Titre: Title:	Contact Detection for Pairs of Ellipses and Ellipsoids: Analysis, Comparisons, and Improvements
Auteur: Author:	Elham Kheradmand Nezhad
Date:	2020
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Kheradmand Nezhad, E. (2020). Contact Detection for Pairs of Ellipses and Ellipsoids: Analysis, Comparisons, and Improvements [Ph.D. thesis, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/5347/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/5347/
Directeurs de recherche: Advisors:	Serge Prudhomme, & Marc Laforest
Programme: Program:	Doctorat en mathématiques

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Contact Detection for Pairs of Ellipses and Ellipsoids: Analysis, Comparisons, and Improvements

ELHAM KHERADMAND NEZHAD
Département de mathématiques et de génie industriel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Mathématiques

Août 2020

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Contact Detection for Pairs of Ellipses and Ellipsoids: Analysis, Comparisons, and Improvements

présentée par **Elham KHERADMAND NEZHAD**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Ahmad SHAKIBAEINIA, président
Serge PRUDHOMME, membre et directeur de recherche
Marc LAFOREST, membre et codirecteur de recherche
David VIDAL, membre
Varvara ROUBTSOVA, membre
Richard BATHURST, membre externe

DEDICATION

To my parents . . .

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor Prof. Serge Prudhomme and my co-supervisor Prof. Marc Laforest for their incredible support and patience. Their door was always open to discuss and advice throughout all stages of my research studies. Their friendly guidance and extensive knowledge were definitely important factors in my achievements and success.

I am grateful to Dr. Varvara Roubtsova, Mohamed Chekired and Dr. Paul Labbé from Institut de recherche d'Hydro-Québec for their invaluable assistance and support during my research.

I thank all the professors, students, and staffs of the Department of Mathematical and Industrial Engineering for providing such a scientific, dynamic and friendly atmosphere.

My special thanks are extended to my friends who went through hard times together, cheered me on, and celebrated each accomplishment.

In addition, I am truly and deeply indebted to Masoud, the love of my life, whose care and support were always on my side.

I owe my lovely family for every single accomplishment in my life. They are the most important reason behind my success because of their unconditional support and love. They mean the world to me.

RÉSUMÉ

Ce travail de recherche offre un cadre théorique pour analyser, comparer et améliorer les algorithmes de détection de contact entre des paires d'ellipses et d'ellipsoïdes. On se concentre surtout sur la catégorie d'algorithmes qui sont les plus efficaces numériquement, qui peuvent produire des estimations de la distance de séparation et de pénétration entre des ellipses et des ellipsoïdes, et qui peuvent définir un point de contact et une direction normale pour calculer les forces, comme nécessaires dans les simulations par éléments discrets. Plus précisément, seules les représentations analytiques des ellipses et ellipsoïdes sont étudiées et la détection de contact entre des ellipsoïdes en mouvement n'est pas traitée ici. La première contribution est d'offrir un cadre mathématique pour étudier ces algorithmes, plus particulèrement les preuves d'existence et d'unicité de solutions pour certaines classes d'algorithmes de détection de contact, pour décrire rigoureusement des paires d'ellipses en contact presque parfait, avec ou sans chevauchement, et pour analyser globalement les contraintes sur les vecteurs normaux. Ce cadre met en valeur le rôle clé joué par les différentes définitions de contact trouvées dans la littérature, indépendamment des stratégies de calcul utilisées pour calculer les distance de séparation ou de pénétration. Plus précisément, on montre que tous les algorithmes étudiés peuvent être exprimés comme des problèmes de minimisation, avec ou en l'absence de contraintes non saturées sur les vecteurs normaux aux points de contact, et que des contraintes additionnelles peuvent être utilisées pour identifier le minimum global parmi les points critiques du problème de minimisation. Une autre contribution de cette recherche, fondée sur le cadre mathématique proposé, est une meilleure classification des algorithmes existants. Ces algorithmes sont comparés sur des cas test et leurs forces et faiblesses sont mises en évidence et expliquées par rapport à cette classification. L'utilité de cette analyse mathématique est illustrée par la présentation d'un algorithme performant combinant de nouvelles idées et d'autres plus anciennes. Cette algorithme appartient à la classe des méthodes de potentiel géométrique, lesquelles considèrent la solution de deux problèmes de minimisation pour déterminer un point de contact entre les particules. L'efficacité de l'algorithme repose sur plusieurs ingrédients, à savoir une transformation qui associe à une paire d'ellipses (ellipsoïdes) une ellipse (ellipsoïde) centrée à l'origine et un cercle (une sphère) unitaire, la construction d'un point initial efficace pour la résolution du problème de minimisation non linéaire, l'utilisation de la méthode de Newton pour le problème de recherche de racine, et l'imposition d'une contrainte supplémentaire pour garantir la convergence vers la racine recherchée. Les résultats de plusieurs exemples numériques montrent que le nouvel algorithme de détection de contact est plusieurs fois plus rapides que les algorithmes existants

à précision comparable. On présente aussi un nouvel algorithme pour générer aléatoirement des paires d'ellipses ou d'ellipsoïdes permettant de comparer la performance et la précision des algorithmes de détection de contact sur de grands volumes de données.

ABSTRACT

This research provides a theoretical framework to analyze, compare, and improve contact detection algorithms for pairs of ellipses and ellipsoids. We focus primarily on the category of algorithms that are the most computationally efficient and can produce estimates of the separation and penetration distance between ellipses and ellipsoids, and can define a contact point and a normal direction to compute forces, as are necessary in Discrete Element Simulations. Specifically, only analytic representations of the ellipses and ellipsoids are considered and contact detection for moving pairs of ellipsoids is not treated. The first contribution is a mathematical framework for the study of these algorithms, most notably with existence and uniqueness proofs for classes of contact detection algorithms, formal descriptions of pairs of ellipses in near-perfect contact, with or without overlap, and a global analysis of constraints on the normals. The framework highlights the key role played by the different definitions of contact found in the literature, independent of the numerical strategies deployed to estimate the separation/penetration distance. Specifically, it is shown that all the studied algorithms can be expressed as minimization problems, with or without non-binding constraints on the normal(s) at the contact point(s), and that constraints can be used to identify the global minima among the critical points in the minimization problem. Another contribution of this research, based on the mathematical framework introduced, is a better classification of the known algorithms. These algorithms are compared on established test problems and their strengths and weaknesses are highlighted and explained in terms of their classification. The usefulness of the new framework is illustrated with the introduction of a very fast algorithm combining some new and old ideas. The algorithm belongs to the class of geometrical potential methods, which consider the solution of two minimization problems in order to determine a contact point between the particles. The efficiency of the algorithm relies on several ingredients, namely, a transformation that maps the pair of ellipses (ellipsoids) into an ellipse (ellipsoid) centered at the origin and a unit circle (sphere), the construction of an effective initial guess for the solution of the nonlinear minimization problem, the use of Newton's method for the root finding problem, and the introduction of an additional constraint to guarantee convergence to the desired root. The results from several numerical examples show that the new contact detection algorithm is several times faster than the existing algorithms for comparable accuracy. A novel algorithm to randomly generate pairs of ellipses or ellipsoids is also described and allows one to compare the performance and accuracy of contact detection algorithms on large data sets.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

1.1 Scientific Context

Human beings are increasingly expanding cities, farms, and infrastructures over the earth due to high demands for energy, food, and shelter. Human-made structures, such as buildings, waterways, and dams are built on various types of soil. Thus, soil stability is essential to the continued growth and prosperity of modern civilization. Moreover, this is particularly true in this era of climate change since a significant portion of the world population live at or near sea level. Water-saturated soil can weaken the integrity of the soil, which is a potential threat for buildings and engineering structures. Interactions between soil and water can initiate a variety of destructive phenomena, such as internal erosion, backward erosion, heave, dispersion, suffusion and liquefaction.

Soil found near the ground surface is often the result of erosion and therefore formed of individual particles with a large void ratio, possibly saturated with water. Under compression and shear, such soil does not behave like homogeneous materials (constant soil properties along the soil profile). Although engineers require predictions at the large scale, the behavior of such soil is intimately related to the particles composing it, and to the dynamics of force transmission at the contacts between particles. In other words, the statistics of the geometry of the particles composing the soil are key to its macroscale behavior.

The aforementioned phenomena are yet to be fully understood and possibly controlled. By modeling and simulating soil-water interactions, we hope to be able to understand their underlying mechanisms in order to predict it and implement procedures to control its damage. The simulation of such phenomena is thus of major interest in different fields of science, such as geology, mechanical engineering and civil engineering.

Models of soil mechanics are generally divided into two categories: continuum models and discrete models. Continuum models are essentially layer-scale models in which the soil layer is considered as a continuum. By contrast, discrete models consider the granular material as an assemblage of discrete particles. Continuum models require phenomenological models of the stress-strain relation that are unable to capture the discrete physics between particles [1, 2]. Unlike continuum models, discrete models can capture the non-linear behavior of granular materials. In addition, they require fewer model parameters than continuum models, which require a large number of phenomenological parameters that need to be evaluated empirically such as initial shear strength, initial tangent shear modulus, permeability, initial tangent

coefficient of volume compressibility, and relative density.

The Discrete Element Method (DEM) is the most common method to solve problems described with a discrete model of particles. Molecular Dynamics (MD) [3] is another discrete method which is used mostly for gases, solids, or crystals. The DEM is a straightforward numerical method that models each particle and their contacts individually, using no more than Newtonian mechanics and Hookian or Hertzian contact models to compute the force between the particles. It was first proposed in 1971 by Cundall and Strack [4,5] to analyze rock mechanics problems, at a time when computer power was extremely limited. The method considers the small displacement and rotation of particles and identifies new contact between particles after each displacement. Particles are assumed rigid but are allowed to overlap in order to account for their deformation.

In practice, DEM can be used to predict transitions from static to hydraulic regimes, often occurring along shear planes. Such phase transitions in granular material is fundamental to the efficient manipulation, transportation, and mixing in the chemical industry, and in particular the large scale food manufacturing. Moreover, it is used to evaluate the mechanical properties of composite materials [6], to study a cement matrix in concrete technology [7], or to model powder based structures [8].

Spheres are the most common particle model for soil simulation due to the simplicity of their collision detection. However, spheres are different from real soil particles, in terms of shape, centers of mass, and hence, the manner by which one rotates around the other. Some studies demonstrate that the particle shape plays a key role in the simulated macroscopic properties of static and dynamic assemblies of particles [9, 10]. More complex geometries have been proposed, such as ellipses [11], ellipsoids [12–14], superquadrics [15,16], and polygons [17,18]. Another approach for particle representation is based on clusters of overlapping [19] or non-overlapping [20] spheres. However, using such representations instead of discs or spheres makes it more difficult to find contact points.

For soil simulation, the most straightforward improvement over spherical particles is using ellipsoidal ones. The main advantage of ellipsoids is that unlike spherical particles, a normal contact force induces a moment to the ellipsoidal particles. This affects the rotation and resistance of particles and gives a better representation of the overall kinetics. This also makes the rotational degrees of freedom easier to excite and enables stacks of aligned flat ellipsoids to be more stable, thereby decreasing void ratio [21].

Identifying contact points for ellipses and ellipsoids is not as straightforward as with circular or spherical particles. Yan and Regueiro [22] reported that the computational time of contact detection for a pair of ellipsoids is an order of magnitude (approximately 50 times) higher

than sphere contact detection. Moreover, contact detection using the existing algorithms is the main computational bottleneck in DEM simulations when dealing with a large number of ellipsoidal particles [13,23]. For example, one study shows that contact detection accounts for 81.4% of the total computation time for a DEM simulation of 2,000 ellipsoids [22]. In comparison, the contact detection for 2,000 spherical particles takes 5.15% of the total computation time. Therefore, an improvement in the computational time of the contact detection translates into a significant saving in the total computational time, the ability to simulate larger assemblies of grain and ultimately approach the macro-scale behavior of soil.

In addition to the computational cost, existing algorithms may become unstable in some conditions [24]. The accuracy and stability of contact detection algorithms in DEM may affect the calculation of the direction and magnitude of normal and tangential contact forces. Errors will propagate both in time and space and may have a significant impact on the short and long term mechanical behavior of assemblies of particles. These issues, along with the low performance of existing algorithms, motivated us to design a new contact detection algorithm to increase the DEM simulation performance while retaining high accuracy.

Another difficulty is that the notion of contact point is not uniquely defined for two overlapping ellipses. In the literature, a variety of methods and algorithms were proposed and developed in order to find contact points for ellipses and ellipsoids. As a result, a review of these algorithms is needed to compare and classify them. For this purpose, it is essential to develop some preliminary definitions to understand the strength and weakness of these algorithms. Therefore, in this research we provide new preliminaries and definitions to study and compare all the existing contact detection algorithms fundamentally.

During DEM simulations, contact detection plays a key role during both the initialization phase and during the dynamic phase. During initialization, the objective is to prepare an assembly of grains in force equilibrium, before compression or shear forces are applied, in a manner that follows established experimental procedures for soil preparation. At the initialization phase, one way is initializing particles randomly in a container with the non-overlapping condition until reaching a desired density, a specific height in the container, or a predefined number of particles. In this case, contact detection for particles is used to evaluate the overlap between particles. This is known as random packing of particles. Although some algorithms are specifically designed to detect the overlapping between two elliptical particles [25–27], elliptical contact detection can be also employed to calculate the penetration or separation distance [28–30]. During the dynamic phase, the required accuracy of the estimates of the contacts will be limited by the accuracy of the time stepping, and hence can be relaxed. Moreover, algorithms should allow one to update the new contact point between

two particles previously in contact in a minimal number of operations in order to be efficient.

Contact detection between two objects is also a fundamental problem in computer graphics, particularly in virtual reality models and in video games. Yet in those applications, simple contact detection algorithms based on planes and circles are usually sufficient because speed trumps accuracy. Contact detection also plays a critical role in the development of autonomous vehicles or in the control of swarms of robots and drones [31]. More sophisticated contact detection algorithms such as elliptical contact detection are applied in obstacle collision detection in the field of robotics [32, 33].

1.2 State of the Art

There are some contact detection methods which employ the approximation of ellipses for contact detection between pairs of elliptical particles. For instance, an ellipse can be approximated by segments of circles [21, 34–36], by grid or polar representation of particles [37], by four-arc approximation [35, 36], as a polyhedral surface [38], or using Non-Uniform Rational Basis Spline (NURBS) [39, 40]. One of the disadvantages of using approximations of actual ellipses is that it can lead to a contact point laying outside of the overlap region. We will not consider these approaches in this work.

Contact detection algorithms based on definitions which rely on the analytical representation of two ellipses and ellipsoids, find a pair of points as a contact pair. The contact point is then defined as the mid point of the contact pair. In this case, the contact point definition is not uniquely introduced. The straightforward approach is to identify the contact pair, if any, belonging to the intersection of the two ellipses. This method was originally developed in 2-D to provide intersection points of two colliding ellipses [41]. It was then extended and modified for ellipsoids in 3-D [42]. The method consists in finding the intersection set between two particles with small overlap. The intersection points in 2D can be found by solving a quartic equation. The method may become unstable and lead to inaccurate solutions as the overlap becomes very small, i.e. as the intersection set reduces to a single point [11]. This issue makes the method unsuitable for DEM simulations.

Another approach to find a contact pair begins with non-overlapping ellipses by identifying the pair of closest points on the ellipses. By introducing a constraint on the normal vectors to the ellipses at the contact pair [43], the notion of minimum distance pair can be extended to overlapping ellipses. This serves as the basis for the algorithm, studied by Wellmann et al. [44]. The problem can be formulated as a coupled minimization problem for the contact pair, which makes the method more computationally expensive than other methods [43],

although it provides a better representation of normal force direction [45]. However, we note that the formulation proposed in [43] is ill-posed and may return an incorrect contact pair for some configurations of particles.

Several researchers have developed algorithms that split the contact detection problem into two decoupled minimization problems, each one consisting in finding the closest point on one ellipse to the other and vice versa. Typically, closeness is measured with respect to the induced norm of the associated ellipse. This approach is the basis of the algorithms proposed in [24,43,46,47], which differ only in the solution process but not in the underlying problem definition. Lin et al. [43] solve the minimization problem using Lagrange multipliers. Cramer's rule [48, 49] was also applied to construct a quartic equation. Ting et al. [46] propose to map two ellipses to a unit circle located at origin and a transformed ellipse. A new constraint is also applied to the problem, in order to define a better conditioned quartic equation. However, the quartic equation degenerates to quadratic equation for some configurations of ellipses, such as when two ellipses are aligned with each other and have the same aspect ratio. Dziugys and Peters [24] claimed to obtain more stable algorithm than Ting. In this algorithm, a quartic equation is derived by transferring two ellipses to one unit circle and an ellipse located at origin with no rotation. In another study, Mustoe and Miyata [47] propose using the parametric equation of ellipse to simplify the quartic equation. In 3-D, the degree of these contact point equations is up to six. To find the contact pair, one approach is choosing the desired point from all solutions according to the aforementioned problem definition. However, finding all the roots of a polynomial function to ultimately keep only one has a non negligible computational cost. Alternatively, one may consider using root-finding algorithms, such as Newton's method, but the difficulty in this case is to determine a suitable initial guess that guarantees one to converge to the root associated with the unique solution of the minimization problem.

The published contact detection algorithms for pairs of ellipses or ellipsoids tend to propose incremental improvements over past methods, offer few comparisons to significantly different algorithms, and rarely distinguish their underlying mathematical problem from their numerical algorithm. As far as we know, the only attempt at a survey of these algorithms has been a dedicated chapter in a monograph [50]. In that survey, algorithms in [43, 46] are covered briefly and no comparison between the algorithms is made. In other studies [24,51], accuracy and performance of some contact detection algorithms are compared on specific test cases. In addition, Lin et al [43] compare their algorithms in terms of their accuracy and performance for $1,000$ ellipsoids. No such detailed review is found to compare all existing algorithms over different test cases for a pair and a large number of ellipses/ellipsoids.

1.3 Scientific Contributions

Contributions from this dissertation work can be summed up as follows:

1. This research describes a unified framework for the analysis and comparison of contact detection algorithms for pairs of ellipses and ellipsoids. Developing the framework is an attempt at bringing to light the common mathematical and computational concepts among the published algorithms. In this mathematical analysis of the contact detection problem, we

    - motivate and highlight different definitions of contact points,
    - recast the contact detection problems as minimization problems, each associated with a specific definition of contact point,
    - prove existence and uniqueness of solutions to the minimization problems,
    - provide a mathematical definition to characterize the notion of small overlap between particles,
    - propose a classification of the existing contact detection algorithms according to the definitions of contact point and determine some connections between them,
    - establish test cases to highlight the strengths and weaknesses of the algorithms,
    - provide comparisons in terms of performance, accuracy, and stability between the most efficient algorithms for each class over a large number of random pairs of ellipses or ellipsoids.

2. One of the main contributions is the development of a fast and robust contact detection algorithm that is computationally more efficient than existing algorithms. The algorithm belongs to the class of geometric potential methods, which consider the solution of two minimization problems in order to determine a contact point between the particles. The algorithm involves an original approach to provide an inexpensive estimate of the solution to one of the two minimization problems that can be used as an initial guess for the root finding iterative method. It also features a specific constraint that allows one to distinguish the global minima among the critical points in the minimization problems.

3. The last contribution deals with the development of a novel algorithm to randomly generate pairs of ellipses or ellipsoids. The algorithm allows one to create a pair of random particles for which the solution to one of the minimization problems in the

geometric potential methods is exactly known. The algorithm was largely used for code verification and for assessing the accuracy and performance of contact detection algorithms.

1.4 Outline

This dissertation is organized as follows. Following this introduction, we provide in Chapter 2 some preliminaries and general notations about the mathematical representations of ellipses and ellipsoids. We also describe the notion of concentric families of ellipses and ellipsoids and establish a new lemma whose purpose is to enumerate several properties of their associated normal vectors.

Chapter 3 describes a detailed mathematical analysis of the contact detection problem thereby laying the foundation results for our description of the various contact detection techniques presented in Chapter 4. We define three separate notions of contact resulting from the following definitions of contact pair: the Intersection Set (IS), the Minimum Distance Pair (MDP), and the Minimum Potential Pair (MPP). We define minimization problems associated with the definitions of these pairs in the case of disjoint ellipses, ellipses in perfect contact, and ellipses with overlap. We establish a new mathematical criterion to make the notion of small overlap precise when dealing with two overlapping ellipses. We show that the contact pair solutions to the minimization problems exist and are unique in the configuration of ellipses in near-perfect contact. In addition, we describe two mapping approaches to normalize a pair of ellipses/ellipsoids into a unit circle/sphere and an ellipse/ellipsoid. We will show how these transformations can help one to simplify the contact detection problem.

We review in Chapter 4 the contact detection algorithms available from the literature and analyze their respective advantages and disadvantages. Following our mathematical analysis of the contact detection problem, we show that all existing algorithms actually belong to one of the three classes of methods associated with the three definitions of contact pair. We will describe how the algorithms within a class differ from each other with respect to specific choices in the solution techniques used to solve the corresponding minimization problem.

We present in Chapter 5 the novel algorithm for contact detection between ellipses and ellipsoids. The algorithm produces the Minimum Potential Pair as solution of the problem. We describe in particular the approach for the calculation of an initial guess point and the additional constraint that we enforce to ensure that the iterative method converges to the global minimum of the objective function. The algorithm is shown to be robust, efficient, and fast when compared to existing algorithms.

In Chapter 6, we present a series of numerical examples to compare the accuracy, stability, and computational cost of the different algorithms. We describe some examples that illustrate how contact points may differ according their definitions. We also describe in the appendix section the algorithm to generate random pairs of ellipses or ellipsoids that are used in the numerical experiments to assess the accuracy of some algorithms.

Finally, we provide in Chapter 7 some concluding remarks and directions for future works.

CHAPTER 2 NOTATION AND PRELIMINARIES ON ELLIPSES AND ELLIPSOIDS

This chapter sets the stage for the comparison between different contact detection algorithms by collecting often recurring notation and definitions. By beginning with a compact but coherent introduction to the terms and expressions, we hope to make the similarities between the different algorithms quickly transparent. All of these notations and definitions are standard and well-known in the literature, except a lemma at the end of this chapter. The lemma provides several properties regarding the normal vectors to ellipses.

2.1 Representation of Ellipses

An ellipse $\mathcal{E}$ is the set of roots $\boldsymbol{x} = [x, y]^T \in \mathbb{R}^2$ of a quadratic polynomial of the form

$$f(\boldsymbol{x}) := (\boldsymbol{x} - \boldsymbol{c})^T \mathcal{Q}(\boldsymbol{x} - \boldsymbol{c}) - 1, \tag{2.1}$$

where $\mathcal{Q}$ is a symmetric positive-definite (SPD) matrix in $\mathbb{R}^{2\times 2}$ and $\boldsymbol{c} = [c_x, c_y]^T \in \mathbb{R}^2$ is the center of the ellipse. Formally written, an ellipse is defined as:

$$\mathcal{E} = \Big\{ \boldsymbol{x} \in \mathbb{R}^2 \,;\, (\boldsymbol{x} - \boldsymbol{c})^T \mathcal{Q}(\boldsymbol{x} - \boldsymbol{c}) - 1 = 0 \Big\}.$$

The coordinates $\boldsymbol{x}$ in which an ellipse is initially given will be referred to as the *global coordinates* but there exists an isometry to a system of coordinates in which the geometry of $\mathcal{E}$ is especially straightforward. Indeed, a fundamental result of linear algebra states that for each SPD matrix $\mathcal{Q}$, there exists an orthogonal matrix $\mathcal{R}$, that is satisfying $\mathcal{R}^{-1} = \mathcal{R}^T$ and therefore in the form

$$\mathcal{R} := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \qquad \theta \in [-\pi, \pi[, \tag{2.2}$$

such that the matrix

$$\mathcal{D} := \mathcal{R}^T \mathcal{Q} \mathcal{R} = \begin{bmatrix} 1/a^2 & 0 \\ 0 & 1/b^2 \end{bmatrix}, \tag{2.3}$$

is diagonal with strictly positive entries, i.e. the eigenvalues of $\mathcal{Q}$. An example of an ellipse is shown in Figure 2.1 under the assumption that $a \geq b$. The axes corresponding to a and b are called the *semi-major axis* and the *semi-minor axis*, respectively. Accordingly, a and b are usually referred to as the *semi-axes* of the ellipse.

Further including a translation to send the center $\boldsymbol{c}$ to the origin, we can introduce the *local*

coordinates $\boldsymbol{\xi} = [\xi, \eta]^T$,

$$\boldsymbol{\xi} = \mathcal{R}^T(\boldsymbol{x} - \boldsymbol{c}), \tag{2.4}$$

with respect to which the ellipse consists in the set of roots of

$$\widehat{f}(\boldsymbol{\xi}) = \boldsymbol{\xi}^T \mathcal{D} \boldsymbol{\xi} - 1, \tag{2.5}$$

which can be recast in the classical form

$$\widehat{f}(\xi, \eta) = \frac{\xi^2}{a^2} + \frac{\eta^2}{b^2} - 1. \tag{2.6}$$

In this dissertation, f will be called the *global geometric potential*, or simply potential, of the ellipse while $\widehat{f}$ will be called the *local potential*. Clearly, the potential will always be a convex function with a minimum at the center $\boldsymbol{c}$ of the ellipse.

Definition 1. [$\mathcal{Q}$**-norm**] *A SPD matrix* $\mathcal{Q}$ *induces the so-called* $\mathcal{Q}$*-norm*

$$\|\boldsymbol{x}\|_{\mathcal{Q}} := \sqrt{\boldsymbol{x}^T \mathcal{Q} \boldsymbol{x}}, \quad \forall \boldsymbol{x} \in \mathbb{R}^2. \tag{2.7}$$

This definition allows one to interpret an ellipse $\mathcal{E}$ as the "circle" satisfying $\|\boldsymbol{x} - \boldsymbol{c}\|_{\mathcal{Q}}^2 = 1$ in the global coordinates. Eventually, when we consider the contact problem for two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$, we may replace the subscript $\mathcal{Q}$ by the index i of the ellipse $\mathcal{E}_i$. Throughout this dissertation, the norm $\|\cdot\|$ written without a subscript will denote the usual Euclidean norm.

We now proceed with an explicit description of an ellipse as defined by (2.1). Let the matrix $\mathcal{Q}$ be explicitly given by

$$\mathcal{Q} = \begin{bmatrix} A & C \\ C & B \end{bmatrix}, \tag{2.8}$$

where positive-definiteness is ensured by the conditions $A > 0$, $B > 0$, and $\det \mathcal{Q} = AB - C^2 > 0$. Then, in the global coordinates, the potential (2.1) is given as

$$f(x, y) = A(x - c_x)^2 + B(y - c_y)^2 + 2C(x - c_x)(y - c_y) - 1. \tag{2.9}$$

For convenience, we provide below the explicit relationships between $\mathcal{D}$ and $\mathcal{R}$, and $\mathcal{Q}$, namely

$$\begin{aligned} A &= \frac{\cos^2\theta}{a^2} + \frac{\sin^2\theta}{b^2}, \\ B &= \frac{\sin^2\theta}{a^2} + \frac{\cos^2\theta}{b^2}, \\ C &= \sin\theta\cos\theta\left(\frac{1}{a^2} - \frac{1}{b^2}\right). \end{aligned} \tag{2.10}$$

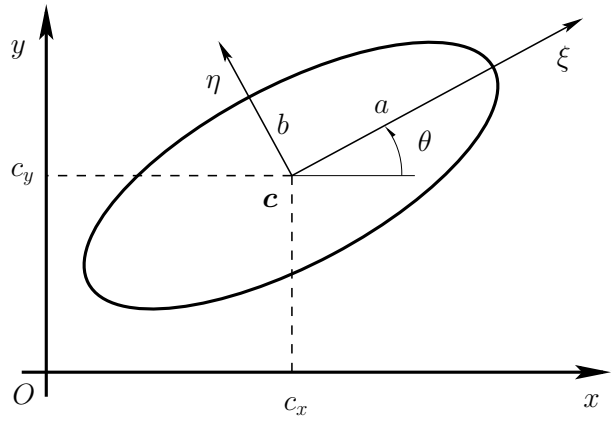


Figure 2.1 Ellipse in global coordinate system (O, x, y) with local coordinate system (O, ξ, η) centered at $\boldsymbol{c}$.

An alternative form in which to express the potential is based on separating out the quadratic, linear, and constant terms. Starting from (2.1), we find that the points $\boldsymbol{x}$ on an ellipse satisfy

$$\begin{aligned} f(\boldsymbol{x}) &= (\boldsymbol{x}-\boldsymbol{c})^T \mathcal{Q}(\boldsymbol{x}-\boldsymbol{c}) - 1 \\ &= \boldsymbol{x}^T\mathcal{Q}\boldsymbol{x} - \boldsymbol{x}^T\mathcal{Q}\boldsymbol{c} - \boldsymbol{c}^T\mathcal{Q}\boldsymbol{x} + \boldsymbol{c}^T\mathcal{Q}\boldsymbol{c} - 1 \\ &= \boldsymbol{x}^T\mathcal{Q}\boldsymbol{x} - \boldsymbol{x}^T(\mathcal{Q}\boldsymbol{c}) - (\mathcal{Q}\boldsymbol{c})^T\boldsymbol{x} + \boldsymbol{c}^T\mathcal{Q}\boldsymbol{c} - 1. \end{aligned} \tag{2.11}$$

Introducing

$$F = \boldsymbol{c}^T\mathcal{Q}\boldsymbol{c} - 1, \quad \begin{bmatrix} D \\ E \end{bmatrix} = -\mathcal{Q}\boldsymbol{c}, \quad \mathcal{P} = \begin{bmatrix} A & C & D \\ C & B & E \\ D & E & F \end{bmatrix}, \quad \boldsymbol{z} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{2.12}$$

the potential f can then be rewritten in augmented matrix form as

$$f(x,y) = \boldsymbol{z}^T\mathcal{P}\boldsymbol{z} = Ax^2 + By^2 + 2Cxy + 2Dx + 2Ey + F. \tag{2.13}$$

Another useful description is based on the parameterization of the ellipse in terms of a parameter $t \in [-\pi, \pi[$ such that all points given by

$$\boldsymbol{\xi}(t) = \begin{bmatrix} \xi(t) \\ \eta(t) \end{bmatrix} = \mathcal{D}^{-1/2} \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} = \begin{bmatrix} a\cos t \\ b\sin t \end{bmatrix}, \tag{2.14}$$

lie on the ellipse. Using the mapping (2.4), the ellipse in the global coordinate system consists

then of the points

$$\boldsymbol{x}(t) = \mathcal{R}\boldsymbol{\xi}(t) + \boldsymbol{c} = \mathcal{R}\mathcal{D}^{-1/2} \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} + \boldsymbol{c}, \qquad t \in [-\pi, \pi[. \tag{2.15}$$

Certain algorithms for contact detection between two ellipses require the outward unit normal vector at a point $\boldsymbol{x}$ or, equivalently, at a point $\boldsymbol{\xi}$, on an ellipse. From the definitions of the potentials f and $\widehat{f}$, see Equations (2.1) and (2.5), respectively, a simple calculation shows that the normal is given in global coordinates as,

$$\boldsymbol{n}(\boldsymbol{x}) = \frac{\nabla f(\boldsymbol{x})}{\|\nabla f(\boldsymbol{x})\|} = \frac{\mathcal{Q}(\boldsymbol{x} - \boldsymbol{c})}{\|\mathcal{Q}(\boldsymbol{x} - \boldsymbol{c})\|}, \tag{2.16}$$

or in local coordinates as,

$$\boldsymbol{n}(\boldsymbol{\xi}) = \frac{\nabla \widehat{f}(\boldsymbol{\xi})}{\|\nabla \widehat{f}(\boldsymbol{\xi})\|} = \frac{\mathcal{D}\boldsymbol{\xi}}{\|\mathcal{D}\boldsymbol{\xi}\|}\,. \tag{2.17}$$

Without delving into the explicit calculations, which can be found in several references [52,53], we note that the minimum radius of curvature along an ellipse is given by

$$\underline{\rho} = \frac{b^2}{a}. \tag{2.18}$$

There exist several alternative descriptions of ellipses. For the sake of completeness, we mention here some of the most important descriptions. The first one will nevertheless motivate an algorithm for finding initial guess points, to be detailed in Section 5.3. We recall that the focal points of an ellipse, $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$, are located on its semi-major axis, at equal distance from the center, and are explicitly given in local coordinates as

$$\boldsymbol{f}_1 = \Big(- \sqrt{a^2 - b^2}, 0 \Big), \qquad \boldsymbol{f}_2 = \Big(+ \sqrt{a^2 - b^2}, 0 \Big). \tag{2.19}$$

The ellipse can then be defined as the set of points $\boldsymbol{x}$ that satisfy

$$\|\boldsymbol{x} - \boldsymbol{f}_1\| + \|\boldsymbol{x} - \boldsymbol{f}_2\| = 2a. \tag{2.20}$$

Moreover, it is possible to show that the normal vector at $\boldsymbol{x}$ generates a line that bisects the angle $\angle \boldsymbol{f}_1 \boldsymbol{x} \boldsymbol{f}_2$. There is also a well-known description of an ellipse in terms of its eccentricity $e = \sqrt{a^2 - b^2}/a \in [0, 1]$, with $e = 0$ corresponding to a circle [54]. Ellipses can be geometrically obtained as the cross-section of the intersection of an inclined plane with a conic section, but this description of a 2-D ellipse requires three dimensions, thus making it

less practical. Ellipses can also be described using mechanical means, such as the Trammel of Archimedes, the Tusi couple, or the ellipsograph of Benjamin Bramer [54–56]. The Steiner method for the construction of an ellipse is quite elegant but requires a discretization, and is therefore not relevant to the continuous contact detection problem. In summary, ellipses possess a wealth of fascinating and unexpected properties but none of these seem to be as useful as (2.1) or (2.5) when one needs to numerically estimate contact points.

