
359                        p(3*j-2:3*j) = p(3*j-2:3*j) - v_tan*(/tangent
                              (1:2),-r(J)/)
360                        tangent = c_tau*INV2_DT*tangent
361                        tan_mat = INV2_DT*tan_mat
362                        AFRIC(1:2,1:2) = TAN_MAT
363                        AFRIC(1:2,  3) = TANGENT*R(I)
364                        AFRIC(  3,1:2) = TANGENT*R(I)
365                        AFRIC(  3,  3) = INV2_DT*C_TAU*R(I)*R(I)
366                        A_II = A_II + AFRIC
367
368                        AFRIC(1:2,  3) =-TANGENT*R(I)
369                        AFRIC(  3,1:2) =-TANGENT*R(I)
370                        A_JI = A_JI - AFRIC
371
372                        AFRIC(1:2,  3) =-TANGENT*R(J)
373                        AFRIC(  3,1:2) =-TANGENT*R(J)
374                        AFRIC(  3,  3) = INV2_DT*C_TAU*R(J)*R(J)
375                        A_IJ = A_IJ - AFRIC
376
377                        AFRIC(1:2,  3) = TANGENT*R(J)
378                        AFRIC(  3,1:2) = TANGENT*R(J)
379                        A_JJ = A_JJ + AFRIC
380                     else
381                        !slip case
382                        EPSY = DSIGN(mu*c_nu,V_TAN) ! SIGN OF THE SLIP
```

```
383                     !print*, 'particle',v_tan, epsy
384                     p(3*i-2:3*i) = p(3*i-2:3*i) +  epsy*tilde_gap
                           *(/tangent(1:2),+R(I)/)
385                     p(3*j-2:3*j) = p(3*j-2:3*j) -  epsy*tilde_gap
                           *(/tangent(1:2),-R(J)/)
386
387                     tan_nor_mat = at * epsy * matmul(reshape(
                           tangent,(/2,1/)),reshape(normal,(/1,2/)))
388
389                     A_II(1:2,1:2) = A_II(1:2,1:2) + tan_nor_mat
390                     A_IJ(1:2,1:2) = A_IJ(1:2,1:2) - tan_nor_mat
391                     A_JI(1:2,1:2) = A_JI(1:2,1:2) - tan_nor_mat
392                     A_JJ(1:2,1:2) = A_JJ(1:2,1:2) + tan_nor_mat
393                   endif
394                   LIN(3*I-2:3*I,NI:NI+2)       =  LIN(3*I-2:3*I,NI:
                         NI+2) + A_II
395                   LIN(3*I-2:3*I,NJ(I):NJ(I)+2) =  A_IJ
396                   ROW(I,NJ(I):NJ(I)+2)     =  (/3*J-2,3*J-1,3*J/)
397                   NJ(I)             = NJ(I) + 3
398                   LIN(3*J-2:3*J,NJ(J):NJ(J)+2) =  A_JI
399                   ROW(J,NJ(J):NJ(J)+2)     = (/3*I-2,3*I-1,3*I/)
400                   NJ(J)             = NJ(J) + 3
401                   DIAG(3*J-2:3*J,1:3)      = DIAG(3*J-2:3*J,1:3) +
                         A_JJ
402                 end if
403               end do
404               ! ASSEMBLY OF THE SPARSE MATRIX
405               if (nl_it.EQ.NL_IT_MAX) then
406                 print*, 'Non convergence - max nl_it ', NL_IT
407                 go to 122
408                 STOP
409               endif
410
411               DO II = 3*I-2,3*I
412                  DO K = 1 , NJ(I)-1
413                     IF (DABS(LIN(II,K))>0.D00) THEN
414                        MAT_VAL(NNZ) = LIN(II,K)
415                        COL_IND(NNZ) = ROW( I,K)
416                        NNZ = NNZ + 1
417                     END IF
418                  END DO
419                  ROW_PTR(II+1) = NNZ
420               END DO
421            END DO
422
423         !ASSEMBLY OF THE R.H.S.
424            SCM = M * (QD_OLD-QD_NEW)/DT + P + F_FREE
425         !-------------------------------------------------
426            !INITIALISATION OF THE RESOLUTION OF THE LINEAR SYSTEM
427            DQ = DQ_A
428         !-------------------------------------------------------
429            !--- RESOLUTION OF THE LINEAR SYSTEM
430            !!-----------------------------------------------------
431            !! Solve the linear system!!!
432            !!  Matrice : MAT_VAL, COL_IND, NNZ, ROW_PTR
433            !!  R.H.S. : SCM
```

```
434             !!  SOLUTION : DQ
435             !!----------------------
436             !! GMRES SOLVER TO SOLVE AX=B
437             alu=0.D00
438             jlu=0
439             ju=0
440             call ilut (ntdl, NZMAX, MAT_VAL,COL_IND,ROW_PTR, lfil, tol
                    , alu, jlu, ju, &
441                              NZMAX, wu, wl, jr, jwl, jwu, ierr )
442             IF (ierr.NE.0) PRINT*, "FACTORISATION PROBLEM : ",ierr,
                    NL_IT
443             !RESOLUTION-Solve preconditioned GMRES
444             call pgmres (ntdl, NZMAX, im, SCM, DQ, vvv, epss, maxits,
                    iout, &
445                           MAT_VAL,COL_IND,ROW_PTR, alu, jlu, ju, ierr
                              )
446             IF (ierr.NE.0) THEN
447                PRINT*, "RESOLUTION PROBLEM : ",ierr, NL_IT
448                STOP
449             END IF
450             !! END OF THE RESOLUTION
451             !! ------------------------------------------------------
452             q_new = q_new + dq
453             !Update Newmark
454             qd_new    = qd_new + INV2_DT*dq
455
456             ENR  = min(norm2(dq),norm2(dq+dq_a))
457             ENRS = norm2(scm)
458             dq_a = dq
459
460             !Condition to stop the loop
461             if ((NL_IT .GT. 1) .AND. (ENR .LT. NL_EPS))  GO TO 10
462          end do
463          !Update for next iteration
464          10  q_old = q_new
465              qd_old = qd_new
466          !print'("Time :",F10.5,"   Number of NL Iterations", I4,"
                 Fin Iterations, Erreur :"D10.5)', it_time*dt,nl_it,enr
467          !write(unit,*) it_time*dt, q_old, qd_old
468
469          ! TEST THE PENETRATION
470          PENET = 0.D00
471          max_pen_loc = 0.D00
472          DO i=1,np
473             DO J = 1, NB_WALL
474                cont_cl = is_cond_limit(q_new(3*i-2:3*i-1),q_old(3*i
                       -2:3*i-1),r(i),gapcl, &
475                            gapcl_old,normal_cl,tangent_cl,WALLS(J,1:2),
                               WALLS(J,3:4))
476                IF (cont_cl) THEN
477                  nb_contact_wall = nb_contact_wall + 1
478                  max_pen_loc = max(max_pen_loc,-gapcl)
479                  PENET(I) = PENET(I) - GAPCL
480                END IF
481             END DO
482
```

```
483             DO j =  i+1,np
484                CONTACT = is_contact(q_new(3*i-2:3*i-1),q_old(3*i-2:3*i
                      -1),q_new(3*j-2:3*j-1), &
485                   q_old(3*j-2:3*j-1),r(i),r(j),gap,gap_old,normal,
                         tangent)
486                IF (CONTACT) THEN
487                   nb_contact_par = nb_contact_par + 1
488                   max_pen_loc = max(max_pen_loc,-gap)
489                   PENET(I) = PENET(I) - GAP
490                   PENET(J) = PENET(J) - GAP
491                END IF
492             END DO
493          END DO
494          max_pen = max(max_pen,max_pen_loc)
495          max_nlit = max(NL_IT,max_nlit)
496
497          !Kinetic energy
498          EC = 0.D00
499          DO i = 1,np
500             ec(i) = ec(i) + 0.5D00*M(3*i-2)*( qd_old(3*i-2)**2)
501             ec(i) = ec(i) + 0.5D00*M(3*i-1)*( qd_old(3*i-1)**2)
502             ec(i) = ec(i) + 0.5D00*M(3*i  )*( qd_old(3*i  )**2)
503          END DO
504
505          !Data K.E , Error, Penetration
506          final_ec = 0.D00
507          DO I = 1,np
508             final_ec = final_ec + ec(I)
509          END DO
510          !write(160, *) it_time*dt, final_ec, nl_it, enr !, enrs
511
512          if ((IT_TIME/500.D00-FLOOR(IT_TIME/500.D00)).EQ.0) THEN
513             DO I = 1,NP
514                !L2 Norm of the forces
515                FC(I) = DSQRT(P(3*I-2)*P(3*I-2)+P(3*I-1)*P(3*I-1))
516            !Force chain
517                Total_fc = Total_fc+fc(i)
518             END DO
519       !write(333, *) it_time*dt, Total_fc
520             call VISU_2D (IT_TIME,WALLS,Q_OLD,QD_OLD,R,PENET,FC,NP,
                    NB_WALL)
521           end if
522
523    total_nl_it = total_nl_it +nl_it
524     END DO
525
526     call cpu_time(finish)
527
528  122 print*, '
       ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++'
529      print'("Alpha =  ",X,f7.1)', alpha
530      print'("c_nu   =  ",f12.1)',  c_nu
531      print'("Imax   =  ",i8)',     imax
532      print'("mu   =  ",f10.2)',    mu
533      print*, 'CPU time =',         finish-start
534      print*, 'INITIAL KINETIC ENERGY OF SYSTEM:',    init_ec
```

```
535       print*, 'KINETIC ENERGY OF SYSTEM:',            final_ec
536       IF (INIT_EC.GT.1.D-10) FINAL_EC = (init_ec-final_ec)/init_ec
537       print*, 'Relative variation of Kinetic energy:', FINAL_EC
538       print*, 'Relative maximal penetration over time:', max_pen/(2.0*
              max_r)
539       !print*, 'Velocity: ',          qd_old
540       print*, 'Number of contact:',   nb_contact_wall+nb_contact_par
541       print*, 'Total NL Iterations:', total_nl_it
542       print*, 'MAX NL Iterations:',   max_nlit
543       print*, '
              ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++'
544
545       close(unit)
546       close(160)
547     contains
548    !+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
549     LOGICAL FUNCTION is_contact(qi,qiold,qj,qjold,ri,rj,dn,dnINC,en,te)
550       Implicit None
551       integer,parameter           :: rk = selected_real_kind(15,307)
552       real(kind=dp),dimension(2) :: qi,qiold,qj,qjold,en,te,D,dINC
553       real(kind=dp)               :: ri,rj,dn,dnINC,rn
554
555       is_contact = .false.
556       D  = QJ-QI
557       RN = NORM2(D)
558       EN = D/RN
559       TE = -[EN(2) , -EN(1)]
560
561       DINC(1:2) = qjold - qiold
562       DN = DOT_PRODUCT(D,EN) - RI - RJ
563       DNINC = DOT_PRODUCT(DINC,EN) - RI - RJ
564       !dn    = dn - RI - RJ
565       !dnINC = dnINC - RI - RJ
566
567       IS_CONTACT = (DN<1.0D-12)
568     END FUNCTION is_contact
569  !+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
570     LOGICAL FUNCTION is_cond_limit(qi,qiold,ri,dn,dnINC,en,te,pt1,pt2)
571        Implicit None
572        real(kind=dp),dimension(2) :: qi,qiold,en,te,PREF,D,dINC,pt1,
               pt2
573        real(kind=dp)               :: ri,dn,dnINC,rn,alp!,dt,dtINC
574
575        is_cond_limit = .false.
576        TE = pt2 - pt1
577        EN = [-TE(2) , TE(1)]
578        RN = NORM2(EN)
579        EN = EN/RN
580        TE = TE/RN
581
582        D = qiold-pt1
583        ALP  = (EN(2)*D(1) - EN(1)*D(2)) / ((EN(2)*TE(1) - EN(1)*TE(2))
               *RN)
584        PREF = ALP*pt2 + (1.D0 - ALP)*pt1
585
586        D  = qi - PREF
```

```
587         DN = DOT_PRODUCT(D,EN) - RI
588
589         DINC  = qiold - PREF
590         DNINC = DOT_PRODUCT(DINC,EN) - RI
591
592               en = -en
593
594               IS_COND_LIMIT = (DN<-1.0D-12)
595           END FUNCTION is_cond_limit
596       !+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
597          SUBROUTINE VISU_2D (IT_TIME,WALLS,POS_NEW,VEL_NEW,RADIUS,
                VAL_PART,FC,NP,NB_WALL)
598             Implicit None
599             integer                                    :: NP,NB_WALL,IT_TIME
600             real(kind=dp),dimension(:),allocatable:: POS_NEW,VEL_NEW,
                   VAL_PART,FC
601             real(kind=dp),dimension(:),allocatable:: RADIUS
602             real(kind=dp),dimension(:,:),allocatable:: WALLS
603             character(len=8)                           :: fmt !format
                   descriptor
604
605             integer           :: P,I,J,NANGLE,INP
606             integer,parameter :: ivtk=10
607             character(len=25) :: x1,filename
608             character         :: buffer*80, lf*1, str1*8, str2*8
609
610             real(kind=4) THETA,PHI,ANGLE,C,S
611             real(kind=4),dimension(3)                  :: U,BILL
612             real(kind=4),dimension(3,3)                :: M
613             real(kind=4),dimension(:,:,:),allocatable:: BILLE
614
615             fmt = '(I6.6)' !an integer of width 5 with zeros at the
                   left
616             WRITE (x1,fmt) IT_TIME !converting integer to string using
                   a 'internal file'
617             filename='./VISU/visuba'//trim(x1)//'.vtk'
618             open(unit=ivtk,file=filename,access='stream',form='
                   unformatted', &
619                       status="replace",action="write",convert='
                          BIG_ENDIAN')
620
621             lf = char(10)
622
623             NANGLE=21
624             ! Bille elementaire
625             ! print*, 'Bille elementaire'
626             allocate(BILLE(1:NANGLE,1:NANGLE-1,1:3))
627             PHI =-ACOS(-1.00)/2.00
628             do I=1,NANGLE
629                THETA=0.00
630                do J=1,NANGLE-1
631                   BILLE(I,J,1) = COS(THETA)*COS(PHI)
632                   BILLE(I,J,2) = SIN(THETA)*COS(PHI)
633                   BILLE(I,J,3) = SIN(PHI)
634                   THETA=THETA+2.00*ACOS(-1.00)/(NANGLE-1.00)
635                end do
```

```
636                PHI=PHI+ACOS(-1.00)/(NANGLE-1.00)
637             end do
638             !! Ecriture dans un fichier
639             BUFFER = '# vtk DataFile Version 3.0'//lf       ; WRITE(IVTK
                   ) TRIM(BUFFER)
640             BUFFER = 'vtk output'//lf                       ; WRITE(IVTK
                   ) TRIM(BUFFER)
641             BUFFER = 'BINARY'//lf                           ; WRITE(IVTK
                   ) TRIM(BUFFER)
642             BUFFER = 'DATASET UNSTRUCTURED_GRID'//lf        ; WRITE(IVTK
                   ) TRIM(BUFFER)
643             WRITE(STR1(1:8),'(I8)') NP*NANGLE*(NANGLE-1)+4*NB_WALL
644             BUFFER = 'POINTS  '//STR1//'  FLOAT'//LF        ; WRITE(IVTK
                   ) TRIM(BUFFER)
645             do P=1,NP
646                ANGLE=SNGL(POS_NEW(3*P))
647                C=COS(ANGLE)
648                S=SIN(ANGLE)
649                U(1)= 0.00 ! C'est pour le 2D
650                U(2)= 0.00
651                U(3)= 1.00
652                M(1,1)=U(1)*U(1)*(1-C)+C
653                M(1,2)=U(1)*U(2)*(1-C)-U(3)*S
654                M(1,3)=U(1)*U(3)*(1-C)+U(2)*S
655                M(2,1)=U(1)*U(2)*(1-C)+U(3)*S
656                M(2,2)=U(2)*U(2)*(1-C)+C
657                M(2,3)=U(2)*U(3)*(1-C)-U(1)*S
658                M(3,1)=U(1)*U(3)*(1-C)-U(2)*S
659                M(3,2)=U(2)*U(3)*(1-C)+U(1)*S
660                M(3,3)=U(3)*U(3)*(1-C)+C
661
662                do I=1,NANGLE
663                   do J=1,NANGLE-1
664                      BILL(1:3)=SNGL(RADIUS(P))*MATMUL(M,BILLE(I,J,1:3))
665                      WRITE(IVTK) SNGL(POS_NEW(3*P-2:3*P-1))+BILL(1:2),
                            BILL(3)
666                   end do
667                end do
668             end do
669             DO P = 1, NB_WALL
670                WRITE(IVTK) SNGL(WALLS(P,1)), SNGL(WALLS(P,2)), -0.3
671                WRITE(IVTK) SNGL(WALLS(P,3)), SNGL(WALLS(P,4)), -0.3
672                WRITE(IVTK) SNGL(WALLS(P,3)), SNGL(WALLS(P,4)), +0.3
673                WRITE(IVTK) SNGL(WALLS(P,1)), SNGL(WALLS(P,2)), +0.3
674             END DO
675
676             WRITE(STR1(1:8),'(I8)')     NP*(NANGLE-1)*(NANGLE-1)+NB_WALL
                   !NUMBER OF CELLS
677             WRITE(STR2(1:8),'(I8)') 5*(NP*(NANGLE-1)*(NANGLE-1)+NB_WALL
                   )
678             BUFFER = LF//'CELLS'//STR1//STR2//LF       ; WRITE(IVTK)
                   TRIM(BUFFER)
679             do P=1,NP
680                INP=(P-1)*(NANGLE)*(NANGLE-1)
681                do J=0,NANGLE-2
682                   do I=(NANGLE-1)*J,(NANGLE-1)*J+NANGLE-3
```

```
683                     WRITE(IVTK) 4, INP+I,INP+I+1,INP+I+NANGLE-1,INP+I+
                           NANGLE
684                  end do
685                  I=(NANGLE-1)*J+NANGLE-2
686                  WRITE(IVTK) 4,INP+I,INP+I-NANGLE+2,INP+I+NANGLE-1,INP
                        +I+1
687               end do
688            end do
689            INP=(NP)*NANGLE*(NANGLE-1)
690            DO P = 1, NB_WALL
691               WRITE(IVTK) 4,INP,INP+1,INP+3,INP+2
692               INP=INP+4
693            END DO
694
695            WRITE(STR1(1:8),'(I8)')    (NP)*(NANGLE-1)*(NANGLE-1)+
                  NB_WALL
696
697            BUFFER = LF//'CELL_TYPES'//STR1//LF      ; WRITE(IVTK) TRIM(
                  BUFFER)
698
699            do I=1,(NP)*(NANGLE-1)*(NANGLE-1)+NB_WALL
700               WRITE(IVTK) 8
701            end do
702
703            BUFFER = LF//'CELL_DATA  '//STR1//LF     ; WRITE(IVTK) TRIM(
                  BUFFER)
704            BUFFER = 'FIELD FieldData 4'//LF         ; WRITE(IVTK) TRIM(
                  BUFFER)
705         !METTRE LE NOMBRE DE doNNEES A AFFICHER PAR PARTICULE (ICI 3)
706
707            BUFFER = 'VELOCITY 2  '//STR1//' FLOAT'//LF   ; WRITE(IVTK)
                  TRIM(BUFFER)
708            do P=1,NP
709               do J=0,NANGLE-2
710                  do I=(NANGLE-1)*J,(NANGLE-1)*J+NANGLE-2
711                     WRITE(IVTK) SNGL(VEL_NEW(3*P-2)),SNGL(VEL_NEW(3*P
                           -1))
712                  END DO
713               END DO
714            END DO
715            DO P = 1, NB_WALL
716               WRITE(IVTK) 0.,0. !,0.,0.,0.,0.,0.
717            END DO
718
719            BUFFER = LF//'PARTICLE_IDs  1  '//str1//'  INT'//LF  ;
                  WRITE(IVTK) TRIM(BUFFER)
720            do P=1,NP
721               do J=0,NANGLE-2
722                  do I=(NANGLE-1)*J,(NANGLE-1)*J+NANGLE-2
723                     WRITE(IVTK) P
724                  END DO
725               END DO
726            END DO
727            DO P = 1, NB_WALL
728               WRITE(IVTK) NP+P !1,NP+2,NP+3,NP+4
729            END DO
```

```
730
731             BUFFER = LF//'Penetration  1  '//str1//'  FLOAT'//LF  ;
                   WRITE(IVTK) TRIM(BUFFER)
732             DO P=1,NP
733                DO J=0,NANGLE-2
734                   DO I=(NANGLE-1)*J,(NANGLE-1)*J+NANGLE-2
735                      WRITE(IVTK) SNGL(VAL_PART(P))
736                   END DO
737                END DO
738             END DO
739             DO P = 1, NB_WALL
740                WRITE(IVTK) 0. !,0.,0.,0.
741             END DO
742
743             BUFFER = LF//'Contact_Impulsions  1  '//str1//'  FLOAT'//LF
                     ; WRITE(IVTK) TRIM(BUFFER)
744             DO P=1,NP
745                DO J=0,NANGLE-2
746                   DO I=(NANGLE-1)*J,(NANGLE-1)*J+NANGLE-2
747                      WRITE(IVTK) SNGL(FC(P))
748                   END DO
749                END DO
750             END DO
751             DO P = 1, NB_WALL
752                WRITE(IVTK) 0. !,0.,0.,0.
753             END DO
754
755             CLOSE (IVTK)
756
757   END SUBROUTINE VISU_2D
758
759   !+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
760   SUBROUTINE VISU_2D_FAST (IT_TIME,WALLS,POS_NEW,VEL_NEW,RADIUS,
         VAL_PART,FC,NP,NB_WALL)
761   ! VISUALISATION RAPIDE DES RESULTATS AVEC UNIQUEMENT POINTS DE GAUSS
         POUR CHAQUE PARTICULE
762    IMPLICIT NONE
763    integer :: NP,NB_WALL,IT_TIME
764
765    real(kind=dp),dimension(:)  ,allocatable:: POS_NEW,VEL_NEW,VAL_PART
         ,FC
766    real(kind=dp),dimension(:)  ,allocatable:: RADIUS
767    real(kind=dp),dimension(:,:),allocatable:: WALLS
768    CHARACTER(LEN=8) :: FMT ! format descriptor
769
770    integer :: P,INP
771    integer, PARAMETER :: IVTK=10
772    CHARACTER(LEN=22) :: X1,FILENAME
773    CHARACTER :: BUFFER*80, lf*1, STR1*8, STR2*8
774
775    FMT = '(I5.5)' ! an integer of width 5 with zeros at the left
776    WRITE (x1,fmt) IT_TIME ! converting integer to string using a '
         internal file'
777    FILENAME='./VISU/visuba'//trim(x1)//'.vtk'
778    OPEN(UNIT=IVTK,FILE=FILENAME,ACCESS='stream',FORM='unformatted', &
779                 STATUS="replace",ACTION="write",CONVERT='BIG_ENDIAN')
```

```
780
781     LF = CHAR(10) ! line feed character
782
783 !! Ecriture dans un fichier
784     BUFFER = '# vtk DataFile Version 3.0'//lf       ; WRITE(IVTK) TRIM(
            BUFFER)
785     BUFFER = 'vtk output'//lf                       ; WRITE(IVTK) TRIM(
            BUFFER)
786     BUFFER = 'BINARY'//lf                           ; WRITE(IVTK) TRIM(
            BUFFER)
787     BUFFER = 'DATASET UNSTRUCTURED_GRID'//lf        ; WRITE(IVTK) TRIM(
            BUFFER)
788     WRITE(STR1(1:8),'(I8)') NP+4*NB_WALL
789     BUFFER = 'POINTS  '//STR1//'  FLOAT'//LF        ; WRITE(IVTK) TRIM(
            BUFFER)
790     DO P=1,NP
791        WRITE(IVTK) SNGL(POS_NEW(3*P-2:3*P-1)),0.
792     END DO
793     DO P = 1, NB_WALL
794        WRITE(IVTK) SNGL(WALLS(P,1)), SNGL(WALLS(P,2)), -0.3
795        WRITE(IVTK) SNGL(WALLS(P,3)), SNGL(WALLS(P,4)), -0.3
796        WRITE(IVTK) SNGL(WALLS(P,3)), SNGL(WALLS(P,4)), +0.3
797        WRITE(IVTK) SNGL(WALLS(P,1)), SNGL(WALLS(P,2)), +0.3
798     END DO
799
800     WRITE(STR1(1:8),'(I8)') NB_WALL    ! NUMBER OF CELLS FOR WALLS !!
801     WRITE(STR2(1:8),'(I8)') 5*NB_WALL
802
803     BUFFER = LF//'CELLS'//STR1//STR2//LF       ; WRITE(IVTK) TRIM(BUFFER
            )
804
805     INP=NP
806     DO P = 1, NB_WALL
807       WRITE(IVTK) 4,INP,INP+1,INP+3,INP+2
808       INP=INP+4
809     END DO
810
811     WRITE(STR1(1:8),'(I8)') NB_WALL
812
813     BUFFER = LF//'CELL_TYPES'//STR1//LF      ; WRITE(IVTK) TRIM(BUFFER)
814     DO P = 1,NB_WALL
815        WRITE(IVTK) 8
816     END DO
817 !
818 !--- POINTS DATA
819     WRITE(STR1(1:8),'(I8)')  NP+4*NB_WALL    ! NUMBER OF CELLS !!
820 !
821     BUFFER = LF//'POINT_DATA  '//STR1//LF     ; WRITE(IVTK) TRIM(BUFFER)
822     BUFFER = 'SCALARS RADIUS FLOAT'//lf       ; WRITE(IVTK) TRIM(BUFFER)
823     BUFFER = 'LOOKUP_TABLE default'//lf       ; WRITE(IVTK) TRIM(BUFFER)
824     DO P = 1, NP
825        WRITE(IVTK) SNGL(RADIUS(P))
826     END DO
827     DO P = 1, 4*NB_WALL
828        WRITE(IVTK) SNGL(RADIUS(1))*0.1
829     END DO
```

```
!
   BUFFER = 'FIELD PointData 4'//LF          ; WRITE(IVTK) TRIM(BUFFER)
!
   BUFFER = 'VELOCITY 2  '//STR1//' FLOAT'//LF   ; WRITE(IVTK) TRIM(
      BUFFER)
   DO P=1,NP
      WRITE(IVTK) SNGL(VEL_NEW(3*P-2)),SNGL(VEL_NEW(3*P-1))
   END DO
   DO P = 1, 4*NB_WALL
      WRITE(IVTK) 0.,0.
   END DO
!
   BUFFER = LF//'PARTICLE_IDs  1  '//STR1//'  INT'//LF   ; WRITE(IVTK)
      TRIM(BUFFER)
   DO P = 1, NP
      WRITE(IVTK) P
   END DO
   DO P = 1, NB_WALL
      WRITE(IVTK) NP+1,NP+2,NP+3,NP+4
   END DO
!
   BUFFER = LF//'Penetration  1  '//STR1//'  FLOAT'//LF   ; WRITE(IVTK)
       TRIM(BUFFER)
   DO P = 1, NP
            WRITE(IVTK) SNGL(VAL_PART(P))
   END DO
   DO P = 1, 4*NB_WALL
      WRITE(IVTK) 0.
   END DO
!
   BUFFER = LF//'Contact_Impulsions  1  '//str1//'  FLOAT'//LF   ;
      WRITE(IVTK) TRIM(BUFFER)
   DO P = 1, NP
      WRITE(IVTK) SNGL(FC(P))
   END DO
   DO P = 1, 4*NB_WALL
      WRITE(IVTK) 0.
   END DO

   CLOSE (IVTK)

END SUBROUTINE VISU_2D_FAST
!============================================================

  SUBROUTINE command_line(r)
     Implicit None
     real(kind=8)         :: r
     integer              :: i
     character(len=1024)  :: arg
     integer              :: stat
     ! --r
     i = 0
     do
        call get_command_argument(i, arg)
        i = i+1
        if (len_TRIM(arg) == 0) EXIT
```

```
882         select case (arg)
883         case ('--r')
884             call get_command_argument(i, arg) ; i = i+1
885             call str2r8(arg,r,stat)
886             !case ('--n')
887             !call get_command_argument(i, arg) ; i = i+1
888             !call str2int(arg,N(1),stat)
889         case default
890         end select
891       end do
892     END SUBROUTINE command_line
893     !=============================================================
894      SUBROUTINE STR2INT(STR,INT,STAT)
895         Implicit None
896         ! ARGUMENTS
897         character(len=*),intent(in) :: STR
898         integer,intent(out)         :: INT
899         integer,intent(out)         :: STAT
900
901         read(STR,*,IOSTAT=STAT)  INT
902     END SUBROUTINE STR2INT
903     !=============================================================
904     SUBROUTINE STR2R8(STR,R8,STAT)
905         Implicit None
906         !ARGUMENTS
907         character(len=*), intent(in) :: STR
908         real(kind=8), intent(out)    :: R8
909         integer, intent(out)         :: STAT
910
911         read(STR,*,IOSTAT=STAT)  R8
912         END SUBROUTINE STR2R8
913     !=====================================================================
914     SUBROUTINE INITIALIZE_one_xy(q_old,qd_old,r,m,np,ntdl,WALLS)
915         Implicit None
916         integer                                          :: np,ntdl
917         real(kind=dp), dimension(:,:), allocatable :: WALLS
918         real(kind=dp), dimension(:), allocatable   :: M,r,q_old,qd_old,
               q2d_old
919         np = 1           !Number of particles
920         ntdl = 3*np      !Number of degrees of freedom (3*np in 2D)
921         allocate( m(ntdl))
922         allocate( r(np))
923         allocate( q2d_old(ntdl), qd_old(ntdl) , q_old(ntdl))
924
925         print*, "1 PARTICLE"
926         r = [0.2D00]                  !Radius
927         M(1:ntdl) = 0.D00             !Mass matrix
928         do l = 1,3*np
929            m(l) = 0.2D00
930         end do
931
932         q_old  = [1.0D00,0.6D00,0.0D00]   !Position
933         qd_old = [-0.8D00,-0.4D00,0.0D00] !Velocity
934         print*, 'Initial position', q_old
935         print*, 'Initial velocity', qd_old
936
```

```
937         NB_WALL = 4
938         NB_CORNER = 4
939         allocate(WALLS(NB_WALL,4))
940 
941         WALLS(1,1:2) = [ 0.0D00,0.D00]
942         WALLS(1,3:4) = [ 1.2D00,0.D00]
943         WALLS(2,1:2) = WALLS(1,3:4)
944         WALLS(2,3:4) = [ 1.2D00, 1.5D00]
945         WALLS(3,1:2) = WALLS(2,3:4)
946         WALLS(3,3:4) = [0.0D00, 1.5D00]
947         WALLS(4,1:2) = WALLS(3,3:4)
948         WALLS(4,3:4) = WALLS(1,1:2)
949     END SUBROUTINE INITIALIZE_one_xy
950     !=================================================================
951     SUBROUTINE INITIALIZE_one_a(q_old,qd_old,r,m,np,ntdl,WALLS)
952         Implicit None
953         integer                                          :: np,ntdl
954         real(kind=dp), dimension(:,:), allocatable :: WALLS
955         real(kind=dp), dimension(:), allocatable   :: M,r,q_old,qd_old
956         np   = 1              !Number of particles
957         ntdl = 3*np           !Number of degrees of freedom
958 
959         allocate(m(ntdl))
960         allocate(r(np))
961         allocate(qd_old(ntdl),q_old(ntdl))
962 
963         print*, "1 PARTICLE"
964         r = [0.20001]         !Radius of each particle
965         M(1:ntdl) = 0.D00     !Mass matrix
966         do l = 1,3
967            m(l) = 1.0D0
968         end do
969 
970         q_old = 0.D00            !Position
971         !q_old = [1.0,1.0,0.0]
972         q_old = [0.0,0.2,0.0]
973         qd_old = 0.D00           !Velocity
974         !qd_old = [-1.,-1.,0.]
975         qd_old = [1.0,0.,0.]
976 
977         NB_WALL = 4
978         NB_CORNER = 4
979         allocate(WALLS(NB_WALL,4))
980 
981         WALLS(1,1:2) = [ -0.3D00, 0.0D00]
982         WALLS(1,3:4) = [  1.0D00, 0.0D00]
983         WALLS(2,1:2) = WALLS(1,3:4)
984         WALLS(2,3:4) = [ 1.0D00, 0.5D00]
985         WALLS(3,1:2) = WALLS(2,3:4)
986         WALLS(3,3:4) = [ -0.3D00, 0.5D00]
987         WALLS(4,1:2) = WALLS(3,3:4)
988         WALLS(4,3:4) = WALLS(1,1:2)
989     END SUBROUTINE INITIALIZE_one_a
990     !=================================================================
991     SUBROUTINE INITIALIZE_one_b(q_old,qd_old,r,m,np,ntdl,WALLS)
992         Implicit None
```