2.2 Representation of Ellipsoids

Similarly to ellipses, an ellipsoid $\mathcal{E} \subset \mathbb{R}^3$ is the set of roots to the potential:

$$f(\boldsymbol{x}) := (\boldsymbol{x} - \boldsymbol{c})^T \mathcal{Q} (\boldsymbol{x} - \boldsymbol{c}) - 1 \tag{2.21}$$

where $\mathcal{Q}$ is a 3×3 SPD matrix and $\boldsymbol{c} \in \mathbb{R}^3$ is the center of the ellipsoid. As in 2-D, there exists an orthogonal change of variable $\mathcal{R} \in \mathbb{R}^{3\times 3}$, $\mathcal{R}^{-1} = \mathcal{R}^T$, which diagonalizes $\mathcal{Q}$ such that $\mathcal{D} = \mathcal{R}^T \mathcal{Q} \mathcal{R}$. The eigenvalues of $\mathcal{Q}$, i.e. the entries of $\mathcal{D}$, are all strictly positive and are denoted by a^{-2}, b^{-2}, and c^{-2}, where the positive constants a, b, and c are assumed to be ordered as $c \leq b \leq a$. Using the change of variable (2.4), with $\boldsymbol{\xi} = [\xi, \eta, \zeta]^T$, one can write the local potential in its so-called local coordinate system (O, ξ, η, ζ) as:

$$\widehat{f}(\boldsymbol{\xi}) = \boldsymbol{\xi}^T \mathcal{D} \boldsymbol{\xi} - 1,$$

or

$$\widehat{f}(\xi, \eta, \zeta) = \frac{\xi^2}{a^2} + \frac{\eta^2}{b^2} + \frac{\zeta^2}{c^2} - 1. \tag{2.22}$$

The positive constants a, b, and c are called the semi-axes of the ellipsoid.

Remark 1. *Unlike in 2-D, the explicit form of the rotation matrix $\mathcal{R}$ can be obtained in several manners. We first note that an arbitrary ellipsoid is defined in terms of nine parameters: the coordinates of its center $\boldsymbol{c} = [c_x, c_y, c_z]^T$ and the six entries of the symmetric matrix $\mathcal{Q}$. However, since the matrix $\mathcal{Q}$ can also be written as $\mathcal{R}\mathcal{D}\mathcal{R}^T$, these six entries can also be identified with the positive eigenvalues a, b, and c appearing as the diagonal elements of $\mathcal{D}$, and the three parameters needed to describe a general orthogonal transformation $\mathcal{R}$. In other words, one needs to introduce three angles, each corresponding to an elementary rotation, and write $\mathcal{R}$ as the composition of the three rotation matrices in order to map the local coordinate system into the principal axes of the ellipsoid in the global coordinate system. The choice of the three angles is actually not unique and depends on the representation considered, for example the Euler rotations [57] or the quaternion rotations [58]. We will not describe*

these methods here and will simply assume that $\mathcal{R}$, if necessary, is provided using one of the methods as

$$\mathcal{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}.$$

An ellipsoid in the local coordinate system can be represented in terms of the spherical coordinates $(u, v) \in [-\pi, \pi[\times[0, \pi[$ as

$$\boldsymbol{\xi}(u, v) = D^{-1/2} \begin{bmatrix} \cos u \sin v \\ \sin u \sin v \\ \cos v \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} \cos u \sin v \\ \sin u \sin v \\ \cos v \end{bmatrix} = \begin{bmatrix} a \cos u \sin v \\ b \sin u \sin v \\ c \cos v \end{bmatrix}. \tag{2.23}$$

Using the mapping $\boldsymbol{x} = \mathcal{R}\boldsymbol{\xi} + \boldsymbol{c}$, the ellipsoid in the global coordinate system is therefore parameterized as

$$\boldsymbol{x}(u, v) = \mathcal{R}\boldsymbol{\xi}(u, v) + \boldsymbol{c} = \mathcal{R}\mathcal{D}^{-1/2} \begin{bmatrix} \cos u \sin v \\ \sin u \sin v \\ \cos v \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix}, \quad \forall (u, v) \in [-\pi, \pi[\times[0, \pi[. \tag{2.24}$$

As in 2-D, the outward unit normal vector at a point $\boldsymbol{x}$ on an ellipsoid is given in global coordinates by

$$\boldsymbol{n}(\boldsymbol{x}) = \frac{\nabla f(\boldsymbol{x})}{\|\nabla f(\boldsymbol{x})\|} = \frac{\mathcal{Q}(\boldsymbol{x} - \boldsymbol{c})}{\|\mathcal{Q}(\boldsymbol{x} - \boldsymbol{c})\|}, \tag{2.25}$$

or in local coordinates by

$$\boldsymbol{n}(\boldsymbol{\xi}) = \frac{\nabla \widehat{f}(\boldsymbol{\xi})}{\|\nabla \widehat{f}(\boldsymbol{\xi})\|} = \frac{\mathcal{D}\boldsymbol{\xi}}{\|\mathcal{D}\boldsymbol{\xi}\|}. \tag{2.26}$$

We conclude by observing that the gradient in 3-D is given by the same formula as (2.18) while the minimum radius of curvature is, assuming $a \geq b \geq c$

$$\underline{\rho} = \frac{c^2}{a}. \tag{2.27}$$

2.3 Family of Concentric Similar Ellipses and Ellipsoids

Let $\mathcal{E}$ be an arbitrary ellipse or ellipsoid with potential

$$f(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{c})^T \mathcal{Q}(\boldsymbol{x} - \boldsymbol{c}) - 1.$$

Then, the family of concentric similar ellipses ($d = 2$) or ellipsoids ($d = 3$) associated with $\mathcal{E}$ consists of the sets

$$\mathcal{E}(r) := \left\{ \boldsymbol{x} \in \mathbb{R}^d \, ; f(\boldsymbol{x}) - (r^2 - 1) = 0 \right\}, \qquad \forall r \geq 0, \tag{2.28}$$

or as the roots of $f_r(\boldsymbol{x}) := f(\boldsymbol{x}) - (r^2 - 1) = (\boldsymbol{x} - \boldsymbol{x})^T \mathcal{Q}(\boldsymbol{x} - \boldsymbol{x}) - r^2$. We note that two ellipses or two ellipsoids within the same family, i.e. $\mathcal{E}(r_1)$ and $\mathcal{E}(r_2)$ with $r_1 \neq r_2$, form a homoeoid, that is, the bounded region between $\mathcal{E}(r_1)$ and $\mathcal{E}(r_2)$. Moreover, $\mathcal{E}(0)$ reduces to the singleton $\{\boldsymbol{c}\}$ while $\mathcal{E}(1)$ corresponds to $\mathcal{E}$. Furthermore, for every point $\boldsymbol{x} \in \mathbb{R}^d$, there exists a unique $r \geq 0$ such that $\boldsymbol{x} \in \mathcal{E}(r)$ and the gradient of f_r at $\boldsymbol{x} \in \mathcal{E}(r)$ is given by

$$\nabla f_r(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) = 2\mathcal{Q}(\boldsymbol{x} - \boldsymbol{c}).$$

In other words, the outward unit normal vector to the ellipse/ellipsoid $\mathcal{E}(r)$ associated with $\mathcal{E}$ at an arbitrary point $\boldsymbol{x} \in \mathbb{R}^d \backslash \{\boldsymbol{c}\}$ is then given by

$$\boldsymbol{n}(\boldsymbol{x}) = \frac{\nabla f_r(\boldsymbol{x})}{\|\nabla f_r(\boldsymbol{x})\|} = \frac{\nabla f(\boldsymbol{x})}{\|\nabla f(\boldsymbol{x})\|} = \frac{\mathcal{Q}(\boldsymbol{x} - \boldsymbol{c})}{\|\mathcal{Q}(\boldsymbol{x} - \boldsymbol{c})\|}. \tag{2.29}$$

We now provide some properties satisfied by the vector field $\boldsymbol{n}(\boldsymbol{x})$ associated to an ellipse $\mathcal{E}$ which will be extensively used in Chapter 3. These properties will be expressed using complex multiplication and elements of projective geometry which we now recall. Let S^1 be the set of points of unit modulus in the complex plane $\mathbb{C}$, which will be used to represent both the unit gradient field $\boldsymbol{n}$ and unit direction $\boldsymbol{w}$. Given the points $e^{i\omega}$ and $e^{i\theta}$ in S^1, then complex multiplication between the two points can be written as

$$e^{i\omega} e^{i\theta} = e^{i(\omega+\theta)},$$

thereby representing the composition of two rotations.

In projective geometry, the real plane $\mathbb{R}^2$ is embedded into the compact space of all directions in $\mathbb{R}^3$ using the association of $[x, y]^T \in \mathbb{R}^2$ to the direction

$$[x : y : 1] := \{ \, [xt, yt, t]^T \in \mathbb{R}^3 \, | \, t \in \mathbb{R}^+ \, \},$$

identified here in *homogeneous coordinates*. The space of all directions

$$[x : y : z] := \{ \, [xt, yt, zt]^T \in \mathbb{R}^3 \setminus \{\mathbf{0}\} \, | \, t \in \mathbb{R}^+ \, \},$$

is called the *projective sphere* SP^2, not to be confused with the projective plane obtained after associating antipodal points in the projective sphere. The *points at infinity* are those corresponding to the directions $[x : y : 0]$, thus forming a circle. In fact, given a first degree map $g : \mathbb{R}^2 \to \mathbb{R}^2$, say

$$g(\boldsymbol{x}) = \mathcal{T}\boldsymbol{x} + \mathbf{b}\,,$$

for $\mathbf{b} \in \mathbb{R}^2$ and $\mathcal{T}$ a 2×2 matrix, then along the segment in direction $\boldsymbol{w} \in S^1$

$$\boldsymbol{x} = \boldsymbol{c} + t\boldsymbol{w}\,, \quad t \in \mathbb{R}^+\,,$$

we can define a limiting direction

$$\lim_{t\to\infty} \frac{1}{t} g(\boldsymbol{x}) = \lim_{t\to\infty} \frac{1}{t} g(\boldsymbol{c} + t\boldsymbol{w}) = \lim_{t\to\infty} \mathcal{T}\boldsymbol{w} + \frac{1}{t}\mathcal{T}\boldsymbol{c} + \frac{1}{t}\mathbf{b} = \mathcal{T}\boldsymbol{w}\,.$$

This association is independent of $\boldsymbol{c}$ and the parametrization t, thus leading to a well defined map $g^\infty : S^1 \to S^1$ according to

$$\boldsymbol{w} \longmapsto \frac{\mathcal{T}\boldsymbol{w}}{\|\mathcal{T}\boldsymbol{w}\|}\,.$$

This map is the restriction at infinity of the extension of g from the projective sphere to itself.

Lemma 1. *Consider an ellipse $\mathcal{E} \subset \mathbb{R}^2$ centered at $\boldsymbol{c} \in \mathbb{R}^2$ whose unit vector associated with the semi-major and semi-minor axes are $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, respectively, oriented counter-clockwise. The vector field $\boldsymbol{n}$ given by* (2.29) *satisfies the following properties:*

i) The vector field $\boldsymbol{n}(\boldsymbol{x})$ is well-defined $\forall \boldsymbol{x} \in \mathbb{R}^2 \setminus \{\boldsymbol{c}\}$.

ii) The relations $\boldsymbol{n}(\boldsymbol{c} \pm t\boldsymbol{\xi}) = \pm\boldsymbol{\xi}$ and $\boldsymbol{n}(\boldsymbol{c} \pm t\boldsymbol{\eta}) = \pm\boldsymbol{\eta}$ hold $\forall t \in \mathbb{R}^+$.

iii) Given $\boldsymbol{w} \in S^1$, $\boldsymbol{n}(\boldsymbol{c} + t\boldsymbol{w})$ is constant $\forall t \in \mathbb{R}^+$.

iv) The map

$$\begin{aligned} \mathcal{N} : S^1 &\longrightarrow S^1 \\ \boldsymbol{w} &\longmapsto \lim_{r\to\infty} \boldsymbol{n}(\boldsymbol{c} + r\boldsymbol{w})\,, \end{aligned} \tag{2.30}$$

is well-defined and satisfies the following properties:

(a) $\pm\boldsymbol{\xi}$ and $\pm\boldsymbol{\eta}$ are fixed points.

(b) $\mathcal{N}$ is bijective and $\mathcal{N}$ is the identity if and only if $\mathcal{E}$ is a circle.

(c) If $\boldsymbol{w} = e^{i\sigma}\boldsymbol{\xi} \in S^1$, $\sigma \in [0, 2\pi[$, *there exists* $\theta \in \,]-\pi/2, \pi/2[$ *such that*

$$\mathcal{N}(\boldsymbol{w}) = e^{i\theta}\boldsymbol{w} = e^{i(\theta+\sigma)}\boldsymbol{\xi}\,, \quad \textit{with} \quad \tan(\theta+\sigma) = (a/b)^2 \tan\sigma\,. \tag{2.31}$$

(d) If $\boldsymbol{x}_0 \neq \boldsymbol{c}$, *there exists* $R = R(\boldsymbol{x}_0) \in \mathbb{R}^+$, *such that for* $r \geq R$ *the estimate*

$$\Big\|\mathcal{N}(\boldsymbol{w}) - \boldsymbol{n}(\boldsymbol{x}_0 + r\boldsymbol{w})\Big\| = \mathcal{O}\Big(r^{-1}\|\boldsymbol{x}_0 - \boldsymbol{c}\|\Big)\,, \tag{2.32}$$

is uniform with respect to $\boldsymbol{w} \in S^1$.

Proof. From (2.29), the vector field $\boldsymbol{n}(\boldsymbol{x})$ is the unit vector field associated with the gradient field

$$\nabla f(\boldsymbol{x}) = 2\mathcal{Q}(\boldsymbol{x} - \boldsymbol{c}). \tag{2.33}$$

Given that $\mathcal{Q}$ is SPD, ∇f only vanishes at $\boldsymbol{x} = \boldsymbol{c}$. This proves property *i)*. To demonstrate property *ii)*, we observe that $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are the eigenvectors of $\mathcal{Q}$ associated with the eigenvalues $1/a^2$ and $1/b^2$, respectively; see (2.3). Hence, substituting $\boldsymbol{x} = \boldsymbol{c} \pm t\boldsymbol{\xi}$ into (2.33) we find

$$2\mathcal{Q}(\boldsymbol{x} - \boldsymbol{c}) = 2\mathcal{Q}(\pm t\boldsymbol{\xi}) = \pm\frac{2t}{a^2}\boldsymbol{\xi},$$

which implies that $\boldsymbol{n}(\boldsymbol{x}) = \pm\boldsymbol{\xi}$. Similarly, substituting $\boldsymbol{x} = \boldsymbol{c} \pm t\boldsymbol{\eta}$ into (2.33) shows that $\boldsymbol{n}(\boldsymbol{x}) = \pm\boldsymbol{\eta}$.

Let $\boldsymbol{w} \in S^1$ and consider the half-line supported by $\boldsymbol{w}$, i.e. the set of points $\boldsymbol{x} = \boldsymbol{c} + t\boldsymbol{w}$ with $t > 0$. The gradients along the half-line

$$\nabla f(\boldsymbol{x}) = 2t\mathcal{Q}\boldsymbol{w} \tag{2.34}$$

are all positive multiples of the same vector $\mathcal{Q}\boldsymbol{w}$. Hence, the vector field $\boldsymbol{n}(\boldsymbol{x})$ is constant along the half-line, which proves property *iii)*. This is illustrated in Figure 2.2.

We now consider the proof of *iv)* by first demonstrating *(d)*. The bound (2.32) will imply that the function

$$\mathcal{N}(\boldsymbol{w}) = \lim_{r\to\infty} \boldsymbol{n}(\boldsymbol{c} + r\boldsymbol{w})$$

is in fact the same as if we had taken $\lim \boldsymbol{n}(\boldsymbol{x}_0 + r\boldsymbol{w})$, and therefore does not depend on the origin $\boldsymbol{x}_0$ of the segment $\boldsymbol{x}_0 + r\boldsymbol{w}$. For any $\boldsymbol{x}_0 \neq \boldsymbol{c}$ and any direction $\boldsymbol{w}$,

$$\nabla f(\boldsymbol{x}_0 + r\boldsymbol{w}) = 2\mathcal{Q}\Big(r\boldsymbol{w} - (\boldsymbol{c} - \boldsymbol{x}_0)\Big) = 2r\mathcal{Q}\boldsymbol{w} - 2\mathcal{Q}(\boldsymbol{c} - \boldsymbol{x}_0).$$

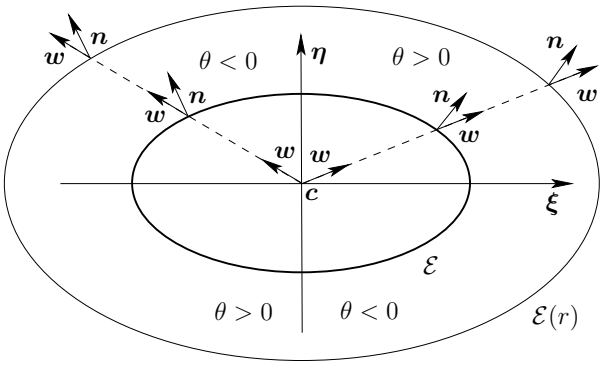


Figure 2.2 Illustration of the property *iii)* of Lemma 1, showing that the vector $\boldsymbol{n}(\boldsymbol{c}+t\boldsymbol{w})$ is constant for a given vector $\boldsymbol{w}$ and $t > 0$. The angle θ is the angle between $\boldsymbol{n}$ and $\boldsymbol{w}$ measured counter-clockwise.

For large and positive r, we have that

$$\boldsymbol{n}(\boldsymbol{x}_0 + r\boldsymbol{w}) = \frac{\mathcal{Q}\boldsymbol{w} - r^{-1}\mathcal{Q}(\boldsymbol{c}-\boldsymbol{x}_0)}{\|\mathcal{Q}\boldsymbol{w} - r^{-1}\mathcal{Q}(\boldsymbol{c}-\boldsymbol{x}_0)\|} = \frac{\mathcal{Q}\boldsymbol{w}}{\|\mathcal{Q}\boldsymbol{w}\|} + \mathcal{O}\Big(r^{-1}\|\boldsymbol{x}_0 - \boldsymbol{c}\|\Big) \approx \mathcal{N}(\boldsymbol{w}).$$

As a matter of fact, this approximation can be made uniform in $\boldsymbol{w}$ for r sufficiently large. In other words, there exists constants R, δ, and C such that $\forall r \geq R$ and $\forall \boldsymbol{x}_0 \in \mathbb{R}^2$ satisfying $\|\boldsymbol{x}_0 - \boldsymbol{c}\| < \delta$, one has

$$\Big\|\mathcal{N}(\boldsymbol{w}) - \boldsymbol{n}(\boldsymbol{x}_0 + r\boldsymbol{w})\Big\| < C\frac{\|\boldsymbol{x}_0 - \boldsymbol{c}\|}{r}, \qquad \forall \boldsymbol{w} \in S^1.$$

Property *iv)-(a)* follows immediately from *ii)*. Property *iv)-(b)* will follow immediately from *iv)-(c)*. In particular, we observe that $\mathcal{N}(\boldsymbol{w})$ is the identity if and only if $\theta = 0$ which according to the relation (2.31) occurs if and only if $a/b = 1$.

Only property *iv)-(c)* still needs to be demonstrated. We will begin by proving it for $\boldsymbol{w} \in [\boldsymbol{\xi}, \boldsymbol{\eta}] \subset S^1$. Consider the parameterization by $\sigma \in [0, \pi/2]$ of every direction $\boldsymbol{w}(\sigma) \in [\boldsymbol{\xi}, \boldsymbol{\eta}]$ according to

$$\sigma \longmapsto \boldsymbol{w}(\sigma) := \cos\sigma\, \boldsymbol{\xi} + \sin\sigma\, \boldsymbol{\eta}, \tag{2.35}$$

and remark that

$$\nabla f\Big(\boldsymbol{c} + r\boldsymbol{w}(\sigma)\Big) = 2r\mathcal{Q}\boldsymbol{w}(\sigma) = 2r\left[\frac{\cos\sigma}{a^2}\,\boldsymbol{\xi} + \frac{\sin\sigma}{b^2}\,\boldsymbol{\eta}\right].$$

From this expression, it is clear that $\mathcal{N}\Big(\boldsymbol{w}(\sigma)\Big)$ belongs between $[\boldsymbol{\xi}, \boldsymbol{\eta}] \subset S^1$, and hence that

there exists an angle $\widehat{\sigma} \in [0, \pi/2]$ such that

$$\mathcal{N}\Big(\boldsymbol{w}(\sigma)\Big) = \boldsymbol{w}(\widehat{\sigma}),$$

In fact, for all $\sigma \in [0, \pi/2[$ and $\widehat{\sigma} \in [0, \pi/2[$, we have the relation

$$\tan \widehat{\sigma} = \frac{a^2}{b^2} \tan \sigma \tag{2.36}$$

announced in (2.31). We remark that the derivative of the map (2.30) in the coordinates (2.35) satisfies

$$\frac{d\widehat{\sigma}}{d\sigma} = \frac{a^2}{b^2} \frac{\cos^2 \widehat{\sigma}}{\cos^2 \sigma} > 0. \tag{2.37}$$

This shows that the map is bijective over $[\boldsymbol{\xi}, \boldsymbol{\eta}]$ and that the map is the identity if and only if the ellipse is a circle (i.e. $b = a$). In the map (2.30), the angle θ is given by

$$\theta = \widehat{\sigma} - \sigma,$$

and because $\widehat{\sigma} > \sigma$ by (2.36) while both angles belong to $[0, \pi/2[$, then $\theta \in [0, \pi/2[$. Since $\mathcal{N}$ has a fixed point at $\boldsymbol{w} = \boldsymbol{\eta}$, that is when $\sigma = \pi/2$, we conclude that $\theta(\pi/2) = 0$. Therefore, for all $\sigma \in [0, \pi/2]$, we have $\theta(\sigma) \in [0, \pi/2[$. In fact, the parameterization (2.35) with $\sigma \in [-\pi/2, 0]$ can be used for $\boldsymbol{w}(\sigma) \in [-\boldsymbol{\eta}, \boldsymbol{\xi}] \subset S^1$, and leads again to the relation (2.36). Applying the same argument (or by symmetry along the $\boldsymbol{\eta}$ axis) over $[\boldsymbol{\eta}, -\boldsymbol{\xi}]$ and $[-\boldsymbol{\xi}, -\boldsymbol{\eta}]$ demonstrates *iv)-(c)*. In all these cases, we have that $\theta \in]-\pi/2, \pi/2[$. This concludes the proof. □

CHAPTER 3 MATHEMATICAL FRAMEWORK FOR PAIRS OF ELLIPSES AND ELLIPSOIDS

The purpose of this chapter is to introduce elementary notions of *contact points* and of their properties for pairs of ellipses and ellipsoids. Lacking a common framework, much of the past work provided little indication of the connections between the different algorithms. This chapter is an attempt at filling this void by presenting a few precise definitions and results which will serve as a common thread in later comparisons of the different contact detection algorithms. Our approach shares the same level of mathematical rigor as that provided by Perram, Wertheim et al. [25, 59] in their development of *Potential Contact Theory*, based on earlier work of Vieillard-Baron [60], and leads to its own definition of separation/penetration distance. We will not be considering the Perram-Wertheim theory because of its inherently high computational cost [61]. A rigorous theory for the continuous contact detection problem has also been developed by Wang and his collaborators [26, 62, 63], but it is too computationally expensive for the quasi-static regime found in DEM, and therefore will not be described. Nevertheless, we will make connections to those theories in Section 3.5. Unfortunately, the most computationally efficient algorithms are not expressed with the same level of rigor, which is what this chapter attempts to correct within the literature. In order to make the presentation more concise, we will focus on the 2-D contact detection problem and will indicate in Section 3.6 how these results could be extended to the 3-D case. At the end of this chapter, we introduce two mapping approaches for a pair of ellipses, which can be straightforwardly adapted to the case of ellipsoids. These mapping are independent from other sections of the current chapter and will be used in some contact detection algorithms which we will describe in Chapter 4.

In practice, every contact detection algorithm for ellipses should provide a single contact point, a single contact normal, and either a separation or a penetration distance. However, given two elliptical particles[1] $\mathcal{E}_i$ and $\mathcal{E}_j \subset \mathbb{R}^2$, and two points judiciously constructed on each particle, say $\boldsymbol{x}_i \in \mathcal{E}_i$ and $\boldsymbol{x}_j \in \mathcal{E}_j$, one could compute the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ as the separation or penetration distance and define the midpoint between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ as the contact point (which should provide reasonable approximations of the contact properties in case of ellipses with small overlap). The contact normal could then be defined in terms of the segment joining $\boldsymbol{x}_i$ to $\boldsymbol{x}_j$. For most of the algorithms we shall describe in Chapter 4, this is precisely how the contact point and contact normal are actually computed.

[1] Note that we have chosen to follow the usual notation $\mathcal{E}_i$ and $\mathcal{E}_j$ for an arbitrary pair of ellipses in order to be consistent with the notation most frequently encountered in the literature.

Estimating the contact point between all possible configurations of pairs of ellipses is in general not necessary in DEM applications or could involve a number of degenerate cases, such as when the center of mass of one of the ellipses is inside the area of the second ellipse or when one ellipse is virtually penetrating completely through the other. See Figure 3.1 for examples of configurations of pairs of ellipses. Our objective is to restrict ourselves to the configurations (a), (b), and (c) that we encounter in DEM applications, thus avoiding the other degenerate contacts. The first few definitions and lemmas below aim at characterizing such configurations, which we will refer to as *near perfect contact*. When pairs of ellipses are in near perfect contact, then Theorem 9 will show that only a few cases need to be studied.

3.1 Intersection of Ellipses

Definition 2. [**Intersection set**] *Let $\mathcal{E}_i$, $\mathcal{E}_j \subset \mathbb{R}^2$ be two ellipses. Their intersection set is defined as:*

$$\mathcal{I}_{ij} = \mathcal{E}_i \cap \mathcal{E}_j. \tag{3.1}$$

Before proceeding with an analysis of the intersection set, observe that, it is at least formally, computable as the solution to a minimization problem, $\mathcal{I}_{ij} \neq \varnothing$. Although not the only possible formulation, it is nevertheless the most obvious.

Lemma 2. *Given two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ such that $\mathcal{I}_{ij} \neq \varnothing$, with potentials f_i and f_j, respectively, then*

$$\mathcal{I}_{ij} = \underset{\boldsymbol{x}\in\mathbb{R}^2}{\operatorname{argmin}} \left[f_i(\boldsymbol{x})^2 + f_j(\boldsymbol{x})^2 \right]. \tag{3.2}$$

Proof. It is obvious that the minimum of the sum of two positive functions occurs where both functions simultaneously vanish, i.e. at the common roots of f_i and f_j. □

Intuitively, it is easy to imagine the different forms that the intersection set (see Figure 3.1) may take but it is less straightforward to give a complete and thorough description. Bézout's Theorem [64,65] applied to the roots of two quadratic bivariate polynomials in $\mathbb{R}^2$ states that the intersection set $\mathcal{I}_{ij}$ can be either

1. empty: the ellipses are disjoint;
2. one point: the ellipses are in perfect contact (see Definition 4);
3. two, three, or four points: the ellipses intersect;
4. or an entire ellipse, if the two ellipses coincide.

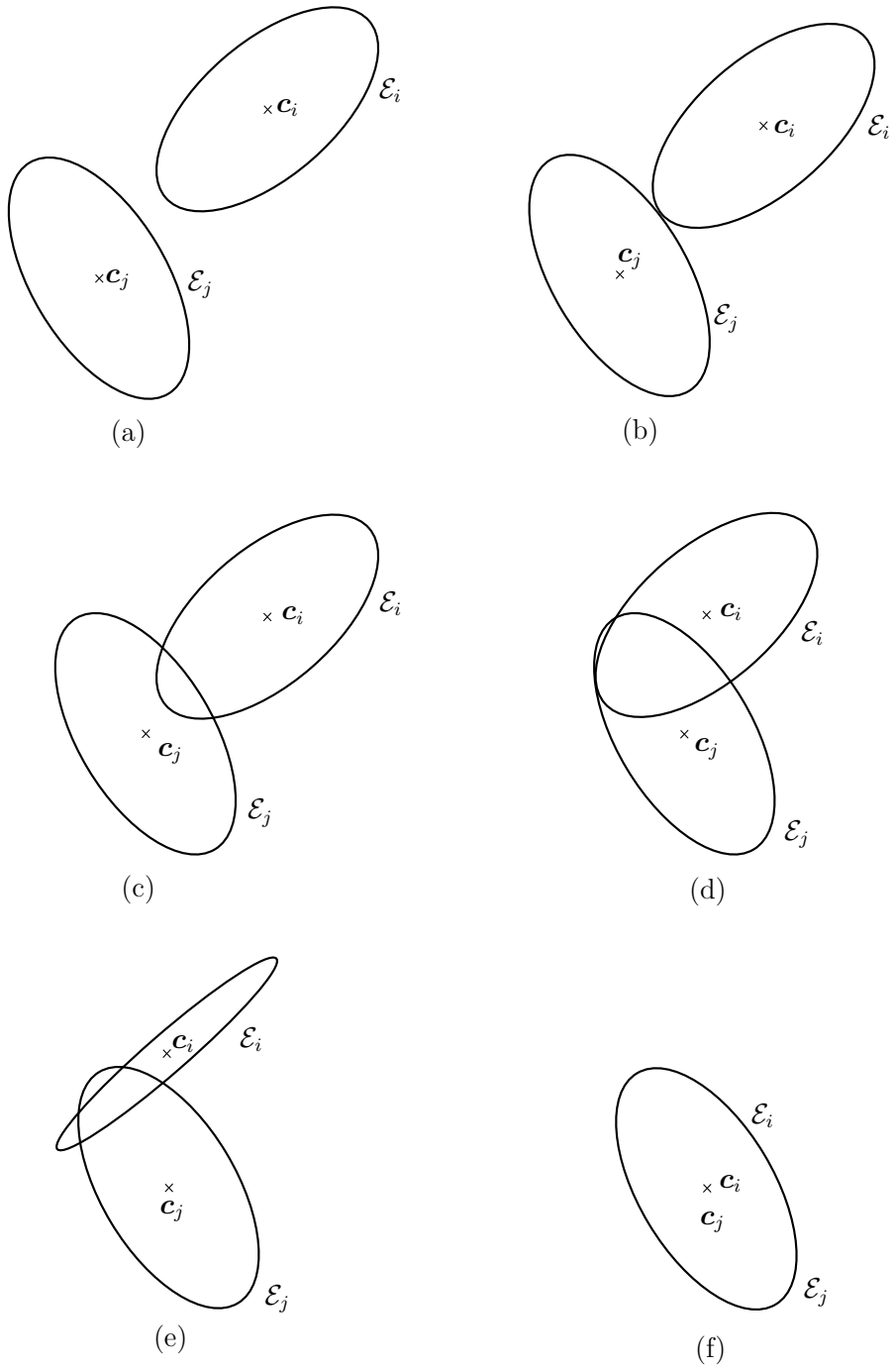


Figure 3.1 Illustration of possible configurations for a pair of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$: (a) disjoint ellipses with no overlap, (b) ellipses in perfect contact (see Definition 4), (c) ellipses with two intersection points, (d) ellipses with three intersection points, (e) ellipses with four intersection points, (f) ellipses coincide. Note that all configurations satisfy non-penetrating CoM (see Definition 3), except (f) and that only the cases (a), (b), and (c) are of interest in DEM applications.

We first state the following definition that will allow us to disregard trivial cases such as when one ellipse lies fully within a second ellipse or when two ellipses coincide.

Definition 3. [Ellipses with non-penetrating centers of mass] *Two ellipses $\mathcal{E}_i$, $\mathcal{E}_j \subset \mathbb{R}^2$ are said to have non-penetrating centers of mass (CoM) if the distances between the centers, evaluated in both the $\mathcal{E}_i$- and $\mathcal{E}_j$-norms* (2.7)*, satisfy*

$$\|\boldsymbol{c}_j - \boldsymbol{c}_i\|_{\mathcal{E}_i} \geq 1, \tag{3.3}$$

$$\|\boldsymbol{c}_i - \boldsymbol{c}_j\|_{\mathcal{E}_j} \geq 1. \tag{3.4}$$

3.2 Case of two Disjoint Ellipses

In this section, we consider the case of two disjoint ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$, i.e. $\mathcal{I}_{ij} = \varnothing$, that satisfy the non-penetrating CoM property, see Definition 3. In this case, there is obviously no contact nor overlap but one can estimate the distance between the two particles. Our objective in doing so is to find formulations of the separation distance that can be extended to the definition of distance, or contact point, when the ellipses are overlapping.

The most obvious and straightforward formulation of the distance between two ellipses, which could naturally be applied to any pair of objects, is characterized in the following lemma.

Lemma 3. [Minimum Distance Pair] *Let $\mathcal{E}_i$ and $\mathcal{E}_j$ be two disjoint ellipses with non-penetrating CoM. Then, there exists a unique pair of points $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ which minimizes the Euclidean norm $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$, i.e.*

$$(\boldsymbol{x}_i, \boldsymbol{x}_j) \;=\; \underset{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j}{\operatorname{argmin}} \|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\| \;=\; \underset{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j}{\operatorname{argmin}} \frac{1}{2}\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\|^2. \tag{3.5}$$

Moreover, at the minimum, the unit normal vectors $\boldsymbol{n}_i(\boldsymbol{x}_i)$ and $\boldsymbol{n}_j(\boldsymbol{x}_j)$ to $\mathcal{E}_i$ and $\mathcal{E}_j$, respectively, are opposite

$$\boldsymbol{n}_i(\boldsymbol{x}_i) + \boldsymbol{n}_j(\boldsymbol{x}_j) = \boldsymbol{0}. \tag{3.6}$$

The pair $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ will be referred to as the minimum distance pair (MDP) of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$.

Proof. The existence of a unique pair $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ that minimizes distance $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ follows from elementary results in linear algebra [66], which we now present.

Consider the convex and compact set $E_k := \left\{\boldsymbol{x} \in \mathbb{R}^2 \,;\; f_k(\boldsymbol{x}) \leq 0\right\}$, $k = i, j$, formed by $\mathcal{E}_k$ and its *interior*. Since $\mathcal{E}_i$ and $\mathcal{E}_j$ are disjoint and have non-penetrating CoM, then $E_i \cap E_j = \varnothing$.

It implies that the set

$$D := \Big\{ \boldsymbol{d} \in \mathbb{R}^2 \,;\; \boldsymbol{d} = \boldsymbol{x}_i - \boldsymbol{x}_j,\ \forall\, (\boldsymbol{x}_i, \boldsymbol{x}_j) \in E_i \times E_j \Big\},$$

is also compact and convex. Hence there exists a unique $\boldsymbol{d} \in D$ minimizing $\|\boldsymbol{d}\|$. In fact, if $\boldsymbol{d} = \boldsymbol{x}_i - \boldsymbol{x}_j$ for $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in E_i \times E_j$, then the pair must belong to $\mathcal{E}_i \times \mathcal{E}_j$. If not, say $\boldsymbol{x}_i \in E_i \setminus \mathcal{E}_i$, then one could always find an ε, $0 < \varepsilon \ll 1$, such that $\boldsymbol{x}_i - \varepsilon \boldsymbol{d} \in E_i$ and the pair $(\boldsymbol{x}_i - \varepsilon \boldsymbol{d}, \boldsymbol{x}_j)$ would define a smaller distance

$$\Big\| (\boldsymbol{x}_i - \varepsilon \boldsymbol{d}) - \boldsymbol{x}_j \Big\| = \Big\| (1 - \varepsilon) \boldsymbol{d} \Big\| < \|\boldsymbol{d}\|.$$

Finally, we observe that the pair $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ must be unique. Indeed, if one can find $(\boldsymbol{y}_i, \boldsymbol{y}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ such that $\boldsymbol{d} = \boldsymbol{x}_i - \boldsymbol{x}_j = \boldsymbol{y}_i - \boldsymbol{y}_j$, then all pairs

$$(1 - \lambda)\Big(\boldsymbol{x}_i, \boldsymbol{x}_j\Big) + \lambda\Big(\boldsymbol{y}_i, \boldsymbol{y}_j\Big), \qquad \forall \lambda \in [0, 1],$$

would also minimize distance

$$\begin{aligned} \Big\| \Big((1 - \lambda)\boldsymbol{x}_i + \lambda \boldsymbol{y}_i \Big) - \Big((1 - \lambda)\boldsymbol{x}_j + \lambda \boldsymbol{y}_j \Big) \Big\| &= \Big\| (1 - \lambda)(\boldsymbol{x}_i - \boldsymbol{x}_j) + \lambda (\boldsymbol{y}_i - \boldsymbol{y}_j) \Big\| \\ &\leq (1 - \lambda)\|\boldsymbol{x}_i - \boldsymbol{x}_j\| + \lambda \|\boldsymbol{y}_i - \boldsymbol{y}_j\| = \|\boldsymbol{d}\|, \end{aligned}$$

and thus, by virtue of the previous result, should lie on the boundaries of $\mathcal{E}_i$ and $\mathcal{E}_j$. However, the points $(1 - \lambda)\boldsymbol{x}_i + \lambda \boldsymbol{y}_i$, (resp. $(1 - \lambda)\boldsymbol{x}_j + \lambda \boldsymbol{y}_j$), $\forall \lambda \in [0, 1]$, form a straight segment and cannot lie on the boundary of $\mathcal{E}_i$ (resp. $\mathcal{E}_j$), since the ellipses are strictly convex. This shows that $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ is unique.

Let f_i and f_j be the global potentials of $\mathcal{E}_i$ and $\mathcal{E}_j$, respectively. In order to show that the unit normal vectors are opposite, we introduce the Lagrangian functional:

$$\mathcal{L}(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j, \lambda_i, \lambda_j) = \frac{1}{2}\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\|^2 - \lambda_i f_i(\hat{\boldsymbol{x}}_i) - \lambda_j f_j(\hat{\boldsymbol{x}}_j),$$

where $\lambda_i \in \mathbb{R}$ and $\lambda_j \in \mathbb{R}$ denote the Lagrange multipliers associated with the constraints $f_i(\hat{\boldsymbol{x}}_i) = 0$ and $f_j(\hat{\boldsymbol{x}}_j) = 0$, i.e. $\hat{\boldsymbol{x}}_i \in \mathcal{E}_i$ and $\hat{\boldsymbol{x}}_j \in \mathcal{E}_j$, respectively, in the minimization problem (3.5). The derivative $\mathcal{L}_{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j)}$ of $\mathcal{L}$ with respect to $(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j)$ is given, $\forall (\boldsymbol{v}_i, \boldsymbol{v}_j) \in \mathbb{R}^2 \times \mathbb{R}^2$, by

$$\mathcal{L}_{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j)}(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j, \lambda_i, \lambda_j, \boldsymbol{v}_i, \boldsymbol{v}_j) = \Big[(\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j) - \lambda_i \nabla f_i(\hat{\boldsymbol{x}}_i) \Big] \cdot \boldsymbol{v}_i + \Big[-(\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j) - \lambda_j \nabla f_j(\hat{\boldsymbol{x}}_j) \Big] \cdot \boldsymbol{v}_j.$$

The solution $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ to (3.5) is a stationary point of $\mathcal{L}$ and must satisfy:

$$\mathcal{L}_{(\boldsymbol{x}_i,\boldsymbol{x}_j)}(\boldsymbol{x}_i, \boldsymbol{x}_j, \lambda_i, \lambda_j, \boldsymbol{v}_i, \boldsymbol{v}_j) = 0, \qquad \forall (\boldsymbol{v}_i, \boldsymbol{v}_j) \in \mathbb{R}^2 \times \mathbb{R}^2,$$

or, equivalently,

$$\begin{aligned}(\boldsymbol{x}_i - \boldsymbol{x}_j) - \lambda_i \nabla f_i(\boldsymbol{x}_i) = \boldsymbol{0},\\ (\boldsymbol{x}_i - \boldsymbol{x}_j) + \lambda_j \nabla f_j(\boldsymbol{x}_j) = \boldsymbol{0}.\end{aligned}$$

Combining those two equations leads to

$$\lambda_i \nabla f_i(\boldsymbol{x}_i) + \lambda_j \nabla f_j(\boldsymbol{x}_j) = \boldsymbol{0},$$

meaning that the gradients $\nabla f_i(\boldsymbol{x}_i)$ and $\nabla f_j(\boldsymbol{x}_j)$ share the same or opposite direction. The fact that $\mathcal{E}_i$ and $\mathcal{E}_j$ are disjoint and satisfy the property of non-penetrating CoM allows one to conclude that (3.6) is satisfied. □

It is worth noting that (3.6) represents a non-binding constraint as it is automatically verified by the solution to the minimization problem. Therefore, (3.5) can be recast as:

$$(\boldsymbol{x}_i, \boldsymbol{x}_j) = \underset{\substack{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j \\ \boldsymbol{n}_i(\hat{\boldsymbol{x}}_i) + \boldsymbol{n}_j(\hat{\boldsymbol{x}}_j) = \boldsymbol{0}}}{\operatorname{argmin}} \|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\| = \underset{\substack{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j \\ \boldsymbol{n}_i(\hat{\boldsymbol{x}}_i) + \boldsymbol{n}_j(\hat{\boldsymbol{x}}_j) = \boldsymbol{0}}}{\operatorname{argmin}} \frac{1}{2}\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\|^2. \tag{3.7}$$

Remark 2. *In the case of two disjoint circles $\mathcal{C}_i$ and $\mathcal{C}_j$, the solution pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ to* (3.5) *are actually aligned with the centers of the circles, $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$. From this observation, one can reformulate the distance $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ as*

$$\|\boldsymbol{x}_i - \boldsymbol{x}_j\| = \|\boldsymbol{x}_i - \boldsymbol{c}_j\| + \|\boldsymbol{x}_j - \boldsymbol{c}_i\| - \|\boldsymbol{c}_i - \boldsymbol{c}_j\|$$

so that the minimization problem (3.5) *can be recast as*

$$\min_{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{C}_i \times \mathcal{C}_j} \Big[\|\hat{\boldsymbol{x}}_i - \boldsymbol{c}_j\| + \|\hat{\boldsymbol{x}}_j - \boldsymbol{c}_i\|\Big] = \min_{\hat{\boldsymbol{x}}_i \in \mathcal{C}_i} \|\hat{\boldsymbol{x}}_i - \boldsymbol{c}_j\| + \min_{\hat{\boldsymbol{x}}_j \in \mathcal{C}_j} \|\hat{\boldsymbol{x}}_j - \boldsymbol{c}_i\|.$$

It follows that the minimization problem can be separated into the fully decoupled minimization problems

$$\begin{aligned}\boldsymbol{x}_i &= \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{C}_i} \|\boldsymbol{x} - \boldsymbol{c}_j\|,\\ \boldsymbol{x}_j &= \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{C}_j} \|\boldsymbol{x} - \boldsymbol{c}_i\|.\end{aligned}$$

In other words, the points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be viewed as the closest points on $\mathcal{C}_i$ and $\mathcal{C}_j$ to

the centers $\boldsymbol{c}_j$ and $\boldsymbol{c}_i$, respectively. Recalling that ellipses can be viewed as circles in their respective $\mathcal{E}$-norms, one can actually introduce similar decoupled minimization problems in the case of ellipses.

Lemma 4. [Minimum Potential Pair] *Let $\mathcal{E}_i$ and $\mathcal{E}_j$ be two disjoint ellipses with non-penetrating CoM with global potentials f_i and f_j, respectively. Then, there exists a unique pair of points $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ satisfying the two problems*

$$\boldsymbol{x}_i = \underset{\boldsymbol{x}\in\mathcal{E}_i}{\operatorname{argmin}} \|\boldsymbol{x} - \boldsymbol{c}_j\|_{\mathcal{E}_j} = \underset{\boldsymbol{x}\in\mathcal{E}_i}{\operatorname{argmin}} f_j(\boldsymbol{x}), \tag{3.8}$$

$$\boldsymbol{x}_j = \underset{\boldsymbol{x}\in\mathcal{E}_j}{\operatorname{argmin}} \|\boldsymbol{x} - \boldsymbol{c}_i\|_{\mathcal{E}_i} = \underset{\boldsymbol{x}\in\mathcal{E}_j}{\operatorname{argmin}} f_i(\boldsymbol{x}). \tag{3.9}$$

Moreover, following the convention (2.29)*, we have*

$$\boldsymbol{n}_i(\boldsymbol{x}_i) + \boldsymbol{n}_j(\boldsymbol{x}_i) = \boldsymbol{0}, \tag{3.10}$$

$$\boldsymbol{n}_i(\boldsymbol{x}_j) + \boldsymbol{n}_j(\boldsymbol{x}_j) = \boldsymbol{0}. \tag{3.11}$$

The unique pair $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ will be referred to as the minimum potential pair (MPP) with respect to the i-norm and the j-norm.

Proof. Since $f_j(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{c}_j\|^2_{\mathcal{E}_j} - 1$, it follows that the two minimization problems in (3.8) are equivalent. The same reasoning implies that the two minimization problems in (3.9) are also equivalent. The demonstration of the existence and uniqueness to the minimization problems (3.8), or (3.9), is similar to the one given in Lemma 3. Consider the problem of minimizing

$$\min_{\boldsymbol{x}\in E_i} \|\boldsymbol{x} - \boldsymbol{c}_j\|_{\mathcal{E}_j}, \tag{3.12}$$

where $E_i = \{\boldsymbol{x} \in \mathbb{R}^2 |\ f_i(\boldsymbol{x}) \leq 0\}$ is compact and strictly convex. Then it is well-known that (3.12) has a unique solution, say $\boldsymbol{x}_i$. As we argued earlier, $\boldsymbol{x}_i$ must in fact belong to the boundary $\mathcal{E}_i$ and is unique because E_i is strictly convex.

The Lagrangian functional associated with the constrained minimization problem (3.8) is given by

$$\mathcal{L}_i(\boldsymbol{x}, \lambda) = f_j(\boldsymbol{x}) - \lambda f_i(\boldsymbol{x}).$$

Since the solution $\boldsymbol{x}_i$ to (3.8) is a stationary point of $\mathcal{L}_i$, it necessarily satisfies

$$\nabla f_j(\boldsymbol{x}_i) - \lambda \nabla f_i(\boldsymbol{x}_i) = 0,$$

which, using the fact that the two ellipses are disjoint and have non-penetrating CoM, implies

that the two normals at $\boldsymbol{x}_i$ are in the same or opposite direction. As we observe in the Figure 3.2, for only $\boldsymbol{x}_i$ which is the solution to (3.10), we have

$$\boldsymbol{n}_i(\boldsymbol{x}_i) + \boldsymbol{n}_j(\boldsymbol{x}_i) = \boldsymbol{0}.$$

The relation (3.11) is shown in the same manner by introducing the Lagrangian functional $\mathcal{L}_j$ associated with the minimization problem (3.9). □

Since the relations (3.10) and (3.11) are satisfied at the points of the MPP $(\boldsymbol{x}_i, \boldsymbol{x}_j)$, they can each be added to the minimization problems (3.8) and (3.9), respectively, as non-binding constraints so that the two problems can be recast as

$$\boldsymbol{x}_i = \underset{\substack{\boldsymbol{x} \in \mathcal{E}_i \\ \boldsymbol{n}_i(\boldsymbol{x}) + \boldsymbol{n}_j(\boldsymbol{x}) = \boldsymbol{0}}}{\operatorname{argmin}} \|\boldsymbol{x} - \boldsymbol{c}_j\|_{\mathcal{E}_j} = \underset{\substack{\boldsymbol{x} \in \mathcal{E}_i \\ \boldsymbol{n}_i(\boldsymbol{x}) + \boldsymbol{n}_j(\boldsymbol{x}) = \boldsymbol{0}}}{\operatorname{argmin}} f_j(\boldsymbol{x}), \tag{3.13}$$

and

$$\boldsymbol{x}_j = \underset{\substack{\boldsymbol{x} \in \mathcal{E}_j \\ \boldsymbol{n}_i(\boldsymbol{x}) + \boldsymbol{n}_j(\boldsymbol{x}) = \boldsymbol{0}}}{\operatorname{argmin}} \|\boldsymbol{x} - \boldsymbol{c}_i\|_{\mathcal{E}_i} = \underset{\substack{\boldsymbol{x} \in \mathcal{E}_j \\ \boldsymbol{n}_i(\boldsymbol{x}) + \boldsymbol{n}_j(\boldsymbol{x}) = \boldsymbol{0}}}{\operatorname{argmin}} f_i(\boldsymbol{x}). \tag{3.14}$$

Before proceeding with the other cases, we will make a few remarks on the solution to Problem (3.8). One classical approach for solving the constrained minimization problem proceeds by means of Lagrange multipliers, as seen earlier. However, the resulting problem could lead to several solutions as the nonlinear Lagrangian functional may have up to four critical points depending on the configuration and size of the ellipses. In other words, the solutions correspond to local minima and maxima of the potential function f_j restricted to $\mathcal{E}_i$. This is exemplified in Figure 3.2. In practice, all of the known methods identify all of the critical points and distinguish the global minimum by explicitly evaluating the distance at each critical point.

It is worth noting here that the non-binding constraint in the minimization problem (3.13) has the added benefit of yielding a Lagrangian functional with a unique critical point. Indeed, the constraint (3.10) is only satisfied at the global minimum. Alternatively, one may enforce the uniqueness of the critical point by considering the inequality constraint $\boldsymbol{n}_i(\boldsymbol{x}) \cdot \boldsymbol{n}_j(\boldsymbol{x}) < 0$ or $\nabla f_i(\boldsymbol{x}) \cdot \nabla f_j(\boldsymbol{x}) < 0$.

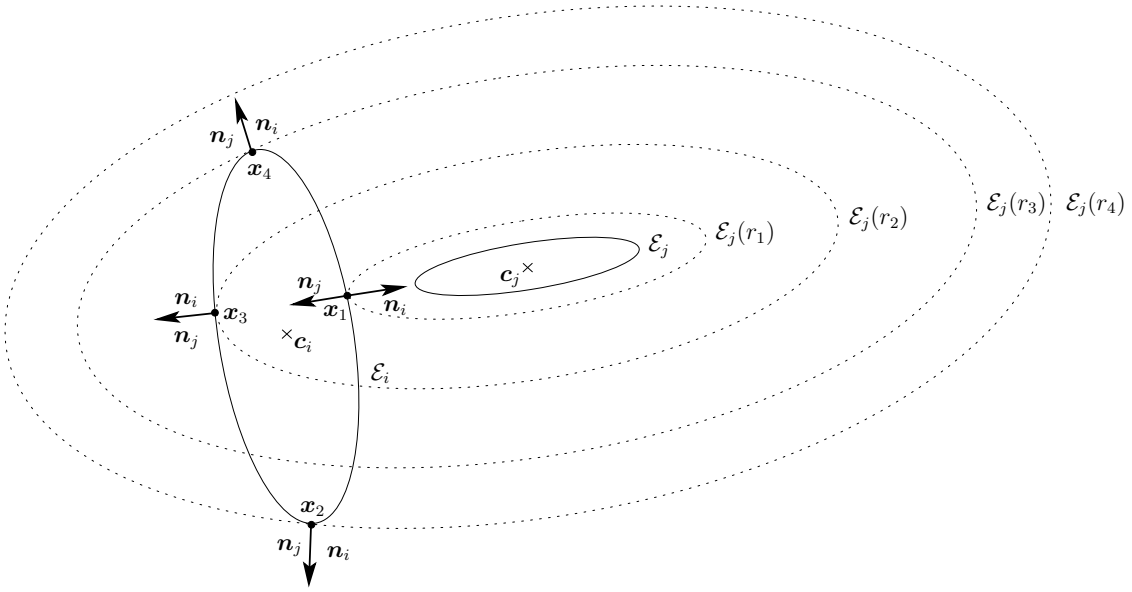


Figure 3.2 Illustration of four critical points of Problem (3.8), with $\boldsymbol{x}_k$, $k = 1, \cdots, 4$. We can observe that $\boldsymbol{n}_i(\boldsymbol{x}_k) + \boldsymbol{n}_j(\boldsymbol{x}_k) = \boldsymbol{0}$ only if $k = 1$.

3.3 Case of two Ellipses in Perfect Contact

The case of perfect contact between ellipses with non-penetrating CoM can be viewed as a limiting case of two disjoint ellipses. Therefore, we shall see that the previous results straightforwardly apply to this particular case.

Definition 4. [Perfect contact point] *Two ellipses $\mathcal{E}_i$, $\mathcal{E}_j \subset \mathbb{R}^2$ with non-penetrating CoM are said to be in perfect contact if $\mathcal{I}_{ij}$ consists of a single point. That point is then called a perfect contact point.*

Lemma 5. *Let $\mathcal{E}_i$ and $\mathcal{E}_j$ be two ellipses in perfect contact at point $\boldsymbol{x}_c$ with non-penetrating CoM. Moreover, let $\boldsymbol{n}_i(\boldsymbol{x}_c)$ and $\boldsymbol{n}_j(\boldsymbol{x}_c)$ denote the outward normal unit vectors at $\boldsymbol{x}_c$ to $\mathcal{E}_i$ and $\mathcal{E}_j$, respectively. Then, the pair $(\boldsymbol{x}_c, \boldsymbol{x}_c)$ is the MDP and MPP of the two ellipses. Moreover, it holds that:*

$$\boldsymbol{n}_i(\boldsymbol{x}_c) + \boldsymbol{n}_j(\boldsymbol{x}_c) = \boldsymbol{0}. \tag{3.15}$$

Proof. We first show that $(\boldsymbol{x}_c, \boldsymbol{x}_c)$ is the MDP. If $\boldsymbol{x}_c$ is a perfect contact point, then for all $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$, such that $(\boldsymbol{x}_i, \boldsymbol{x}_j) \neq (\boldsymbol{x}_c, \boldsymbol{x}_c)$, the distance $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| > 0$. Hence $(\boldsymbol{x}_c, \boldsymbol{x}_c)$ is the unique solution to (3.5). To show that $(\boldsymbol{x}_c, \boldsymbol{x}_c)$ is the MPP, we observe that for any point $\boldsymbol{x} \in \mathcal{E}_i \setminus \{\boldsymbol{x}_c\}$ then $\boldsymbol{x} \notin \mathcal{E}_j$ and therefore $f_j(\boldsymbol{x}) > 0$. Hence

$$1 = \|\boldsymbol{x}_c - \mathbf{c}_j\|_{\mathcal{E}_j} < \|\boldsymbol{x} - \mathbf{c}_j\|_{\mathcal{E}_j}.$$

This shows that $\boldsymbol{x}_c$ is the unique solution to (3.8), and, in a similar manner, $\boldsymbol{x}_c$ is also the unique solution to (3.9). Finally, the relation (3.15) is clearly a consequence of (3.10) and (3.11) when $\boldsymbol{x}_c = \boldsymbol{x}_i = \boldsymbol{x}_j$. □

3.4 Case of two Ellipses with Overlap

The case of overlapping ellipses with non-penetrating CoM is more difficult to analyze than the previous two cases for the simple reason that the intersection set $\mathcal{I}_{ij}$ may consist of two, three, or even four points. However, in applications dealing with assemblies of ellipses, one is usually concerned with pairs of particles whose configurations can be viewed as perturbations of particles in perfect contact. In DEM applications, for instance, ellipses are only allowed to slightly overlap, meaning that the $\mathcal{I}_{ij}$ would consist of only two points. The main goal in this section is to rigorously define the notion of small overlaps in order to clearly discard the other two cases where the intersection set $\mathcal{I}_{ij}$ consists of three or four points.

The analysis in the previous two sub-sections has highlighted the importance of the relationship that the gradients of the potential functions associated with ellipses/ellipsoids satisfy at the contact point, in case of perfect contact, or at the MPP, in case of disjoint particles. We are now in a position to introduce an important locus in $\mathbb{R}^2$, which we shall refer to as the co-gradient locus. The locus is actually equivalent to the *locus of common slope*, which was first introduced, to the best of our knowledge, in [46, 67]. However, the main issue with the locus of common slope is that it does not straightforwardly extend to the 3-D case, while the one given below does.

Definition 5. [Co-gradient function and co-gradient locus] *Given two ellipses $\mathcal{E}_i$, $\mathcal{E}_j \subset \mathbb{R}^2$, the co-gradient function is defined as*

$$\boldsymbol{H}(\boldsymbol{x}) := \nabla f_i(\boldsymbol{x}) \times \nabla f_j(\boldsymbol{x}). \tag{3.16}$$

The associated co-gradient locus is the set of all roots of the co-gradient function, i.e.

$$\mathcal{H}_{ij} := \Big\{ \boldsymbol{x} \in \mathbb{R}^2 \,;\; \boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{0} \Big\}. \tag{3.17}$$

In 2-D, the cross product defining the co-gradient function is interpreted as a cross-product in 3-D between the gradients in the 2-D plane. It therefore results in a 3-D vector with only one non-zero component along the z-axis, which is given by the scalar function

$$H(\boldsymbol{x}) := \det \Big[\nabla f_i(\boldsymbol{x}) \quad \nabla f_j(\boldsymbol{x}) \Big] = \partial_x f_i(\boldsymbol{x}) \partial_y f_j(\boldsymbol{x}) - \partial_y f_i(\boldsymbol{x}) \partial_x f_j(\boldsymbol{x}). \tag{3.18}$$

We shall consider this definition of the co-gradient function when dealing with ellipses rather than the vector-valued $\boldsymbol{H}$ in (3.16). Introducing the anti-symmetric matrix,

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{3.19}$$

the co-gradient function can also be written as:

$$H(\boldsymbol{x}) = (\nabla f_i(\boldsymbol{x}))^T A \nabla f_j(\boldsymbol{x}) = 4(\boldsymbol{x} - \boldsymbol{c}_i)^T \mathcal{Q}_i A \mathcal{Q}_j (\boldsymbol{x} - \boldsymbol{c}_j), \tag{3.20}$$

where we have used the fact that $\mathcal{Q}_i$ is symmetric. We immediately observe that H is a quadratic polynomial in $\boldsymbol{x}$ and that the centers $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ of the ellipses belong to $\mathcal{H}_{ij}$. If the product $\mathcal{Q}_i A \mathcal{Q}_j$ was symmetric, then the determinant of the product could immediately tells us the geometry of the co-gradient locus. Unfortunately, the detailed characterization of $\mathcal{H}_{ij}$ presented in Theorem 6 will require significantly more work.

A second characterization can be made by normalizing the gradients in (3.16). Recalling the definition of the unit normal vectors (2.16), i.e.

$$\boldsymbol{n}_k(\boldsymbol{x}) = \frac{\nabla f_k(\boldsymbol{x})}{\|\nabla f_k(\boldsymbol{x})\|}, \qquad k = i, j, \tag{3.21}$$

then the *normalized co-gradient function* is

$$\hat{\boldsymbol{H}}(\boldsymbol{x}) = \boldsymbol{n}_i(\boldsymbol{x}) \times \boldsymbol{n}_j(\boldsymbol{x}). \tag{3.22}$$

The scalar component in the z-direction of $\hat{\boldsymbol{H}}(\boldsymbol{x}) \in \mathbb{R}^3$ is equal to $\sin \eta_{ij}(\boldsymbol{x})$ where $\eta_{ij}(\boldsymbol{x})$ is the angle between $\boldsymbol{n}_i$ and $\boldsymbol{n}_j$, well-defined modulo 2π. Mimicking the definition (3.20) of the z-component of $\hat{\boldsymbol{H}}$, the co-gradient locus in 2-D is simply the set of roots of

$$\hat{H}(\boldsymbol{x}) = \sin \eta_{ij}(\boldsymbol{x}). \tag{3.23}$$

The angle $\eta_{ij}(\boldsymbol{x})$ can also be defined by identifying $\boldsymbol{n}_i$ and $\boldsymbol{n}_j$ with unitary complex numbers, so that, using complex multiplication

$$\boldsymbol{n}_j(\boldsymbol{x}) = e^{i\eta_{ij}(\boldsymbol{x})} \boldsymbol{n}_i(\boldsymbol{x}). \tag{3.24}$$

The roots of $\hat{H}$ correspond to $\eta_{ij}(\boldsymbol{x}) = m\pi$, $m \in \mathbb{Z}$. We note that at infinity, the relation (3.24) can be rewritten as $\mathcal{N}_j(\boldsymbol{w}) = e^{i\eta_{ij}(\boldsymbol{w})} \mathcal{N}_i(\boldsymbol{w})$ using Equation (2.32), that is the angle $\eta_{ij} = \eta_{ij}(\boldsymbol{w})$ only depends on the direction $\boldsymbol{w}$. Eventually, we will show that the angle η_{ij}

must belong to $]-\pi,\pi[$, and hence is well-defined.

We now characterize the co-gradient locus in the case of arbitrary pairs of ellipses with non-penetrating CoM.

Theorem 6. [Co-gradient locus] *Let $\mathcal{E}_i$, $\mathcal{E}_j \subset \mathbb{R}^2$ be two ellipses with non-penetrating CoM. Then the co-gradient locus $\mathcal{H}_{ij}$ is a hyperbola and the two centers of the ellipses belong to only one branch of the hyperbola. The portion of $\mathcal{H}_{ij}$ between $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ can be parameterized by a smooth injection $\boldsymbol{\gamma}_{ij} : [0,1] \longrightarrow \mathcal{H}_{ij}$ satisfying*

$$\begin{cases} \boldsymbol{\gamma}_{ij}(0) = \boldsymbol{c}_i, \\ \boldsymbol{\gamma}_{ij}(1) = \boldsymbol{c}_j. \end{cases} \tag{3.25}$$

Moreover, there exists a unique pair of parameters $t_k \in [0,1]$, for $k = i, j$, such that

$$\boldsymbol{\gamma}_{ij}(t_k) \in \mathcal{E}_k. \tag{3.26}$$

Proof. The proof will show that the co-gradient function H, which the formula (3.20) shows is a quadratic function, possesses four roots at infinity. This will imply that the roots $\mathcal{H}_{ij} \subset \mathbb{R}^2$ form a hyperbola because ellipses, parabolas and hyperbolas possesses respectively $0, 2$ and 4 roots at infinity on the projective sphere. Afterwards, we will argue that a single branch of the hyperbola must cross both centers of the ellipses, thereby justifying the existence of the parametrization.

The majority of the analysis will be performed on a pair of ellipses in a *generic* configuration but this will require us to begin the proof with a lengthy study of different *degenerate* configurations. Bivariate quadratic polynomials have roots that can degenerate to either of the following configurations: two intersecting lines, two parallel lines, a line with a second line at infinity, two coincident lines, and a single point. The last option will never occur because H already vanishes at the centers $\boldsymbol{c}_i \neq \boldsymbol{c}_j$. The analysis below will show that $\mathcal{H}_{ij}$ always contains at least four points at infinity, and hence cannot be formed of two parallel lines or two coincident lines.

The first configuration we study assumes that $\boldsymbol{c}_j$ belongs to the axis $\boldsymbol{\xi}_i$ and that the principal axes of $\mathcal{E}_j$ are aligned with those of $\mathcal{E}_i$, although the argument will also work if $\boldsymbol{c}_j$ belongs to the axes $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_i = \boldsymbol{\xi}_j$. Under these conditions, for all $t, s \in \mathbb{R}$, property *ii)* of Lemma 1 shows that

$$\boldsymbol{n}_i(\boldsymbol{c}_i + t\boldsymbol{\xi}_i) = \operatorname{sign}(t)\boldsymbol{\xi}_i = \pm\boldsymbol{\xi}_j = \pm\boldsymbol{n}_j(\boldsymbol{c}_j + s\boldsymbol{\xi}_j)\,.$$

Hence every point of the axis $\boldsymbol{c}_i + t\boldsymbol{\xi}_i$ belongs to $\mathcal{H}_{ij}$. Furthermore, the fact that the axis are

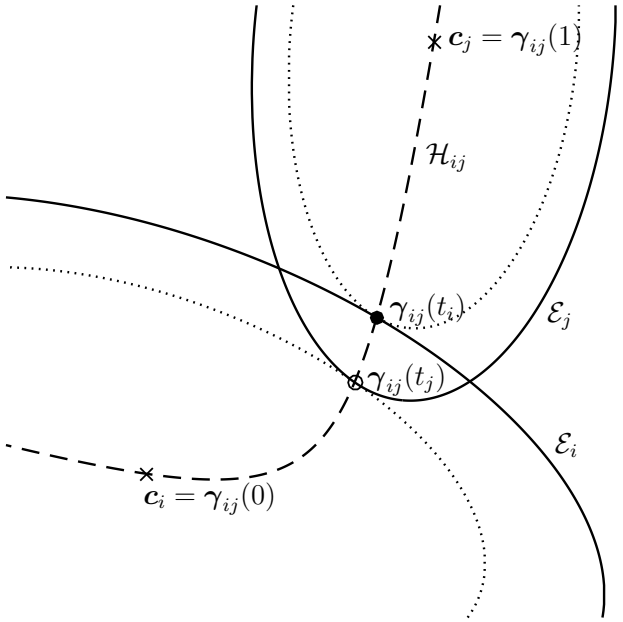


Figure 3.3 Illustration of the smooth injection $\boldsymbol{\gamma}_{ij}$ onto the gradient locus $\mathcal{H}_{ij}$, as formulated in Theorem 6. The smooth injection $\boldsymbol{\gamma}_{ij}$ from Theorem 6 is a single component of the hyperbola $\mathcal{H}_{ij}$ which passes through both centers $\boldsymbol{c}_i$ ad $\boldsymbol{c}_j$.

aligned and property *iv)-(a)* of Lemma 1 implies that

$$\mathcal{N}_i(\boldsymbol{\eta}_i) = \mathcal{N}_j(\boldsymbol{\eta}_j)\,,$$

or in other words H possesses roots at infinity in the direction $\boldsymbol{w} = \pm\boldsymbol{\eta}_i = \pm\boldsymbol{\eta}_j$. In this configuration, the co-gradient locus has four roots at infinity and contains a line, and hence is either two intersecting lines, or a line with a second line at infinity. If the two ellipses have the same aspect ratio, then the relation (2.31) of Lemma 1 states that $\mathcal{N}_i(\boldsymbol{w}) = \mathcal{N}_j(\boldsymbol{w})$, for all $\boldsymbol{w} \in S^1$. In other words, $\mathcal{H}_{ij}$ contains the line at infinity. On the other hand, when the aspect ratios are different then the same relation shows that $\mathcal{N}_i \neq \mathcal{N}_j$ and the co-gradient locus is formed of two intersecting lines.

Consider now the case where $\boldsymbol{c}_j$ does not belong to either principal axes of $\mathcal{E}_i$ but continue to assume that the principal axes of both ellipses are aligned, say $\boldsymbol{\xi}_i = \boldsymbol{\xi}_j$ and $\boldsymbol{\eta}_i = \boldsymbol{\eta}_j$. Property *iv)-(a)* of Lemma 1 tells us that

$$\mathcal{N}_i(\pm\boldsymbol{\xi}_i) = \mathcal{N}_j(\pm\boldsymbol{\xi}_j)\,, \qquad \mathcal{N}_i(\pm\boldsymbol{\eta}_i) = \mathcal{N}_j(\pm\boldsymbol{\eta}_j)\,,$$

hence there are at least four roots at infinity. Again, if the ellipses have the same aspect ratio, then $\mathcal{N}_i = \mathcal{N}_j$ and the co-gradient locus is degenerate and contains a line at infinity.