```
993         integer                                          :: np,ntdl
994         real(kind=dp), dimension(:,:), allocatable :: WALLS
995         real(kind=dp), dimension(:), allocatable   :: M,r,q_old,qd_old
996         !real(kind=dp) :: RAD
997         np = 1            !Number of particles
998         ntdl = 3*np       !Number of degrees of freedom (3*np in 2D)
999         allocate( m(ntdl))
1000        allocate( r(np))
1001        allocate( qd_old(ntdl),q_old(ntdl))
1002        print*, "1 PARTICLE"
1003        r = [0.20001D0]           !Radius
1004        M(1:ntdl) = 0.D00         !Mass matrix
1005        do l = 1,3*np
1006           m(l) = 1.D00
1007        end do
1008
1009        q_old = [0.80D00,1.0D00,0.0D00] !Position
1010        qd_old = [0.D00,-1.D00,0.D00]   !Velocity
1011
1012        NB_WALL = 4                      !BOX
1013        NB_CORNER = 4
1014        allocate(WALLS(NB_WALL,4))
1015
1016        WALLS(1,1:2) = [0.D00,0.D00]
1017        WALLS(1,3:4) = [ 1.0D0,0.D00]
1018        WALLS(2,1:2) = WALLS(1,3:4)
1019        WALLS(2,3:4) = [ 1.0D0, 2.0D00]
1020        WALLS(3,1:2) = WALLS(2,3:4)
1021        WALLS(3,3:4) = [0.D00, 2.0D00]
1022        WALLS(4,1:2) = WALLS(3,3:4)
1023        WALLS(4,3:4) = WALLS(1,1:2)
1024    END SUBROUTINE INITIALIZE_one_b
1025    !=================================================================
1026     SUBROUTINE INITIALIZE_bounce(q_old,qd_old,r,m,np,ntdl,WALLS)
1027         Implicit None
1028         integer                                          :: np,ntdl
1029         real(kind=dp), dimension(:,:), allocatable :: WALLS
1030         real(kind=dp), dimension(:), allocatable   :: M,r, q_old,
                qd_old
1031         np = 1            !Number of particles
1032         ntdl = 3*np       !Number of degrees of freedom (3*np in 2D)
1033         allocate( m(ntdl))
1034         allocate( r(np))
1035         allocate( qd_old(ntdl) , q_old(ntdl))
1036
1037         print*, "1 PARTICLE"
1038
1039         r = [0.2]
1040         M(1:ntdl) = 0.D00
1041         do l = 1,3
1042            m(l) = 1.0
1043         end do
1044
1045         q_old = 0.D00
1046         q_old = [1.0,1.0,0.0]
1047
```

```
1048          qd_old = 0.D00
1049          qd_old = [0.,0.,0.]
1050
1051           NB_WALL = 4
1052           NB_CORNER = 4
1053           allocate(WALLS(NB_WALL,4))
1054           WALLS(1,1:2) = [ 0.D00, 0.0D00]
1055           WALLS(1,3:4) = [ 2.D00, 0.0D00]
1056           WALLS(2,1:2) = WALLS(1,3:4)
1057           WALLS(2,3:4) = [ 2.D00, 3.0D00]
1058           WALLS(3,1:2) = WALLS(2,3:4)
1059           WALLS(3,3:4) = [ 0.D00, 3.0D00]
1060           WALLS(4,1:2) = WALLS(3,3:4)
1061           WALLS(4,3:4) = WALLS(1,1:2)
1062    END SUBROUTINE INITIALIZE_bounce
1063    !=============================================================
1064    SUBROUTINE INITIALIZE_two_a(q_old,qd_old,r,m,np,ntdl,WALLS)
1065        Implicit None
1066        integer                                          :: np,ntdl
1067        real(kind=dp), dimension(:,:), allocatable :: WALLS
1068        real(kind=dp), dimension(:), allocatable   :: M,r, q_old,
              qd_old
1069
1070        np = 2            !Number of particles
1071        ntdl = 3*np       !Number of degrees of freedom (3*np in 2D)
1072
1073        allocate( m(ntdl))
1074        allocate( r(np))
1075        allocate( qd_old(ntdl) , q_old(ntdl))
1076
1077        print*, "2 PARTICLES"
1078        r = [0.1,0.1]             !Radius
1079        M(1:ntdl) = 0.D00         !Mass matrix
1080        do l = 1,3*np
1081           M(l) = 0.1
1082        end do
1083
1084        q_old = 0.D00             !Position
1085        q_old = [0.,0.0,0.0,1.0,1.0,0.0]
1086        qd_old = 0.D00            !Velocity
1087        qd_old = [1.,1.,0.,-1.,-1.,0.]
1088
1089        NB_WALL = 4               !BOX
1090        NB_CORNER = 4
1091        allocate(WALLS(NB_WALL,4))
1092        WALLS(1,1:2) = [-2.D00,-1.0D00]
1093        WALLS(1,3:4) = [ 2.D00,-1.0D00]
1094        WALLS(2,1:2) = WALLS(1,3:4)
1095        WALLS(2,3:4) = [ 2.D00, 1.5D00]
1096        WALLS(3,1:2) = WALLS(2,3:4)
1097        WALLS(3,3:4) = [-2.D00, 1.5D00]
1098        WALLS(4,1:2) = WALLS(3,3:4)
1099        WALLS(4,3:4) = WALLS(1,1:2)
1100    END SUBROUTINE INITIALIZE_two_a
1101   !=============================================================
1102    SUBROUTINE INITIALIZE_two_b(q_old,qd_old,r,m,np,ntdl,WALLS)
```

```
1103        Implicit None
1104        integer                                          :: np,ntdl
1105        real(kind=dp), dimension(:,:), allocatable :: WALLS
1106        real(kind=dp), dimension(:), allocatable   :: M,r, q_old,
               qd_old
1107        !real(kind=dp) :: RAD
1108        np = 2            !Number of particles
1109        ntdl = 3*np       !Number of degrees of freedom (3*np in 2D)
1110        allocate( m(ntdl))
1111        allocate( r(np))
1112        allocate(  qd_old(ntdl) , q_old(ntdl))
1113        print*, "2 PARTICLES"
1114
1115        r = [0.2D00,0.2D00]         !Radius
1116        M(1:ntdl) = 0.D00           !Mass matrix
1117        do l = 1,3*np
1118           m(l) = 1.D00
1119        end do
1120
1121        q_old = [0.D00,0.20D00,0.0D00,0.5D00,0.2D00,0.0D00] !Position
1122        qd_old = [1.D00,0.D00,0.D00,-1.D00,0.D00,0.D00]    !Velocity
1123
1124        NB_WALL = 4                                        !BOX
1125        NB_CORNER = 4
1126        allocate(WALLS(NB_WALL,4))
1127        WALLS(1,1:2) = [-1.D00,0.D00]
1128        WALLS(1,3:4) = [ 2.D00,0.D00]
1129        WALLS(2,1:2) = WALLS(1,3:4)
1130        WALLS(2,3:4) = [ 2.D00, 1.0D00]
1131        WALLS(3,1:2) = WALLS(2,3:4)
1132        WALLS(3,3:4) = [-1.D00, 1.D00]
1133        WALLS(4,1:2) = WALLS(3,3:4)
1134        WALLS(4,3:4) = WALLS(1,1:2)
1135    END SUBROUTINE INITIALIZE_two_b
1136    !=================================================================
1137    SUBROUTINE INITIALIZE_three(q_old,qd_old,r,m,np,ntdl,WALLS)
1138         Implicit None
1139         integer                                          :: np,ntdl
1140         real(kind=dp), dimension(:,:), allocatable :: WALLS
1141         real(kind=dp), dimension(:), allocatable   :: M,r, q_old,
                qd_old
1142         np = 3              !Number of particles
1143         ntdl = 3*np         !Number of degrees of freedom
1144
1145         allocate( M(ntdl))
1146         allocate( r(np))
1147         allocate( qd_old(ntdl) , q_old(ntdl))
1148         print*, "3 PARTICLES"
1149
1150         r = [0.5D00,0.5D00,0.5D00]  !Radius of each particle
1151         M(1:ntdl) = 0.D00           !Mass matrix
1152         do l = 1,3*np
1153            m(l) = 1.0D00
1154         end do
1155
1156         q_old = 0.D00                          !Position
```

```
1157        q_old(1:ntdl:3) = [-0.5,0.5,1.5]
1158        qd_old = 0.D00                          !Velocitiy
1159        qd_old(1:ntdl:3) = [1.,-1.,.5]
1160
1161        NB_WALL = 4                             !Box
1162        NB_CORNER = 4
1163        allocate(WALLS(NB_WALL,4))
1164        WALLS(1,1:2) = [-2.D00,-0.5D00]
1165        WALLS(1,3:4) = [ 2.D00,-0.5D00]
1166        WALLS(2,1:2) = WALLS(1,3:4)
1167        WALLS(2,3:4) = [ 2.D00, 1.5D00]
1168        WALLS(3,1:2) = WALLS(2,3:4)
1169        WALLS(3,3:4) = [-2.D00, 1.5D00]
1170        WALLS(4,1:2) = WALLS(3,3:4)
1171        WALLS(4,3:4) = WALLS(1,1:2)
1172    END SUBROUTINE INITIALIZE_three
1173    !=================================================================
1174    SUBROUTINE INITIALIZE_four(q_old,qd_old,r,m,np,ntdl,WALLS)
1175        Implicit None
1176        integer                                         :: np,ntdl,P
1177        real(kind=dp), dimension(:,:), allocatable :: WALLS
1178        real(kind=dp), dimension(:), allocatable    :: M,r, q_old,
               qd_old
1179
1180        np = 4                  !Number of particles
1181        ntdl = 3*np             !Number of degrees of freedom (3*np in 2D
               )
1182        allocate( m(ntdl))
1183        allocate( r(np))
1184        allocate(  qd_old(ntdl) , q_old(ntdl))
1185
1186        print*, "4 PARTICLES"
1187        r = 0.4D00              !Radius of each particles
1188        M(1:ntdl) = 1.D00       !Mass matrix
1189        !q_old = 0.D00          !particles 1 to 4
1190        P = 1
1191        q_old (3*P-2:3*P) = [ 0.D0, 0.D0,0.D0]
1192        qd_old(3*P-2:3*P) = [ 1.D0, 1.D0,-1.D0]
1193        P = 2
1194        q_old (3*P-2:3*P) = [ 1.D0, 0.D0,0.D0]
1195        qd_old(3*P-2:3*P) = [-1.D0, 1.D0,0.D0]
1196        P = 3
1197        q_old (3*P-2:3*P) = [ 1.D0, 1.D0,0.D0]
1198        qd_old(3*P-2:3*P) = [-1.D0,-1.D0,0.D0]
1199        P = 4
1200        q_old (3*P-2:3*P) = [ 0.D0, 1.D0,0.D0]
1201        qd_old(3*P-2:3*P) = [ 1.D0,-1.D0,0.D0]
1202
1203       NB_WALL = 4
1204       NB_CORNER = 4
1205       allocate(WALLS(NB_WALL,4))
1206       WALLS(1,1:2) = [-1.D00,-1.D00]
1207       WALLS(1,3:4) = [ 2.D00,-1.D00]
1208       WALLS(2,1:2) = WALLS(1,3:4)
1209       WALLS(2,3:4) = [ 2.D00, 2.0D00]
1210       WALLS(3,1:2) = WALLS(2,3:4)
```

```
1211        WALLS(3,3:4) = [-1.D00, 2.0D00]
1212        WALLS(4,1:2) = WALLS(3,3:4)
1213        WALLS(4,3:4) = WALLS(1,1:2)
1214    END SUBROUTINE INITIALIZE_four
1215    !=============================================================
1216    SUBROUTINE INITIALIZE_newton(q_old,qd_old,r,m,np,ntdl,WALLS)
1217        Implicit None
1218        integer                                          :: np,ntdl
1219        real(kind=dp), dimension(:,:), allocatable :: WALLS
1220        real(kind=dp), dimension(:), allocatable   :: M,r,q_old,qd_old
1221        np = 4                !Number of particles
1222        ntdl = 3*np           !Number of degrees of freedom (3*np in 2D)
1223        allocate( m(ntdl))
1224        allocate( r(np))
1225        allocate(  qd_old(ntdl) , q_old(ntdl))
1226        print*, "4 PARTICLES"
1227        r = 0.2D00            !Radius of each particles
1228        M(1:ntdl) = 0.D00     !Mass matrix
1229        do l = 1,3*np
1230           m(l)=2.0D00
1231        end do
1232
1233        q_old = 0.D00         ! particule 1 & 2
1234        q_old(1:ntdl:3) = [0.0,1.0,1.4,1.8]
1235        qd_old = 0.D00        ! vitesse initiale
1236        qd_old(1:ntdl:3) = [1.,0.,0.,0.]
1237
1238        NB_WALL = 4
1239        NB_CORNER = 4
1240        allocate(WALLS(NB_WALL,4))
1241        WALLS(1,1:2) = [-1.0D00,-r]
1242        WALLS(1,3:4) = [ 4.D00,-r]
1243        WALLS(2,1:2) = WALLS(1,3:4)
1244        WALLS(2,3:4) = [ 4.D00, 1.D00]
1245        WALLS(3,1:2) = WALLS(2,3:4)
1246        WALLS(3,3:4) = [-1.D00, 1.D00]
1247        WALLS(4,1:2) = WALLS(3,3:4)
1248        WALLS(4,3:4) = WALLS(1,1:2)
1249    END SUBROUTINE INITIALIZE_newton
1250    !=============================================================
1251    SUBROUTINE INITIALIZE_81_free(q_old,qd_old,r,m,np,ntdl,WALLS)
1252        integer                                        :: np,ntdl,i,j,p,l,k
1253        real(kind=dp),dimension(:,:),allocatable :: WALLS
1254        real(kind=dp),dimension(:),allocatable   :: M,r,q_old,qd_old
1255        real(kind=dp)                                  :: RAD,xc,yc,pi,RHO
1256
1257        k = 9       !construct a matrix k*k for initial particles
1258        np = k*k    !Number of particles
1259        ntdl = 3*np !Number of degrees of freedom (3*np in 2D)
1260
1261        allocate( m(ntdl))
1262        allocate( r(np))
1263        allocate( qd_old(ntdl) , q_old(ntdl))
1264
1265        print*, "SIMULATION:",np," PARTICLES" !T=6s
1266        NB_WALL = 4
```

```
1267        NB_CORNER = 4
1268        allocate(WALLS(NB_WALL,4))
1269        WALLS(1,1:2) = [ 0.D00, 0.D00]
1270        WALLS(1,3:4) = [ 4.D00, 0.D00]
1271        WALLS(2,1:2) = [ 4.D00, 0.D00]
1272        WALLS(2,3:4) = [ 4.D00, 4.50D00]
1273        WALLS(3,1:2) = [ 4.D00, 4.50D00]
1274        WALLS(3,3:4) = [ 0.D00, 4.50D00]
1275        WALLS(4,1:2) = [ 0.D00, 4.50D00]
1276        WALLS(4,3:4) = [ 0.D00, 0.D00]
1277        P = 1
1278        RAD = 0.16D0
1279        XC = 3.6D00
1280        q_old = 0.D0
1281        qd_old = 0.D0
1282        r = 0.16D0
1283        PI  = DACOS(-1.0D0)
1284        rho = 2600.D0
1285        M(1:ntdl) = 0.D00
1286        do l = 1,3*np
1287          !m(l,l)= 4.D00* (RHO*PI*rad**3)/3.D00
1288           m(l) = 1.0D0
1289        end do
1290
1291        do I=1,k
1292           YC = 3.6D00
1293           do J=1,k
1294              q_old(p)   = XC - 0.05D00 + 1.D-01!RAND()
1295              q_old(p+1) = YC - 0.05D00 + 1.D-01!RAND()
1296              q_old(p+2) = 0.0D0
1297              qd_old(p)   = 0.D-00!*RAND()
1298              qd_old(p+1) = 0.D-00 !-1.D00!*RAND()
1299              qd_old(p+2) = 0.0D00
1300              P = P+3
1301              YC = YC - 2.5D00*RAD
1302           end do
1303           XC = XC - 2.5D00*RAD
1304        end do
1305     END SUBROUTINE INITIALIZE_81_free
1306    !=============================================================
1307     SUBROUTINE INITIALIZE_81(q_old,qd_old,r,m,np,ntdl,WALLS)
1308           integer                                    :: np,ntdl,i,j,p,l,k
1309           real(kind=dp),dimension(:,:),allocatable :: WALLS
1310           real(kind=dp),dimension(:),allocatable   :: M,r,q_old,qd_old
1311           real(kind=dp)                              :: RAD,xc,yc,pi,RHO
1312
1313           k = 9       !construct a matrix k*k for initial particles
1314           np = k*k    !Number of particles
1315           ntdl = 3*np !Number of degrees of freedom (3*np in 2D)
1316
1317           allocate( m(ntdl))
1318           allocate( r(np))
1319           allocate( qd_old(ntdl) , q_old(ntdl))
1320           print*, "SIMULATION:",np," PARTICLES"
1321
1322           NB_WALL = 4                   !BOX
```