Otherwise, the co-gradient locus is a hyperbola, which or may not be degenerate; see Corollary 7 for more on this issue.

The general configuration on which we will focus the remainder of our attention assumes that the principal axes of the two ellipses are not aligned, whether or not either center belongs to the axis of its brethren. In this case, we observe that the straight line $\boldsymbol{c}_j + t\boldsymbol{\xi}_j$, $t \in \mathbb{R}$, for $|t|$ sufficiently large, belongs to two opposing quadrants; either $[\boldsymbol{\xi}_i, \boldsymbol{\eta}_i]$ and $[-\boldsymbol{\xi}_i, -\boldsymbol{\eta}_i]$ or $[\boldsymbol{\eta}_i, -\boldsymbol{\xi}_i]$ and $[-\boldsymbol{\eta}_i, \boldsymbol{\xi}_i]$. When the second case occurs, then the line $t\boldsymbol{\xi}_i$ crosses the opposing quadrants $[\boldsymbol{\xi}_j, \boldsymbol{\eta}_j]$ and $[-\boldsymbol{\xi}_j, -\boldsymbol{\eta}_j]$. Hence, the second case can be brought into the first configuration by translating $\mathcal{E}_j$ to the origin and exchanging the indices i and j. The Figure 3.4 illustrates this configuration, after assuming a translation and a rotation sending $\boldsymbol{c}_i$ to the origin and the axes of $\mathcal{E}_i$ over to the usual Cartesian axes.

The map (2.30) associates to each direction $\boldsymbol{w} \in S^1$, the unique normals $\mathcal{N}_i(\boldsymbol{w})$ and $\mathcal{N}_j(\boldsymbol{w})$ on the line at infinity. We may then measure the angle $\eta_{ij} = \eta_{ij}(\boldsymbol{w})$ between the normals at infinity using the relation (3.24), rewritten here using complex multiplication as

$$\mathcal{N}_j(\boldsymbol{w}) = e^{i\eta_{ij}(\boldsymbol{w})}\mathcal{N}_i(\boldsymbol{w}), \tag{3.27}$$

In this last identity, positive or negative angles correspond respectively to a counter-clockwise or clockwise rotation when rotating $\mathcal{N}_i$ towards $\mathcal{N}_j$. It is essential to observe that the angle η_{ij} is well-defined within $]-\pi, \pi[$ because property *iv)-(c)* of Lemma 1 states that the rotation θ from $\boldsymbol{w}$ to either $\mathcal{N}_i(\boldsymbol{w})$ or $\mathcal{N}_j(\boldsymbol{w})$ is strictly bounded $|\theta| < \pi/2$, and hence the rotation from $\mathcal{N}_i$ to $\mathcal{N}_j$ must be by an angle η_{ij} strictly less than π in absolute value. As $\boldsymbol{w}$ moves counter-clockwise around S^1 starting at $\boldsymbol{\xi}_i$, the configuration we have chosen, as shown in Figure 3.4, implies that we will encounter in order the directions $\boldsymbol{\xi}_i$, $\boldsymbol{\xi}_j$, $\boldsymbol{\eta}_i$, $\boldsymbol{\eta}_j$, $-\boldsymbol{\xi}_i$, $-\boldsymbol{\xi}_j$, $-\boldsymbol{\eta}_i$, $-\boldsymbol{\eta}_j$, $\boldsymbol{\xi}_i$. We will focus on demonstrating that η_{ij} possesses a root inside the arc $[\boldsymbol{\xi}_j, \boldsymbol{\eta}_i] \subset S^1$, but similar arguments will show that there are at least three other roots, one in each of the three arcs $[\boldsymbol{\eta}_j, -\boldsymbol{\xi}_i]$, $[-\boldsymbol{\xi}_j, -\boldsymbol{\eta}_i]$, and $[-\boldsymbol{\eta}_j, -\boldsymbol{\xi}_i]$. Each root of η_{ij} corresponds to equal normals and hence to a root of H, thereby demonstrating that $\mathcal{H}_{ij}$ is a hyperbola.

Consider the principal axes for $\mathcal{E}_j$ centered at $\boldsymbol{c}_j \neq \boldsymbol{c}_i$, i.e. take $\boldsymbol{w} = \boldsymbol{\xi}_j$. Then the estimate (2.32) and property *iv)-(d)* of Lemma 1 show that the unit normal $\boldsymbol{n}_j(\boldsymbol{c}_i + r\boldsymbol{\xi}_j)$ converges to $\mathcal{N}_j(\boldsymbol{\xi}_j) = \boldsymbol{\xi}_j$ as the radius r increases. Furthermore, if $\boldsymbol{\xi}_j = e^{i\sigma_j}\boldsymbol{\xi}_i$ for $\sigma_j \in [0, \pi/2[$, then property *iv)-(c)* of Lemma 1 states that

$$\mathcal{N}_i(\boldsymbol{\xi}_j) = e^{i\theta_i}\boldsymbol{\xi}_j\,,$$

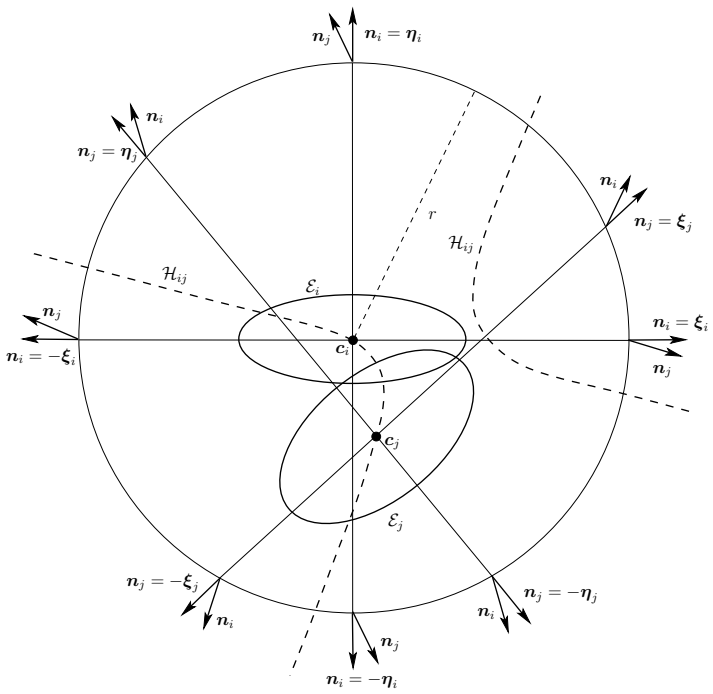


Figure 3.4 Illustration of the proof of Theorem 6. The ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ are in a configuration with $\boldsymbol{c}_i$ at the origin and $\mathcal{E}_i$ is aligned with horizontal axis. The circle with radius r is large enough that all normals are external on the ellipses, i.e. $\boldsymbol{n}_i \cdot \boldsymbol{n}_j > 0$.

with

$$\tan(\theta_i + \sigma_j) = \left(\frac{a_i}{b_i}\right)^2 \tan \sigma_j \,.$$

Given that $\tan \sigma_j > 0$ and $a_i/b_i > 1$, we find that $\theta_i > 0$ and therefore must belong to $[0, \pi/2[$. Using these facts and the estimate (2.32), we find

$$\mathcal{N}_i(\boldsymbol{\xi}_j) = e^{i\theta_i}\boldsymbol{\xi}_j = e^{i\theta_i}\mathcal{N}_j(\boldsymbol{\xi}_j) = e^{i\theta_i}e^{i\eta_{ij}}\mathcal{N}_i(\boldsymbol{\xi}_j).$$

This implies that $\eta_{ij}(\boldsymbol{\xi}_j) = -\theta_i \leq 0$. On the other hand, we have $\mathcal{N}_i(\boldsymbol{\eta}_i) = \boldsymbol{\eta}_i$ while Inequality (2.31) states that

$$\mathcal{N}_j(\boldsymbol{\eta}_i) = e^{i\theta_j}\boldsymbol{\eta}_i = e^{i\theta_j}\mathcal{N}_i(\boldsymbol{\eta}_i) = e^{i\theta_j}e^{-i\eta_{ij}}\mathcal{N}_j(\boldsymbol{\eta}_i)\,,$$

with $\theta_j \in [0, \pi/2[$ following the previous argument. Hence, $\eta_{ij}(\boldsymbol{\eta}_i) = \theta_j \geq 0$. Since $\eta_{ij}(\boldsymbol{w}) \in \,]-\pi, \pi[$ and changes sign as the direction $\boldsymbol{w}$ varies from $\boldsymbol{\xi}_j$ to $\boldsymbol{\eta}_i$, there exists a direction $\boldsymbol{w}$ in $]\boldsymbol{\xi}_j, \boldsymbol{\eta}_i[$ where η_{ij} vanishes and both normals are equal. The same argument applied to the three other intervals, shows that there are four directions at infinity where the normals coincide, i.e. the co-gradient function H has four roots at infinity.

To conclude the proof, we need to show that a single connected component of the hyperbola crosses the centers of the two ellipses. It is easy to verify that $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ belong to $\mathcal{H}_{ij}$ by substituting directly into (3.20). Consider the function $\alpha(\boldsymbol{x}) = \boldsymbol{n}_i(\boldsymbol{x}) \cdot \boldsymbol{n}_j(\boldsymbol{x})$ for all $\boldsymbol{x}$ belonging to the smooth affine variety $\mathcal{H}_{ij}$. By construction, α only takes on the values ± 1 and is ill-defined at both centers. Since α is continuous on $\mathcal{H}_{ij} \setminus \{\boldsymbol{c}_i, \boldsymbol{c}_j\}$, it must have constant values over each connected component of $\mathcal{H}_{ij} \setminus \{\boldsymbol{c}_i, \boldsymbol{c}_j\}$. Furthermore, for points $\boldsymbol{x}$ far from the centers, the normals are related by $\boldsymbol{n}_j = e^{i\eta_{ij}}\boldsymbol{n}_i$ where $\eta_{ij} \in]-\pi, \pi[$ and therefore the dot product must take positive values, i.e. $\alpha(\boldsymbol{x}) = +1$. If the points $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ belong to different branches of the hyperbola $\mathcal{H}_{ij}$, then each connected component of $\mathcal{H}_{ij} \setminus \{\boldsymbol{c}_i, \boldsymbol{c}_j\}$ reaches infinity and α must be equal to $+1$ everywhere. Yet, in the neighborhood of a center, say $\boldsymbol{c}_i$, the normal $\boldsymbol{n}_j$ varies smoothly, and along the tangent to $\mathcal{H}_{ij}$ at $\boldsymbol{c}_i$, property *ii)* of Lemma 1 states that the normals $\boldsymbol{n}_i$ are equal and opposite on both sides of $\boldsymbol{c}_i$. Hence, the function α must take opposite values, i.e. $+1$ and -1, when passing through the center $\boldsymbol{c}_i$ along the branch of the hyperbola. This would contradict the conclusion that α is identically $+1$ on $\mathcal{H}_{ij} \setminus \{\boldsymbol{c}_i, \boldsymbol{c}_j\}$, following from the hypothesis that $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ belong to different branches of $\mathcal{H}_{ij}$. The existence of the parameterization γ_{ij} is a trivial consequence of the fact that both centers belong to the same branch of the hyperbola. □

Corollary 7. *Two ellipses $\mathcal{E}_i, \mathcal{E}_j \subset \mathbb{R}^2$ with non-penetrating CoM are parameterized by an open subset of $\left(\mathbb{R}^2 \times \mathrm{SPD}_2(\mathbb{R})\right)^2$, a manifold of dimension 10. The hyperbola $\mathcal{H}_{ij}$ only degenerates for a subset of codimension 3. When the hyperbola $\mathcal{H}_{ij}$ degenerates, then it can only be either two intersecting lines, or a line with a second line at infinity. The line at infinity can only appear if the principal axes of both ellipses are aligned and have the same aspect ratio.*

Proof. The analysis in Lemma 6 has already shown that the degenerate quadratic cannot be formed of a single point ($\boldsymbol{c}_i, \boldsymbol{c}_j \in \mathcal{H}_{ij}$), two parallel lines or two coincident lines ($\mathcal{H}_{ij}$ at infinity always has ≥ 4 points). The only two remaining possibilities are those mentioned in the statement of the Corollary. To complete the proof we will first deduce the conditions required for $\mathcal{H}_{ij}$ to contain a line at infinity. Afterwards, we will identify the conditions under which the co-gradient locus contains at least one line.

If the ellipses have aligned axes, then the relation (2.31) clearly implies that the co-gradient locus possesses a line at infinity if and only if the aspect ratios are the same. Suppose now that the axes are not aligned but that the co-gradient locus still possesses a line at infinity. We will show that these hypothesis lead to a contradiction. Considering the axis directions as elements of S^1, suppose that $\boldsymbol{\xi}_j \in]\boldsymbol{\xi}_i, \boldsymbol{\eta}_i]$ although similar arguments will apply if it belonged

to $]\boldsymbol{\eta}_i, -\boldsymbol{\xi}_i[$. Given that $\mathcal{H}_{ij}$ possesses a line at infinity, we have that $\mathcal{N}_i = \mathcal{N}_j$ and in particular

$$\mathcal{N}_i(\boldsymbol{\xi}_j) = \mathcal{N}_j(\boldsymbol{\xi}_j) = \boldsymbol{\xi}_j\,, \qquad \mathcal{N}_j(\boldsymbol{\xi}_i) = \mathcal{N}_i(\boldsymbol{\xi}_i) = \boldsymbol{\xi}_i\,. \tag{3.28}$$

The map $\mathcal{N}_i$ cannot be the identity map, or else both ellipses would be circles and hence have aligned axes. Hence, for $\boldsymbol{w} = e^{i\sigma}\boldsymbol{\xi}_i$ with $\sigma \in]0, \pi/2[$, there exists $\theta \in]0, \pi/2[$ satisfying

$$\mathcal{N}(e^{i\sigma}\boldsymbol{\xi}_i) = e^{i(\theta+\sigma)}\boldsymbol{\xi}_i\,, \qquad \text{with } \tan(\theta+\sigma) = (a_i/b_i)^2 \tan\sigma\,.$$

The first relation in (3.28) for $\boldsymbol{\xi}_j = e^{i\sigma}\boldsymbol{\xi}_i$ gives us

$$\boldsymbol{\xi}_j = \mathcal{N}_i(\boldsymbol{\xi}_j) = \mathcal{N}_i(e^{i\sigma}\boldsymbol{\xi}_i) = e^{i(\theta+\sigma)}\boldsymbol{\xi}_i = e^{i\theta}\boldsymbol{\xi}_j\,,$$

that is $\theta = 0$. The implicit relationship (2.31) between θ and σ shows that θ can vanish only if $a_i/b_i = 1$. Repeating the same argument with the second relation in (3.28) proves that $a_j/b_j = 1$. In conclusion, if the axes are not aligned but $\mathcal{H}_{ij}$ possesses a line at infinity then both ellipses are circles.

We now attempt to determine the most general conditions under which the co-gradient locus can degenerate to a pair of intersecting lines. If we examine the values of the continuous function $\alpha(\boldsymbol{x}) = \boldsymbol{n}_i(\boldsymbol{x}) \cdot \boldsymbol{n}_j(\boldsymbol{x})$ along $\mathcal{H}_{ij}$, we notice that it only takes on the values ± 1. During the proof of Theorem 6, we observed that α changes sign as we cross either center but that α on co-gradient locus always took the value $\alpha(\boldsymbol{x}) = 1$ when $\|\boldsymbol{x}\|$ was sufficiently large. This implies that if $\mathcal{H}_{ij}$ degenerates to two intersecting lines, then the two centers cannot belong to different lines, and when they do, the second line cannot cross the first between the two centers.

Consider the line connecting both centers, which can be parametrized as $\boldsymbol{c}_i + t\boldsymbol{w}$, for $t \in]0, \tau[$, with $\boldsymbol{w} = (\boldsymbol{c}_j - \boldsymbol{c}_i)/\|\boldsymbol{c}_j - \boldsymbol{c}_i\|$ so that $\tau = \|\boldsymbol{c}_j - \boldsymbol{c}_i\|$. Along this segment, property *iii)* of Lemma 1 states that the normals are constant. Using properties *ii)* and *iv)-(c)* of Lemma 1 only on the portion of the line between both centers, we compute

$$\begin{aligned}
&\boldsymbol{n}_i(\boldsymbol{c}_i + t\boldsymbol{w}) = \mathcal{N}_i(\boldsymbol{w}) = e^{i\theta_i}\boldsymbol{w},\\
&\boldsymbol{n}_j(\boldsymbol{c}_i + t\boldsymbol{w}) = \boldsymbol{n}_j(\boldsymbol{c}_j + (t-\tau)\boldsymbol{w}) = \mathcal{N}_j(-\boldsymbol{w}) = -e^{i\theta_j}\boldsymbol{w} = -e^{i\theta_j}e^{-i\theta_i}\mathcal{N}_i(\boldsymbol{w})\,.
\end{aligned}$$

In the previous identities, the angles were functions $\theta_k = \theta_k(\sigma_k)$ for $\boldsymbol{w} = e^{i\sigma_k}\boldsymbol{\xi}_k$, $k = i, j$, according to the implicit relations

$$\tan\Big(\theta_k + \sigma_k\Big) = \left(\frac{a_k}{b_k}\right)^2 \tan\sigma_k, \qquad k = i, j. \tag{3.29}$$

Since $\mathcal{N}_i(\boldsymbol{w}) = -\mathcal{N}_j(-\boldsymbol{w})$, we conclude that

$$\theta_i = \theta_j. \tag{3.30}$$

Each ellipse is parameterized in a space of dimension 5, two for each center $\boldsymbol{c}_k$ and three for each matrix $\mathcal{Q}_k$. These 10 parameters determine a_k, b_k, σ_k, and θ_k, hence the three identities (3.30) and (3.29) determine a subspace of co-dimension 3 where $\mathcal{H}_{ij}$ is degenerate. In other words, Sard's Theorem [68, 69] states that $\mathcal{H}_{ij}$ degenerates only over a subset of measure zero in the space of parameters for $\mathcal{E}_i$ and $\mathcal{E}_j$. □

Remark 3. *We will present a degenerate co-gradient locus formed of two intersecting lines but for a pair of ellipses whose principal axes are not aligned. This example is instructive in that it goes beyond the degenerate examples identified during the proof of Theorem 6 and demonstrates that Conditions* (3.29) *and* (3.30) *are non-empty. Furthermore, this example is new to the literature and could be used for code verification.*

Consider the ellipse $\mathcal{E}_i$ described by its geometric potential

$$f_i(\boldsymbol{x}) = \frac{x^2}{3} + y^2 - 1\,. \tag{3.31}$$

The point $\boldsymbol{x}_i = [\sqrt{3/2}, 1/\sqrt{2}]^T$ belongs to $\mathcal{E}_i$ and the line ℓ through the origin and $\boldsymbol{x}_i$ forms an angle of $\pi/6$ with the horizontal axis because

$$\tan(\pi/6) = 1/\sqrt{3}\,.$$

Using the formula (2.16) *shows that the gradient to $\mathcal{E}_i$ at $\boldsymbol{x}_i$ is given by*

$$\nabla f_i(\boldsymbol{x}_i) = [\sqrt{3/2}, \sqrt{2}]^T\,.$$

A simple calculation shows that the angle measured counter-clockwise between $\boldsymbol{x}_i$ and $\nabla f_i(\boldsymbol{x}_i)$ is also $\pi/6$. This conforms with the relation (2.31)*,*

$$\tan\left(\frac{\pi}{6} + \frac{\pi}{6}\right) = \left(\sqrt{3}\right)^2 \tan\left(\frac{\pi}{6}\right)\,.$$

Our objective is to construct an ellipse $\mathcal{E}_j$ whose center $\mathbf{c}_j$ belongs to the line ℓ and whose normal $\mathcal{N}_j$ in the direction $-\boldsymbol{x}_i$ is opposite to the normal $\mathcal{N}_i$ in the direction $\boldsymbol{x}_i$. To keep this

as simple as possible, we will describe the ellipse in its local coordinates

$$\widehat{f}_j(\boldsymbol{x}) = \frac{\xi^2}{2+\sqrt{3}} + \eta^2 - 1\,, \tag{3.32}$$

and we introduce the point

$$\widehat{\boldsymbol{x}}_j = \left[-\sqrt{\frac{2+\sqrt{3}}{3+\sqrt{3}}}, -\sqrt{\frac{2+\sqrt{3}}{3+\sqrt{3}}} \right]^T ,$$

on the ellipse. With respect to the $\boldsymbol{\xi}_j$ axis and measured counter-clockwise, the line through the origin and $\widehat{\boldsymbol{x}}_j$ forms an angle of $5\pi/4$. The gradient to $\mathcal{E}_j$ gradient at $\widehat{\boldsymbol{x}}_j$ is

$$\nabla \widehat{f}_j(\widehat{\boldsymbol{x}}_j) = 2 \left[-\frac{1}{\sqrt{(2+\sqrt{3})(3+\sqrt{3})}}, -\sqrt{\frac{2+\sqrt{3}}{3+\sqrt{3}}} \right]^T ,$$

and it is easy to see that it forms an angle of $\pi/6$ with $\widehat{\boldsymbol{x}}_j$. Hence, the relation (2.31) *is again satisfied*

$$\tan\left(\frac{\pi}{6} + \frac{5\pi}{4}\right) = \left(2+\sqrt{2}\right) \tan\left(\frac{5\pi}{4}\right) ,$$

using the fact that $\tan(5\pi/12) = 2+\sqrt{3}$.

The ellipse $\mathcal{E}_j$ can be translated so that $\boldsymbol{c}_j$ is located along the line ℓ, say in the first quadrant, and then it can be rotated by $-\pi/12 = \pi/6 - \pi/4$ so that the point $\widehat{\boldsymbol{x}}_j$ also falls on the line. Afterwards, the axes of the two ellipses will no longer be aligned but their normals will be opposite everywhere on the line between the two centers. The line ℓ will therefore be a part of the co-gradient locus. We conclude by remarking that this example clearly satisfies the three conditions for degeneracy of Corollary 7.

As mentioned at the beginning of Section 3.1, the intersection of two ellipses can lead to four different types of intersection sets. Yet in practice, the estimation of the distance between two ellipses is usually of interest when they are *close*, which is intuitive but contradictory statement. How can one say two objects are close without estimating their distance? We begin by studying the intersection of two circles and presenting a condition under which the intersection of two overlapping circles contains only two points.

Lemma 8. *Consider a pair of overlapping circles $\mathcal{C}_i, \mathcal{C}_j$ with non-penetrating CoM. If D_i and D_j are the closed discs bounded by the respective circles $\mathcal{C}_i$ and $\mathcal{C}_j$, then the intersection can*

be split into closed domains

$$D_i \cap D_j = A_{ij}^+ \cup A_{ij}^-,$$

whose intersection is $A_{ij}^+ \cap A_{ij}^- = \mathcal{H}_{ij} \cap (D_i \cap D_j)$*, both of which are diffeomorphic to a triangle*

$$T = \left\{ (\lambda_i, \lambda_j) \in \mathbb{R}^2 \,\middle|\, \widehat{\lambda}_j(\lambda_i) \leq \lambda_j \leq 1 \right\},$$

defined by the constraint $\widehat{\lambda}_j : [0,1] \to \mathbb{R}$*, with the help of the maps*

$$\begin{aligned} \varphi^{\pm} : A_{ij}^{\pm} &\longrightarrow T \\ \boldsymbol{x} &\longmapsto \Big(\lambda_i(\boldsymbol{x}), \lambda_j(\boldsymbol{x})\Big). \end{aligned} \tag{3.33}$$

Proof. We begin by simplifying the geometry through a translation and a rotation sending the center $\boldsymbol{c}_i$ to the origin, and the second center $\boldsymbol{c}_j$ to $(c_j, 0)$ along the positive x-axis. Corollary 7 states that the co-gradient locus is the x-axis with a second branch at infinity. The potential for the circles are

$$\begin{aligned} f_i(x,y) =& \left(\frac{x}{r_i}\right)^2 + \left(\frac{y}{r_i}\right)^2 - 1, \\ f_j(x,y) =& \left(\frac{x - c_j}{r_j}\right)^2 + \left(\frac{y}{r_j}\right)^2 - 1, \end{aligned}$$

and the normalized distance to the centers as

$$\begin{aligned} \lambda_i(x,y) =& \Big(f_i(x,y) + 1\Big)^{1/2}, \\ \lambda_j(x,y) =& \Big(f_j(x,y) + 1\Big)^{1/2}. \end{aligned}$$

Notice that the notation $\boldsymbol{c}_j = (c_j, 0)$ might induce some confusion.

We now proceed to detail the intersection between the two discs. The circles are overlapping, hence $c_j < r_i + r_j$. The non-penetrating CoM implies the additional condition

$$\{r_i, r_j\} < c_j \,.$$

The intersection can be characterized as

$$D_i \cap D_j = \left\{ \boldsymbol{x} \in \mathbb{R}^2 \,\middle|\, \lambda_i(\boldsymbol{x}), \lambda_j(\boldsymbol{x}) \leq 1 \right\} = A_{ij}^+ \cup A_{ij}^-,$$

where

$$A_{ij}^{+} = \Big\{ \boldsymbol{x} = (x,y) \in D_i \cap D_j \Big| x \geq 0 \Big\}, \qquad A_{ij}^{-} = \Big\{ \boldsymbol{x} = (x,y) \in D_i \cap D_j \Big| x \leq 0 \Big\}.$$

It is clear that $A_{ij}^{+} \cap A_{ij}^{-}$ is the portion of $\mathcal{H}_{ij}$ at the intersection of the two discs. For a point $\boldsymbol{x} \in D_i \cap D_j$, the smallest value that can be attained by $\lambda_j(\boldsymbol{x})$ occurs along the x-axis and is $(c_j - r_i\lambda_i(\boldsymbol{x}))/r_j$. For each value of λ_i, the value of λ_j therefore ranges between $[\widehat{\lambda}_j, 1]$ where

$$\widehat{\lambda}_j(\lambda_i) := \frac{c_j - r_i\lambda_i}{r_j}.$$

This allows us to define the triangular domain

$$T := \left\{ (\lambda_i, \lambda_j) \in \mathbb{R}^2 \,\middle|\, (c_j - r_j)/r_i \leq \lambda_i \leq 1,\ \widehat{\lambda}_j(\lambda_i) \leq \lambda_j \leq 1 \right\}.$$

Our objective is now to show that the maps $\varphi^{\pm}$ are bijective diffeomorphisms. Given a point $(\lambda_i, \lambda_j) \in T$, we will find a point $\boldsymbol{x} \in A_{ij}^{+}$, or A_{ij}^{-}, such that

$$\varphi^{+}(\boldsymbol{x}) = (\lambda_i, \lambda_j), \text{ or } \varphi^{-}(\boldsymbol{x}) = (\lambda_i, \lambda_j).$$

For $\boldsymbol{x} = (x,y)$, the map is

$$\begin{aligned} \lambda_i^2 &= \left(\frac{x}{r_i}\right)^2 + \left(\frac{y}{r_i}\right)^2, \\ \lambda_j^2 &= \left(\frac{x - c_j}{r_j}\right)^2 + \left(\frac{y}{r_j}\right)^2. \end{aligned}$$

If we isolate the x coordinate, we find

$$r_j^2\lambda_j^2 - r_i^2\lambda_i^2 = (x - c_j)^2 - x^2 = -2xc_j + c_j^2 \quad \Longrightarrow \quad x = \frac{1}{2c_j}\Big(r_i^2\lambda_i^2 - r_j^2\lambda_j^2 + c_j^2\Big).$$

Substituting back into the equation for λ_i^2 and simplifying

$$y^2 = r_i^2\lambda_i^2 - x^2 = r_i^2\lambda_i^2 - \frac{1}{4c_j^2}\Big(r_i^2\lambda_i^2 - r_j^2\lambda_j^2 + c_j^2\Big)^2.$$

This equation possesses two solutions for y, corresponding to either the map φ^{+} or φ^{-}. Taking $\lambda_i = \lambda_j = 1$, we can quickly verify that there are only two solutions at the intersection of the two circles.

The previous calculation shows that $\varphi^{\pm}$ are bijective between $A_{ij}^{\pm}$ and T. To show that they

are diffeomorphisms, we compute the differential by first observing that

$$\nabla f_k = \nabla \lambda_k = 2\lambda_k \nabla \lambda_k,$$

hence

$$d\varphi^{\pm}(\boldsymbol{x}) = \begin{bmatrix} \nabla \lambda_i \\ \nabla \lambda_j \end{bmatrix} = \begin{bmatrix} \dfrac{1}{2\lambda_i} \nabla f_i \\ \dfrac{1}{2\lambda_j} \nabla f_j \end{bmatrix}.$$

If one recalls the definition (3.20) of H, once can conclude that

$$\det\left(d\varphi^{\pm}(\boldsymbol{x})\right) = \frac{H(\boldsymbol{x})}{4\lambda_i(\boldsymbol{x})\lambda_j(\boldsymbol{x})}.$$

In the case of two circles, a simple calculation shows that

$$\det\left(d\varphi^{\pm}(\boldsymbol{x})\right) = \frac{yc_j}{\lambda_i(\boldsymbol{x})\lambda_j(\boldsymbol{x})r_i^2 r_j^2},$$

which only vanishes along the co-gradient locus. □

The analysis of the overlap between two circles has demonstrated that the condition of non-penetrating CoM is sufficient to characterize closeness. The next definition is a condition for ellipses that generalizes the previous condition, but relies on the intrinsic notion of co-gradient locus. Such a description can be used to clarify the convergence analysis of contact detection algorithms, much in the same way that a Taylor series is used, thereby avoiding the study of degenerate cases. To the best of the our knowledge, no such precise description exists in the literature.

Definition 6. [Ellipses in near perfect contact] *Two ellipses $\mathcal{E}_i$, $\mathcal{E}_j \subset \mathbb{R}^2$ with non-penetrating CoM are said to be in near perfect contact if*

$$\frac{|t_i - t_j| \cdot \|\boldsymbol{\gamma}'_{ij}(t_i)\|}{\min_{k=i,j}\ 2\underline{\rho}_k |\boldsymbol{\gamma}'_{ij}(t_k) \cdot \boldsymbol{n}_k(\boldsymbol{\gamma}_{ij}(t_k))|} \ll 1, \tag{3.34}$$

where $\boldsymbol{\gamma}_{ij}$, t_i, t_j are defined as in Theorem 6, $\boldsymbol{n}_k(\boldsymbol{\gamma}_{ij}(t_k))$ is the outward normal unit vector to $\mathcal{E}_k$ at point $\boldsymbol{\gamma}_{ij}(t_k)$ on $\mathcal{E}_k$, and $\underline{\rho}_k = b_k^2/a_k$ (2.18) is the smallest radius of curvature on $\mathcal{E}_k$, $k = i, j$.

Remark 4. *The proof of the next theorem will show that this analytic condition implies that the pair of ellipses satisfy two geometric properties. To explain these conditions, we introduce $D_k^+(r)$ the disc of radius r tangent to $\mathcal{E}_k$ at $\boldsymbol{\gamma}_{ij}(t_k)$ but whose interior is disjoint*

from $E_k := \{\boldsymbol{x} \mid f_k(\boldsymbol{x}) \leq 0\}$. *Similarly, let* $D_k^-(r)$ *be the disc of radius* r *tangent to* $\mathcal{E}_k$ *whose interior lies inside* E_k. *Intuitively,* $D_k^+(r)$ *(*$D_k^-(r)$*) is the disc tangent to* $\mathcal{E}_k$ *placed on the outside (inside) of* $\mathcal{E}_k$, *see Figure 3.5 (a). We will also be using* $\underline{\rho}_k$ *and* $\overline{\rho}_k$ *the minimum and the maximum of the radius of curvature of the ellipse* $\mathcal{E}_k$, *respectively.*

Condition 6 implies that

1) *the discs* $D_i^-(\underline{\rho}_i)$ *and* $D_j^-(\underline{\rho}_j)$ *have non-penetrating CoM; and*
2) *the portion of* $\mathcal{H}_{ij}$ *inside the intersection of the two ellipses is entirely inside the intersection* $D_i^-(\underline{\rho}_i) \cap D_j^-(\underline{\rho}_j)$.

Theorem 9. *Consider two ellipses* $\mathcal{E}_i$, $\mathcal{E}_j \subset \mathbb{R}^2$ *in near perfect contact, and let* t_i *and* t_j *be defined as in Theorem 6. Then the intersection* $\mathcal{I}_{ij} = \mathcal{E}_i \cap \mathcal{E}_j$ *is one of three options:*

1) *if* $t_i < t_j$, *then* $\mathcal{I}_{ij} = \varnothing$, *i.e. the ellipses are disjoint;*
2) *if* $t_i = t_j$, *then* $\mathcal{I}_{ij}$ *consists of a singleton, i.e. the ellipses are in perfect contact;*
3) *if* $t_i > t_j$, *then* $\mathcal{I}_{ij}$ *consists of two distinct points, i.e. the ellipses have small overlap.*

Proof. This will not be a complete proof of Theorem 9 but will attempt to explain how the condition is related to the two geometrical properties in the previous remark, as well as the maps $\varphi^{\pm}$ describing the intersection. We begin by studying the neighborhood of the point $\boldsymbol{\gamma}_{ij}(t_i) \in \mathcal{E}_i$. The tangent to the curve $\boldsymbol{\gamma}_{ij}$ at $\boldsymbol{\gamma}_{ij}(t_i)$ is not necessarily in the same direction as the normal $\boldsymbol{n}_i(\boldsymbol{\gamma}_{ij}(t_i))$, but certainly not perpendicular, and we note the defect as

$$\eta = \arccos\left(\frac{\boldsymbol{n}_i(\boldsymbol{\gamma}_{ij}(t_i)) \cdot \boldsymbol{\gamma}'_{ij}(t_i)}{\|\boldsymbol{\gamma}'_{ij}(t_i)\|}\right) \in]-\pi/2, \pi/2[\,.$$

We now establish a condition under which the tangent line to $\boldsymbol{\gamma}_{ij}$ at $\boldsymbol{\gamma}_{ij}(t_i)$ remains inside $D_i^-(\underline{\rho}_i) \cup D_i^+(\underline{\rho}_i)$. By translating $\boldsymbol{\gamma}_{ij}(t_i)$ to the origin, then applying a rotation to send the tangent line to $\mathcal{E}_i$ at t_i to the horizontal x-axis, see Figure 3.5 (b). We observe that we can reduce the analysis to showing that a curve crossing the origin at an angle η with respect to the vertical axis remains inside the discs tangent to the x-axis centered at the origin.