```
1323          NB_CORNER = 4
1324          allocate(WALLS(NB_WALL,4))
1325          WALLS(1,1:2) = [ 0.D00, 0.D00]
1326          WALLS(1,3:4) = [ 5.3D00, 0.D00]
1327          WALLS(2,1:2) = WALLS(1,3:4)
1328          WALLS(2,3:4) = [ 5.3D00, 6.00D00]
1329          WALLS(3,1:2) = WALLS(2,3:4)
1330          WALLS(3,3:4) = [ 0.D00, 6.00D00]
1331          WALLS(4,1:2) = WALLS(3,3:4)
1332          WALLS(4,3:4) = WALLS(1,1:2)
1333
1334          P = 1
1335          RAD = 0.2D0
1336          XC = 4.6D00
1337          q_old = 0.D0
1338          qd_old = 0.D0
1339          r = 0.2D0
1340          PI  = DACOS(-1.0D0)
1341          rho = 2600.D0
1342          M(1:ntdl) = 0.D00
1343          !Mass matrix
1344          do l = 1,3*np
1345            m(l) = 1.0D0
1346          end do
1347
1348          do I=1,k
1349             YC = 4.32D00
1350             do J=1,k
1351                q_old(p)   = XC - 0.05D00 + 1.D-01*RAND()
1352                q_old(p+1) = YC - 0.05D00 + 1.D-01*RAND()
1353                q_old(p+2) = 0.0D0
1354
1355                qd_old(p)   = (rand(0)*(0.5-(-0.5))+(-0.5))*1.D00
1356                qd_old(p+1) = -0.5D00*RAND()
1357                qd_old(p+2) = 0.0D00
1358                p = p+3
1359                YC = YC - 2.5D00*RAD
1360             end do
1361             XC = XC - 2.5D00*RAD
1362          enddo
1363       END SUBROUTINE INITIALIZE_81
1364      !=============================================================
1365      SUBROUTINE INITIALIZE_1225(q_old,qd_old,r,m,np,ntdl,WALLS)
1366         integer                                    :: np,ntdl,i,j,p,l,k
1367         real(kind=dp),dimension(:,:),allocatable :: WALLS
1368         real(kind=dp),dimension(:),allocatable   :: r,q_old,qd_old,m
1369         real(kind=dp)                              :: RAD,xc,yc,pi,RHO
1370
1371         k = 35      !construct a matrix k*k for initial particles
1372         np = k*k    !Number of particles
1373         ntdl = 3*np !Number of degrees of freedom (3*np in 2D)
1374
1375         allocate( m(ntdl))
1376         allocate( r(np))
1377         allocate( qd_old(ntdl) , q_old(ntdl))
1378
```

```
1379        print*, "SIMULATION:",np," PARTICLES"
1380
1381        NB_WALL = 4                    !BOX
1382        NB_CORNER = 4
1383        allocate(WALLS(NB_WALL,4))
1384        WALLS(1,1:2) = [ 0.D00, 0.D00]
1385        WALLS(1,3:4) = [ 17.D00, 0.D00]
1386        WALLS(2,1:2) = WALLS(1,3:4)
1387        WALLS(2,3:4) = [ 17.D00, 17.D00]
1388        WALLS(3,1:2) = WALLS(2,3:4)
1389        WALLS(3,3:4) = [ 0.D00, 17.D00]
1390        WALLS(4,1:2) = WALLS(3,3:4)
1391        WALLS(4,3:4) = WALLS(1,1:2)
1392
1393        P = 1
1394        RAD = 0.17D0
1395        XC = 16.5D00
1396        q_old = 0.D0
1397        qd_old = 0.D0
1398        PI  = DACOS(-1.0D0)
1399        rho = 2600.D0
1400
1401        do l = 1,np
1402           r(l) = RAD    -  0.015D0*rand(0)
1403           print*, l, r(l)
1404        end do
1405
1406        M(1:ntdl) = 0.D00
1407        do l = 1,3*np
1408          m(l) = 1.0D0
1409        end do
1410
1411        !Initial positions, velocities
1412        do I=1,k
1413           YC = 16.50D00
1414           do J=1,k
1415              q_old(p)   = XC - 0.025D00 + 1.25D-01*RAND()
1416              q_old(p+1) = YC - 0.025D00 + 1.25D-01*RAND()
1417              q_old(p+2) = 0.0D0
1418
1419              qd_old(p)   = 0.0D00
1420              qd_old(p+1) = 0.0D00
1421              qd_old(p+2) = 0.0D00
1422              p = p+3
1423              YC = YC - 2.75D00*RAD
1424           end do
1425           XC = XC - 2.75D00*RAD
1426        enddo
1427     END SUBROUTINE INITIALIZE_1225
1428     !=====================================================
1429     SUBROUTINE INITIALIZE_25(q_old,qd_old,r,m,np,ntdl,WALLS)
1430        integer                                    :: np,ntdl,i,j,p,l,k
1431        real(kind=dp),dimension(:,:),allocatable   :: WALLS
1432        real(kind=dp),dimension(:),allocatable     :: M,r,q_old,qd_old
1433        real(kind=dp)                              :: RAD,xc,yc,pi,RHO
1434        k = 5         !construct a matrix k*k for initial particles
```

```
1435        np = k*k      !Number of particles
1436        ntdl = 3*np   !Number of degrees of freedom (3*np in 2D)
1437
1438        allocate( m(ntdl))
1439        allocate( r(np))
1440        allocate( qd_old(ntdl) , q_old(ntdl))
1441        print*, "SIMULATION:",np," PARTICLES"
1442
1443        NB_WALL = 4                    !BOX
1444        NB_CORNER = 4
1445        allocate(WALLS(NB_WALL,4))
1446        WALLS(1,1:2) = [ 0.D00, 0.D00]
1447        WALLS(1,3:4) = [ 3.5D00, 0.D00]
1448        WALLS(2,1:2) = WALLS(1,3:4)
1449        WALLS(2,3:4) = [ 3.5D00, 4.00D00]
1450        WALLS(3,1:2) = WALLS(2,3:4)
1451        WALLS(3,3:4) = [ 0.0D00, 4.00D00]
1452        WALLS(4,1:2) = WALLS(3,3:4)
1453        WALLS(4,3:4) = WALLS(1,1:2)
1454
1455        P = 1
1456        RAD = 0.25D0
1457        XC = 3.0D00
1458        q_old = 0.D0
1459        qd_old = 0.D0
1460        do l = 1,np
1461           r(l) = RAD  -  0.025D0*rand(0)
1462           print*, l, r(l)
1463        end do
1464
1465        PI  = DACOS(-1.0D0)
1466        rho = 2600.D0
1467        M(1:ntdl) = 0.D00
1468        do l = 1,3*np
1469           m(l) = rad
1470        end do
1471
1472        do I=1,k
1473           YC = 2.90D00
1474           do J=1,k
1475              q_old(p)   = XC + 0.12D00*rand(0)
1476              q_old(p+1) = YC - 0.05D00 + 1.2D-01*RAND(0)
1477              q_old(p+2) = 0.0D0
1478              p = p+3
1479              YC = YC - 2.5D00*RAD
1480           end do
1481           XC = XC - 2.5D00*RAD
1482        enddo
1483        qd_old = 0.0D0
1484     END SUBROUTINE INITIALIZE_25
1485     !=====================================================
1486     SUBROUTINE INITIALIZE_100(q_old,qd_old,r,m,np,ntdl,WALLS)
1487           integer                                    :: np,ntdl,i,j,p,l,k
1488           real(kind=dp),dimension(:,:),allocatable   :: WALLS
1489           real(kind=dp),dimension(:),allocatable     :: M,r,q_old,qd_old
1490           real(kind=dp)                              :: RAD,xc,yc,pi,RHO
```

```
1491
1492        k = 10       !construct a matrix k*k for initial particles
1493        np = k*k     !Number of particles
1494        ntdl = 3*np !Number of degrees of freedom (3*np in 2D)
1495
1496        allocate( m(ntdl))
1497        allocate( r(np))
1498        allocate( qd_old(ntdl) , q_old(ntdl))
1499
1500        print*, "SIMULATION:",np," PARTICLES"
1501        NB_WALL = 4      !BOX
1502        NB_CORNER = 4
1503        allocate(WALLS(NB_WALL,4))
1504        WALLS(1,1:2) = [ 0.D00, 0.D00]
1505        WALLS(1,3:4) = [ 5.3D00, 0.D00]
1506        WALLS(2,1:2) = WALLS(1,3:4)
1507        WALLS(2,3:4) = [ 5.3D00, 5.3D00]
1508        WALLS(3,1:2) = WALLS(2,3:4)
1509        WALLS(3,3:4) = [ 0.D00, 5.3D00]
1510        WALLS(4,1:2) = WALLS(3,3:4)
1511        WALLS(4,3:4) = WALLS(1,1:2)
1512
1513        P = 1
1514        RAD = 0.21D0
1515        XC = 5.00D00
1516        q_old = 0.D0
1517        qd_old = 0.D0
1518        PI  = DACOS(-1.0D0)
1519        rho = 2600.D0
1520
1521        !Radius of each particles
1522        do l = 1,np
1523           r(l) = RAD   -  0.025D0*rand(0)
1524           print*, l, r(l)
1525        end do
1526
1527        M(1:ntdl) = 0.D00
1528        !Mass matrix
1529        do l = 1,3*np
1530          m(l) = rad
1531        end do
1532
1533        do I=1,k
1534           YC = 5.0D00
1535           do J=1,k
1536              q_old(p)   = XC - 0.05D00 + 1.D-01*RAND()
1537              q_old(p+1) = YC - 0.05D00 + 1.D-01*RAND()
1538              q_old(p+2) = 0.0D0
1539
1540              qd_old(p)   = 0.0D00!(rand(0)*(0.5-(-0.5))+(-0.5))*1.
                    D00
1541              qd_old(p+1) = 0.0D00  !-0.1D00*RAND()
1542              qd_old(p+2) = 0.0D00
1543              p = p+3
1544              YC = YC - 2.5D00*RAD
1545           end do
```

```
1546            XC = XC - 2.5D00*RAD
1547         enddo
1548       END SUBROUTINE INITIALIZE_100
1549       !=============================================
1550       SUBROUTINE INITIALIZE_529(q_old,qd_old,r,m,np,ntdl,WALLS)
1551             integer                                    :: np,ntdl,i,j,p,l
                  ,k
1552             real(kind=dp),dimension(:,:),allocatable:: WALLS
1553             real(kind=dp),dimension(:),allocatable  :: M,r,q_old,
                  qd_old
1554             real(kind=dp)                              :: RAD,xc,yc,pi,
                  RHO
1555
1556             k = 23      !construct a matrix k*k for initial particles
1557             np = k*k    !Number of particles
1558             ntdl = 3*np !Number of degrees of freedom (3*np in 2D)
1559
1560             allocate( m(ntdl))
1561             allocate( r(np))
1562             allocate( qd_old(ntdl) , q_old(ntdl))
1563             print*, "SIMULATION:",np," PARTICLES"
1564
1565             NB_WALL = 4                    !BOX
1566             NB_CORNER = 4
1567             allocate(WALLS(NB_WALL,4))
1568             WALLS(1,1:2) = [ 0.D00, 0.D00]
1569             WALLS(1,3:4) = [ 12.D00, 0.D00]
1570             WALLS(2,1:2) = WALLS(1,3:4)
1571             WALLS(2,3:4) = [ 12.D00, 12.D00]
1572             WALLS(3,1:2) = WALLS(2,3:4)
1573             WALLS(3,3:4) = [ 0.D00, 12.D00]
1574             WALLS(4,1:2) = WALLS(3,3:4)
1575             WALLS(4,3:4) = WALLS(1,1:2)
1576
1577             p = 1
1578             rad = 0.2D0
1579             r = 0.2D0
1580             XC = 11.5D00
1581             q_old = 0.D0
1582             qd_old = 0.D0
1583             m(1:ntdl) = 0.D00
1584             do l = 1,np
1585                r(l) = RAD    -  0.025D0*rand(0)
1586             end do
1587             DO l = 1, 3*np
1588                 m(l) = 1.0D0
1589             END DO
1590
1591             DO i = 1, k
1592                YC = 11.5D00
1593                DO j = 1, k
1594                  q_old(p)   = XC - 0.05D00 + 1.D-01*RAND()
1595                  q_old(p+1) = YC - 0.05D00 + 1.D-01*RAND()
1596                  q_old(p+2) = 0.0D0
1597
```

```
1598                    qd_old(p)    = 0.0D00!(rand(0)*(0.5-(-0.5))+(-0.5))*1.
                           D00
1599                    qd_old(p+1) = 0.0D00  !-0.1D00*RAND()
1600                    qd_old(p+2) = 0.0D00
1601                    p = p+3
1602                    YC = YC - 2.5D00*RAD
1603                 end do
1604                 XC = XC - 2.5D00*RAD
1605              enddo
1606              ! print*, q_old
1607              ! stop
1608        END SUBROUTINE initialize_529
1609
1610  !=========================================================
1611     SUBROUTINE NS_CONJUGATE_GRADIENT(MAT_VAL, COL_IND, ROW_PTR,B,X,X_0,
          NDIM,EPS)
1612  ! =========================================================
1613  !    CONJUGATE GRADIENT METHOD TO SOLVE A^T A X = A^B !! Has to be done
        !!!
1614  ! =========================================================
1615    IMPLICIT NONE
1616    integer,PARAMETER :: dp = selected_real_kind(15, 307)
1617
1618    integer :: K,NDIM
1619    real(kind=dp) :: R2,R3,ALPHA,BETA,EPS,EPS1
1620    real(kind=dp), dimension(:) :: B,X,X_0
1621    real(kind=dp), dimension(:), allocatable :: R,P,Q,W
1622    real(kind=dp), dimension(:), allocatable :: MAT_VAL
1623    integer,       dimension(:), allocatable :: COL_IND, ROW_PTR
1624
1625    allocate(R(1:NDIM))
1626    allocate(P(1:NDIM))
1627    allocate(Q(1:NDIM))
1628    allocate(W(1:NDIM))
1629
1630    X(1:NDIM) = X_0(1:NDIM)
1631    K   = 0
1632    R(1:NDIM) = B(1:NDIM) - SP_MATMUL(MAT_VAL, COL_IND, ROW_PTR,X(1:NDIM
          ),NDIM)
1633    R(1:NDIM) = SP_MATMUL_TRANSP(MAT_VAL, COL_IND, ROW_PTR,R(1:NDIM),
          NDIM)
1634
1635    P(1:NDIM)  = R(1:NDIM)
1636    R2 = DOT_PRODUCT(R(1:NDIM),R(1:NDIM))
1637    R3 = R2
1638  !  PRINT*, 'CONJUGATE GRADIENT'
1639  !  print*, 'R2=',R2
1640  !  print*, 'B=',B(1:NDIM)
1641  !  print*, 'X=',X(1:NDIM)
1642  !  PRINT*, 'A=',A
1643  !  print*, 'R=',R(1:NDIM)
1644    EPS1 = EPS * DOT_PRODUCT(B(1:NDIM),B(1:NDIM))
1645  !  print*, 'EPSILON=',EPS1
1646
1647    IF (R2>EPS1) THEN
1648       DO
```

```
1649          K = K+1
1650          Q(1:NDIM) = SP_MATMUL(MAT_VAL, COL_IND, ROW_PTR,P(1:NDIM),NDIM
                 )
1651          W(1:NDIM) = SP_MATMUL_TRANSP(MAT_VAL, COL_IND, ROW_PTR,Q(1:
                 NDIM),NDIM)
1652          ALPHA     = R2/DOT_PRODUCT(Q(1:NDIM),Q(1:NDIM))
1653          X(1:NDIM) = X(1:NDIM) + ALPHA * P(1:NDIM)
1654          R(1:NDIM) = R(1:NDIM) - ALPHA * W(1:NDIM)
1655          R3= DOT_PRODUCT(R(1:NDIM),R(1:NDIM))
1656          IF ((R3 < EPS1).OR.(K>NDIM)) THEN
1657 !             PRINT*, 'CONJUGATE GRADIENT: K=',K,'R3=',R3,'X=',X
1658 !             PRINT*, 'B+',B(1:NDIM)
1659              EXIT
1660          ENDIF
1661          BETA = R3/R2
1662          P(1:NDIM) = R(1:NDIM) + BETA * P(1:NDIM)
1663          R2 = R3
1664       END DO
1665    END IF
1666    IF (R3 > EPS1) THEN
1667 !      WRITE(33,'(D20.13, 3X, D20.13)'),R3,EPS1
1668       PRINT*, "NON CONVERGENCE : ",R3,EPS1
1669 !      PRINT*, 'A= ',A
1670 !      PRINT*, 'B= ',B
1671 !      PRINT*, 'X= ',X
1672 !      STOP
1673    END IF
1674 !  PRINT*, 'CONJUGATE GRADIENT: ITERATION=',K,'EPSILON=',R3!,'ERROR',
       NORM2(MATMUL(A,X)-B)
1675    END SUBROUTINE NS_CONJUGATE_GRADIENT
1676 !=================================================================
1677
1678    FUNCTION SP_MATMUL(MAT_VAL, COL_IND, ROW_PTR,B,NDIM) RESULT(V)
1679 ! ================================================================
1680 !    MULPLICATION MATRIX VECTOR WITH A SPARSE MATRIX (CSR)
1681 ! ================================================================
1682    IMPLICIT NONE
1683    integer,PARAMETER :: dp = selected_real_kind(15, 307)
1684
1685    integer :: I,NNZ,NDIM
1686    real(kind=dp), dimension(:) :: B
1687    real(kind=dp), dimension(:), allocatable :: MAT_VAL,V
1688    integer,       dimension(:), allocatable :: COL_IND, ROW_PTR
1689
1690    allocate (V(NDIM))
1691
1692    DO I = 1, NDIM
1693       V(I) = 0.D00
1694       DO NNZ = ROW_PTR(I), ROW_PTR(I+1)-1
1695          V(I) = V(I) + MAT_VAL(NNZ) * B(COL_IND(NNZ))
1696       END DO
1697    END DO
1698
1699    END FUNCTION
1700
```

```
1701   FUNCTION SP_MATMUL_TRANSP(MAT_VAL, COL_IND, ROW_PTR,B,NDIM) RESULT(V
       )
1702 ! ==========================================================
1703 !   MULPLICATION TRANSPOSE OF A SPARSE MATRIX (CSR)  WITH A VECTOR
1704 ! ==========================================================
1705   IMPLICIT NONE
1706   integer,PARAMETER :: dp = selected_real_kind(15, 307)
1707
1708   integer :: I,NNZ,NDIM
1709   real(kind=dp), dimension(:) :: B
1710   real(kind=dp), dimension(:), allocatable :: MAT_VAL,V
1711   integer,       dimension(:), allocatable :: COL_IND, ROW_PTR
1712
1713   allocate (V(NDIM))
1714
1715   V = 0.D00
1716   DO I = 1, NDIM
1717      DO NNZ = ROW_PTR(I), ROW_PTR(I+1)-1
1718         V(COL_IND(NNZ)) = V(COL_IND(NNZ)) + MAT_VAL(NNZ) * B(I)
1719      END DO
1720   END DO
1721   END FUNCTION
1722 !=============================================================
1723
1724   SUBROUTINE lusol0 ( n, ncc, y, x, alu, jlu, ju )
1725 !! LUSOL0 performs a forward followed by a backward solve
1726 ! for LU matrix as produced by  ILUT
1727 !  Modified:
1728 !    07 January 2004
1729 !  Author:
1730 !    Youcef Saad
1731 !  Parameters:
1732 !    Input, integer ( kind = 4 ) N, the order of the matrix.
1733 !    Input, real Y(N), the right hand side of the linear system.
1734 !    Output, real X(N), the solution.
1735 !    ALU, JLU, JU, ...
1736
1737   implicit none
1738
1739   integer ( kind = 4 ) n
1740   integer ( kind = 4 ) ncc
1741   real ( kind = dp ) alu(ncc)
1742   integer ( kind = 4 ) i
1743   integer ( kind = 4 ) jlu(ncc)
1744   integer ( kind = 4 ) ju(ncc)
1745   integer ( kind = 4 ) k
1746   real ( kind = dp ) x(n)
1747   real ( kind = dp ) y(n)
1748
1749 !  Forward solve
1750   do i = 1, n
1751     x(i) = y(i)
1752     do k = jlu(i), ju(i)-1
1753       x(i) = x(i) - alu(k) * x(jlu(k))
1754     end do
1755   end do
```

```
1756
1757 !  Backward solve.
1758   do i = n, 1, -1
1759     do k = ju(i), jlu(i+1)-1
1760       x(i) = x(i) - alu(k) * x(jlu(k))
1761     end do
1762     x(i) = alu(i) * x(i)
1763   end do
1764
1765   return
1766 end subroutine
1767 !=============================================================
1768  ! ILU
1769  SUBROUTINE ilut ( n, ncc, a, ja, ia, lfil, tol, alu, jlu, ju, iwk, wu
        , wl, jr, &
1770   jwl, jwu, ierr )
1771 ! ILUT is an ILUT preconditioner.
1772 !  Discussion:
1773 !    This routine carries out incomplete LU factorization with dual
1774 !    truncation mechanism.  Sorting is done for both L and U.
1775 !
1776 !  The dual drop-off strategy works as follows:
1777 !    1) Theresholding in L and U as set by TOL.  Any element whose
      size
1778 !       is less than some tolerance (relative to the norm of current
1779 !       row in u) is dropped.
1780 !    2) Keeping only the largest lenl0+lfil elements in L and the
1781 !       largest lenu0+lfil elements in U, where lenl0=initial number
1782 !       of nonzero elements in a given row of lower part of A
1783 !       and lenlu0 is similarly defined.
1784 !
1785 !    Flexibility: one can use tol=0 to get a strategy based on keeping
       the
1786 !    largest elements in each row of L and U. Taking tol /= 0 but lfil
      =n
1787 !    will give the usual threshold strategy (however, fill-in is then
1788 !    unpredictible).
1789 !
1790 !    A must have all nonzero diagonal elements.
1791 !
1792 !  Modified:
1793 !    21 January 2015
1794 !
1795 !  Author:
1796 !    Youcef Saad
1797 !
1798 !  Reference:
1799 !    Youcef Saad,
1800 !    Sparsekit: a basic tool kit for sparse matrix computations,
1801 !    Technical Report, Computer Science Department,
1802 !    University of Minnesota, June 1994
1803 !
1804 !  Parameters:
1805 !    Input, integer ( kind = 4 ) N, the order of the matrix.
1806 !    Input, real ( kind = 8 ) A(*), integer ( kind = 4 ) JA(*), IA(N
      +1),
```

```
1807 !    the matrix in Compressed Sparse Row (CSR) format.
1808 !    Input, integer ( kind = 4 ) LFIL, the fill-in parameter.  Each
     row of L
1809 !    and each row of U will have a maximum of LFIL elements in
     addition to
1810 !    the original number of nonzero elements.  Thus storage can be
1811 !    determined beforehand.
1812 !    0 <= LFIL.
1813 !
1814 !    Input, real ( kind = 8 ) TOL, the tolerance.
1815 !    Output, real ( kind = 8 ) ALU(*), integer ( kind = 4 ) JUL(*),
1816 !    the matrix stored in Modified Sparse Row (MSR) format, containing
1817 !    the L and U factors together. The diagonal (stored in alu(1:n) )
     is
1818 !    inverted. Each I-th row of the ALU, JLU matrix contains the I-th
     row
1819 !    of L (excluding the diagonal entry=1) followed by the I-th row of
      U.
1820 !
1821 !    Output, integer ( kind = 4 ) JU(N), pointers to the beginning of
     each
1822 !    row of U in the matrix ALU, JLU.
1823 !    Input, integer ( kind = 4 ) IWK, the minimum length of arrays ALU
      and
1824 !    JLU.
1825 !
1826 !    Workspace, real ( kind = 8 ) WU(N+1), WL(N).
1827 !
1828 !    Workspace, integer ( kind = 4 ) JR(N), JWL(N), JWU(N).
1829 !
1830 !    Output, integer IERR. Error message with the following meaning.
1831 !    ierr  = 0    --> successful return.
1832 !    ierr > 0  --> zero pivot encountered at step number ierr.
1833 !    ierr  = -1   --> Error. input matrix may be wrong.
1834 !                     (The elimination process has generated a
1835 !                     row in L or U whose length is >  n.)
1836 !    ierr  = -2   --> The matrix L overflows the array al.
1837 !    ierr  = -3   --> The matrix U overflows the array alu.
1838 !    ierr  = -4   --> Illegal value for lfil.
1839 !    ierr  = -5   --> zero pivot encountered.
1840 !
1841   implicit none
1842
1843   integer ( kind = 4 ) n
1844   integer ( kind = 4 ) ncc
1845
1846   real ( kind = dp ) a(ncc)
1847   real ( kind = dp ) alu(ncc)
1848   real ( kind = dp ) fact
1849   integer ( kind = 4 ) ia(n+1)
1850   integer ( kind = 4 ) idiag
1851   integer ( kind = 4 ) ierr
1852   integer ( kind = 4 ) ii
1853   integer ( kind = 4 ) iwk
1854   integer ( kind = 4 ) j
1855   integer ( kind = 4 ) j1
```

```
1856    integer ( kind = 4 ) j2
1857    integer ( kind = 4 ) ja(ncc)
1858    integer ( kind = 4 ) jj
1859    integer ( kind = 4 ) jlu(ncc)
1860    integer ( kind = 4 ) jpos
1861    integer ( kind = 4 ) jr(ncc)
1862    integer ( kind = 4 ) jrow
1863    integer ( kind = 4 ) ju(ncc)
1864    integer ( kind = 4 ) ju0
1865    integer ( kind = 4 ) jwl(n)
1866    integer ( kind = 4 ) jwu(n)
1867    integer ( kind = 4 ) k
1868    integer ( kind = 4 ) len
1869    integer ( kind = 4 ) lenl
1870    integer ( kind = 4 ) lenl0
1871    integer ( kind = 4 ) lenu
1872    integer ( kind = 4 ) lenu0
1873    integer ( kind = 4 ) lfil
1874    integer ( kind = 4 ) nl
1875    real ( kind = dp ) s
1876    real ( kind = dp ) t
1877    real ( kind = dp ) tnorm
1878    real ( kind = dp ) tol
1879    real ( kind = dp ) wl(n)
1880    real ( kind = dp ) wu(n+1)
1881
1882    if ( lfil < 0 ) then
1883      ierr = -4
1884      return
1885    end if
1886  !
1887  !  Initialize JU0 (points to next element to be added to ALU, JLU)
1888  !  and pointer.
1889  !
1890    ju0 = n + 2
1891    jlu(1) = ju0
1892  !
1893  !  integer ( kind = 4 ) double pointer array.
1894  !
1895    jr(1:n) = 0
1896  !
1897  !  The main loop.
1898  !
1899    do ii = 1, n
1900
1901      j1 = ia(ii)
1902      j2 = ia(ii+1) - 1
1903      lenu = 0
1904      lenl = 0
1905
1906      tnorm = 0.0D+00
1907      do k = j1, j2
1908        tnorm = tnorm + abs ( a(k) )
1909      end do
1910      tnorm = tnorm / real ( j2-j1+1, kind = 8 )
1911  !
```

```
!  Unpack L-part and U-part of row of A in arrays WL, WU.
!
     do j = j1, j2

       k = ja(j)
       t = a(j)

       if ( tol * tnorm <= abs ( t ) ) then

         if ( k < ii ) then
           lenl = lenl + 1
           jwl(lenl) = k
           wl(lenl) = t
           jr(k) = lenl
         else
           lenu = lenu+1
           jwu(lenu) = k
           wu(lenu) = t
           jr(k) = lenu
         end if

       end if

     end do

     lenl0 = lenl
     lenu0 = lenu
     jj = 0
     nl = 0
!
!  Eliminate previous rows.
!
150  continue

     jj = jj + 1

     if ( lenl < jj ) then
       go to 160
     end if
!
!  In order to do the elimination in the correct order we need to
!  exchange the current row number with the one that has
!  smallest column number, among JJ, JJ+1, ..., LENL.
!
     jrow = jwl(jj)
     k = jj
!
!  Determine the smallest column index.
!
     do j = jj+1, lenl
        if ( jwl(j) < jrow ) then
           jrow = jwl(j)
           k = j
        end if
     end do
!
```

```
!  Exchange in JWL.
!
     j = jwl(jj)
     jwl(jj) = jrow
     jwl(k) = j
!
!  Exchange in JR.
!
     jr(jrow) = jj
     jr(j) = k
!
!  Exchange in WL.
!
     s = wl(k)
     wl(k) = wl(jj)
     wl(jj) = s

     if ( ii <= jrow ) then
       go to 160
     end if
!
!  Get the multiplier for row to be eliminated: JROW.
!
     fact = wl(jj) * alu(jrow)
     jr(jrow) = 0

     if ( abs ( fact ) * wu(n+2-jrow) <= tol * tnorm ) then
       go to 150
     end if
!
!  Combine current row and row JROW.
!
     do k = ju(jrow), jlu(jrow+1)-1
        s = fact * alu(k)
        j = jlu(k)
        jpos = jr(j)
!
!  If fill-in element and small disregard.
!
        if ( abs ( s ) < tol * tnorm .and. jpos == 0 ) then
          cycle
        end if

        if ( ii <= j ) then
!
!  Dealing with upper part.
!
          if ( jpos == 0 ) then
!
!  This is a fill-in element.
!
            lenu = lenu + 1

            if ( n < lenu ) then
              go to 995
            end if
```

```
2024
2025                jwu(lenu) = j
2026                jr(j) = lenu
2027                wu(lenu) = - s
2028             else
2029 !
2030 !  No fill-in element.
2031 !
2032                wu(jpos) = wu(jpos) - s
2033             end if
2034          else
2035 !
2036 !  Dealing with lower part.
2037 !
2038             if ( jpos == 0 ) then
2039 !
2040 !  This is a fill-in element.
2041 !
2042                lenl = lenl + 1
2043
2044                if ( n < lenl ) then
2045                  go to 995
2046                end if
2047
2048                jwl(lenl) = j
2049                jr(j) = lenl
2050                wl(lenl) = -s
2051             else
2052 !
2053 !  No fill-in element.
2054 !
2055                wl(jpos) = wl(jpos) - s
2056             end if
2057          end if
2058
2059    end do
2060
2061      nl = nl + 1
2062      wl(nl) = fact
2063      jwl(nl) = jrow
2064    go to 150
2065 !
2066 !  Update the L matrix.
2067 !
2068   160 continue
2069
2070      len = min ( nl, lenl0 + lfil )
2071
2072      call bsort2 ( wl, jwl, nl, len )
2073
2074      do k = 1, len
2075
2076         if ( iwk < ju0 ) then
2077           ierr = -2
2078           return
2079         end if
```

```
2080
2081          alu(ju0) =  wl(k)
2082          jlu(ju0) =  jwl(k)
2083          ju0 = ju0 + 1
2084
2085       end do
2086  !
2087  !    Save pointer to beginning of row II of U.
2088  !
2089       ju(ii) = ju0
2090  !
2091  !    Reset double pointer JR to zero (L-part - except first
2092  !    JJ-1 elements which have already been reset).
2093  !
2094     do k = jj, lenl
2095       jr(jwl(k)) = 0
2096     end do
2097  !
2098  !    Be sure that the diagonal element is first in W and JW.
2099  !
2100       idiag = jr(ii)
2101
2102       if ( idiag == 0 ) then
2103         go to 900
2104       end if
2105
2106       if ( idiag /= 1 ) then
2107
2108          s = wu(1)
2109          wu(j) = wu(idiag)
2110          wu(idiag) = s
2111
2112          j = jwu(1)
2113          jwu(1) = jwu(idiag)
2114          jwu(idiag) = j
2115
2116       end if
2117
2118       len = min ( lenu, lenu0 + lfil )
2119
2120       call bsort2 ( wu(2), jwu(2), lenu-1, len )
2121  !
2122  !  Update the U-matrix.
2123  !
2124       t = 0.0D+00
2125
2126       do k = 2, len
2127
2128          if ( iwk < ju0 ) then
2129            ierr = -3
2130            return
2131          end if
2132
2133          jlu(ju0) = jwu(k)
2134          alu(ju0) = wu(k)
2135          t = t + abs ( wu(k) )
```