Recall the formula (2.18) for the smallest radius of curvature $\underline{\rho}_i := b_i^2/a_i$, and the parameterization of the points on the boundary of $D_i^-(\underline{\rho}_i)$, centered at $(0, -\underline{\rho}_i)$ below the horizontal axis:

$$\big(\underline{\rho}_i \cos(\theta), \underline{\rho}_i \sin(\theta) - \underline{\rho}_i\big), \quad \forall \theta \in [-\pi, \pi[.$$

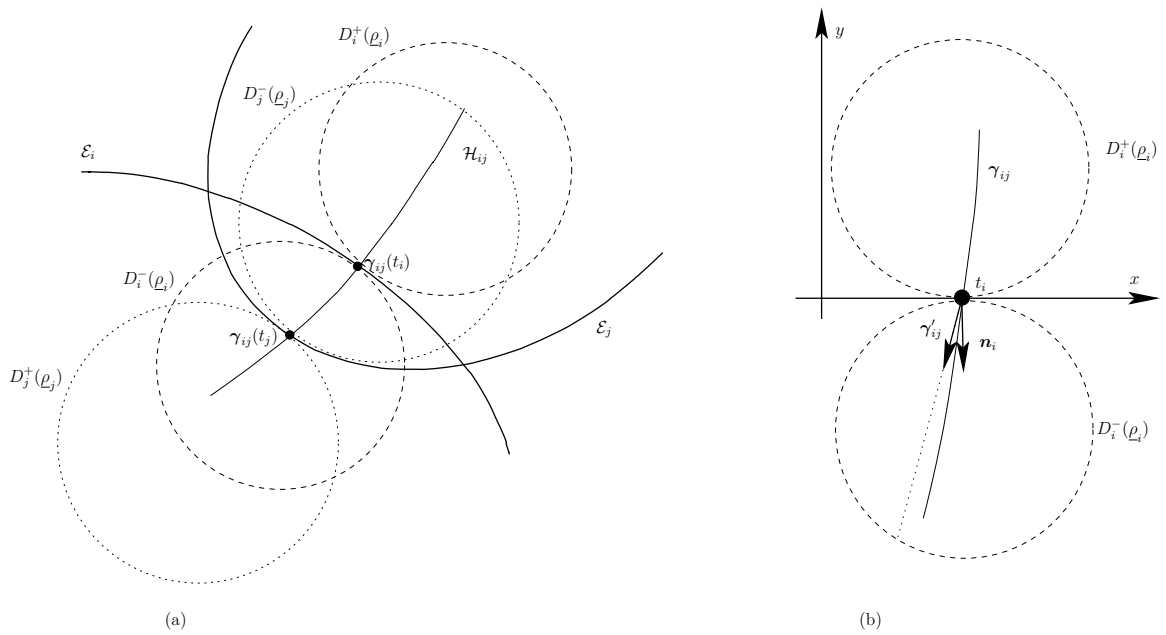


Figure 3.5 (a) Illustration of disks at points $\boldsymbol{\gamma}_{ij}(t_k)$ with $k = i, j$ for a pair of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$. (b) Transformation of two disks $D_i^{\pm}(\underline{\rho}_i)$ where $\boldsymbol{\gamma}_{ij}(t_i)$ is at the origin and the tangent line to $\mathcal{E}_i$ at t_i is aligned with the horizontal x-axis.

The smooth curve $\boldsymbol{\gamma}_{ij}$ remains close to its tangent line $s\boldsymbol{\gamma}'_{ij}(t_i)$, $\forall s \in \mathbb{R}$. We now compute the length of the portion of the tangent line inside $D_i^-(\underline{\rho}_i)$, which by symmetry will be of the same length inside $D_i^+(\underline{\rho}_i)$. The tangent line forms an angle η with the vertical axis and let $\mathbf{p} \neq \mathbf{0}$ be the unique point on $D_i^-(\underline{\rho}_i)$ where the tangent line crosses. We observe that an equilateral triangle is formed between the origin $\mathbf{0}$, the center $(0, -\underline{\rho}_i)$, and the point $\mathbf{p}$ with two sides of length $\underline{\rho}_i$ and two angles of measure η. The length we are looking for is the length of the side of this equilateral triangle opposite to the center $(0, -\underline{\rho}_i)$. Based on this geometry, it is easy to verify that the length of the portion of the tangent line inside $D_i^-(\underline{\rho}_i)$ is

$$2\underline{\rho}_i \cos(\eta) = 2\underline{\rho}_i \boldsymbol{n}_i\Big(\boldsymbol{\gamma}_{ij}(t_i)\Big) \cdot \frac{\boldsymbol{\gamma}'_{ij}(t_i)}{\|\boldsymbol{\gamma}'_{ij}(t_i)\|}.$$

The curve $\boldsymbol{\gamma}_{ij}(t)$ will remain close to its tangent line as long as the parameter t stays close to t_i, with the deviation being proportional to $|t - t_i|^2$ and the curvature of $\boldsymbol{\gamma}_{ij}$. Requiring that the distance $|t_i - t_j|$ along the curve $\boldsymbol{\gamma}_{ij}$ be small when compared to the length above implies that the curve cannot exit $D_i^+ \cap D_i^-$ and hence the portion of $\mathcal{H}_{ij}$ with t close to both t_i and t_j crosses $\mathcal{E}_i$ only once. The minimum over i and j in (3.34) also constrains the co-gradient locus, at least the part inside $D_j^-(\underline{\rho}_j) \cup D_j^+(\underline{\rho}_j)$, to contain only a single point from $\mathcal{E}_j$. It is also clear that by taking $t_j - t_i$ sufficiently small and negative, then the two discs $D_i^-(\underline{\rho}_i)$

and $D_j^-(\underline{\rho}_j)$ will satisfy the non-penetrating CoM condition.

We can now distinguish the following three cases:

- If $t_i = t_j$ then the definition of the co-gradient locus implies that both ellipses have opposing normals and that the two ellipses are in perfect contact.

- Suppose now that both $t_i < t_j$ and (3.34) are satisfied. The portion of the co-gradient locus between $\boldsymbol{\gamma}_{ij}(t_i)$ and $\boldsymbol{\gamma}_{ij}(t_j)$ must be inside $D_i^+(\underline{\rho}_i) \cup D_j^+(\underline{\rho}_j)$, which is outside of both ellipses. There exists a unique ellipse $\mathcal{E}_j(r_j)$ which is tangent to $\mathcal{E}_i$ at $\boldsymbol{\gamma}_{ij}(t_i)$ and because $\boldsymbol{\gamma}_{ij}(t_i)$ is outside of E_j, r_j must be greater than 1. Since both $\mathcal{E}_i$ and $\mathcal{E}_j(r_j)$ are convex and tangent, then E_i is disjoint from $\mathcal{E}_j(r_j)$ and the set E_j strictly bounded by $\mathcal{E}_j(r_j)$. These remarks imply that the ellipses are disjoint when $t_i < t_j$.

- Suppose now that both $t_j < t_i$ and (3.34) hold. The objective is to show that $\mathcal{E}_i \cap \mathcal{E}_j$ contains exactly two points, which is fundamentally no longer a question only about a neighborhood of $\boldsymbol{\gamma}_{ij}(t)$ and $t \in [t_i, t_j]$. The question will be answered by providing a detailed description of the region at the intersection $E_i \cap E_j$. The characterization of the intersection $\mathcal{E}_i \cap \mathcal{E}_j$ will be a simple corollary of the description of $E_i \cap E_j$. We shall show that there exists a diffeomorphism of a triangle towards each half of the domain $E_i \cap E_j \setminus \mathcal{H}_{ij}$.

  We begin by identifying a subdivision of $E_i \cap E_j$. Following the definition of H in (3.18), we can construct

  $$A_{ij}^+ = \Big\{ \boldsymbol{x} \in E_i \cap E_j \,;\, H(\boldsymbol{x}) \geq 0 \Big\},$$
  $$A_{ij}^- = \Big\{ \boldsymbol{x} \in E_i \cap E_j \,;\, H(\boldsymbol{x}) \leq 0 \Big\},$$

  and deduce that

  $$A_{ij}^+ \cup A_{ij}^- = E_i \cap E_j \,.$$

  Intuitively, the normalized co-gradient function (3.23) can be used to uniquely define the angle $\eta_{ij}(\boldsymbol{x})$ between the two normals in a neighborhood of $\mathcal{H}_{ij}$ according to

  $$\hat{H}(\boldsymbol{x}) = \sin \eta_{ij}(\boldsymbol{x}) = 0 \quad \Longleftrightarrow \quad \eta_{ij}(\boldsymbol{x}) = \pi,$$

  and the convention that we measure increasing η_{ij} when rotating counter-clockwise from $\boldsymbol{n}_i$ to $\boldsymbol{n}_j$. Hence, for all $\boldsymbol{x}$ in a neighborhood of $\mathcal{H}_{ij} = A_{ij}^+ \cup A_{ij}^-$, there is a well-defined angle $\eta_{ij}(\boldsymbol{x})$ with values near π. Since the scalar function H is the z-component of the cross-product $\boldsymbol{n}_i \times \boldsymbol{n}_j$, the two regions A_{ij}^+ and A_{ij}^- correspond to the regions where

$\eta_{ij}(\boldsymbol{x})$ is respectively less than π and greater than π, see Figure 3.6 (a). As introduced earlier in Section 2.3, each point $\boldsymbol{x} \in \mathbb{R}^2$ belongs to unique scaled ellipses $\mathcal{E}_i(\lambda_i(\boldsymbol{x}))$ and $\mathcal{E}_j(\lambda_j(\boldsymbol{x}))$ where the real valued functions

$$\lambda_k(\boldsymbol{x}) = \sqrt{f_k(\boldsymbol{x}) + 1}, \quad k = i, j,$$

are smooth. The gradient of these functions are multiples of the gradient because

$$\nabla f_k = \nabla \lambda_k^2 = 2\lambda_k \nabla \lambda_k.$$

The values of λ_i and λ_j along the co-gradient locus range between

$$\lambda_i \in [r_i, 1], \qquad \lambda_j \in [r_j, 1],$$

where $r_i := \lambda_i(\boldsymbol{\gamma}_{ij}(t_j))$ and $r_j := \lambda_j(\boldsymbol{\gamma}_{ij}(t_i))$. More precisely, along the segment $\boldsymbol{\gamma}_{ij}([t_i, t_j])$ the parameter $\lambda_i(\boldsymbol{\gamma}_{ij}(t))$ is increasing and $\lambda_j(\boldsymbol{\gamma}_{ij}(t))$ is decreasing and therefore we can parameterize λ_j as a function of λ_i along the segment, say $\widehat{\lambda}_j(\lambda_i)$. From this, we can define a triangular domain

$$\mathcal{T} = \Big\{(s_1, s_2) \in [r_i, 1] \times [r_j, 1] \,\Big|\, s_2 \geq \widehat{\lambda}_j(s_1)\Big\},$$

and two smooth maps

$$\begin{aligned} \boldsymbol{\varphi}^{\pm} : A_{ij}^{\pm} &\longrightarrow \mathcal{T} \\ \boldsymbol{x} &\longmapsto \Big(\lambda_i(\boldsymbol{x}), \lambda_j(\boldsymbol{x})\Big). \end{aligned}$$

The determinants of the differentials of these maps are

$$\det\Big(d\boldsymbol{\varphi}^{\pm}\Big) = \begin{vmatrix} (\nabla\lambda_i)^T \\ (\nabla\lambda_j)^T \end{vmatrix} = \frac{1}{4\lambda_i(\boldsymbol{x})\lambda_j(\boldsymbol{x})} \begin{vmatrix} (\nabla f_i)^T \\ (\nabla f_j)^T \end{vmatrix} = \frac{H(\boldsymbol{x})}{4\lambda_i(\boldsymbol{x})\lambda_j(\boldsymbol{x})},$$

hence the maps $\boldsymbol{\varphi}^{\pm}$ are of rank 2 except along $\mathcal{H}_{ij}$. This implies that each map is locally bijective, at least in the interior of $A_{ij}^{\pm}$.

The analysis presented so far shows that $\varphi^{\pm}$ is a diffeomorphism only in a neighborhood of $\mathcal{H}_{ij}$. A complete proof would require a proof that λ_j is monotone along the boundary of $\mathcal{E}_i$ in $A_{ij}^{\pm}$, and vice versa for λ_i. This would demontrate that $\varphi^{\pm}$ is bijective along the boundaries. Given that $D_k^-(\underline{\rho}_k)$ is tangent to $\mathcal{E}_k$, it is intuitively clear that if Lemma 8 holds for both discs, then the maps $\varphi^{\pm}$ for ellipses should also characterize

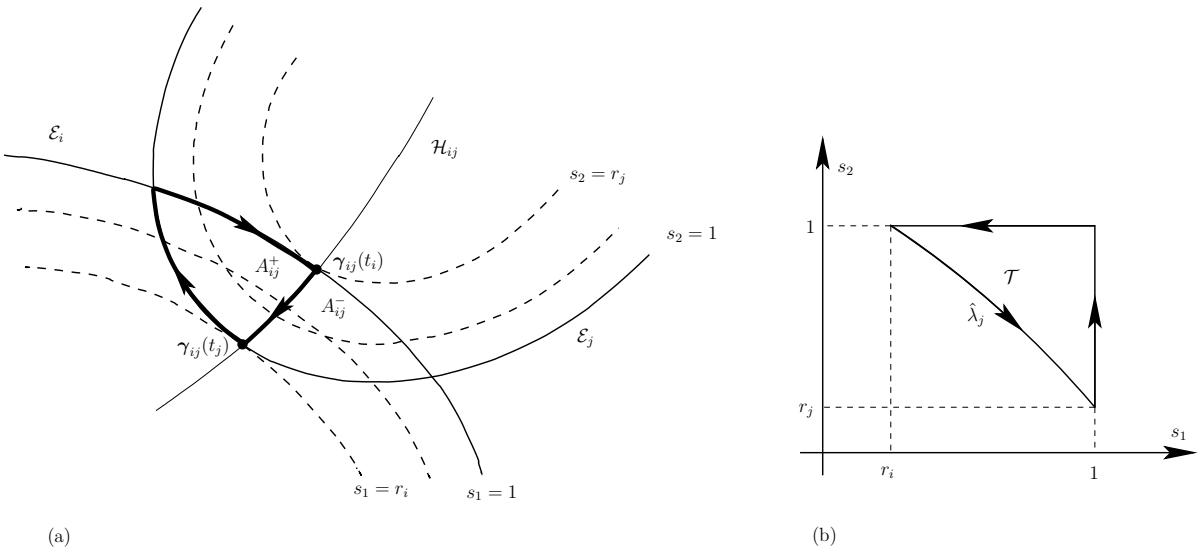


Figure 3.6 Illustration of the proof of Theorem 9. (a) The ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ are in near perfect contact with overlap. The region A_{ij}^+ is illustrated with bold boundary. (b) The region $\mathcal{T}$ which is mapped by function $\boldsymbol{\varphi}$ from region A_{ij}^+. The curve $\hat{\lambda}_j$ which is mapped from the curve $A_{ij}^+ \cap A_{ij}^-$.

the intersection $E_i \cap E_j$.

□

Remark 5. *If we consider MDP as a pair of points $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ without satisfying the condition of opposite outward unit normal vectors $\boldsymbol{n}_i(\boldsymbol{x}_i)$ and $\boldsymbol{n}_j(\boldsymbol{x}_j)$ to $\mathcal{E}_i$ and $\mathcal{E}_j$, respectively, we have:*

$$(\boldsymbol{x}_i, \boldsymbol{x}_j) \;=\; \operatorname*{argmin}_{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j} \|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\| \;=\; \operatorname*{argmin}_{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j} \frac{1}{2}\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\|^2. \tag{3.35}$$

The solution of above problem is the intersection pair for ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ with small overlap. Moreover, if we add opposite outward unit normal constraint to Problem (3.35)*, then we find a unique solution as MDP pair:*

$$(\boldsymbol{x}_i, \boldsymbol{x}_j) \;=\; \operatorname*{argmin}_{\substack{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j \\ \boldsymbol{n}_i(\hat{\boldsymbol{x}}_i) + \boldsymbol{n}_j(\hat{\boldsymbol{x}}_j) = \boldsymbol{0}}} \|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\| \;=\; \operatorname*{argmin}_{\substack{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j \\ \boldsymbol{n}_i(\hat{\boldsymbol{x}}_i) + \boldsymbol{n}_j(\hat{\boldsymbol{x}}_j) = \boldsymbol{0}}} \frac{1}{2}\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\|^2. \tag{3.36}$$

On the other hand, the MPP $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ for two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ with small overlap

is the unique solution to problem

$$\boldsymbol{x}_i = \underset{\boldsymbol{x}\in\mathcal{E}_i}{\operatorname{argmin}} \|\boldsymbol{x}-\boldsymbol{c}_j\|_{\mathcal{E}_j} = \underset{\boldsymbol{x}\in\mathcal{E}_i}{\operatorname{argmin}} f_j(\boldsymbol{x}), \tag{3.37}$$

$$\boldsymbol{x}_j = \underset{\boldsymbol{x}\in\mathcal{E}_j}{\operatorname{argmin}} \|\boldsymbol{x}-\boldsymbol{c}_i\|_{\mathcal{E}_i} = \underset{\boldsymbol{x}\in\mathcal{E}_j}{\operatorname{argmin}} f_i(\boldsymbol{x}), \tag{3.38}$$

and the constraints

$$\boldsymbol{n}_i(\boldsymbol{x}_i) + \boldsymbol{n}_j(\boldsymbol{x}_i) = \boldsymbol{0}, \tag{3.39}$$

$$\boldsymbol{n}_i(\boldsymbol{x}_j) + \boldsymbol{n}_j(\boldsymbol{x}_j) = \boldsymbol{0}. \tag{3.40}$$

are still non-binding.

Therefore, the unique MPP $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ *is also the solution to problems*

$$\boldsymbol{x}_i = \underset{\substack{\boldsymbol{x}\,\in\,\mathcal{E}_i \\ \boldsymbol{n}_i(\boldsymbol{x})+\boldsymbol{n}_j(\boldsymbol{x})=\boldsymbol{0}}}{\operatorname{argmin}} \|\boldsymbol{x}-\boldsymbol{c}_j\|_{\mathcal{E}_j} = \underset{\substack{\boldsymbol{x}\,\in\,\mathcal{E}_i \\ \boldsymbol{n}_i(\boldsymbol{x})+\boldsymbol{n}_j(\boldsymbol{x})=\boldsymbol{0}}}{\operatorname{argmin}} f_j(\boldsymbol{x}), \tag{3.41}$$

$$\boldsymbol{x}_j = \underset{\substack{\boldsymbol{x}\,\in\,\mathcal{E}_j \\ \boldsymbol{n}_i(\boldsymbol{x})+\boldsymbol{n}_j(\boldsymbol{x})=\boldsymbol{0}}}{\operatorname{argmin}} \|\boldsymbol{x}-\boldsymbol{c}_i\|_{\mathcal{E}_i} = \underset{\substack{\boldsymbol{x}\,\in\,\mathcal{E}_j \\ \boldsymbol{n}_i(\boldsymbol{x})+\boldsymbol{n}_j(\boldsymbol{x})=\boldsymbol{0}}}{\operatorname{argmin}} f_i(\boldsymbol{x}). \tag{3.42}$$

and

$$\boldsymbol{x}_i = \underset{\boldsymbol{x}\in\mathcal{E}_i\cap\mathcal{H}_{ij}}{\operatorname{argmin}} f_j(\boldsymbol{x}), \tag{3.43}$$

$$\boldsymbol{x}_j = \underset{\boldsymbol{x}\in\mathcal{E}_j\cap\mathcal{H}_{ij}}{\operatorname{argmin}} f_i(\boldsymbol{x}). \tag{3.44}$$

3.5 Relationship to Time-dependent Contact Detection

For pairs of rapidly moving and/or rotating ellipses/ellipsoids, estimating the contact point and the penetration distance requires anticipating future positions. More specifically, the future contact point might be very different from the MDP at a given time when the relative velocity of the two particles is large and their surfaces are close. In this section, we show that the constraint of the co-gradient locus $\mathcal{H}_{ij}$ appearing in the MPP can be interpreted as enforcing displacements of the ellipses along a specific trajectory. In other words, we can interpret $\mathcal{H}_{ij}$ as a specific particle trajectory bringing the two ellipses in contact. The purpose of this section is not to study time-dependent collision detection for which there exists particular techniques, such as the *continuous contact detection method* of Wang et al. [63].

Consider two ellipses $\mathcal{E}_i$, $\mathcal{E}_j$ that are disjoint but in near perfect contact. Consider the parameterization $\boldsymbol{\gamma}_{ij}$ of the curve $\mathcal{H}_{ij}$ going from $\boldsymbol{c}_i = \boldsymbol{\gamma}_{ij}(0)$ to $\boldsymbol{c}_j = \boldsymbol{\gamma}_{ij}(1)$, and fix $t \in]t_i, t_j[$ corresponding to $\boldsymbol{\gamma}_{ij}(t)$ in the exterior of both ellipses. Along the directed line $\overline{\boldsymbol{c}_j\boldsymbol{\gamma}_{ij}(t)}$, there is one unique intersection point $\boldsymbol{p}_j(t) \in \mathcal{E}_j$, which depends analytically on t, and the translation is defined by

$$\begin{aligned} T_t :\ \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ \boldsymbol{x} &\longmapsto \boldsymbol{x} + \Big(\boldsymbol{\gamma}_{ij}(t) - \boldsymbol{p}_j(t)\Big). \end{aligned}$$

Remark that $\boldsymbol{\gamma}_{ij}(t) - \boldsymbol{c}_j$ and $\boldsymbol{\gamma}_{ij}(t) - \mathbf{p}_j(t)$ are colinear, hence the displaced ellipse $T_t\mathcal{E}_j$ has the same normal along the line $T_t\boldsymbol{c}_j$ to $T_t\mathbf{p}_j(t)$ as it had along the line connecting $\boldsymbol{c}_j$ to $\boldsymbol{\gamma}_{ij}(t)$. This implies that T_t will map $\mathbf{p}_j(t_i)$ to $\boldsymbol{\gamma}_j(t_i)$ and that at $\boldsymbol{\gamma}_{ij}(t_i)$, the normal to $T_t\mathcal{E}_j$ will be the same as $\boldsymbol{n}_j$ for $\mathcal{E}_j$. The definition of $\mathcal{H}_{ij}$ then implies that the image of $T_{t_i}\mathcal{E}_j$ will be in perfect contact with $\mathcal{E}_i$.

We have therefore shown that as t goes from t_j to t_i, the transformation T_t brings $\mathcal{E}_j$ in contact with $\mathcal{E}_i$ at $\boldsymbol{\gamma}_{ij}(t)$. This transformation does not correspond to a freefall trajectory for $\mathcal{E}_j$, because on one hand, each translation T_t preserves the orientation and therefore the physical displacement could not have applied torque. Yet, the displacement of the center of mass $T_t\boldsymbol{c}_j$ is determined (indirectly) by the points on a hyperbola in any rotated frame of reference, and therefore does not follow a parabola. Hence T_t does not describe a physical displacement.

We conclude this section by highlighting a relationship between the MPP and the Perram-Wertheim theory of contact detection [25, 59]. The construction described above can be symmetrized by redefining the translation T_t as $T_t^{(j)}$ and introducing a second one

$$\begin{aligned} T_t^{(i)} : \mathbb{R}^2 &\longrightarrow \mathbb{R}^2 \\ \boldsymbol{x} &\longmapsto \boldsymbol{x} + \Big(\boldsymbol{\gamma}_{ij}(t) - \boldsymbol{p}_i(t)\Big) \end{aligned}$$

where $\mathbf{p}_i(t)$ is the unique intersection point of the directed line $\overline{\boldsymbol{c}_i\boldsymbol{\gamma}_{ij}(t)}$ with $\mathcal{E}_i$. In this case, for every $t \in [t_i, t_j]$ the translations $T_t^{(i)}$ and $T_t^{(j)}$ send $\mathcal{E}_i$ and $\mathcal{E}_j$ respectively to ellipses that are in perfect contact at $\boldsymbol{\gamma}_{ij}(t)$. This is similar to the construction of Perram and Wertheim that identifies the family of *scalings* of $\mathcal{E}_i$ and $\mathcal{E}_j$ such that the scaled ellipses come in perfect contact. A notion of distance is then defined from this construction, just as $|t_i - t_j|$ serves as a proxy for distance. It is also interesting to note that the Perram-Wertheim theory is also expressed as a basic unconstrained minimization problem.

3.6 Extension to Ellipsoids

One of our objective is to review and compare algorithms for the rapid and accurate estimation of separation/penetration distance for pairs of ellipses and ellipsoids in the quasi-static regime. The main mathematical results are Lemma 1, Theorem 6 and Theorem 9 which were all stated and demonstrated in 2-D, so it is natural to enquire as to what can be said in three-space dimension. In this section, we will briefly suggest how these three could be extended to 3-D. Before proceeding, we remark that Definitions 4 and 5 have obvious extensions, that Lemmas 3 and 4 and Theorem 5 have identical statements and proofs in 3-D.

The most difficult result to extend to 3-D is Theorem 6 and it should be studied with the techniques of algebraic geometry [64,70]. Lemma 1 was only a preliminary result for the proof of Theorem 6 and provided tools to circumvent real projective geometry, hence we will analyze only Theorem 6. As far as Theorem 9 is concerned, the extension to 3-D is straightforward assuming that $\mathcal{H}_{ij}$ has been characterized. In fact, our initial proof of Theorem 9 was carried out in 3-D and the key map $\boldsymbol{\varphi}$ is slightly easier to study because the intersection $E_i \cap E_j$ is path-connected (but not simply connected).

The statement in 3-D of Theorem 6 concerns the co-gradient locus $\mathcal{H}_{ij}$ which is now the set of common roots of each of the three components of the cross-product (3.16), each of which can be written roughly in a quadratic form similar to (3.20). Each component defines a 2-dimensional subvariety that intersects the sphere at infinity in real projective space of three dimension $\mathbb{R}P^3$ along a curve. We conjecture that, as we did in 2-D, we can characterize the common zeros in $\mathbb{R}^3$ by identifying the isolated intersection points of the three 1-D curves at infinity. Each ellipse will define three planes corresponding to each pair of its orthogonal axis, where the normals take on known values. These six curves on the sphere at infinity will provide a triangular subdivision of the sphere and the existence of a root to (3.16) in each triangle will be determined by examining the signs of the components of $\boldsymbol{H}$ along the edges and nodes of the triangle.

We deemed that complete proofs in 3-D would have distracted readers from the focus on the definitions of contact points and on the comparison between the different algorithms. Nevertheless, such extensions should be studied and we encourage other researchers more familiar with the necessary tools to address these questions.

3.7 Mapping of $(\mathcal{E}_i, \mathcal{E}_j)$

Many of the algorithms to be presented in Chapters 4 and 5 share a common feature that the problem can be simplified by introducing normalized coordinates where one of the ellipses has

become a circle. In this section, we review two mapping steps, see sections 3.7.1 and 3.7.2. The mapping will simplify the later presentation of the algorithms, and one of our proposed techniques for guessing the contact point, see Section 5.3, is done in one of the normalized coordinate systems we introduce below. We remark that the estimation of the contact point is rarely addressed in the literature and, to the best of our knowledge, the *focal point* estimate presented in Section 5.3 is new. It is worth noting that the following mappings are easily extendable to ellipsoids.

3.7.1 Mapping of $(\mathcal{E}_i, \mathcal{E}_j)$ into a Unit Circle $\widehat{\mathcal{C}}_i$ Centered at Origin and an Ellipse $\widehat{\mathcal{E}}_j$

The first mapping was suggested by Ting et al. in [46]. Let $\mathcal{E}_i$ and $\mathcal{E}_j$ be two arbitrary ellipses defined by

$$f_i(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{c}_i)^T \mathcal{Q}_i (\boldsymbol{x} - \boldsymbol{c}_i) - 1 = (\boldsymbol{x} - \boldsymbol{c}_i)^T \mathcal{R}_i \mathcal{D}_i \mathcal{R}_i^T (\boldsymbol{x} - \boldsymbol{c}_i) - 1 = 0, \tag{3.45}$$

$$f_j(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{c}_j)^T \mathcal{Q}_j (\boldsymbol{x} - \boldsymbol{c}_j) - 1 = (\boldsymbol{x} - \boldsymbol{c}_j)^T \mathcal{R}_j \mathcal{D}_j \mathcal{R}_j^T (\boldsymbol{x} - \boldsymbol{c}_j) - 1 = 0, \tag{3.46}$$

where the diagonal matrices $\mathcal{D}_i$ and $\mathcal{D}_j$ and the rotation matrices $\mathcal{R}_i$ and $\mathcal{R}_j$ are as defined in (2.3) and (2.2), respectively.

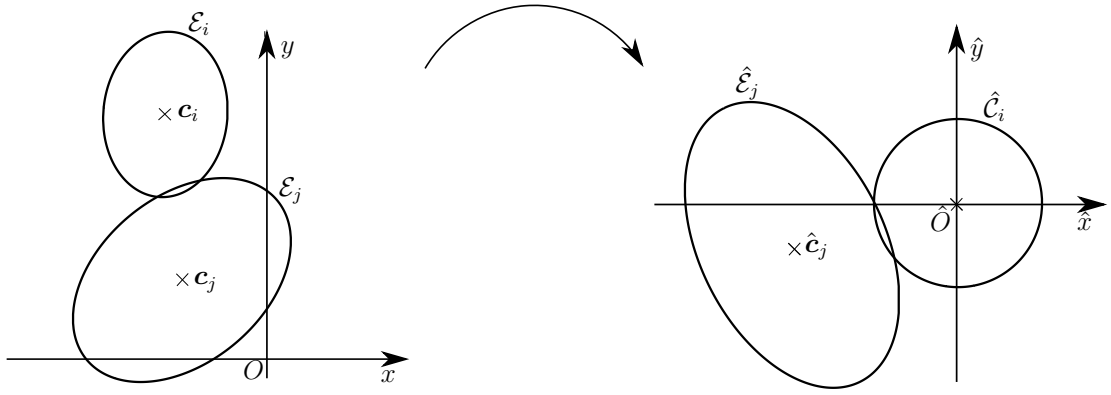


Figure 3.7 Mapping of $(\mathcal{E}_i, \mathcal{E}_j)$ into a unit circle $\hat{\mathcal{C}}_i$ centered at origin and an ellipse $\hat{\mathcal{E}}_j$.

The mapping consists in transforming one of the two ellipses, say $\mathcal{E}_i$, into the unit circle $\hat{\mathcal{C}}_i$ centered at the origin and the other ellipse $\mathcal{E}_j$ into the ellipse $\hat{\mathcal{E}}_j$, see Figure 3.7. The mapping that transforms $\hat{\mathcal{C}}_i$ into $\mathcal{E}_i$ consists of the transformation $\mathcal{D}_i^{-1/2}$, followed by the rotation $\mathcal{R}_i$ of angle θ_i and the translation $\boldsymbol{c}_i$. We thus have

$$\boldsymbol{x} = \mathcal{R}_i \mathcal{D}_i^{-1/2} \hat{\boldsymbol{x}} + \boldsymbol{c}_i. \tag{3.47}$$

Indeed, it is straightforward to check that the equation of the circle $\hat{\mathcal{C}}_i$ in the coordinate system $(\hat{O}, \hat{x}, \hat{y})$, using (3.47) in (3.45), thus reads:

$$\hat{f}_i(\hat{\boldsymbol{x}}) = \hat{\boldsymbol{x}}^T \hat{\boldsymbol{x}} - 1 = 0. \tag{3.48}$$

The equation of the new ellipse $\hat{\mathcal{E}}_j$, following the mapping of ellipse $\mathcal{E}_j$, is obtained by substituting (3.47) in (3.46). We find that $\hat{\mathcal{E}}_j$ in the $(\hat{O}, \hat{x}, \hat{y})$ coordinates are the roots of

$$\hat{f}_j(\hat{\boldsymbol{x}}) = (\hat{\boldsymbol{x}} - \hat{\boldsymbol{c}}_j)^T \mathcal{D}_i^{-1/2} \mathcal{R}_i^T \mathcal{Q}_j \mathcal{R}_i \mathcal{D}_i^{-1/2} (\hat{\boldsymbol{x}} - \hat{\boldsymbol{c}}_j) - 1 = 0,$$

where we have introduced $\hat{\boldsymbol{c}}_j$ satisfying

$$\mathcal{R}_i \mathcal{D}_i^{-1/2} \hat{\boldsymbol{c}}_j = \boldsymbol{c}_j - \boldsymbol{c}_i, \tag{3.49}$$

The equation for $\hat{f}_j$ reduces to

$$\hat{f}_j(\hat{\boldsymbol{x}}) = (\hat{\boldsymbol{x}} - \hat{\boldsymbol{c}}_j)^T \hat{\mathcal{Q}}_j (\hat{\boldsymbol{x}} - \hat{\boldsymbol{c}}_j) - 1 = 0, \tag{3.50}$$

where

$$\hat{\mathcal{Q}}_j = \hat{\mathcal{R}}_j \hat{\mathcal{D}}_j \hat{\mathcal{R}}_j^T \equiv \mathcal{D}_i^{-1/2} \mathcal{R}_i^T \mathcal{Q}_j \mathcal{R}_i \mathcal{D}_i^{-1/2} = \mathcal{D}_i^{-1/2} \mathcal{R}_i^T \mathcal{R}_j \mathcal{D}_j \mathcal{R}_j^T \mathcal{R}_i \mathcal{D}_i^{-1/2}.$$

For the sake of completeness, we remark that the coefficients of $\hat{\mathcal{Q}}_j$, i.e.

$$\hat{\mathcal{Q}}_j = \begin{bmatrix} \hat{A}_j & \hat{C}_j \\ \hat{C}_j & \hat{B}_j \end{bmatrix}$$

can be found explicitly as

$$\begin{aligned}
\hat{A}_j &= a_i^2 \left(\frac{\cos^2(\theta_j - \theta_i)}{a_j^2} + \frac{\sin^2(\theta_j - \theta_i)}{b_j^2} \right), \\
\hat{B}_j &= b_i^2 \left(\frac{\sin^2(\theta_j - \theta_i)}{a_j^2} + \frac{\cos^2(\theta_j - \theta_i)}{b_j^2} \right), \\
\hat{C}_j &= a_i b_i \left(\frac{1}{a_j^2} - \frac{1}{b_j^2} \right) \cos(\theta_j - \theta_i) \sin(\theta_j - \theta_i),
\end{aligned}$$

and the parameters $\{\hat{a}_j, \hat{b}_j, \hat{\theta}_j\}$ associated with ellipse $\hat{\mathcal{E}}_j$ can be recovered by identification from the Formulas (2.10) or by computing the eigenvalues and eigenvectors of $\hat{\mathcal{Q}}_j$.