```
2136          ju0 = ju0 + 1
2137        end do
2138  !
2139  !  Save norm in WU (backwards). Norm is in fact average absolute value
          .
2140  !
2141        wu(n+2-ii) = t / real ( len + 1, kind = 8 )
2142  !
2143  !  Store inverse of diagonal element of U.
2144  !
2145        if ( wu(1) == 0.0D+00 ) then
2146          ierr = -5
2147          return
2148        end if
2149
2150        alu(ii) = 1.0D+00 / wu(1)
2151  !
2152  !  Update pointer to beginning of next row of U.
2153  !
2154    jlu(ii+1) = ju0
2155  !
2156  !  Reset double pointer JR to zero (U-part).
2157  !
2158    do k = 1, lenu
2159      jr(jwu(k)) = 0
2160    end do
2161    end do
2162
2163    ierr = 0
2164
2165    return
2166  !
2167  !  Zero pivot :
2168  !
2169   900    ierr = ii
2170      return
2171  !
2172  !  Incomprehensible error. Matrix must be wrong.
2173  !
2174   995    ierr = -1
2175      return
2176  end subroutine
2177
2178  !=====================================================================
2179  SUBROUTINE bsort2 ( w, ind, n, ncut )
2180  !! BSORT2 returns the NCUT largest elements of an array, using bubble
       sort.
2181  !
2182  !  Discussion:
2183  !    This routine carries out a simple bubble sort for getting the
       NCUT largest
2184  !    elements in modulus, in array W.  IND is sorted accordingly.
2185  !    (Ought to be replaced by a more efficient sort especially
2186  !    if NCUT is not that small).
2187  !
2188  !  Modified:
```

```
!    07 January 2004
!
!  Author:
!    Youcef Saad
!
!  Parameters:
  implicit none
  integer ( kind = 4 ) n
  integer ( kind = 4 ) i
  integer ( kind = 4 ) ind(*)
  integer ( kind = 4 ) iswp
  integer ( kind = 4 ) j
  integer ( kind = 4 ) ncut
  logical test
  real ( kind = 8 ) w(n)
  real ( kind = 8 ) wswp

  i = 1
  do
    test = .false.
    do j = n-1, i, -1
      if ( abs ( w(j) ) < abs ( w(j+1) ) ) then
!
!  Swap.
!
        wswp = w(j)
        w(j) = w(j+1)
        w(j+1) = wswp
!
!  Reorder the original ind array accordingly.
!
        iswp = ind(j)
        ind(j) = ind(j+1)
        ind(j+1) = iswp
!
!  Set indicator that sequence is still unsorted.
!
        test = .true.

      end if
    end do

    i = i + 1

    if ( .not. test .or. ncut < i ) then
      exit
    end if
  end do

  return
end subroutine
!=============================================================
! Preconditioner GMRES
subroutine pgmres ( n, ncc, im, rhs, sol, vv, eps, maxits, iout, &
  aa, ja, ia, alu, jlu, ju, ierr )
!
```

```
2245 !! PGMRES is an ILUT - Preconditioned GMRES solver.
2246 !
2247 !  Discussion:
2248 !    This is a simple version of the ILUT preconditioned GMRES
     algorithm.
2249 !    The ILUT preconditioner uses a dual strategy for dropping
     elements
2250 !    instead  of the usual level of-fill-in approach. See details in
     ILUT
2251 !    subroutine documentation. PGMRES uses the L and U matrices
     generated
2252 !    from the subroutine ILUT to precondition the GMRES algorithm.
2253 !    The preconditioning is applied to the right. The stopping
     criterion
2254 !    utilized is based simply on reducing the residual norm by epsilon
     .
2255 !    This preconditioning is more reliable than ilu0 but requires more
2256 !    storage. It seems to be much less prone to difficulties related
     to
2257 !    strong nonsymmetries in the matrix. We recommend using a nonzero
     tol
2258 !    (tol=.005 or .001 usually give good results) in ILUT. Use a large
2259 !    lfil whenever possible (e.g. lfil = 5 to 10). The higher lfil the
2260 !    more reliable the code is. Efficiency may also be much improved.
2261 !    Note that lfil=n and tol=0.0 in ILUT  will yield the same factors
      as
2262 !    Gaussian elimination without pivoting.
2263 !
2264 !    ILU(0) and MILU(0) are also provided for comparison purposes
2265 !    USAGE: first call ILUT or ILU0 or MILU0 to set up preconditioner
     and
2266 !    then call pgmres.
2267 !
2268 !  Modified:
2269 !    07 January 2004
2270 !
2271 !  Author:
2272 !    Youcef Saad
2273 !
2274 !  Parameters:
2275 !    Input, integer ( kind = 4 ) N, the order of the matrix.
2276 !    Input, integer ( kind = 4 ) IM, the size of the Krylov subspace.
      IM
2277 !    should not exceed 50 in this version.  This restriction can be
     reset by
2278 !    changing the parameter command for KMAX below.
2279 !    Input/output, real RHS(N), on input, the right hand side vector.
2280 !    On output, the information in this vector has been destroyed.
2281 !
2282 ! sol   == real vector of length n containing an initial guess to the
2283 !          solution on input. approximate solution on output
2284 !
2285 ! eps   == tolerance for stopping criterion. process is stopped
2286 !          as soon as ( ||.|| is the euclidean norm):
2287 !          || current residual||/||initial residual|| <= eps
2288 !
```

```
2289 ! maxits== maximum number of iterations allowed
2290 !
2291 ! iout   == output unit number number for printing intermediate results
2292 !           if (iout <= 0) nothing is printed out.
2293 !
2294 !     Input, real AA(*), integer ( kind = 4 ) JA(*), IA(N+1), the
         matrix in CSR
2295 !     Compressed Sparse Row format.
2296 !
2297 !
2298 ! alu,jlu== A matrix stored in Modified Sparse Row format containing
2299 !           the L and U factors, as computed by routine ilut.
2300 !
2301 ! ju     == integer ( kind = 4 ) array of length n containing the
         pointers to
2302 !           the beginning of each row of U in alu, jlu as computed
2303 !           by routine ILUT.
2304 !
2305 ! on return:
2306 !
2307 ! sol    == contains an approximate solution (upon successful return).
2308 ! ierr   == integer ( kind = 4 ). Error message with the following
         meaning.
2309 !           ierr = 0 --> successful return.
2310 !           ierr = 1 --> convergence not achieved in itmax iterations.
2311 !           ierr =-1 --> the initial guess seems to be the exact
2312 !                        solution (initial residual computed was zero)
2313 !
2314 ! work arrays:
2315 !
2316 ! vv     == work array of length  n x (im+1) (used to store the Arnoli
2317 !           basis)
2318 !
2319   implicit none
2320   integer ( kind = 4 ), parameter :: kmax = 50 !1500
2321   integer ( kind = 4 ) n
2322   integer ( kind = 4 ) ncc
2323   !integer ( kind = 4 ) im
2324   real ( kind = 8 ) aa(ncc)
2325   real ( kind = 8 ) alu(ncc)
2326   real ( kind = 8 ) c(kmax)
2327   real ( kind = 8 ) ddot
2328   real ( kind = 8 ) eps
2329   real ( kind = 8 ) eps1
2330   real ( kind = 8 ), parameter :: epsmac = 1.0D-16
2331   real ( kind = 8 ) gam
2332   real ( kind = 8 ) hh(kmax+1,kmax)
2333   integer ( kind = 4 ) i
2334   integer ( kind = 4 ) i1
2335   integer ( kind = 4 ) ia(n+1)
2336   integer ( kind = 4 ) ierr
2337   integer ( kind = 4 ) ii
2338   integer ( kind = 4 ) im
2339   integer ( kind = 4 ) iout
2340   integer ( kind = 4 ) its
2341   integer ( kind = 4 ) j
```

```
2342    integer ( kind = 4 ) ja(ncc)
2343    integer ( kind = 4 ) jj
2344    integer ( kind = 4 ) jlu(ncc)
2345    integer ( kind = 4 ) ju(ncc)
2346    integer ( kind = 4 ) k
2347    integer ( kind = 4 ) k1
2348    integer ( kind = 4 ) maxits
2349    integer ( kind = 4 ) n1
2350    real ( kind = 8 ) rhs(n)
2351    real ( kind = 8 ) ro
2352    real ( kind = 8 ) rs(kmax+1)
2353    real ( kind = 8 ) s(kmax)
2354    real ( kind = 8 ) sol(n)
2355    real ( kind = 8 ) t
2356    real ( kind = 8 ) vv(n,im+1)
2357  !
2358  !  Arnoldi size should not exceed KMAX=50 in this version.
2359  !  To reset modify parameter KMAX accordingly.
2360  !
2361    n1 = n + 1
2362    its = 0
2363  !
2364  !  Outer loop starts here.
2365  !  Compute initial residual vector.
2366  !
2367    call ope ( n, sol, vv, aa, ja, ia )
2368
2369    vv(1:n,1) = rhs(1:n) - vv(1:n,1)
2370
2371    do
2372
2373      ro = sqrt ( ddot ( n, vv, 1, vv, 1 ) )
2374
2375      if ( 0 < iout .and. its == 0 ) then
2376        write(iout, 199) its, ro
2377      end if
2378
2379      if ( ro == 0.0D+00 ) then
2380        ierr = -1
2381        exit
2382      end if
2383
2384      t = 1.0D+00 / ro
2385      vv(1:n,1) = vv(1:n,1) * t
2386
2387      if ( its == 0 ) then
2388        eps1 = eps * ro
2389      end if
2390  !
2391  !  Initialize first term of RHS of Hessenberg system.
2392  !
2393       rs(1) = ro
2394       i = 0
2395
2396   4   continue
2397
```

```
      i = i + 1
      its = its + 1
      i1 = i + 1
      call lusol0 ( n, ncc, vv(1,i), rhs, alu, jlu, ju )
      call ope ( n, rhs, vv(1,i1), aa, ja, ia )
!
!  Modified Gram - Schmidt.
!
      do j = 1, i
        t = ddot ( n, vv(1,j), 1, vv(1,i1), 1 )
        hh(j,i) = t
        call daxpy ( n, -t, vv(1,j), 1, vv(1,i1), 1 )
      end do

      t = sqrt ( ddot ( n, vv(1,i1), 1, vv(1,i1), 1 ) )
      hh(i1,i) = t

      if ( t /= 0.0D+00 ) then
        t = 1.0D+00 / t
        vv(1:n,i1) = vv(1:n,i1) * t
      end if
!
!  Update factorization of HH.
!
     if ( i == 1 ) then
       go to 121
     end if
!
!  Perform previous transformations on I-th column of H.
!
     do k = 2, i
        k1 = k-1
        t = hh(k1,i)
        hh(k1,i) = c(k1) * t + s(k1) * hh(k,i)
        hh(k,i) = -s(k1) * t + c(k1) * hh(k,i)
     end do

121  continue

     gam = sqrt ( hh(i,i)**2 + hh(i1,i)**2 )
!
!  If GAMMA is zero then any small value will do.
!  It will affect only residual estimate.
!
     if ( gam == 0.0D+00 ) then
       gam = epsmac
     end if
!
!  Get the next plane rotation.
!
     c(i) = hh(i,i) / gam
     s(i) = hh(i1,i) / gam
     rs(i1) = -s(i) * rs(i)
     rs(i) = c(i) * rs(i)
!
!  Determine residual norm and test for convergence.
```

```
2454 !
2455       hh(i,i) = c(i) * hh(i,i) + s(i) * hh(i1,i)
2456       ro = abs ( rs(i1) )
2457 !131 format(1h ,2e14.4)
2458 
2459       if ( 0 < iout ) then
2460         write(iout, 199) its, ro
2461       end if
2462 
2463       if ( i < im .and. eps1 < ro ) then
2464         go to 4
2465       end if
2466 !
2467 !  Now compute solution.  First solve upper triangular system.
2468 !
2469       rs(i) = rs(i) / hh(i,i)
2470 
2471       do ii = 2, i
2472         k = i - ii + 1
2473         k1 = k + 1
2474         t = rs(k)
2475         do j = k1, i
2476           t = t - hh(k,j) * rs(j)
2477         end do
2478         rs(k) = t / hh(k,k)
2479       end do
2480 !
2481 !  Form linear combination of V(*,i)'s to get solution.
2482 !
2483       t = rs(1)
2484       rhs(1:n) = vv(1:n,1) * t
2485 
2486       do j = 2, i
2487         t = rs(j)
2488         rhs(1:n) = rhs(1:n) + t * vv(1:n,j)
2489       end do
2490 !
2491 !  call preconditioner.
2492 !
2493       call lusol0 ( n, ncc, rhs, rhs, alu, jlu, ju )
2494 
2495       sol(1:n) = sol(1:n) + rhs(1:n)
2496 !
2497 !  Restart outer loop when necessary.
2498 !
2499       if ( ro <= eps1 ) then
2500         ierr = 0
2501         exit
2502       end if
2503 
2504       if ( maxits < its ) then
2505         ierr = 1
2506         exit
2507       end if
2508 !
2509 !  Else compute residual vector and continue.
```

```
!
    do j = 1, i
      jj = i1 - j + 1
      rs(jj-1) = -s(jj-1) * rs(jj)
      rs(jj) = c(jj-1) * rs(jj)
    end do

    do j = 1, i1
      t = rs(j)
      if ( j == 1 ) then
        t = t - 1.0D+00
      end if
      call daxpy ( n, t, vv(1,j), 1,  vv, 1 )
    end do

  199 format(' its =', i4, ' res. norm =', G14.6)
  end do

  return
end subroutine
!=====================================================================
SUBROUTINE ope ( n, x, y, a, ja, ia )
!! OPE sparse matrix * vector multiplication
!
!  Modified:
!    07 January 2004
!
!  Author:
!    Youcef Saad
!
!  Parameters:
!    Input, integer ( kind = 4 ) N, the order of the matrix.
!
!    Input, real X(N), the vector to be multiplied.
!
!    Output, real Y(N), the product A * X.
!
!    Input, real A(*), integer ( kind = 4 ) JA(*), IA(N+1), the matrix
    in CSR
!    Compressed Sparse Row format.
!
  implicit none
  integer ( kind = 4 ) n
  real ( kind = 8 ) a(*)
  integer ( kind = 4 ) i
  integer ( kind = 4 ) ia(n+1)
  integer ( kind = 4 ) ja(*)
  integer ( kind = 4 ) k
  integer ( kind = 4 ) k1
  integer ( kind = 4 ) k2
  real ( kind = 8 ) x(n)
  real ( kind = 8 ) y(n)

  do i = 1, n
    k1 = ia(i)
    k2 = ia(i+1) -1
```

```
2565      y(i) = 0.0D+00
2566      do k = k1, k2
2567        y(i) = y(i) + a(k) * x(ja(k))
2568      end do
2569    end do
2570
2571    return
2572   end subroutine
2573  END PROGRAM INC
```

Listing A.1: Newmark-PDAS-INC-Sparse+Gmres

# Bibliography


[1] Vo Anh Thuong Nguyen, Stéphane Abide, Mikaël Barboteu, Serge Dumont, *An Improved Normal Compliance Method for Non-Smooth Contact Dynamics*, Banach Center Publications 127 (2024), 191-217. (DOI: 10.4064/bc127-9)

[2] Stéphane Abide, Mikaël Barboteu, Serge Dumont, Vo Anh Thuong Nguyen, *Enhanced Moreau-Yosida Regularization Method Compared to Non-Smooth Contact Dynamics: Nonlinear Gauss-Seidel Approach for Granular Media*, in preparation.

[3] Florent Nacry, Vo Anh Thuong Nguyen, Lionel Thibault. *Farthest distance function to strongly convex sets. Journal of Convex Analysis*, In press, 30 (4), pp.1217-1240. (hal-04193961)

[4] Florent Nacry, Vo Anh Thuong Nguyen, Juliette Venel. *Metric subregularity and $\omega(\cdot)$-normal regularity properties*. 2023. (hal-04418792)

[5] P. A. Cundall, *The measurement and analysis of accelerations in rock slopes* (1971).

[6] P. A. Cundall, *A computer model for simulating progressive large scale movements of blocky rock systems*, In Proceedings of the symposium of the international society of rock mechanics, volume 1, pages 132–150 (1971).

[7] P. A. Cundall and O. D. L. Strack, *A discrete numerical model for granular assemblies*, geotechnique, 29(1): 47-65 (1979).

[8] M. Schatzman, *A class of nonlinear differential equations of second order in time*, Nonlinear Anal. Theory Methods Appl. 2(3), 355–373 (1978).

[9] A. Jourani and E. Vilches. *Moreau-Yosida regularization of state-dependent sweeping processes with nonregular sets*, Journal of Optimization Theory and Applications, 173 (2017).

[10] F. Dubois, V. Acary, and M. Jean, *The contact dynamics method: A nonsmooth story*, Comptes Rendus Mécanique, 346(3): 247-262 (2018).

[11] M. Kunze, M. Monteiro Marques, *An introduction to Moreau's sweeping process*, Impacts in mechanical systems, Springer, Berlin, Heidelberg (2000).

[12] M. Kunze, M. Monteiro Marques, *On parabolic quasi-variational inequalities and state-dependent sweeping processes*, Topol. Methods Nonlinear Anal. 12(1), 179–191 (1998)

[13] T. Haddad , T. Haddad, *State-dependent sweeping process with perturbation*, In: Anastassiou, G.A., Duman, O. (eds.) Advances in Applied Mathematics and Approximation Theory, Springer Proc. Math. Stat., vol. 41, pp. 273–281. Springer, New York (2013)

[14] M. Bounkhel, C. Castaing, *State dependent sweeping process in p-uniformly smooth and q-uniformly convex Banach spaces*, Set-Valued Var. Anal. 20(2), 187–201 (2012).

[15] N. Chemetov, M. Monteiro-Marques, *Non-convex quasi-variational differential inclusions*, Set-Valued Anal. 15(3), 209–221 (2007).

[16] N. Chemetov, M. Monteiro-Marques, U. Stefanelli, *Ordered non-convex quasi-variational sweeping processes*, J. Convex Anal. 15(2), 201–214 (2008)

[17] C. Castaing, A.G. Ibrahim, M. Yarou, *Some contributions to nonconvex sweeping process*, J. Nonlinear Convex Anal. 10(1), 1–20 (2009).

[18] D. Azzam-Laouir, S. Izza, L. Thibault, *Mixed semicontinuous perturbation of nonconvex statedependent sweeping process*, Set-Valued Var. Anal. 22(1), 271–283 (2014).

[19] T. Haddad, I. Kecis, L. Thibault, *Reduction of state dependent sweeping process to unconstrained differential inclusion*, J. Global Optim. 62(1), 167–182 (2015).

[20] J. Noel, *Inclusions différentielles d'évolution associées à des ensembles sous lisses*, Ph.D. thesis, Université Montpellier II (2013).

[21] J. Noel, L. Thibault, *Nonconvex sweeping process with a moving set depending on the state*, Vietnam J. Math. 42(4), 595–612 (2014).

[22] V. Acary, B. Brogliato, D. Goeleven, *Higher order Moreau's sweeping process: mathematical formulation and numerical simulation*, Math. Program. 113, 133–217 (2008).

[23] V. Acary, B. Brogliato, *Numerical Methods for Nonsmooth Dynamical Systems: Applications in Mechanics and Electronics*, (2008).

[24] F. Dubois, V. Acary, M. Jean, *The Contact Dynamics method: A nonsmooth story*. Comptes Rendus Mécanique, 346, 247-262 (2017).

[25] V. Acary, *Time-Integration methods for nonsmooth contact dynamics with friction and impact* (2020).

[26] H. Hilber, T. Hughes and R. Taylor, *Improved numerical dissipation for time integration algorithms in structural dynamics*, Earth. Engrg. Struc. Dyna. 5 (1977) 283-292.

[27] R. Dzonou, M. Monteiro-Marques, L. Paoli, *Algorithme de type "sweeping process" pour un problème de vibro-impact avec un opérateur d'inertie non trivial*, C.R. Acad. Sci. Paris, Série IIb - Mécanique, Vol. 335, pp 56-60 (2007).

[28] B. Brogliato, A.A. ten Dam, L. Paoli, F. Génot, M. Abadie, *Numerical simulation of finite dimensional multibody nonsmooth mechanical systems*, ASME Applied Mechanics Reviews, Vol. 55, no. 2, pp 107-149 (2002).

[29] J. Venel, *A numerical scheme for a class of sweeping processes*, Numer. Math. 118, 367–400 (2011).

[30] B. Maury, J. Venel, *A discrete contact model for crowd motion.* ESAIM Math. Model. Numer. Anal. 45 (2011), 145-168.

[31] T. Kawai, *Some considerations on the finite element method.* Int. J. Numer. Meth. Engrg., 16: 81–120 (1980).

[32] Y. Kishino, *Disk model analysis of granular media.* Michromecanics of Granular Materials, pages 143-152 (1988). (eds. M. Satake and J.T. Jenkins), Elsevier.

[33] Y. Kishino, *Granular flow simulation by granular element method.* In Slow Dynamics in Complex Systems, November (2003). Sendai, Japan.

[34] M. Renouf, *Optimisation numérique et calcul parallèle pour l'étude des milieux divisés bi et tridimensionnels*, Thèse de doctorat dirigée par Alart, Pierre Mécanique. Génie mécanique. Génie civil Montpellier 2 (2004).

[35] B.J. Alder, T.E. Wainwright, *Studies in Molecular Dynamics. II. Behavior of a Small Number of Elastic Spheres*, Journal of Chemical Physics 33 (1960): 1439-1451.

[36] Ilya Prigogine, *Proceedings of the International Symposium on Transport Processes in Statistical Mechanics : held in Brussels*, August 27-31,1956. (1958).

[37] Roger W. Alder, Paul R. Allen, Drahomír Hnyk, David W. H. Rankin, Heather E. Robertson, Bruce A. Smart, Ronald J. Gillespie, Ian Bytheway, *Molecular Structure of 3,3-Diethylpentane (Tetraethylmethane) in the Gas Phase as Determined by Electron Diffraction and ab initio Calculations*, Journal of Organic Chemistry 64 (1999): 4226-4232.

[38] M. Jean, J.J. Moreau, *Unilaterality and dry friction in the dynamics of rigid body collections*, In: Curnier A. (Ed.) Contact Mechanics International Symposium. Presses Polytechn. et Universit. Romandes, Lausanne, 31-48 (1992).

[39] J.J. Moreau, *Décomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires*, Comptes Rendus de l'Académie des Sciences de Paris Série A, vol. 255, pp. 238–240 (1962).

[40] J.J. Moreau, *Sur la fonction polaire d'une fonction semi-continue supérieurement*, Comptes Rendus de l'Académie des Sciences de Paris Série A, vol. 258, pp. 1128–1130 (1964).

[41] J.J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bulletin de la S.M.F., tome 93 (1965), p. 273-299.

[42] J.J. Moreau, *Rafle par un convexe variable. I, In Travaux du Seminaire d'Analyse Convexe*, Montpellier (1971).

[43] J.J. Moreau, *Application of convex analysis to some problems of dry friction*, In Trends in applications of pure mathematics to mechanics, pages 263–280. Pitman (1977).

[44] J.J. Moreau, *Unilateral contact and dry friction in finite freedom dynamics*, In: Moreau, J.J., Panagiotopoulos, P.D. (eds.) Nonsmooth Mechanics and Applications, pp. 1–82. CISM, Courses and Lectures, vol. 302. Springer, Berlin Heidelberg New York (1988).

[45] J.J. Moreau, *Unilateral Contact and Dry Friction in Finite Freedom Dynamics* (1988).

[46] J.J. Moreau, *Some numerical methods in multibody dynamics: application to granular materials*, European J. of Mech. A/Solids, 4, (1994).

[47] J.J. Moreau, *Numerical aspects of the sweeping process*, Computer Methods in Applied Mechanics and Engineering, 177, 329-349 (1999).

[48] M. Jean, *The nonsmooth contact dynamics method*, Computer Methods in Applied Mechanics and Engineering, 177: 235-257 (1999). Special issue on computational modeling of contact and friction, J.A.C. Martins and A. Klarbring, editors.

[49] M. Renouf, F. Dubois, P. Alart, *A parallel version of the non smooth contact dynamics algorithm applied to the simulation of granular media*, Journal of Computational and Applied Mathematics, 168, 375-382 (2004).

[50] M. Renouf, P. Alart, *Conjugate gradient type algorithms for frictional multicontact problems: applications to granular materials*, Computer Methods in Applied Mechanics and Engineering, 194, 2019-2041 (2005).

[51] G.D. Saxcé, Z. Feng, *New Inequality and Functional for Contact with Friction: The Implicit Standard Material Approach*, Mechanics of Structures and Machines, 19, 301-325 (1991).

[52] J. Fortin, O. Millet, G.D. Saxcé, *Numerical simulation of granular materials by an improved discrete element method*, International Journal for Numerical Methods in Engineering, 62 (2005).

[53] S. Dumont, *On enhanced descent algorithms for solving frictional multicontact problems: application to the discrete element method*, International Journal for Numerical Methods in Engineering, 93 (2013).

[54] P. Joli, Z. Feng, *Uzawa and Newton algorithms to solve frictional contact problems within the bipotential framework*, International Journal for Numerical Methods in Engineering, 73 (2008).

[55] M. Hintermüller, K. Ito, K. Kunisch, *The Primal-Dual Active Set Strategy as a Semismooth Newton Method*, SIAM J. Optim., 13, 865-888 (2002).

[56] S. Hüeber, B.I. Wohlmuth, *A primal-dual active set strategy for non-linear multibody contact problems*, Computer Methods in Applied Mechanics and Engineering, 194, 3147-3166 (2005).

[57] P. Hauret and P. Le Tallec, *Energy-controlling time integration methods for A elastodynamicsand low-velocity impact*, Computer Methods in Applied Mechanics and Engineering, 195(37-40): 4890–4916 (2006).

[58] S. Hüeber, G. Stadler, B. I. Wohlmuth, *A primal-dual active set algorithm for three-dimensional contact problems with coulomb friction*, SIAM J.Scientific Computing 30 (2008).

[59] M. Barboteu, S. Dumont, *A primal-dual active set method for solving multi-rigid-body dynamic contact problems*, Mathematics and Mechanics of Solids, 23, 489-503 (2018).

[60] S. Abide, M. Barboteu, S. Cherkaoui, and S. Dumont, *A semi-smooth newton and primal-dual active set method for non-smooth contact dynamics*. Comput Methods Appl Mech Eng, 387 (2021).

[61] S. Abide, M. Barboteu, S. Cherkaoui, S. Dumont, *Unified primal-dual active set method for dynamic frictional contact problems*, Fixed Point Theory and Algorithms for Sciences and Engineering, 1-22 (2022).

[62] M. Hintermüller, Kovtunenko, K. Kunisch, *Semismooth Newton methods for a class of unilaterally constrained variational problems* (2004).

[63] K. Kunisch, G. Stadler, *Generalized Newton methods for the 2D-Signorini contact problem with friction in function space*, ESAIM Math. Model. Numer. Anal., Vol 39 (2005).

[64] N. Newmark, *A method of computation for structural dynamics*, Journal of Engineering Mechanics, Div. ASCE, 67-94 (1959).

[65] Haim Brezis, *Operateurs maximaux monotones: et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland mathematics studies, Elsevier Science, Burlington, MA (1973).

[66] Bernard Brogliato, *Impacts in mechanical systems*, Springer Berlin Heidel-berg (2000).

[67] Bernard Cambou, Michel Jean, *Micro mecanique des materiaux granulaires* (2001).

[68] M. Monteiro Marques, *Differential inclusions in nonsmooth mechanical problems*, Birkhauser Basel (1993).

[69] J.J. Moreau, *Evolution problem associated with a moving convex set in a hilbert space*, Journal of Differential Equations 26 (1977), no. 3, 347–374.

[70] E. Youngs, D. Hillel, *Fundamentals of soil physics*, The Journal of Applied Ecology 19 (1982), 307.

[71] I. Vardoulakis, *Mechanics of discrete granular media: With applications to granular media*, pp. 151–179, 01 (2019).

[72] D. Huet, M. Jalaal, R. Beek, D. Meer, and A. Wachs, *Granular avalanches of entangled rigid particles*, Physical Review Fluids 6 (2021).

[73] S. Nineb, P. Alart, D. Dureisseix, *Domain decomposition ap-proach for non-smooth discrete problems, example of a tensegrity structure*, vol. 85,05 (2007).

[74] A. Taboada, N. Estrada, *Rock-and-soil avalanches: Theory and simulation*, Journal of Geophysical Research 114 (2009).

[75] J. Noel, F. Chevoir, *Simulation numerique discrete et comportement mecanique des materiaux granulaires* (2009).

[76] G. Saussine, C. Cholet, P.E. Gautier, F. Dubois, C. Bohatier,J. Moreau, *Modelling ballast under cyclic loading using discrete element method*, pp. 649–658, 07 (2004).

[77] S. Hüeber, G. Stadler, B. I. Wohlmuth, *A primal-dual active set algorithm for three-dimensional contact problems with coulomb friction*, SIAM J.Scientific Computing 30 (2008).

[78] M. Renouf, Y. Berthier, *Modélisation numérique de la génération et transmission de chaleur*, Hermes. Modélisation numérique discrète des matériaux granulaires, Lavoisier, Chap. 13 (2010).

[79] M. Renouf, Y. Berthier, *Numerical Modeling of Heat Production and Transmission*, In F. Radjai and F. Dubois, editors, Discrete numerical modeling of granular materials, page 448. Wiley (2011).

[80] A. Lefebvre, *Numerical simulation of gluey particles*, M2AN, 43:53-80 (2009).

[81] H. Martin, M. Peruzzetto, S. Viroulet, A. Mangeney, P.-Y. Lagrée, S. Popinet, B. Maury, A. Lefebvre, Y. Maday, F. Bouchut, *Numerical simulations of granular dam break : comparison between discrete element, Navier-Stokes and thin-layer models.* Physical Review E, 108.5 (2023), pp.054902-054927

[82] N.E. Gibbs, W.G. Poole, P.K. Stockmeyer, *An algorithm for reducing the Bandwidth and profile of a sparse matrix.* SIAM Journal on Numerical Analysis, 13(2), 236–250 (1976).

[83] S. Adly, F. Nacry, L. Thibault, *Prox-regular sets and Legendre-Fenchel transform related to separation properties.* Optimization 71 (2022), 2097-2129.

[84] S. Adly, F. Nacry, L. Thibault, *New metric properties for prox-regular sets.* Math. Program. 189 (2021), Ser. B, 7-36.

[85] G. Allaire, C. Dapogny, P. Frey, *Shape optimization with a level set based mesh evolution method.* Comput. Methods Appl. Mech. Engrg. 282 (2014), 22–53.

[86] L. Ambrosio, *Geometric evolution problems, distance function and viscosity solutions.* Calculus of Variations and Partial Differential Equations, 5–93, Springer, Berlin (2000).

[87] J.P. Aubin, A. Cellina, *Differential inclusions. Set-Valued Maps and Viability Theory.* Grundlehren der mathematischen Wissenschaften, 264. Springer-Verlag, Berlin, 1984.

[88] M.V. Balashov, *Weak convexity of the distance function.* J. Convex Anal, 20 (2013), 93-106.

[89] M.V. Balashov, M.O. Golubev, *Weak concavity of the antidistance function.* J. Convex Anal. 21 (2014), 951-964.

[90] M.V. Balashov, G.E. Ivanov, *On farthest points of sets.* Mathematical Notes, 80 (2006), 159-166.

[91] M.V. Balashov, G.E. Ivanov, *Weakly convex and proximally smooth sets in Banach spaces.* Izv. Math. 73 (2009), 455–499.

[92] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert spaces.* CMS Books in Mathematics, Springer, New York (2011).

[93] F. Bernard, L. Thibault, N. Zlateva, *Prox-regular sets and epigraphs in uniformly convex Banach spaces: various regularities and other properties.* Trans. Amer. Math. Soc. 363 (2011), 2211-2247.

[94] A. Canino, *On p-convex sets and geodesics.* J. Differential Equations 75 (1988), 118-157.

[95] P. Cannarsa, C. Sinestrari, *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control.* Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser Boston, Inc., Boston, MA, 2004.