The following lemma explains that the minimization problem in the new coordinates is related to a minimization in the original coordinates.

Lemma 10. *In the coordinate system $(\hat{O}, \hat{x}, \hat{y})$, the Euclidean distance is the same as the distance with respect to the $\mathcal{E}_i$-norm in the original coordinates (O, x, y).*

Proof. For every $\boldsymbol{x}$, its image $\hat{\boldsymbol{x}}$ is computed as

$$\mathcal{D}^{1/2}\mathcal{R}_i^T(\boldsymbol{x} - \boldsymbol{c}_i) = \hat{\boldsymbol{x}}.$$

We then have, for arbitrary $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$,

$$\begin{aligned}
\left\|\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\right\|^2 &= \left(\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\right)^T\left(\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\right) \\
&= \left(\mathcal{D}_i^{1/2}\mathcal{R}_i^T(\boldsymbol{x} - \boldsymbol{c}_i) - \mathcal{D}_i^{1/2}\mathcal{R}_i^T(\boldsymbol{y} - \boldsymbol{c}_i)\right)^T\left(\mathcal{D}_i^{1/2}\mathcal{R}_i^T(\boldsymbol{x} - \boldsymbol{c}_i) - \mathcal{D}_i^{1/2}\mathcal{R}_i^T(\boldsymbol{y} - \boldsymbol{c}_i)\right) \\
&= \left(\left(\boldsymbol{x} - \boldsymbol{c}_i\right) - \left(\boldsymbol{y} - \boldsymbol{c}_i\right)\right)^T\left(\mathcal{D}_i^{1/2}\mathcal{R}_i^T\right)^T\mathcal{D}_i^{1/2}\mathcal{R}_i^T\left(\left(\boldsymbol{x} - \boldsymbol{c}_i\right) - \left(\boldsymbol{y} - \boldsymbol{c}_i\right)\right) \\
&= \left(\boldsymbol{x} - \boldsymbol{y}\right)^T\mathcal{R}_i\mathcal{D}_i\mathcal{R}_i^T\left(\boldsymbol{x} - \boldsymbol{y}\right) \\
&= \left\|\boldsymbol{x} - \boldsymbol{y}\right\|_{\mathcal{Q}_i}^2.
\end{aligned}$$

□

3.7.2 Mapping of $(\mathcal{E}_i, \mathcal{E}_j)$ into a Unit Circle $\bar{\mathcal{C}}_i$ and an Ellipse $\bar{\mathcal{E}}_j$ Centered at Origin

Džiugys and Peters [24] suggested another mapping such that the ellipse $\hat{\mathcal{E}}_j$ introduced above is now positioned in its local reference system denoted here as $(\bar{O}, \bar{x}, \bar{y})$. We will describe this new mapping as the previous mapping followed by a rotation and a translation sending the previous ellipse $\hat{\mathcal{E}}_j$ to the origin with axes aligned with $(\bar{O}, \bar{x}, \bar{y})$. The circle $\hat{\mathcal{C}}_i$, previously at the origin under the mapping of Section 3.7.1, is now shifted around the ellipse centered at the origin, see Figure 3.8.

The mapping amounts to considering the transformation that maps $\bar{\boldsymbol{x}}$ into $\hat{\boldsymbol{x}}$ in $\mathbb{R}^2$ by the rotation $\hat{\mathcal{R}}_j$ and translation $\hat{\boldsymbol{c}}_j$:

$$\hat{\boldsymbol{x}} = \hat{\mathcal{R}}_j\bar{\boldsymbol{x}} + \hat{\boldsymbol{c}}_j, \tag{3.51}$$

so that circle $\hat{\mathcal{C}}_i$, located at the origin, is now mapped into circle $\bar{\mathcal{C}}_i$ with center $\bar{\boldsymbol{c}}_i$ as

$$\bar{\boldsymbol{c}}_i = -\hat{\mathcal{R}}_j^T\hat{\boldsymbol{c}}_j.$$

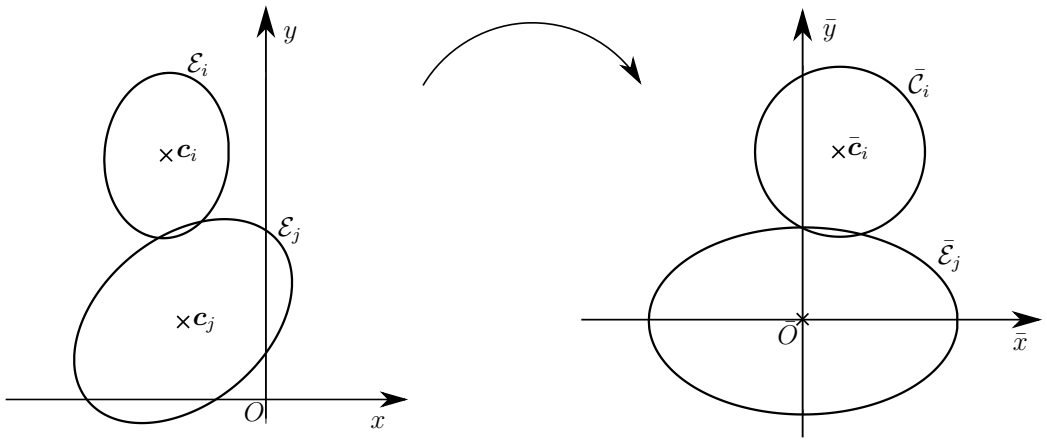


Figure 3.8 Mapping of $(\mathcal{E}_i, \mathcal{E}_j)$ into a unit circle $\bar{\mathcal{C}}_i$ and an ellipse $\bar{\mathcal{E}}_j$ centered at origin.

The equations of the two ellipses in the new coordinate system $(\bar{O}, \bar{x}, \bar{y})$ are then given by

$$\begin{aligned} \bar{f}_i(\bar{\boldsymbol{x}}) &= (\bar{\boldsymbol{x}} - \bar{\boldsymbol{c}}_i)^T(\bar{\boldsymbol{x}} - \bar{\boldsymbol{c}}_i) - 1 = 0, \\ \bar{f}_j(\bar{\boldsymbol{x}}) &= \bar{\boldsymbol{x}}^T \hat{\mathcal{D}}_j \bar{\boldsymbol{x}} - 1 = 0. \end{aligned} \tag{3.52}$$

In other words, the original ellipse $\mathcal{E}_j$ is now transformed into the ellipse $\bar{\mathcal{E}}_j$ centered at the origin and the ellipse $\mathcal{E}_i$ is mapped into the unit circle $\bar{\mathcal{C}}_i$ centered at $\bar{\boldsymbol{c}}_i$ using the global transformation obtained by combining (3.47) and (3.51)

$$\boldsymbol{x} = \mathcal{R}_i \mathcal{D}_i^{-1/2}(\hat{\mathcal{R}}_j \bar{\boldsymbol{x}} + \hat{\mathbf{c}}_j) + \boldsymbol{c}_i = \mathcal{R}_i \mathcal{D}_i^{-1/2} \hat{\mathcal{R}}_j \bar{\boldsymbol{x}} + \left(\mathcal{R}_i \mathcal{D}_i^{-1/2} \hat{\mathbf{c}}_j + \boldsymbol{c}_i\right) = \mathcal{R}_i \mathcal{D}_i^{-1/2} \hat{\mathcal{R}}_j \bar{\boldsymbol{x}} + \boldsymbol{c}_j, \tag{3.53}$$

where, we have used (3.49) to obtain the last expression.

Remark 6. *The transformation of an arbitrary pair of ellipses into an ellipse in its local coordinate system and a circle can be useful to simplify the mathematical analysis of the pair of ellipses. As an example, we can easily show that the co-gradient locus is a hyperbola. Recalling the co-gradient function* (3.20)*, the equation of the co-gradient locus reads*

$$H(\boldsymbol{x}) = 4(\boldsymbol{x} - \boldsymbol{c}_i)^T \mathcal{Q}_i A \mathcal{Q}_j (\boldsymbol{x} - \boldsymbol{c}_j) = 0,$$

where in this particular case

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \mathcal{Q}_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \mathcal{Q}_j = \begin{bmatrix} 1/a_j^2 & 0 \\ 0 & 1/b_j^2 \end{bmatrix}, \qquad \boldsymbol{c}_i = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \qquad \boldsymbol{c}_j = \mathbf{0}.$$

The equation can be rewritten as

$$\boldsymbol{x}^T M(\boldsymbol{x} - \boldsymbol{c}_i) = \boldsymbol{x}^T M\boldsymbol{x} - \boldsymbol{x}^T M\boldsymbol{c}_i = 0,$$

where matrix M *is given by*

$$M = \begin{bmatrix} 0 & -1/a_j^2 \\ 1/b_j^2 & 0 \end{bmatrix}.$$

Developing the above equation leads to

$$(a_j^2 - b_j^2)xy + b_j^2 c_y x - a_j^2 c_x y = 0. \tag{3.54}$$

This is actually the equation of a hyperbola in the case that $a_i \neq b_i$*. Indeed, using classical formulas, the center of the hyperbola, is given by*

$$x_h = \frac{a_j^2}{a_j^2 - b_j^2} c_x,$$
$$y_h = \frac{b_j^2}{b_j^2 - a_j^2} c_y,$$

and, using the change of variables $\xi = x - x_h$ *and* $\eta = y - y_h$*, the equation can be recast as*

$$\xi\eta = \frac{a_j^2 b_j^2}{(a_j^2 - b_j^2)^2} c_x c_y.$$

In the case that $a_j = b_j$*, the locus reduces to a straight line passing through the origin (i.e. the center of ellipse* $\mathcal{E}_j$*, which is a circle here) and the center of circle* $\mathcal{C}_i$*.*

CHAPTER 4 CONTACT DETECTION ALGORITHMS

The main goal of this chapter is to review the main contact detection algorithms for pairs of ellipses that have been proposed in the literature and, more specifically, recast the methods as minimization problems such as those satisfied by the MDP and MPP introduced in Chapter 3. Many of the algorithms to be discussed were not explicitly defined as minimization problems (with or without explicit constraints) and they were often categorized differently by the researchers themselves. Since it is ultimately our hope to better highlight the similarities and differences between the published algorithms, it is incumbent on us to introduce a new classification which may conflict with those found in the literature. Whenever possible, we will indicate those conflicts in naming and justify the new terms.

The framework will consist of a pair of ellipses $\mathcal{E}_i$, $\mathcal{E}_j \subset \mathbb{R}^2$ in near perfect contact, with or without overlap, as defined in Definition 6, but we shall discuss, when deemed necessary, the behavior of the algorithms for other configurations of the ellipses. A common feature of all algorithms presented here is that they compute two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ on the ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$, respectively. Then the *contact point* is defined as the midpoint between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$

$$\boldsymbol{x}_c = \frac{1}{2}\Big(\boldsymbol{x}_i + \boldsymbol{x}_j\Big), \tag{4.1}$$

and allows one to compute a penetration (or separation) distance $d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$. Moreover, one identifies the contact normal as unit vector of the average vector $\boldsymbol{n}_c$ computed from the normal vector $\boldsymbol{n}_i(\boldsymbol{x}_i)$ and the opposite normal vector $\boldsymbol{n}_j(\boldsymbol{x}_j)$

$$\boldsymbol{n}_c(\boldsymbol{x}_c) = \frac{\boldsymbol{n}_i(\boldsymbol{x}_i) - \boldsymbol{n}_j(\boldsymbol{x}_j)}{\|\boldsymbol{n}_i(\boldsymbol{x}_i) - \boldsymbol{n}_j(\boldsymbol{x}_j)\|}, \tag{4.2}$$

or

$$\boldsymbol{n}_c(\boldsymbol{x}_c) = \frac{\boldsymbol{n}_i(\boldsymbol{x}_c) - \boldsymbol{n}_j(\boldsymbol{x}_c)}{\|\boldsymbol{n}_i(\boldsymbol{x}_c) - \boldsymbol{n}_j(\boldsymbol{x}_c)\|}. \tag{4.3}$$

In the case of finding the intersection of two ellipses, the normal is defined as a normal to the line passes through the intersection set [41]. Moreover, the computation of normal vector is proposed as a vector passes through the centers of two tangent circles at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ [47]. The tangent line is then defined as the line perpendicular to $\boldsymbol{n}_c$. As the relative positions of the ellipses approach that of perfect contact, then the points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ coalesce, the distance d_{ij} vanishes, and the tangent line approaches to those of both $\mathcal{E}_i$ and $\mathcal{E}_j$.

4.1 Intersection Algorithm (IA)

The Intersection Algorithm was introduced by Rothenburg et al. in [41], and relies on estimating the points in the intersection set $\mathcal{I}_{ij} = \mathcal{E}_i \cap \mathcal{E}_j$ introduced in Definition 2. It is perhaps the most intuitive algorithm. We believe that it needs to be included in this chapter, despite some of its drawbacks and limitations described below.

The Intersection Algorithm can be cast as the minimization problem (3.2), which, from a practical point of view, consists in solving the system of equations:

$$\begin{cases} f_i(x,y) = A_i x^2 + B_i y^2 + 2C_i xy + 2D_i x + 2E_i y + F_i = 0, \\ f_j(x,y) = A_j x^2 + B_j y^2 + 2C_j xy + 2D_j x + 2E_j y + F_j = 0, \end{cases} \tag{4.4}$$

where f_i and f_j from Equation (2.13) are the global geometric potentials of $\mathcal{E}_i$ and $\mathcal{E}_j$, respectively. A priori, we make no assumptions concerning the two ellipses. If $\mathcal{E}_i$ and $\mathcal{E}_j$ do not coincide, the system of equations can naturally be reduced into a single quartic equation in the first coordinate x (alternatively, in the second coordinate y):

$$\sum_{k=0}^{4} a_k x^k = 0. \tag{4.5}$$

where the coefficients a_k, $k = 0, \ldots, 4$, of the polynomial can be explicitly given in terms of the coefficients of (4.4), see [41] or Appendix A. The quartic equation (4.5) admits at most four real roots x_ℓ, $\ell = 1, \ldots, 4$. Moreover, for each root x_ℓ, one can use an explicit formula (see [41]) to compute the corresponding coordinate y_ℓ.

Different configurations of the ellipses will provide a different number of solutions to the quartic equation (4.5):

1. There are no real root. This is the case when the two ellipses are disjoint, i.e. $\mathcal{I}_{ij} = \varnothing$, whether they are disjoint with non-penetrating CoM or one ellipse lies inside the other (penetrating CoM). In this case, the algorithm does not provide information about the penetration distance.

2. There is one real root of multiplicity two. This is the case when there is a point at which the two ellipses are in perfect contact or when the x-coordinate of two distinct intersection points coincide. In the latter case, one cannot compute the y-coordinate of the two distinct points using the explicit formula in [41]. Instead, given the common x-coordinate x_ℓ, one should solve the quadratic equation in y using the global potential

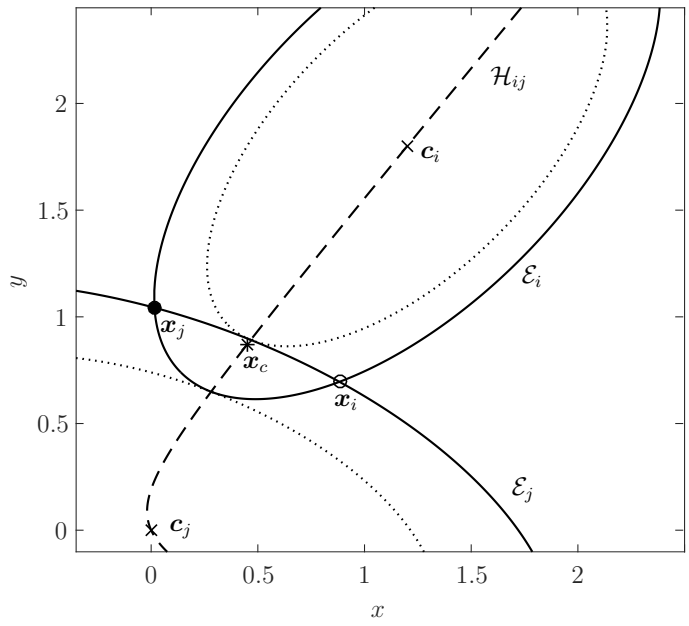


Figure 4.1 The points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are the intersection points of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$, i.e. $\mathcal{I}_{ij} = \{\boldsymbol{x}_i, \boldsymbol{x}_j\}$ from Definition 2. The contact point $\boldsymbol{x}_c$ between the ellipses is obtained by the Intersection Algorithm (IA).

of ellipse $\mathcal{E}_i$,

$$B_i y^2 + (2C_i x_\ell + 2E_i)y + (A_i x_\ell^2 + 2D_i x_\ell + F_i) = 0.$$

Alternatively, one could consider solving the quartic equation in y to obtain two distinct roots y_k. If a single root in x corresponds to a single point (x, y), then the separation distance and the normal can be computed easily.

3. There are two real distinct roots only. This is the most common configuration if the ellipses are in near perfect contact with overlap.

   If $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are two intersection points, the length $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ provides an approximation of the length of the overlap. However, additional work needs to be done to provide an estimate of the penetration, see more details in [41].

4. The four roots of the polynomial are all real, either all distinct, or two distinct roots and one root of multiplicity two, or two roots of multiplicity two. In practice, these cases are unlikely to occur in DEM applications since all are indicative of relatively large overlaps between ellipses.

Drawbacks in the Intersection Algorithm are that one needs to compute all real roots of the

quartic polynomial and handle all specific cases depending on the number of roots and their multiplicity. A major issue though is that the algorithm is prone to numerical instabilities [41] in the case of very small overlaps between ellipses when estimating the two points $\{\boldsymbol{x}_i, \boldsymbol{x}_j\}$ and the normal direction. For these reasons, it is usually not the recommended approach for contact detection.

Moreover, the extension of the Intersection Algorithm to 3-D is increasingly more elaborate since the intersection set $\mathcal{I}_{ij}$ consists of 2-D ellipses on the surface of the ellipsoids. Given the weaknesses of this approach, we will not be providing further details on its extension to 3-D. However, we refer the reader to Ouadfel and Rothenburg [42] who proposed an algorithm in 3-D to find the contact point and contact normal.

4.2 Geometric Potential Algorithm (GPA)

The Geometric Potential Algorithm was first described by Ng et al. [43, 71] and has been further improved by Ting et al. [46], Mustoe and Miyata [47], and Džiugys and Peters [24]. The GPA is based on the symmetric pair of minimization problems (3.8)-(3.9) that we recall here for convenience

$$\boldsymbol{x}_i = \underset{\boldsymbol{x}\in\mathcal{E}_i}{\operatorname{argmin}}\, f_j(\boldsymbol{x}), \tag{4.6}$$

$$\boldsymbol{x}_j = \underset{\boldsymbol{x}\in\mathcal{E}_j}{\operatorname{argmin}}\, f_i(\boldsymbol{x}). \tag{4.7}$$

For any pair of ellipses in near perfect contact the two above problems have unique solutions, see Lemma 4.

As the numerical experiments will show, the GPA is quite robust, even for pairs of ellipses with high aspect ratios, and at a computational cost competitive with other methods. However, two distinct problems must be solved and each problem generates up to four critical points from which the global minimum must be found. Instabilities in the method may appear during the root-finding step, although they have a geometric source.

We present several numerical implementations of the GPA, including Lagrangian and parametric formulations. Penalization has not been found to be an effective means of solving (3.8)-(3.9) because the wide range of values of volume and aspect ratio make the choice of the penalization parameter difficult. In practice, the parametric implementation of the GPA is fastest but requires slightly more analytic effort by the user. In GPA algorithm, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are found by solving two distinct problems in the same way. Therefore, in the following, we present only finding the point $\boldsymbol{x}_j$. The point $\boldsymbol{x}_i$ is found in a similar fashion.

For convenience, the various algorithms and the problems in GPA are summarized in Table

4.1. We also include in the table the Constrained GPA described in Section 4.3.

Table 4.1 Geometric potential algorithms with associated minimization problems

Algorithm	Minimization Problem	Reference
Lagrangian GPA (L-GPA)	$\boldsymbol{x}_i = \underset{\boldsymbol{x}\in\mathcal{E}_i}{\operatorname{argmin}} f_j(\boldsymbol{x})$ $\boldsymbol{x}_j = \underset{\boldsymbol{x}\in\mathcal{E}_j}{\operatorname{argmin}} f_i(\boldsymbol{x})$	Lin and NG, 1995 [43]
Parametric GPA (P-GPA)	$t_i = \underset{t\in[-\pi,\pi[}{\operatorname{argmin}} \widehat{f}_j(t)$ $t_j = \underset{t\in[-\pi,\pi[}{\operatorname{argmin}} \widehat{f}_i(t)$	Mustoe and Miyata, 2001 [47]
Mapped GPA (M-GPA)	$\bar{\boldsymbol{x}}_i = \underset{\bar{\boldsymbol{x}}\in\bar{\mathcal{E}}_i}{\operatorname{argmin}} \|\bar{\boldsymbol{x}} - \bar{\boldsymbol{c}}_j\|^2$ $\bar{\boldsymbol{x}}_j = \underset{\bar{\boldsymbol{x}}\in\bar{\mathcal{E}}_j}{\operatorname{argmin}} \|\bar{\boldsymbol{x}} - \bar{\boldsymbol{c}}_i\|^2$	Džiugys and Peters, 2001 [24]
Constrained GPA (C-GPA)	$\hat{\boldsymbol{x}}_i = \underset{\hat{\boldsymbol{x}}\in\hat{\mathcal{C}}_i\cap\mathcal{H}_{ij}}{\operatorname{argmin}} \hat{f}_j(\boldsymbol{x})$ $\hat{\boldsymbol{x}}_j = \underset{\hat{\boldsymbol{x}}\in\hat{\mathcal{C}}_j\cap\mathcal{H}_{ij}}{\operatorname{argmin}} \hat{f}_i(\boldsymbol{x})$	Ting et al., 1993 [46]

4.2.1 Lagrangian GPA (L-GPA)

We describe the solution method proposed in [43] for the constrained minimization problem (4.7) to find $\boldsymbol{x}_j$. We thus introduce

$$\mathcal{L}_j(\boldsymbol{x},\lambda) = f_i(\boldsymbol{x}) - \lambda f_j(\boldsymbol{x}), \qquad \forall \boldsymbol{x}\in\mathbb{R}^2,\ \forall\lambda\in\mathbb{R},$$

where the constraint $\boldsymbol{x}\in\mathcal{E}_j$, i.e. $f_j(\boldsymbol{x}) = 0$, is enforced via the Lagrange multiplier λ. Using the representation (2.1) for f_k, $k = i, j$, i.e.

$$\mathcal{L}_j(\boldsymbol{x},\lambda) = (\boldsymbol{x}-\boldsymbol{c}_i)^T\mathcal{Q}_i(\boldsymbol{x}-\boldsymbol{c}_i) - \lambda(\boldsymbol{x}-\boldsymbol{c}_j)^T\mathcal{Q}_j(\boldsymbol{x}-\boldsymbol{c}_j),$$

the stationary points $(\boldsymbol{x},\lambda)$ of the Lagrangian $\mathcal{L}_i$ satisfy the system:

$$0 = \partial_{\boldsymbol{x}}\mathcal{L}_i(\boldsymbol{x},\lambda) = 2\mathcal{Q}_i(\boldsymbol{x}-\boldsymbol{c}_i) - 2\lambda\mathcal{Q}_j(\boldsymbol{x}-\boldsymbol{c}_j), \tag{4.8}$$

$$0 = \partial_{\lambda}\mathcal{L}_i(\boldsymbol{x},\lambda) = (\boldsymbol{x}-\boldsymbol{c}_j)^T\mathcal{Q}_j(\boldsymbol{x}-\boldsymbol{c}_j) - 1. \tag{4.9}$$

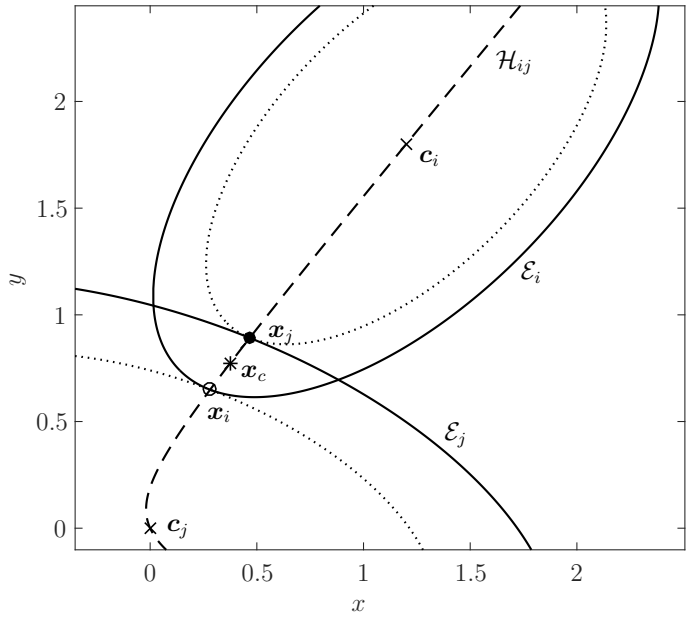


Figure 4.2 The points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are obtained by the Geometric Potential Algorithm which provides MPP, i.e. $(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Note that the contact point $\boldsymbol{x}_c$ does not necessary belong to gradient locus $\mathcal{H}_{ij}$.

Isolating $\boldsymbol{x}$ as a function of λ from (4.8), we find

$$\boldsymbol{x}(\lambda) = (\mathcal{Q}_i - \lambda \mathcal{Q}_j)^{-1}(\mathcal{Q}_i \boldsymbol{c}_i - \lambda \mathcal{Q}_j \boldsymbol{c}_j). \tag{4.10}$$

Substituting (4.10) for $\boldsymbol{x}(\lambda)$ in (4.9) produces a quartic polynomial in λ:

$$\sum_{k=1}^{4} a_k \lambda^k = 0, \tag{4.11}$$

whose coefficients a_k can be computed explicitly and can be found in [43] or Appendix B.1.1. Solving (4.11) provides at least two and at most four real roots λ_ℓ, which in turn yields four candidate points $\boldsymbol{x}(\lambda_\ell)$, $\ell = 1, \ldots, 4$, according to (4.10). Then the point $\boldsymbol{x}_j$ is selected as the point that minimizes $f_i(\boldsymbol{x})$ among the points $\boldsymbol{x}(\lambda_\ell)$. Extension of the Lagrangian approach to the case of ellipsoids is straightforward and results in a root-finding problem equivalent to (4.11), but involving a polynomial of degree 6 (see [43] or Appendix B.1.2 for details).

4.2.2 Parametric GPA (P-GPA)

We briefly describe here the approach proposed in [47] to solve the minimization problem (4.7). The main idea is to use the parametric representation (2.15) of an ellipse in order to eliminate the constraint from the minimization problem. This technique works in both two and three dimensions. Let the local potential $\widehat{f}_i$ of $\mathcal{E}_i$ be given as in (2.5), i.e.

$$\widehat{f}_i(\boldsymbol{\xi}) = \boldsymbol{\xi}^T \mathcal{D}_i \boldsymbol{\xi} - 1, \tag{4.12}$$

where

$$\mathcal{D}_i = \begin{bmatrix} 1/a_i^2 & 0 \\ 0 & 1/b_i^2 \end{bmatrix}.$$

The points $\boldsymbol{x}$ on $\mathcal{E}_j$ can be parameterized in the local coordinate system by

$$\boldsymbol{\zeta}(t) = (a_j \cos t, b_j \sin t),$$

for $t \in [-\pi, \pi[$. Using (2.4), the coordinates of $\boldsymbol{x}$ in the local reference system (O, ξ, η) associated with $\mathcal{E}_i$ are then given by:

$$\boldsymbol{\xi}(t) = \mathcal{R}_i^T \Big[(\mathcal{R}_j \boldsymbol{\zeta}(t) + \boldsymbol{c}_j) - \boldsymbol{c}_i \Big] = \mathcal{R}_i^T \mathcal{R}_j \boldsymbol{\zeta}(t) + \boldsymbol{\xi}_0 \,, \tag{4.13}$$

with $\boldsymbol{\xi}_0 = (\xi_0, \eta_0) = \mathcal{R}_i^T(\boldsymbol{c}_j - \boldsymbol{c}_i)$. Replacing $\boldsymbol{\xi}(t)$ in (4.12) by (4.13), one can then express $\widehat{f}_i$ as a function of parameter t only, i.e.

$$\widehat{f}_i(t) = \boldsymbol{\xi}(t)^T D_i \boldsymbol{\xi}(t) - 1 = [\mathcal{R}_i^T \mathcal{R}_j \boldsymbol{\zeta}(t) + \boldsymbol{\xi}_0]^T \mathcal{D}_i [\mathcal{R}_i^T \mathcal{R}_j \boldsymbol{\zeta}(t) + \boldsymbol{\xi}_0] - 1,$$

which can be reduced to

$$\widehat{f}_i(t) = \frac{\Big(a_j \cos\theta_{ij} \cos t - b_j \sin\theta_{ij} \sin t + \xi_0\Big)^2}{a_i^2} + \frac{\Big(b_j \cos\theta_{ij} \sin t + a_j \sin\theta_{ij} \cos t + \eta_0\Big)^2}{b_i^2} - 1.$$

where $\theta_{ij} = \theta_j - \theta_i$, with θ_k for $k = i, j$, being the angle of rotation of $\mathcal{R}_k$, as defined in (2.2). It follows that the constrained minimization problem (3.9) for $\boldsymbol{x}_j$ can be recast into the unconstrained minimization problem in one variable

$$t_j = \underset{t \in [-\pi, \pi[}{\operatorname{argmin}} \widehat{f}_i(t), \tag{4.14}$$

from which one would obtain the point $\boldsymbol{x}_j \in \mathcal{E}_j$ by the change of variable (2.4):

$$\boldsymbol{x}_j = \mathcal{R}_j \boldsymbol{\zeta}(t_j) + \boldsymbol{c}_j. \tag{4.15}$$

The nonlinear functions $\widehat{f}_k(t)$, $k = i, j$, may have several extrema, which makes root-finding algorithms for $(\widehat{f}_k)'(t) = 0$ difficult. For example, in Figure 4.4, we plot $\widehat{f}_i(t)$ and $\widehat{f}_j(t)$ associated with the pair of ellipses $\mathcal{E}_i$ and ellipse $\mathcal{E}_j$ shown in Figure 4.3. In particular, we observe that the function $\widehat{f}_j(t)$ has two minima and two maxima. In order words, without any additional constraint, one needs to search for all extrema in order to find the global minimum.

4.2.3 Mapped GPA (M-GPA)

Džiugys and Peters [24] proposed an alternative approach by introducing the mapping described in Section 3.7.2, which transforms $\mathcal{E}_i$ into a unit circle $\hat{\mathcal{C}}_i$ and the other ellipse $\mathcal{E}_j$ into an ellipse $\hat{\mathcal{E}}_j$ in its local reference system by the same mapping. Assuming the map has been applied and dropping the hat symbols, the circle and ellipse are then given by

$$\begin{aligned} f_i(\boldsymbol{x}) &= (\boldsymbol{x} - \boldsymbol{c}_i)^T (\boldsymbol{x} - \boldsymbol{c}_i) - 1 = 0, \\ f_j(\boldsymbol{x}) &= \boldsymbol{x}^T \mathcal{D}_j \boldsymbol{x} - 1 = 0. \end{aligned}$$

We first observe that the $\mathcal{E}_i$-norm is simply the Euclidean norm as the ellipse $\mathcal{E}_i$ is now reduced to the circle $\mathcal{C}_i$; see Lemma 10. It follows that the minimization problem (4.7) now reads

$$\boldsymbol{x}_j = \underset{\boldsymbol{x} \in \mathcal{E}_j}{\operatorname{argmin}} f_i(\boldsymbol{x}) = \underset{\boldsymbol{x} \in \mathcal{E}_j}{\operatorname{argmin}} \|\boldsymbol{x} - \boldsymbol{c}_i\|^2. \tag{4.16}$$

In other words, the problem is to find the point $\boldsymbol{x}_j$ on ellipse $\mathcal{E}_j$ that is the closest to $\boldsymbol{c}_i$ in Euclidean norm.

Džiugys and Peters [24, 72] proposed to solve the minimization problem (4.16) using two different approaches, one based on an iteration method and the other based on partial analytical results. We will present only the latter approach below. Introducing the distance ρ_i from the center of circle $\mathcal{C}_i$ to any point $\boldsymbol{x} \in \mathcal{E}_j$

$$\rho_i(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{c}_i\|, \tag{4.17}$$

they enforce the constraint $\boldsymbol{x} \in \mathcal{E}_j$ explicitly by rewriting ρ_i as a function of x only, using

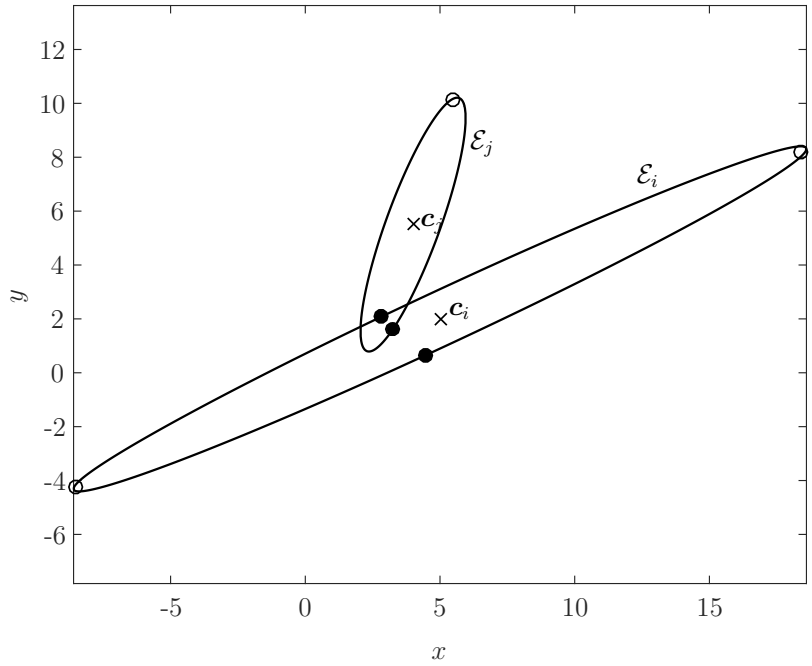


Figure 4.3 The configuration of two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ in which the minimization Problem (4.14) has non-unique minimum, which is illustrated in Figure 4.4. The ellipses are $\{a_i = 15, b_i = 1, \boldsymbol{c}_i = (5, 2), \theta_i = 0.4363\}$ and $\{a_j = 5, b_j = 1, \boldsymbol{c}_j = (4, 5.5), \theta_j = 1.2217\}$. The points which are corresponding to minimum and maximum are shown by a dot ($\bullet$) and a disc($\circ$), respectively.

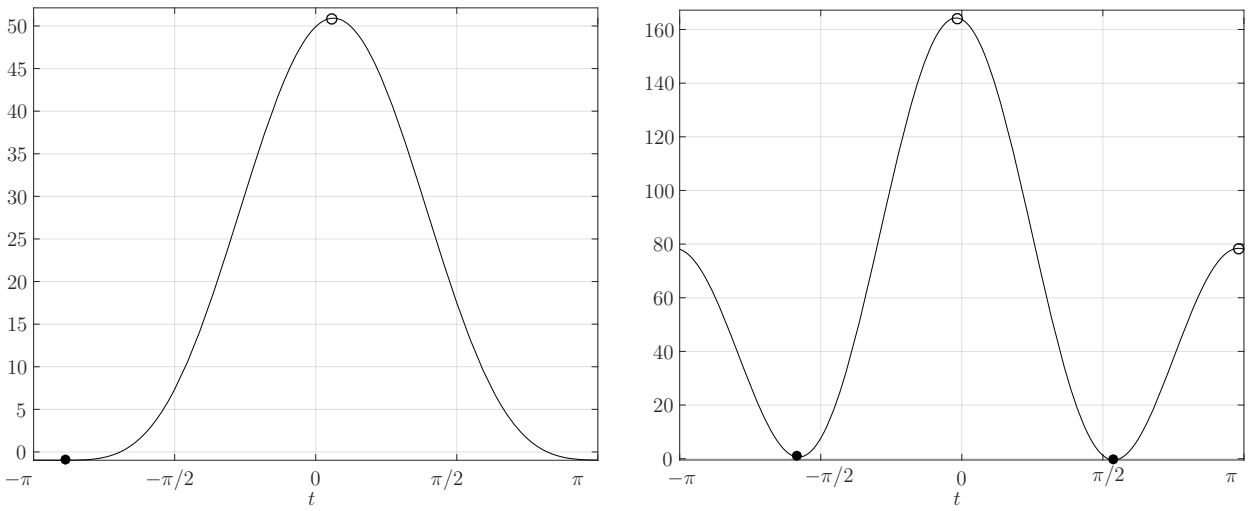


Figure 4.4 The graphs of functions $\widehat{f}_i(t)$ (left) and $\widehat{f}_j(t)$ (right) are associated with the ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ shown in Figure 4.3. The function $\widehat{f}_i(t)$ has exactly one minimum and one maximum. However, the function $\widehat{f}_j(t)$ has two local minima and maxima for $t \in [-\pi, \pi[$. In both graphs, the minima and maxima are represented by a dot ($\bullet$) and a disc($\circ$), respectively.

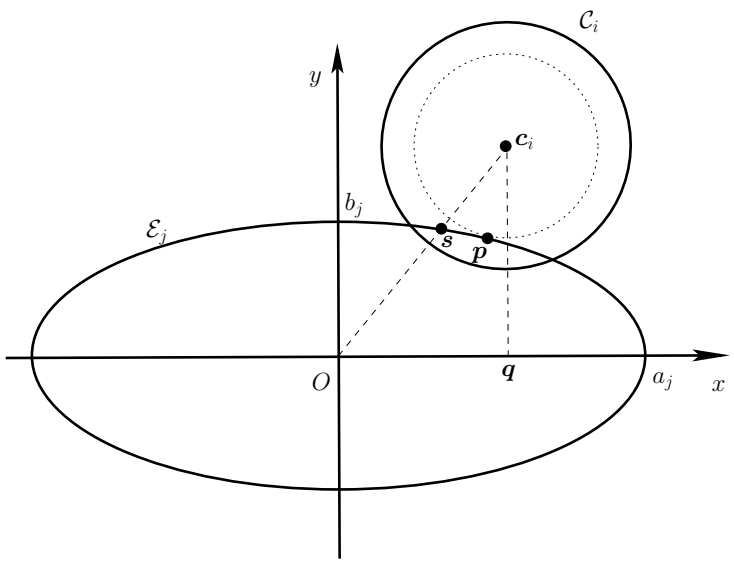


Figure 4.5 The initial point $\boldsymbol{p}$ is any point which satisfies conditions 4.24.

the equation of ellipse $\mathcal{E}_j$

$$\rho_i^2(x) = (x - c_x)^2 + (\pm\kappa\sqrt{a_j^2 - x^2} - c_y)^2, \qquad |x| \le a_j, \tag{4.18}$$

where $\kappa = b_j/a_j$ and $\boldsymbol{c}_i = (c_x, c_y)$. The goal is therefore to find the global minimizer of the functional ρ_i^2, i.e.

$$x_j = \underset{|x|\le a_j}{\operatorname{argmin}}\, \rho_i^2(x).$$

Finding the critical points of the minimization problem leads to solving a quartic equation in x:

$$\sum_{k=0}^{4} a_k x^k = 0, \quad \text{with} \quad \begin{cases} a_4 = (1-\kappa^2)^2, \\ a_3 = -2c_x(1-\kappa^2), \\ a_2 = c_x^2 + \kappa^2 c_y^2 - a_j^2(1-\kappa^2)^2, \\ a_1 = 2a_j^2 c_x(1-\kappa^2), \\ a_0 = -a_j^2 c_x^2. \end{cases} \tag{4.19}$$

The quartic equation has a maximum of four roots, with possibly some conjugate complex roots, which can be found using an iterative nonlinear solver. Džiugys and Peters [24] also suggested an approach in which the solution procedure could be reduced to solving a cubic equation after introducing a special change of variable. Once the roots x_ℓ, for $\ell = 1, \ldots, 4$ are found, one can compute the corresponding y_ℓ coordinate, and the point $\boldsymbol{x}_j$ is selected among the solutions $\boldsymbol{x}_\ell = (x_\ell, y_\ell)$ such that it minimizes $\rho_i(\boldsymbol{x})$ in (4.17).

Džiugys and Peters [24] proposed using iterative approaches and introduced some conditions for the initial point $\boldsymbol{p}$ as following, see Figure 4.5. If the center of circle $\boldsymbol{c}_i = (c_x, c_y)$, then the point q is defined as $(c_x, 0)$. The initial point $\boldsymbol{p}$ should be located inside the triangular $\boldsymbol{cOq}$. At the same time, $\boldsymbol{p}$ is on the ellipse $\mathcal{E}_j$ between s, the intersection of $\boldsymbol{cO}$, and the x-axis. In other words, the coordinate of x_p and y_p of point $\boldsymbol{p}$ should satisfy the following conditions

$$\operatorname{sign}(x_p) = \operatorname{sign}(c_x), \tag{4.20}$$

$$\operatorname{sign}(y_p) = \operatorname{sign}(c_y), \tag{4.21}$$

$$|x_p| \leq \min(|c_x|, a_j), \tag{4.22}$$

$$|x_p| \geq \frac{|c_x|}{\sqrt{c_x^2/{a_j}^2 + c_y^2/{b_j}^2}}, \tag{4.23}$$

$$|y_p| \leq \frac{|c_y|}{\sqrt{c_x^2/{a_j}^2 + c_y^2/{b_j}^2}}. \tag{4.24}$$

We have described how to solve the minimization problem (4.16) for $\boldsymbol{x}_j \in \mathcal{E}_j$, but applying the map of Section 3.7.2 by switching $\mathcal{E}_i$ and $\mathcal{E}_j$, we can then solve for $\boldsymbol{x}_i \in \mathcal{E}_i$ using the same procedure as above.

4.3 Constrained Geometric Potential Algorithm (C-GPA)

The constrained geometric potential algorithm was introduced by Ting and his collaborators [46, 67, 73]. It can be viewed as an extension to the GPA with the additional constraints on the normals (3.10) and (3.11) after applying the mapping described in Section 3.7.1. More specifically, it was shown in Lemma 4 that those two constraints are always satisfied at the GPA and in Theorem 6 this constraint defines a hyperbola, the so-called co-gradient locus. Although the additional constraint on the normals are non-binding and could simply be ignored, it does significantly modify how the problems are solved and should further stabilize the problems.

The point $\boldsymbol{x}_j$ (resp. $\boldsymbol{x}_i$) is then defined as the closest point with respect to the $\mathcal{E}_i$-norm (resp. $\mathcal{E}_j$-norm) to the center of ellipse $\mathcal{E}_i$ (resp. $\mathcal{E}_j$) that intersects the co-gradient locus $\mathcal{H}_{ij}$ and the ellipse $\mathcal{E}_j$ (resp. $\mathcal{E}_i$). Formally, the problems can be recast as the constrained minimization

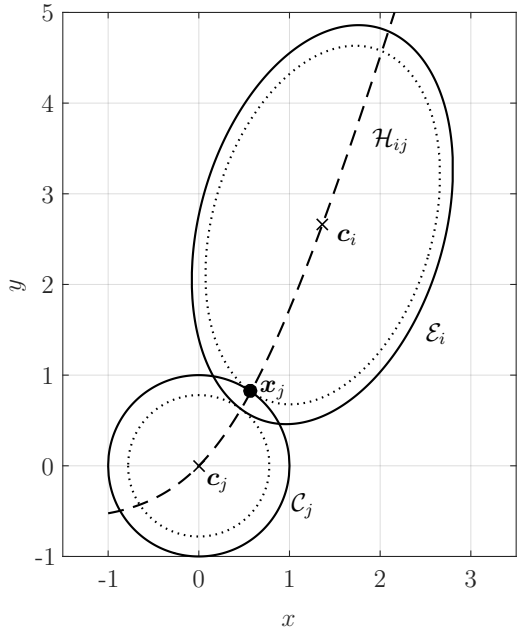


Figure 4.6 The two steps of the C-GPA are illustrated above. First, the mapping of Section 3.7.1 is applied to transform ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ to ellipse $\mathcal{E}_i$ the unit circle $\mathcal{C}_j$ with center at the origin. Second, the solution of Problem (4.26) provides the point $\boldsymbol{x}_j$ as the closest point to ellipse $\mathcal{E}_i$ among the points that intersect co-gradient locus $\mathcal{H}_{ij}$ and circle $\mathcal{C}_j$.

problems:

$$\boldsymbol{x}_i = \underset{\boldsymbol{x}\in\mathcal{E}_i\cap\mathcal{H}_{ij}}{\operatorname{argmin}} f_j(\boldsymbol{x}), \tag{4.25}$$

$$\boldsymbol{x}_j = \underset{\boldsymbol{x}\in\mathcal{E}_j\cap\mathcal{H}_{ij}}{\operatorname{argmin}} f_i(\boldsymbol{x}). \tag{4.26}$$

The problems above are nevertheless different from the minimization problems (3.13) and (3.14). Indeed, the fact that a point $\boldsymbol{x}$ belongs to the co-gradient locus $\mathcal{H}_{ij}$ does not necessarily imply that $\boldsymbol{n}_i(\boldsymbol{x}) + \boldsymbol{n}_j(\boldsymbol{x}) = \boldsymbol{0}$ as one could also have $\boldsymbol{n}_i(\boldsymbol{x}) = \boldsymbol{n}_j(\boldsymbol{x})$.

The method proposed by Ting et al. in [46, 67] to solve Problems (4.25) and (4.26) aims at finding the intersection points between each ellipse and the locus $\mathcal{H}_{ij}$. Since the two branches of the hyperbola may cross twice each ellipse, the solutions $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are chosen as those that minimize f_i and f_j (equivalently, the $\mathcal{E}_i$- and $\mathcal{E}_j$-norms), respectively. In order to simplify the analysis to search for $\boldsymbol{x}_i$, Ting et al. in [46, 67] proposed to consider the transformation of Section 3.7.1 that maps $\mathcal{E}_i$ into the unit circle $\mathcal{C}_i$ centered at the origin. Similarly, the second point $\boldsymbol{x}_j$ is found by transforming the other ellipse $\mathcal{E}_j$ into the unit circle $\mathcal{C}_j$ while mapping the ellipse $\mathcal{E}_i$ into a new ellipse under the same transformation, see Figure 4.6.

We briefly describe the algorithm in the case of a unit circle $\mathcal{C}_j$ and an ellipse $\mathcal{E}_i$ given by

$$\begin{aligned} f_i(\boldsymbol{x}) &= (\boldsymbol{x}-\boldsymbol{c}_i)^T \mathcal{Q}_i (\boldsymbol{x}-\boldsymbol{c}_i) - 1 = 0, \\ f_j(\boldsymbol{x}) &= \boldsymbol{x}^T \boldsymbol{x} - 1 = 0. \end{aligned}$$

The equation of the co-gradient locus $\mathcal{H}_{ij}$, see Equation (3.20), is given in that case by

$$H(\boldsymbol{x}) = 4\boldsymbol{x}^T A \mathcal{Q}_i (\boldsymbol{x}-\boldsymbol{c}_i) = 0,$$

where A is the anti-symmetric matrix (3.19). Using the notation of Chapter 2, the co-gradient function can be written as

$$H(x,y) = 4C_i x^2 - 4C_i y^2 + 4(B_i - A_i)xy + 4E_i x - 4D_i y = 0. \tag{4.27}$$

The intersection of $\mathcal{H}_{ij}$ and $\mathcal{C}_j$ can be found by combining this equation with $x^2+y^2-1=0$ to obtain a single quartic equation in x

$$\sum_{k=0}^{4} a_k x^k = 0, \quad \text{with} \quad \begin{cases} a_4 = (A_i - B_i)^2 + 4C_i^2, \\ a_3 = 2(A_i - B_i)D_i + 4C_i E_i, \\ a_2 = -(A_i - B_i)^2 - 4C_i^2 + D_i^2 + E_i^2, \\ a_1 = -2(A_i - B_i)D_i - 2C_i E_i, \\ a_0 = C_i^2 - D_i^2. \end{cases} \tag{4.28}$$

Solving for (4.28) leads a maximum of four real roots x_ℓ, $\ell = 1, \ldots, 4$. Ting et al. [46, 67] compute the second coordinate y_ℓ from x_ℓ using the formula derived from the equations of the circle and ellipse:

$$y_\ell = \frac{C_i(2x_\ell^2 - 1) + E_i x_\ell}{(A_i - B_i)x_\ell + D_i}, \qquad \ell = 1, 2, \tag{4.29}$$

as long as the denominator does not vanish for x_ℓ. There may be an issue when the solution of (4.28) yields one real root of multiplicity two, which is the case when the two intersection points between the circle and the ellipse have the same coordinate x_ℓ. Similarly to IA, the computation of the y-coordinate using (4.29) would result in one intersection point. A remedy would be to use the equation of the circle to solve for y_ℓ

$$y_\ell = \pm\sqrt{1 - x_\ell^2}.$$

Actually, one could use this equation in all cases, which would provide two or four points and choose $\boldsymbol{x}_j$ that minimizes $f_i(\boldsymbol{x})$, i.e. the closest point to the center of ellipse $\mathcal{E}_i$ with respect

to the $\mathcal{E}_i$-norm.

4.4 Common Normal Algorithm (CNA)

The Common Normal Algorithm (CNA), first studied by Lin et al. [43], rewrites the condition on the normals (3.6), that is usually a consequence of the MDP definition (3.5), into a system of solvable equations for points $\boldsymbol{x}_i \in \mathcal{E}_i$ and $\boldsymbol{x}_j \in \mathcal{E}_j$. The condition on the normals, namely

$$\boldsymbol{n}_i(\boldsymbol{x}_i) + \boldsymbol{n}_j(\boldsymbol{x}_j) = \boldsymbol{0}, \tag{4.30}$$

is an obvious property of the minimum distance pair of closest points $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$, when the ellipses are disjoint, but it is interesting to remark that the condition is relevant even when the ellipses are in near perfect contact, while the minimization problem (3.5) is no longer applicable. Below, we will re-interpret the original formulation of the CNA into a minimization problem, before discussing how it can be implemented.

Let $\mathcal{E}_i$ and $\mathcal{E}_j$ be two arbitrary ellipses. For any given point $\boldsymbol{x}_i$ on $\mathcal{E}_i$, one can always identify a unique point $\boldsymbol{x}_j$ on $\mathcal{E}_j$ such that $\boldsymbol{n}_j(\boldsymbol{x}_j) + \boldsymbol{n}_i(\boldsymbol{x}_i) = \boldsymbol{0}$. This implies that the set of constraints $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathcal{E}_i \times \mathcal{E}_j$ and $\boldsymbol{n}_i(\boldsymbol{x}_i) + \boldsymbol{n}_j(\boldsymbol{x}_j) = \boldsymbol{0}$ still provides an underdetermined system and, more precisely one additional scalar constraint is required. In order to determine candidate pairs of points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, one needs to set additional constraints. Lin and Ng [43] proposed to consider that the unit vector going from $\boldsymbol{x}_i$ towards $\boldsymbol{x}_j$ be also equal to the normal vectors $\boldsymbol{n}_j(\boldsymbol{x}_j)$ and $-\boldsymbol{n}_i(\boldsymbol{x}_i)$. Introducing the unit vector

$$\boldsymbol{n}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|},$$

the problem of finding $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ would then consist in the following set of equations

$$\begin{cases} f_i(\boldsymbol{x}_i) = 0, \\ f_j(\boldsymbol{x}_j) = 0, \\ \boldsymbol{n}_j(\boldsymbol{x}_j) + \boldsymbol{n}_i(\boldsymbol{x}_i) = \boldsymbol{0}, \\ \boldsymbol{n}_j(\boldsymbol{x}_j) - \boldsymbol{n}_i(\boldsymbol{x}_i) = 2\boldsymbol{n}(\boldsymbol{x}_i, \boldsymbol{x}_j). \end{cases} \tag{4.31}$$

Unfortunately, the problem above presents a few issues, which can be clearly described in the case of two circles: 1) if the two circles are disjoint, the problem has a unique solution, but the distance between the two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ reaches a maximum rather than a minimum, meaning that it is not really the pair of points that one is looking for; 2) if the two circles

overlap, the problem admits two solutions, one solution for which the distance between the two points is a minimum and the other for which the distance is a maximum; 3) moreover, if the overlap between the two circles becomes very small, the calculation of the vector $\boldsymbol{n}$ becomes problematic and the problem is ill-posed in the limit case of perfect contact.

Finally, System (4.31) consists of six equations for the four variables $\boldsymbol{x}_i = (x_i, y_i)$ and $\boldsymbol{x}_j = (x_j, y_j)$.We illustrate the existence of two solutions in the system of the equations (4.31) in Figure 4.9. The method proposed by Lin and Ng [43] in 3-D would translate in the 2-D case into arbitrarily selecting the following four nonlinear equations only

$$\begin{cases} f_i(\boldsymbol{x}_i) = 0, \\ f_j(\boldsymbol{x}_j) = 0, \\ \dfrac{x_j - x_i}{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|} = -\dfrac{1}{\|\nabla f_i\|}\dfrac{\partial f_i}{\partial x}(\boldsymbol{x}_i), \\ \dfrac{x_j - x_i}{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|} = +\dfrac{1}{\|\nabla f_j\|}\dfrac{\partial f_j}{\partial x}(\boldsymbol{x}_j). \end{cases} \tag{4.32}$$

Such an arbitrary selection of equations introduce additional solutions, as seen in Figure 4.8. This is due to the fact that one should consider the direction of the normal vectors rather than simply equating their components in the x-direction. For all these reasons, the common normal algorithm is deemed inappropriate for finding contact points between ellipses or ellipsoids. As a final remark, Problem (4.31) is equivalent to the constrained minimization problem:

$$(\boldsymbol{x}_i, \boldsymbol{x}_j) = \underset{\substack{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j \\ \boldsymbol{n}_i(\hat{\boldsymbol{x}}_i) + \boldsymbol{n}_j(\hat{\boldsymbol{x}}_j) = \boldsymbol{0}}}{\operatorname{argmin}} \|\boldsymbol{n}_j(\hat{\boldsymbol{x}}_j) - \boldsymbol{n}_i(\hat{\boldsymbol{x}}_i) - 2\boldsymbol{n}(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j)\|^2. \tag{4.33}$$

We note that above problem is actually similar to the minimization problem (3.7) except for the choice of the minimization functional. We now elaborate on the closest co-normal algorithm.

4.5 Closest Co-Normal Algorithm (CCA)

The algorithm was proposed in [44] as a variation of the Common Normal Algorithm. The problem is formulated as the closest co-normal minimization problem (3.7) that we recall here for convenience:

$$(\boldsymbol{x}_i, \boldsymbol{x}_j) = \underset{\substack{(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j) \in \mathcal{E}_i \times \mathcal{E}_j \\ \boldsymbol{n}_i(\hat{\boldsymbol{x}}_i) + \boldsymbol{n}_j(\hat{\boldsymbol{x}}_j) = \boldsymbol{0}}}{\operatorname{argmin}} \|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\|^2. \tag{4.34}$$

We know from Lemmas 3, 5, and Remark 5 that above problem admits a unique solution

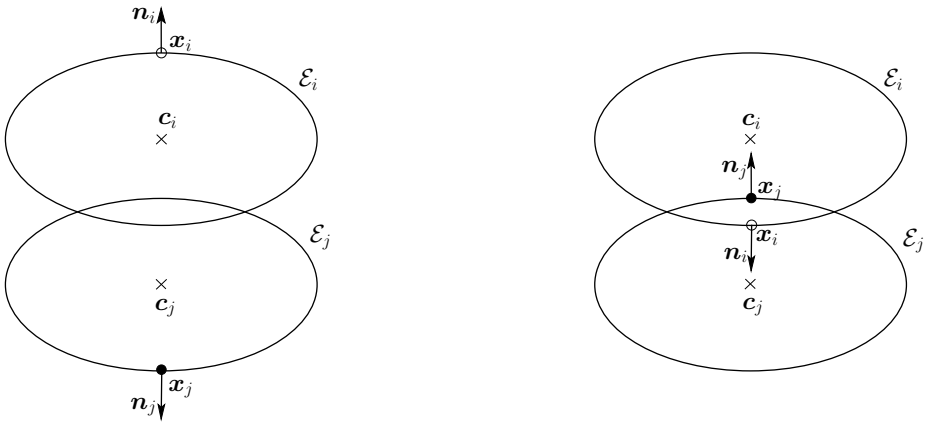


Figure 4.7 Equations (4.31) and (4.32) may prove non-unique pairs of solution. For the same pair of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$. The figure at the left presents a pair of $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ which has a maximum distance. The figure at the right presents the pair with minimum distance.

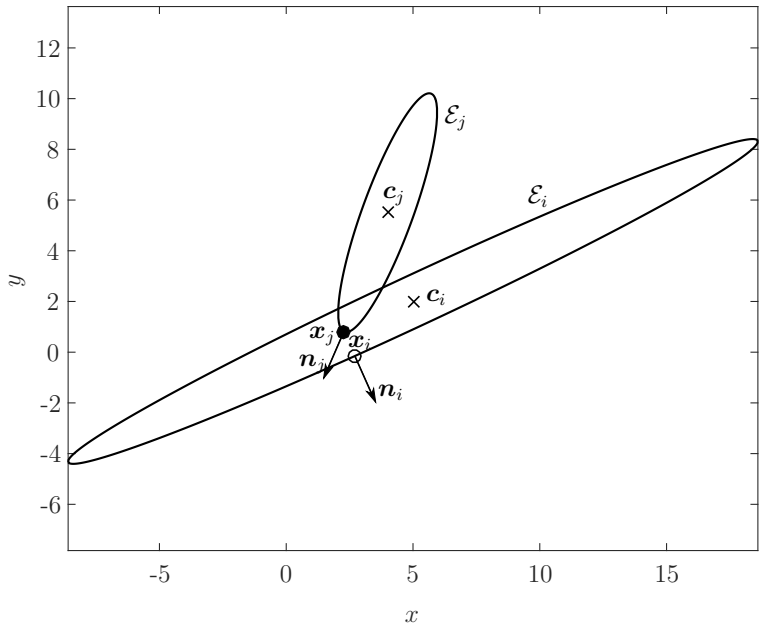


Figure 4.8 The pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is found by solving system of equations (4.32). Ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ are the same ellipses as defined in the Figure 4.3. The normal vectors are obtained as $\boldsymbol{n}_i = (0.410, -0.912)$ and $\boldsymbol{n}_j = (-0.410, -0.912)$. We see that Equation (4.31) is only satisfied with respect to x-component. This shows that the system of equations (4.32) leads to a wrong solution, since the y-component of the normal vectors are equal rather than being opposite.

for all configurations of ellipses in near-perfect contact. We note that such a problem can be solved by various minimization methods, e.g. Lagrangian method. However, the authors in [44] proposed an original approach by parameterizing the distance $\|\boldsymbol{x}_j - \boldsymbol{x}_i\|$ in terms of a single parameter $t \in [-\pi, \pi[$. The main idea relies on the fact that there exists a bijection between a point on an ellipse and the normal vector to the ellipse at that point. Indeed, let each ellipse $\mathcal{E}_k$, $k = i$ or j, be defined in its own local reference system as:

$$\widehat{f}_k(\boldsymbol{\xi}) = \boldsymbol{\xi}^T \mathcal{D}_k \boldsymbol{\xi} - 1 = 0. \tag{4.35}$$

The outward unit normal vector at $\boldsymbol{\xi}$ to $\mathcal{E}_k$ is then given by:

$$\hat{\boldsymbol{n}}_k(\boldsymbol{\xi}) = \frac{\nabla \widehat{f}_k}{\|\nabla \widehat{f}_k\|} = \frac{2\mathcal{D}_k \boldsymbol{\xi}}{\|\nabla \widehat{f}_k\|},$$

which implies that:

$$\boldsymbol{\xi} = \frac{\|\nabla \widehat{f}_k\|}{2} \mathcal{D}_k^{-1} \hat{\boldsymbol{n}}_k(\boldsymbol{\xi}). \tag{4.36}$$

Since the normal vector $\boldsymbol{n}_k$ in the global reference system is related to $\hat{\boldsymbol{n}}_k$ by the rotation matrix $\mathcal{R}_k$ of angle θ_k (2.2) as:

$$\hat{\boldsymbol{n}}_k = \mathcal{R}_k^T \boldsymbol{n}_k, \tag{4.37}$$

we get using (2.5):

$$\boldsymbol{x}_k = \mathcal{R}_k \boldsymbol{\xi} + \boldsymbol{c}_k = \frac{\|\nabla \widehat{f}_k\|}{2} \mathcal{R}_k \mathcal{D}_k^{-1} \mathcal{R}_k^T \boldsymbol{n}_k + \boldsymbol{c}_k. \tag{4.38}$$

Moreover, substituting (4.36) for $\boldsymbol{\xi}$ in (4.35) provides the new expression for the norm of the gradient:

$$\frac{1}{\|\nabla \widehat{f}_k\|^2} = \frac{1}{4} \hat{\boldsymbol{n}}_k^T \mathcal{D}_k^{-1} \hat{\boldsymbol{n}}_k = \frac{1}{4} \|\mathcal{D}_k^{-1/2} \hat{\boldsymbol{n}}_k\|^2 = \frac{1}{4} \|\mathcal{D}_k^{-1/2} \mathcal{R}_k^T \boldsymbol{n}_k\|^2 \tag{4.39}$$

that is:

$$\|\nabla \widehat{f}_k\| = \frac{2}{\|\mathcal{D}_k^{-1/2} \mathcal{R}_k^T \boldsymbol{n}_k\|},$$

so that:

$$\boldsymbol{x}_k = \frac{\mathcal{R}_k \mathcal{D}_k^{-1} \mathcal{R}_k^T \boldsymbol{n}_k}{\|\mathcal{D}_k^{-1/2} \mathcal{R}_k^T \boldsymbol{n}_k\|} + \boldsymbol{c}_k, \quad k = i, j. \tag{4.40}$$

The minimization problem (4.34) can thus be recast as that of finding the value $t_{ij} \in [-\pi, \pi[$ such that:

$$t_{ij} = \underset{t \in [-\pi, \pi[}{\operatorname{argmin}} \|\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)\|^2. \tag{4.41}$$

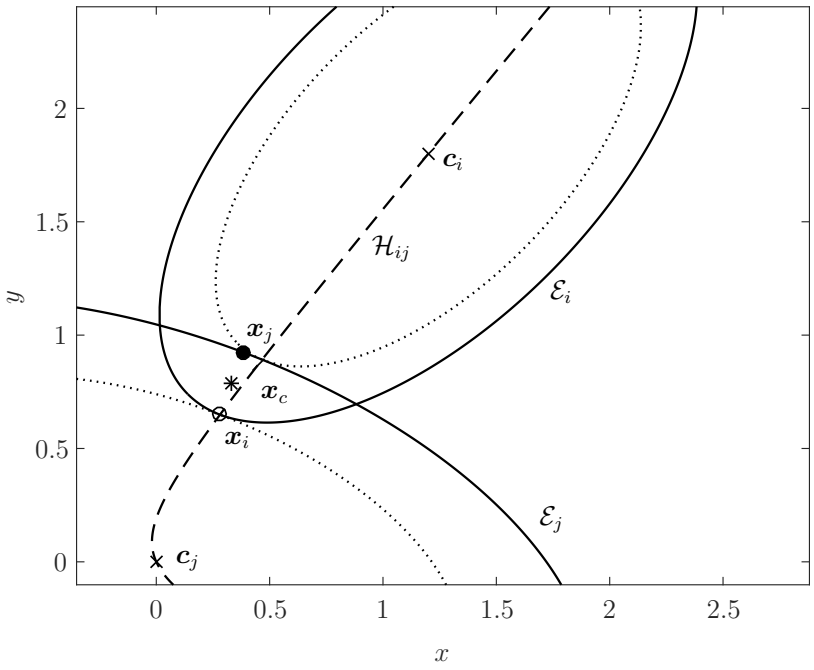


Figure 4.9 The points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are reached the global minimum of Equation (4.41) for the pair of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$. We see that the points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are not necessary located on $\mathcal{H}_{ij}$.

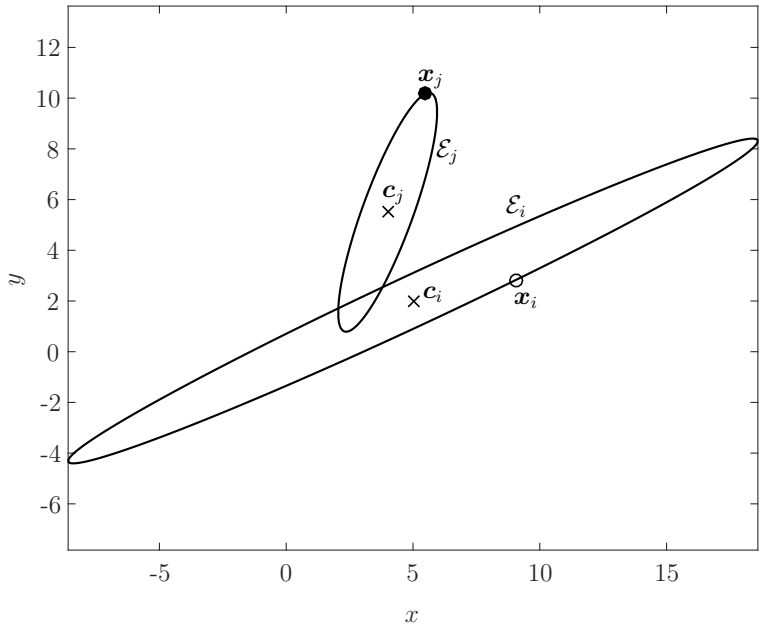


Figure 4.10 The points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are the local minimum of Problem (4.41) for the pair of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ from Figure 4.3. The function $d(t) = \|\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)\|$ and the location of the local minimum is shown in Figure 4.11.