[96] C. Castaing, L. Thibault, *Various perturbations and relaxations of sweeping process.* Submitted.

[97] F.H. Clarke, *Optimization and Nonsmooth Analysis.* Canadian Mathematical Society Series of Monographs and Advanced Texts. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1983.

[98] F.H. Clarke, Y.S. Ledyaev, R.J. Stern, P.R. Wolenski, *Nonsmooth Analysis and Control Theory.* Graduate Texts in Mathematics, 178. Springer-Verlag, New York, 1998.

[99] F.H. Clarke, R.J. Stern, P.R. Wolenski, *Proximal smoothness and the lower-$C^2$ property.* J. Convex Anal. (1995), 117-144.

[100] G. Colombo, L. Thibault, *Prox-regular sets and applications.* Handbook of nonconvex analysis and applications, 99-182 (2010), Int. Press, Somerville, MA.

[101] M.C. Delfour, J.P. Zolésio, *Shape analysis via oriented distance functions.* J. Funct. Anal. 123 (1994), 129–201.

[102] F. Deutsch, *Best Approximation in Inner Product Spaces.* CMS Books in Mathematics, 7, Springer-Verlag, New-York, 2001.

[103] H. Federer, *Curvature measures.* Trans. Amer. Math. Soc. 93 (1959), 418-491.

[104] H. Frankowska, C. Olech, *R-convexity of the integral of the set-valued functions.* Contributions to Analysis and Geometry, John Hopkins Univ. Press, Baltimore, Md., 1981, pp. 117–129.

[105] V.V. Goncharov, G.E. Ivanov, *Strong and weak convexity of closed sets in a Hilbert space.* Operations research, engineering, and cyber security, 259-297, Springer Optim. Appl., 113, Springer, Cham, 2017.

[106] D.R. Luke, *Finding best approximation pairs relative to a convex and prox-regular set in a Hilbert space.* SIAM J. Optim. 19 (2008), 714-739.

[107] A.D. Ioffe, *Variational Analysis of Regular Mappings. Theory and Applications.* Springer Monographs in Mathematics. Springer, Cham, 2017.

[108] G.E. Ivanov, *Weakly Convex Sets and Functions: Theory and Applications.* Fizmatlit, Moscow (2006) (in Russian).

[109] G.E. Ivanov, *Farthest points and strong convexity of sets.* Math. Notes 87, 355-366 (2010).

[110] G.E. Ivanov, E. S. Polovinkin, *On strongly convex differential games.* Differential Equations 31 (1995), 1603-1612.

[111] E.S. Levitin, B.T. Polyak, *Constrained minimization methods.* U.S.S.R. Computational Math. and Math. Phys. 6 (1966), 787-823.

[112] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation. I. Basic Theory*, Grundlehren der mathematischen Wissenschaften, 330, Springer-Verlag, Berlin 2006.

[113] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation. II. Applications.* Grundlehren der mathematischen Wissenschaften, 331, Springer-Verlag, Berlin, 2006.

[114] F. Nacry, L. Thibault, *Regularization of sweeping process: old and new.* Pure Appl. Funct. Anal. 4 (2019), 59-117.

[115] F. Nacry, L. Thibault, *Distance function associated to a prox-regular set.* Set-Valued Var. Anal 30 (2022), 731-750.

[116] J.P. Penot, *Calculus Without Derivatives.* Graduate Texts in Mathematics, 266. Springer, New-York, 2013.

[117] B.T. Polyak, *Existence theorems and convergence of minimizing sequences in extremum problems with restrictions.* Soviet Math. Dokl. 7 (1966), 72-75.

[118] R.A. Poliquin, R.T. Rockafellar, L. Thibault, *Local differentiability of distance functions.* Trans. Amer. Math. Soc. 352 (2000), 5231-5249.

[119] E.S. Polovinkin, M.V. Balashov, *Elements of Convex and Strongly Convex Analysis.* Fizmatlit, Moscow 2004. (Russian).

[120] R.T. Rockafellar, R. J-B. Wets, *Variational Analysis.* Grundlehren der mathematischen Wissenschaften, 317. Springer, Berlin, 1998.

[121] A. Shapiro, *Existence and differentiability of metric projections in Hilbert spaces.* SIAM J. Optimization 4 (1994), 130-141.

[122] I. Singer, *Duality for Nonconvex Approximation and Optimization.* CMS Books in Mathematics, 24. Springer, New York, 2006.

[123] L. Thibault, *Sweeping process with regular and nonregular sets.* J. Differential Equations 193 (2003), 1–26.

[124] L. Thibault, *Unilateral Variational Analysis in Banach Spaces.* World Scientific Publishing Co Pte Ltd, to appear (2022).

[125] J.P. Vial, *Strong and weak convexity of sets and functions.* Math. Oper. Res. 8 (1983), 231-259.

[126] A.Weber, G. Reißig, *Local characterization of strongly convex sets.* J. Math. Anal. Appl. 400 (2013), 743–750.

[127] S. Adly, F. Nacry, L. Thibault, *Preservation of prox-regularity of sets with applications to constrained optimization*, SIAM J. Optim. 26 (2016), 448-473.

[128] S. Adly, F. Nacry, L. Thibault, *Prox-regularity approach to generalized equations and image projection*, ESAIM Control Optim. Calc. Var. 24 (2018), 677-708.

[129] D. Aussel, A. Daniilidis, L. Thibault *Subsmooth sets: functional characterizations and related concepts*, Trans. Amer. Math. Soc. 357 (2005), 1275-1301.

[130] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Classics Appl. Math., 5 Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.

[131] A.L. Dontchev, R.T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer Ser. Oper. Res. Financ. Eng. Springer, New York (2014).

[132] L.M. Graves, *Some mapping theorems*, Duke Math. J. 17 (1950), 111-114.

[133] G.E. Ivanov, *Nonlinear images of sets. I: Strong and weak convexity.*, J. Convex Anal. 27 (2020), 363–382.

[134] A.D. Ioffe, *Metric regularity-a survey Part 1. Theory*, J. Aust. Math. Soc. 101 (2016), 188-243.

[135] A.D. Ioffe, *Metric regularity-a survey Part 2. Applications*, J. Aust. Math. Soc. 101 (2016), 376-414.

[136] A. Jourani, *Open mapping theorem and inversion theorem for $\gamma$-paraconvex multivalued mappings and applications*, Studia Mathematica, 117 (1996), 123-136.

[137] A. Jourani, E. Vilches, *Positively $\alpha$-far sets and existence results for generalized perturbed sweeping processes*, J. Convex Anal. 23 (2016), 775-821.

[138] H. Huang, R.X. Li, *Global error bounds for $\gamma$-paraconvex multifunctions*, Set-Valued Var. Anal. 19 (2011), 487-504.

[139] L.A. Lyusternik, *On conditional extrema of functionals*, Mat. Sb. 41 (1934), 390-401.

[140] X.Y. Zheng, K.F. Ng, *Metric subregularity and calmness for nonconvex generalized equations in Banach spaces*, SIAM J. Optim. 20 (2010), 2119-2136.

[141] S. Rolewicz, *On paraconvex multifunctions*, Operations Res. Verfahren 31 (1979), 539-546.

[142] S.M. Robinson, *Regularity and stability for convex multivalued functions*, Math. Oper. Res., 1 (1976), 130-143.

[143] L. Thibault, *Unilateral Variational Analysis in Banach Spaces. Part II, Special Classes of Functions and Sets*, World Scientific (2023).

[144] C. Ursescu, *Multifunctions with closed convex graphs*, Czechoslovak Math. J. 25 (1975), 438-411.

[145] X.Y. Zheng, K.F. Ng, *Metric subregularity for proximal generalized equations in Hilbert spaces*, Nonlinear Anal. 75 (2012), 1686-1699.

[146] X.Y. Zheng, Q.H. He, *Characterization for metric regularity for $\sigma$-subsmooth multifunctions*, Nonlinear Anal. 100 (2014), 111-121.

[147] F. Bernicot, A. Lefebvre-Lepot, *Existence results for non-smooth second order differential inclusions, Convergence result for a numerical scheme and applications for modelling inelastic collisions* (2010).

[148] J.A. Meijerink, H.A. Van Der Vorst, *An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrices*, Math. Comp., 31(137):148–162, 1977.

[149] Jocelyne Erhel, Nabil Nassif, and Bernard. Philippe, *Calcul matriciel et systèmes linéaires*, cours. 2004.

[150] Youcef Saad, *Sparsekit: a basic tool kit for sparse matrix computations*, Technical Report, Computer Science Department, University of Minnesota, June 1994.

[151] Guoming Hu, Zhenyu Hu, Bin Jian, Liping Liu, and Hui Wan, *On the determination of the damping coefficient of non-linear spring-dashpot system to model hertz contact for simulation by discrete element method*, In 2010 WASE International Conference on Information Engineering, volume 3, pages 295–298. IEEE (2010).

[152] J. Venel, *A numerical scheme for a class of sweeping processes*, Numer. Math. 118 (2011), 367-400.

[153] V. V. Goncharov, G. E. Ivanov, Strong and weak convexity of closed sets in a Hilbert space, Operations research, engineering, and cyber security, 259-297, Springer Optim. Appl., 113, Springer, Cham (2017).

[154] J. P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser, Boston (1990).

[155] M. Bounkhel, *Regularity concepts in nonsmooth analysis. Theory and applications*. Springer Optimization and Its Applications, 59. Springer,New York (2012).

[156] B. Jessen, *Two theorems on convex point sets*, Mat. Tidsskr. B., pages 66-70 (1940).

[157] S. Fitzpatrick, *Metric projections and the differentiability of distance functions*, Bull. Austral. Math. Soc., 22(2): 291-312 (1980).

[158] T. S. Motzkin, E. G. Straus, and F. A. Valentine, *The number of farthest points*, Pacific J. Math., 3:221–232 (1953).

[159] A. Jourani, L. Thibault, D. Zagrodny, $C^{1,\omega(\cdot)}$*-regularity and Lipschitz-like properties of subdifferential*, Proc. Lond. Math. Soc. (3) 105(2012), 189-223.

[160] A. S. Lewis, D. R. Luke, J. Malick, *Local linear convergence of alternating and averaged projections*, Found. Comput. Math., 9 (2009), 485-513.

[161] A. Daniilidis, D.R. Luke, M. Tam, *Characterizations of super-regularity and its variants*, Splitting algorithms, modern operator theory, and applications, 137-152, Springer, Cham (2019).

[162] B.S. Mordukhovich, N.M. Nam, *Subgradient of distance functions with applications to Lipschitzian stability*, Math. Program. 104 (2005), 635-668.

[163] N.M. Nam, B.S. Mordukhovich, *Subgradients of distance functions at out-of-set points*, Taiwanese J. Math. 10 (2006), 299-326.

[164] W. D. Collins, *Dual extremum principles and Hilbert space decompositions*, in: Duality and Complementarity in Mechanics of Solids, pp. 351-418. Ossolineum, Wroc law (1979).

[165] S.P. Han, O.L. Mangasarian, *Conjugate cone characterization of positive definite and semidefinite matrices*, Linear Algebra and Its Applications, vol. 56, 89-103 (1984).

[166] R. T. Rockafellar, *Moreau's proximal mappings and convexity in Hamilton-Jacobi theory*, in: Nonsmooth Mechanics and Analysis, pp. 3-12. SpringerVerlag, New York (2006).

[167] P. L. Combettes and J.-C. Pesquet, *Proximal thresholding algorithm for minimization over orthonormal bases*, SIAM Journal on Optimization, vol. 18, pp. 1351–1376 (2007).

[168] P. L. Combettes, Dinh Dung, B. C. Vu, *Dualization of signal recovery problems, Set-Valued and Variational Analysis*, vol. 18, pp. 373–404 (2010).

[169] S. Marcellin, L. Thibault, *Evolution Problems Associated with Primal Lower Nice Functions*, Journal of Convex Analysis, 13:385–421 (2006).

# Abstract

Based on the growing importance of understanding granular media and their behaviors in various industries, this work proposes a new method different from the Discrete Element Method (DEM) and the Non-Smooth Contact Dynamics (NSCD) approach to model granular dynamics. We focus on a discontinuous Moreau second-order sweeping process for modeling contact dynamics, incorporating the Moreau-Yosida regularization with parameter $\alpha$ to develop a regular contact model. We propose the **Improved Normal Compliance** (INC) method to ensure energy conservation and employ a combination of the Newmark method and **Primal-Dual Active Set** (PDAS) to address nonlinearity. This study aims to compare the efficiency of our approach with other numerical modeling techniques, such as DEM and NSCD via Nonlinear Gauss Seidel method (NLGS), focusing on improving energy conservation and computational cost. Furthermore, related to the Moreau sweeping process, to establish the existence of a solution, it is quite natural to extend beyond convexity with prox-regularity or the dual class associated with the prox-regular set, which is a strongly convex set. Additionally, metric regularity serves as a useful tool for studying the weak convexity of specific subsets. More precisely, we will focus on two main ideas. First, we consider the concepts of **Farthest distance function to strongly convex sets in Hilbert spaces**. On one hand, we show that the strong convexity of a set is equivalent to the semiconcavity of its associated farthest distance function. On the other hand, we establish that the farthest distance of a point to a strongly convex set is the minimum farthest distance to the given point from suitable closed balls separating the set and the point. Second, we concentrate on **Metric subregularity and $\omega(\cdot)$-normal regularity properties**. We establish through an openness condition the metric subregularity of a multimapping with normal $\omega(\cdot)$-regularity of either the graph or values. Various preservation results for prox-regular and subsmooth sets are also provided.



**Keywords**: Unilateral constraint, Normal Compliance, Friction, Newmark methods, Non Smooth Contact Dynamics, Semi-Smooth Newton method, Primal-Dual Active Set, Sweeping process; Variational analysis, strong convexity, prox-regularity, farthest distance function, semiconvexity, normal regularity, subsmoothness, metric regularity, metric subregularity.

# Resume

Basé sur l'importance croissante de comprendre les médias granulaires et leurs comportements dans diverses industries, ce travail propose une nouvelle méthode différente de la Méthode des Éléments Discrets (Discrete Element Method-DEM) et de l'approche de la Dynamique de Contact Non Lisse (Non-Smooth Contact Dynamics-NSCD) pour modéliser la dynamique granulaire. Nous nous concentrons sur un processus de balayage discontinu de Moreau de second ordre pour modéliser la dynamique de contact, en incorporant la régularisation de Moreau-Yosida avec le paramètre $\alpha$ pour développer un modèle de contact régulier. Nous proposons la méthode **Improved Normal Compliance** (INC) pour assurer la conservation de l'énergie et employons une combinaison de la méthode de Newmark et de **Primal-Dual Active Set** (PDAS) pour traiter la non-linéarité. Cette étude vise à comparer l'efficacité de notre approche avec d'autres techniques de modélisation numérique, telles que DEM et NSCD via la méthode de Gauss-Seidel Non Linéaire (Nonlinear Gauss Seidel method-NLGS), en se concentrant sur l'amélioration de la conservation de l'énergie et du coût computationnel. En outre, en relation avec le processus de balayage de Moreau, pour établir l'existence d'une solution, il est assez naturel d'aller au-delà de la convexité avec la prox-régularité ou la classe duale associée à l'ensemble prox-régulier qui est un ensemble fortement convexe. De plus, la régularité métrique sert d'outil utile pour étudier la faible convexité de sous-ensembles spécifiques. Plus précisément, nous nous concentrerons sur deux idées principales. Premièrement, nous considérons les concepts de **Fonction de distance la plus éloignée aux ensembles fortement convexes dans les espaces de Hilbert**. D'une part, nous montrons que la forte convexité d'un ensemble est équivalente à la semi-concavité de sa fonction de distance la plus éloignée associée. D'autre part, nous établissons que la distance la plus éloignée d'un point à un ensemble fortement convexe est la distance la plus éloignée minimale au point donné à partir de boules fermées appropriées séparant l'ensemble et le point. Deuxièmement, nous nous concentrons sur **la sous-régularité métrique et les propriétés de régularité $\omega(\cdot)$-normales**. Nous établissons par une condition d'ouverture la sous-régularité métrique d'une multi-application avec la régularité $\omega(\cdot)$-normale soit du graphe, soit des valeurs. Divers résultats de préservation pour les ensembles prox-réguliers et sous-lisses sont également fournis.

**Mots-clés**: Contrainte unilatérale, Normal Compliance, Friction, Méthodes de Newmark, Dynamique de Contact Non Lisse, Méthode de Newton Semi-Lisse, Primal-Dual Active Set, Processus de balayage; Analyse variationnelle, forte convexité, prox-régularité, fonction de distance la plus éloignée, semi-convexité, régularité normale, sous-lissité, régularité métrique, sous-régularité métrique.