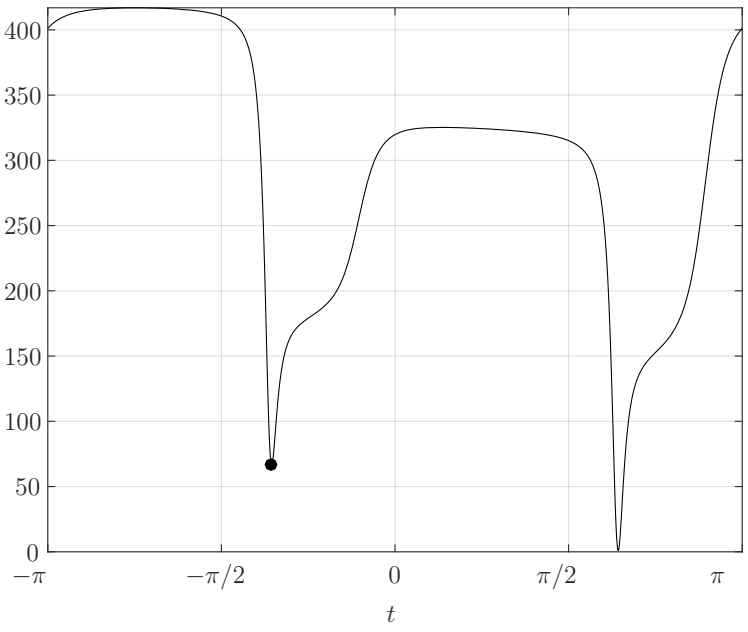


Figure 4.11 Plot of function $d(t) = \|\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)\|$ for the pairs of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ of Figure 4.10. The location of local minimum of Problem (4.41) is shown with a dot (•).

Using above formula, one can now express the difference $\boldsymbol{x}_j - \boldsymbol{x}_i$ as:

$$\boldsymbol{x}_j - \boldsymbol{x}_i = \left[\frac{\mathcal{R}_j \mathcal{D}_j^{-1} \mathcal{R}_j^T \boldsymbol{n}_j}{\|\mathcal{D}_j^{-1/2} \mathcal{R}_j^T \boldsymbol{n}_j\|} - \frac{\mathcal{R}_i \mathcal{D}_i^{-1} \mathcal{R}_i^T \boldsymbol{n}_i}{\|\mathcal{D}_i^{-1/2} \mathcal{R}_i^T \boldsymbol{n}_i\|} \right] + (\boldsymbol{c}_j - \boldsymbol{c}_i).$$

In order to apply the constraint $\boldsymbol{n}_i + \boldsymbol{n}_j = \boldsymbol{0}$, one can introduce the common direction $\boldsymbol{n}$, parameterized with respect to $t \in [-\pi, \pi[$, i.e.

$$\boldsymbol{n}(t) = \begin{bmatrix} \cos t \\ \sin t \end{bmatrix}$$

such that $\boldsymbol{n}(t) = \boldsymbol{n}_j = -\boldsymbol{n}_i$. Therefore:

$$\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t) = \left[\frac{\mathcal{R}_j \mathcal{D}_j^{-1} \mathcal{R}_j^T}{\|\mathcal{D}_j^{-1/2} \mathcal{R}_j^T \boldsymbol{n}(t)\|} + \frac{\mathcal{R}_i \mathcal{D}_i^{-1} \mathcal{R}_i^T}{\|\mathcal{D}_i^{-1/2} \mathcal{R}_i^T \boldsymbol{n}(t)\|} \right] \boldsymbol{n}(t) + (\boldsymbol{c}_j - \boldsymbol{c}_i).$$

The original minimization problem with multiple constraints in the four variables $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is now replaced by the unconstrained minimization problem (4.41) in the only variable t. However, the distance $d(t) = \|\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)\|$ can be, for some configurations of ellipses,

difficult to minimize as it may exhibit very flat regions and multiple extrema (up to two local minima and two local maxima), as shown in Figures 4.10 and 4.11. The efficiency of the method is also dependent on the choice of an initial guess t_0. If such a choice is not available, one has then to find all local minima and select the one that actually minimizes the distance function.

Finally, we note that the algorithm extends straightforwardly to the case of ellipsoids by considering the following parameterization of the normal vector:

$$\boldsymbol{n}(u,v) = \begin{bmatrix} \cos u \cos v \\ \sin u \cos v \\ \sin v \end{bmatrix}, \quad u \in [-\pi, \pi[, \ v \in [0, \pi]. \tag{4.42}$$

CHAPTER 5 NEW CONTACT DETECTION ALGORITHM

We introduce in this chapter a new contact detection method in the case of two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ with non-penetrating CoM. Moreover, we further assume that the two ellipses are in near perfect contact, meaning that the ellipses are either fully disjoint, or are in perfect contact, or exhibit a small overlap. This ensures that the intersection set between $\mathcal{E}_i$ and $\mathcal{E}_j$ either is the empty set, or reduces to a single point, the contact point, or is exactly formed of two points, respectively.

The proposed method intrinsically belongs to the family of the constrained geometrical potential methods. We recall that the point $\boldsymbol{x}_i$ on ellipse $\mathcal{E}_i$ is defined as the closest point to the center $\boldsymbol{c}_j$ of ellipse $\mathcal{E}_j$, with respect to the $\mathcal{E}_j$-norm, while the point $\boldsymbol{x}_j$ on $\mathcal{E}_j$ is the closest point to the center $\boldsymbol{c}_i$ of $\mathcal{E}_i$, with respect to the $\mathcal{E}_i$-norm. The attractive feature of this approach is that the pair of points, referred here as the contact pair, are solutions to uncoupled minimization problems that can be solved separately. Moreover, the solution to each problem exists and is unique. Furthermore, following Definition (5), the solution $\boldsymbol{x}_i$ to (3.8) belongs to $\mathcal{H}_{ij}$. Using the same reasoning, we show that the solution $\boldsymbol{x}_j$ to (3.9) is also on $\mathcal{H}_{ij}$. Moreover, the two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ actually belong to the same branch of the hyperbola $\mathcal{H}_{ij}$ as that passing through the centers. The properties satisfied by the contact pair can be added as non-binding constraints to the minimization problems (4.25) and (4.26). We note that these minimization problems are structurally similar. The proposed method to solve these problems is described below considering only one of them, say (4.26) to find $\boldsymbol{x}_j$, in the case of a unit circle and an ellipse at origin without rotation, which is described in Section 3.7.2. Then, our objective is now to solve Problem (4.26) in this new configuration of the ellipses and to transfer the solution back to the original coordinate system using the transformation (3.53).

5.1 Solution Method

In this section, we consider Problem (4.26) in the particular case where the pair of ellipses consists of the ellipse $\mathcal{E}_j$ defined in its local coordinate system and the unit circle $\mathcal{C}_i$ centered at $\boldsymbol{c}_i = (c_x, c_y)$, as shown in Figure 5.1. The parameters of the problem are reduced to the semi-axes a_j and b_j of ellipse $\mathcal{E}_j$ and the center $\boldsymbol{c}_i$ of circle $\mathcal{C}_i$.

We begin by providing a geometrical interpretation of the solution $\boldsymbol{x}_j \in \mathcal{E}_j$ to Problem (4.26). Supposing first that ellipse $\mathcal{E}_j$ and circle $\mathcal{C}_i$ are in perfect contact at $\boldsymbol{x}_j$, i.e. $\mathcal{E}_j \cap \mathcal{C}_i = \{\boldsymbol{x}_j\}$, then

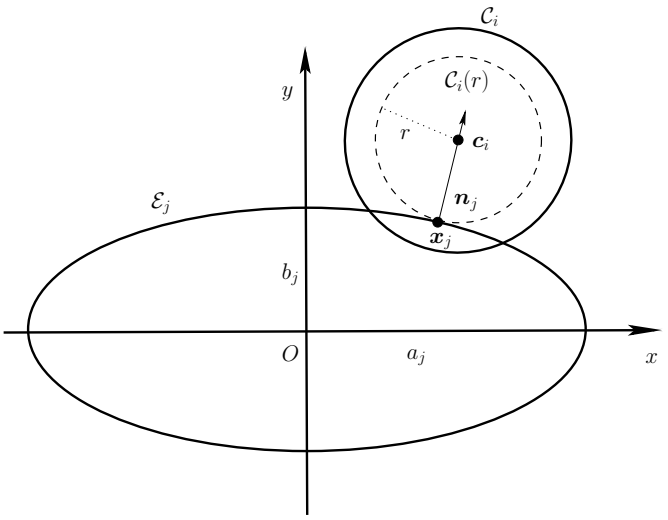


Figure 5.1 Ellipse $\mathcal{E}_j$ and circle $\mathcal{C}_i$ with overlap. The circle $\mathcal{C}_i(r)$ of radius r is constructed as the smallest circle centered at $\boldsymbol{c}_i$ such that $\mathcal{E}_j$ and $\mathcal{C}_i(r)$ are in perfect contact. The point $\boldsymbol{x}_j$, defined as the intersection point between $\mathcal{E}_j$ and $\mathcal{C}_i(r)$, is the solution to Problem (4.26).

it is clear that $\boldsymbol{x}_j$ is the closest point on $\mathcal{E}_j$ to the center $\boldsymbol{c}_i$ of $\mathcal{C}_i$ such that $\boldsymbol{n}_i(\boldsymbol{x}_i)+\boldsymbol{n}_j(\boldsymbol{x}_i) = \boldsymbol{0}$. We also observe that the line supported by the normal vector to $\mathcal{E}_j$ at $\boldsymbol{x}_j$ passes through $\boldsymbol{c}_i$. In the case $\mathcal{E}_j$ and $\mathcal{C}_i$ are not in perfect contact, one can construct the smallest circle $\mathcal{C}_i(r)$ of radius r centered at $\boldsymbol{c}_i$ such that $\mathcal{C}_i(r)$ is in perfect contact with $\mathcal{E}_j$, see Figure 5.1. The intersection point is actually the unique solution $\boldsymbol{x}_j$ to Problem (4.26): indeed, $\boldsymbol{x}_j \in \mathcal{E}_j$, it is the closest point to $\boldsymbol{c}_i$ with respect to the Euclidean norm (which is the same as the $\mathcal{C}_i$-norm since $\mathcal{Q}_i$ is here the identity matrix), and it satisfies $\boldsymbol{n}_i(\boldsymbol{x}_j) + \boldsymbol{n}_j(\boldsymbol{x}_j) = \boldsymbol{0}$, or simply

$$\nabla f_i(\boldsymbol{x}_j) \times \nabla f_j(\boldsymbol{x}_j) = \boldsymbol{0}, \tag{5.1}$$

meaning that $\boldsymbol{x}_j \in \mathcal{H}_{ij}$.

We proceed with the fact that the solution to Problem (4.26) belongs to the ellipse $\mathcal{E}_j$ and the co-gradient locus $\mathcal{H}_{ij}$ (see Definition 5). Recalling (3.54), $\boldsymbol{x} = (x, y) \in \mathcal{E}_i \cap \mathcal{H}_{ij}$ if it satisfies the following system of quadratic equations in x and y:

$$\begin{aligned} b_j^2 x^2 + a_j^2 y^2 - a_j^2 b_j^2 &= 0, \\ (a_j^2 - b_j^2)xy + b_j^2 c_y x - a_j^2 c_x y &= 0. \end{aligned} \tag{5.2}$$

One could formally eliminate one of the variables x or y, say y, by combining the above two equations to obtain an algebraic equation of order four in x. This means that there could be at most four roots of the polynomial function depending on the configuration of the ellipse

and circle. In other words, we can find at most four points that belong to the set $\mathcal{E}_j \cap \mathcal{H}_{ij}$. This result could be expected since $\mathcal{H}_{ij}$ is known to be a hyperbola. Indeed, one of the two branches of $\mathcal{H}_{ij}$ passes through the centers of $\mathcal{E}_j$ and $\mathcal{C}_i$ and thus necessarily intersects the ellipse $\mathcal{E}_j$ at two points. The other branch, depending of the shape of the ellipse and the respective position of the circle, could intersect the ellipse at no point, at one point, or at two points. In other words, the set $\mathcal{E}_j \cap \mathcal{H}_{ij}$ consists of two, three, or four points. In order to select the solution to Problem (4.26), one would need to find the closest point to $\boldsymbol{c}_i$ among those points. This approach is not necessarily the most efficient as it requires to compute all real roots of the polynomial of degree four when one needs to find only the one that provides the closest point to $\boldsymbol{c}_i$. We propose below a different approach based on the parametrization of $\mathcal{E}_j$. The points $\boldsymbol{x}$ on $\mathcal{E}_j$ can be described in terms of a single parameter $t \in [-\pi, \pi[$ as

$$\boldsymbol{x}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_j \cos t \\ b_j \sin t \end{bmatrix}. \tag{5.3}$$

It follows, using (3.54), that the points in $\mathcal{E}_i \cap \mathcal{H}_{ij}$ can be obtained from the roots of the nonlinear function:

$$h(t) = (a_j^2 - b_j^2) \cos t \sin t + b_j c_y \cos t - a_j c_x \sin t. \tag{5.4}$$

The scalar function $h(t)$ is clearly 2π-periodic and is continuous on $t \in [-\pi, \pi[$. Moreover, it may have up to four roots. The function is obviously non-convex. If one wants to use the second-order Newton method to find the root of $h(t)$, one needs to define an accurate initial guess and possibly an additional constraint to ensure that the method converges to the desired root t_j, which will provide the actual solution $\boldsymbol{x}_j$ to Problem (4.26). These will be presented in the next two sections.

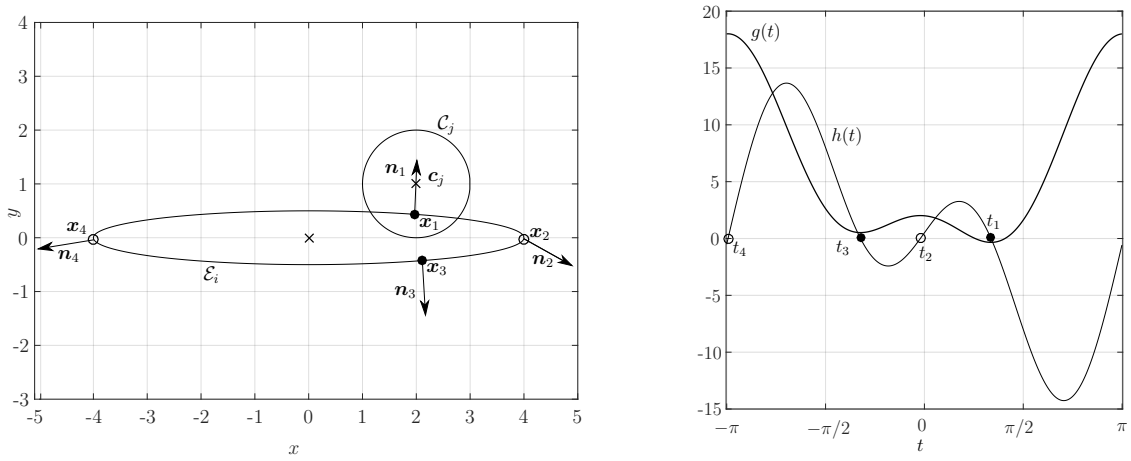
Before proceeding further, we show that the roots of $h(t)$ are actually the critical points of $f_i(\boldsymbol{x})$ when subjected to the constraint that $\boldsymbol{x} \in \mathcal{E}_j$. Indeed, using the parametric form of the ellipse, we introduce

$$g(t) := \frac{1}{2} f_i(x(t)) = \frac{1}{2}\Big[(x(t) - c_x)^2 + (y(t) - c_y)^2 - 1\Big] = \frac{1}{2}\Big[(a_j \cos t - c_x)^2 + (b_j \sin t - c_y)^2 - 1\Big],$$

so that:

$$\begin{aligned} g'(t) &= -a_j \sin t (a_j \cos t - c_x) + 2 b_j \cos t (b_j \sin t - c_y) \\ &= -\Big[(a_j^2 - b_j^2) \cos t \sin t - a_j c_x \sin t + b_j c_y \cos t\Big] \\ &= -h(t). \end{aligned}$$

In other words, the roots of $h(t)$ also satisfy $g'(t) = 0$, meaning that the points in $\mathcal{E}_j \cap \mathcal{H}_{ij}$



(a) (b)

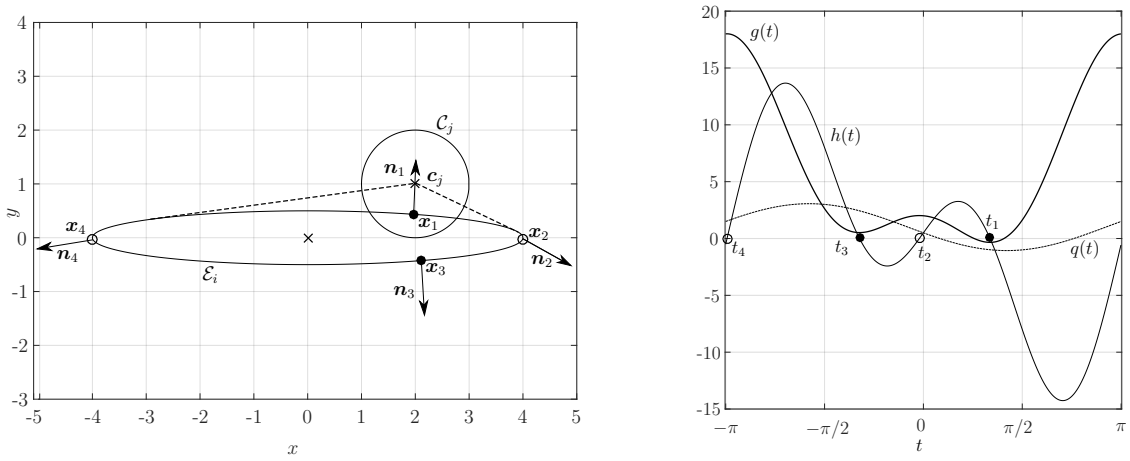
Figure 5.2 (a) Example of a circle $\mathcal{C}_i$ and an ellipse $\mathcal{E}_j$ such that the set $\mathcal{E}_j \cap \mathcal{H}_{ij}$ consists of four points $\boldsymbol{x}_k$, $k = 1, \ldots, 4$. The point $\boldsymbol{x}_1$ is the closest point among the four points to the center $\boldsymbol{c}_i$ of $\mathcal{C}_i$ and corresponds to the solution $\boldsymbol{x}_j$ to Problem (4.26). (b) Plot of the corresponding functions $h(t)$ and $g(t)$. The four roots of $h(t)$ are denoted by t_k, $k = 1, \ldots, 4$. The function $g(t)$ reaches two local minima, at t_1 and t_3, and two local maxima, at t_2 and t_4, in $t \in [-\pi, \pi[$. It reaches the global minimum at t_1, which corresponds to the point $\boldsymbol{x}_1$.

are either at a minimal or maximal distance from $\boldsymbol{c}_i$ when traveling along the ellipse $\mathcal{E}_j$. The point that corresponds to the global minimal distance is thus the point $\boldsymbol{x}_j$. We also note that the lines spanned by the unit outward normal vector to $\mathcal{E}_j$ at the points in $\mathcal{E}_j \cap \mathcal{H}_{ij}$ all pass through the center $\boldsymbol{c}_i$ of the circle.

We show in Figure 5.2(a) an example of an ellipse $\mathcal{E}_j$ and a circle $\mathcal{C}_i$ for which the set $\mathcal{E}_j \cap \mathcal{H}_{ij}$ consists of four points $\boldsymbol{x}_k$, $k = 1, \ldots, 4$. The point $\boldsymbol{x}_1$ in this case is the solution $\boldsymbol{x}_j$ to Problem (4.26) and corresponds to the global minimum of the function $g(t)$. We observe in Figure 5.2(b) that the roots of $h(t)$ corresponds to two local minima and two local maxima of $g(t)$.

Remark 7. *After applying the mapping, if ellipse $\mathcal{E}_j$ is also a circle, i.e. a_j and b_j are equal, then we can directly find the point $\boldsymbol{x}_j$*

$$\boldsymbol{x}_j = a_j \frac{\boldsymbol{c}_i}{\|\boldsymbol{c}_i\|}. \tag{5.5}$$



(a) (b)

Figure 5.3 (a) The two dash lines originating from $\boldsymbol{c}_i$ are constructed such that they are tangent to the ellipse $\mathcal{E}_j$. The intersection point between each line and $\mathcal{E}_j$ satisfies $\nabla f_i(\boldsymbol{x}) \cdot \nabla f_j(\boldsymbol{x}) = 0$. Those two points determine the end points of the set $\mathcal{S}$. We note that the point $\boldsymbol{x}_1$ satisfies the constraint $\nabla f_i(\boldsymbol{x}_1) \cdot \nabla f_j(\boldsymbol{x}_1) < 0$. (b) The function $q(t)$ is negative only at t_1 among the roots $t_1, \ldots, t_4$ of $h(t)$.

5.2 Additional Constraint

We have seen so far that the solution $\boldsymbol{x}_j$ to Problem (4.26) is one of the points in $\mathcal{E}_j \cap \mathcal{H}_{ij}$. If one wants to find the unique point $\boldsymbol{x}_j$, instead of estimating every root of $h(t)$, one needs to consider an additional constraint in order to enforce the unicity of the solution. We observe from Figure 5.2(a) that a property that distinguishes $\boldsymbol{x}_1$, the actual solution of the minimization problem, from the other points $\boldsymbol{x}_k$, $k = 2, \ldots, 4$, is that the normal vector $\boldsymbol{n}_1 = \boldsymbol{n}_j(\boldsymbol{x}_1)$ to $\mathcal{E}_j$ at $\boldsymbol{x}_1$ is the only vector that points in the opposite direction to the vectors $\boldsymbol{x}_k - \boldsymbol{c}_i$, $k = 1, \ldots, 4$. In other words, we have:

$$\begin{aligned} &\nabla f_i(\boldsymbol{x}_1) \cdot \nabla f_j(\boldsymbol{x}_1) < 0, \\ &\nabla f_i(\boldsymbol{x}_k) \cdot \nabla f_j(\boldsymbol{x}_k) > 0, \ k = 2, \ldots, 4. \end{aligned}$$

This motivates us to introduce the function $q(t)$ defined on $[-\pi, \pi[$ as:

$$q(t) = \frac{1}{4}\nabla f_i(\boldsymbol{x}(t)) \cdot \nabla f_j(\boldsymbol{x}(t)) = \frac{x(t)}{a_j^2}(x(t) - c_x) + \frac{y(t)}{b_j^2}(y(t) - c_y) = 1 - \frac{c_x}{a_j}\cos t - \frac{c_y}{b_j}\sin t \quad (5.6)$$

and the constraint set $\mathcal{S}$:

$$\mathcal{S} = \{t \in [-\pi, \pi[; \ q(t) < 0\}. \quad (5.7)$$

We see in Figure 5.3(b), which corresponds to the example of Figure 5.2, that t_1 is the only root of $h(t)$ such that $t_1 \in \mathcal{S}$. The following lemma states that the solution $\boldsymbol{x}_j$ to Problem (4.26) is the unique point in $\mathcal{E}_j \cap \mathcal{H}_{ij}$ that satisfies the constraint $\nabla f_i(\boldsymbol{x}_j) \cdot \nabla f_j(\boldsymbol{x}_j) < 0$. Another way of stating this is as follows, $\boldsymbol{x}_j = \boldsymbol{x}(t_j)$ is the solution to Problem (4.26), where t_j is the only root of $h(t)$ that satisfies the constraint $t_j \in \mathcal{S}$.

Lemma 11. *Let $\mathcal{C}_i$ be the unit circle centered at $\mathbf{c}_i$ and let $\mathcal{E}_j$ be an arbitrary ellipse centered at the origin and defined in its local coordinate system such that $\mathcal{C}_i$ and $\mathcal{E}_j$ are in near perfect contact. Then there exists one and only one point $\boldsymbol{x}_j$ in $\mathcal{E}_j \cap \mathcal{H}_{ij}$ that satisfies the condition $\nabla f_i(\boldsymbol{x}_j) \cdot \nabla f_j(\boldsymbol{x}_j) < 0$.*

Proof. Let the ellipse $\mathcal{E}_j$ be such that $a_j \neq b_j$. The proof relies on the fact that the co-gradient locus is a hyperbola and on the analysis of the scalar function $\alpha(\boldsymbol{x}) = \boldsymbol{n}_i(\boldsymbol{x}) \cdot \boldsymbol{n}_j(\boldsymbol{x})$ along the two branches of the hyperbola. We provide here only a sketch of the proof as more details can be found in the proof of Theorem 6.

We first note that $\nabla f_i(\boldsymbol{x}) \neq \mathbf{0}$ everywhere in $\mathbb{R}^2$ except at the center of the ellipse $\boldsymbol{c}_i = \mathbf{0}$. Similarly, $\nabla f_j(\boldsymbol{x}) \neq \mathbf{0}$, $\forall \boldsymbol{x} \in \mathbb{R}^2 \backslash \{\boldsymbol{c}_j\}$. Therefore, the function $\alpha(\boldsymbol{x})$ is defined everywhere in $\mathbb{R}^2$ except at the two centers. Moreover, it is continuous along the branches of the hyperbola, apart from the points $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$, where it could possibly be discontinuous. In fact, we remark that $\alpha(\boldsymbol{x})$ can only take the values $+1$ or -1 for all $\boldsymbol{x} \in \mathcal{H}_{ij} \backslash \{\boldsymbol{c}_i, \boldsymbol{c}_j\}$. Indeed, $\boldsymbol{x} \in \mathcal{H}_{ij}$ if $\nabla f_i(\boldsymbol{x}) \times \nabla f_j(\boldsymbol{x}) = \mathbf{0}$; this implies that $\boldsymbol{n}_i(\boldsymbol{x}) \times \boldsymbol{n}_j(\boldsymbol{x}) = \mathbf{0}$ for all $\boldsymbol{x} \in \mathcal{H}_{ij} \backslash \{\boldsymbol{c}_i, \boldsymbol{c}_j\}$, meaning that the unit vectors $\boldsymbol{n}_i(\boldsymbol{x})$ and $\boldsymbol{n}_j(\boldsymbol{x})$ are either in the same direction or in the opposite direction.

Let us consider a circle $\mathcal{C}_{ij}$ centered at $\boldsymbol{c}_j = \mathbf{0}$ with a sufficiently large radius that $\mathcal{C}_{ij}$ surrounds $\mathcal{C}_i$, $\mathcal{E}_j$, and a portion of both branches of $\mathcal{H}_{ij}$, as shown in Figure 5.4. Then, it is clear that the hyperbola $\mathcal{H}_{ij}$ must intersect $\mathcal{C}_{ij}$ at four points $\boldsymbol{x}_k$, $k = 1, \ldots, 4$. Moreover, at each of these points, $\boldsymbol{n}_i(\boldsymbol{x}_k) = \boldsymbol{n}_j(\boldsymbol{x}_k)$, which implies that $\alpha(\boldsymbol{x}_k) = +1$, $k = 1, \ldots, 4$. Let us concentrate for now on the branch of $\mathcal{H}_{ij}$ that passes through the center $\boldsymbol{c}_i$ of $\mathcal{C}_i$ and the center $\boldsymbol{c}_j$ of $\mathcal{E}_j$. The branch can intersect the ellipse $\mathcal{E}_j$ at only one point when one travels along the branch between $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$. The intersection point is in fact the solution $\boldsymbol{x}_j$ to Problem (4.26), for which we know that $\boldsymbol{n}_i(\boldsymbol{x}_j) + \boldsymbol{n}_j(\boldsymbol{x}_j) = \mathbf{0}$. It follows that $\alpha(\boldsymbol{x}_j) = -1$, so that $\nabla f_i(\boldsymbol{x}_j) \cdot \nabla f_j(\boldsymbol{x}_j) < 0$. In other words, the function $\alpha(\boldsymbol{x})$ along the branch takes the value $+1$ when one goes from $\mathcal{C}_{ij}$ to $\boldsymbol{c}_i$, is discontinuous at $\boldsymbol{c}_i$, takes the value -1 when one goes from $\boldsymbol{c}_i$ to $\boldsymbol{c}_j$, is again discontinuous at $\boldsymbol{c}_j$, and finally takes the value $+1$ when one goes from $\boldsymbol{c}_j$ to $\mathcal{C}_{ij}$. For the other branch, there is no singular point, which implies that $\alpha(x) = +1$ for all points of the branch.

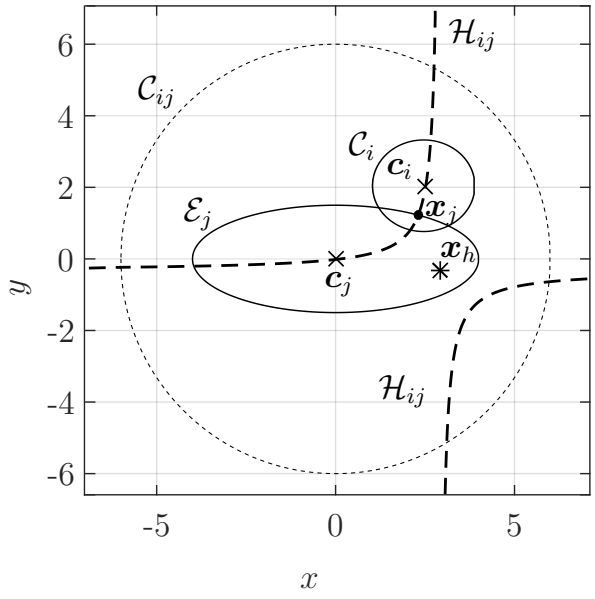


Figure 5.4 Example of a circle $\mathcal{C}_{ij}$ that surrounds the ellipse $\mathcal{E}_j$, the circle $\mathcal{C}_i$, and a portion of both branches of the hyperbola $\mathcal{H}_{ij}$.

In summary, we have shown that the set $\mathcal{E}_j \cap \mathcal{H}_{ij}$ is non empty, as it contains at least the point $\boldsymbol{x}_j$. Moreover, $\boldsymbol{x}_j$ is the only point of $\mathcal{E}_j \cap \mathcal{H}_{ij}$ for which $\alpha(\boldsymbol{x}_j) = -1$, or equivalently $\nabla f_i(\boldsymbol{x}_j) \cdot \nabla f_j(\boldsymbol{x}_j) < 0$.

In the case when $\mathcal{E}_j$ is a circle, i.e. $a_j = b_j$, the set $\mathcal{H}_{ij}$ degenerates into the line passing through the centers $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$. It is then trivial to show that there exists for that case one and only one point $\boldsymbol{x}_j$ in $\mathcal{E}_j \cap \mathcal{H}_{ij}$ that satisfies the condition $\nabla f_i(\boldsymbol{x}_j) \cdot \nabla f_j(\boldsymbol{x}_j) < 0$. □

The above lemma allows us to conclude that Problem (4.26) can be solved by searching for the unique root t_j of $h(t)$ that satisfies the constraint $t_j \in \mathcal{S}$. In addition to $q(t)$, we also introduce, for reasons that will become clearer below, the function $\hat{q}(t)$ defined on $[-\pi, \pi[$:

$$\hat{q}(t) = \boldsymbol{n}_i(\boldsymbol{x}(t)) \cdot \boldsymbol{n}_j(\boldsymbol{x}(t)) = \cos \eta(t), \tag{5.8}$$

where $\eta(t)$ is the angle between the unit normal vectors $\boldsymbol{n}_i$ and $\boldsymbol{n}_j$. It is obvious that $\hat{q}(t) < 0$ for all $t \in \mathcal{S}$. Moreover, $\hat{q}$ reaches its minimum -1 at the unique t_j and its maximum $+1$ at the other roots of $h(t)$. It is explicitly given by

$$\hat{q}(t) = \frac{\nabla f_i(\boldsymbol{x}(t)) \cdot \nabla f_j(\boldsymbol{x}(t))}{\|\nabla f_i(\boldsymbol{x}(t))\| \, \|\nabla f_j(\boldsymbol{x}(t))\|} = \frac{a_j b_j - c_x b_j \cos t - c_y a_j \sin t}{\sqrt{(b_j \cos t)^2 + (a_j \sin t)^2}\sqrt{(a_j \cos t - c_x)^2 + (b_j \sin t - c_y)^2}}.$$

Although the function $\hat{q}$ exhibits attractive features, the function itself as well as its first

derivative are computationally more expensive to evaluate than function $q(t)$ and function $h(t)$, respectively.

We propose to use the Newton's method to find the root $t_j \in \mathcal{S}$ of $h(t)$, see Equation (5.4), since it is a second-order method. However, the function $h(t)$ is non-convex on $[-\pi, \pi[$. Thus, there is always a risk that the initial guess point will be outside of the basin of attraction of t_j, in which case the method will converge to another root of $h(t)$. In order to circumvent this issue, we check that each new iterate belongs to $\mathcal{S}$. If the constraint is not satisfied, we then use one iteration of the line search method to approach the minimizer t_j of $\hat{q}(t)$. We also propose below an approach that allows one to compute an initial guess point for the Newton's method that is reasonably close to the desired root t_j and that, most of the time, prevents one from resorting to the line search method.

5.3 Initial Point Algorithm

In this section, we propose an algorithm to compute an initial guess point t_0 that approximates the root t_j of $h(t)$. The approach is motivated by the fact that, given a point $\boldsymbol{x} \in \mathcal{E}_i$, the normal vector $-\boldsymbol{n}_j(\boldsymbol{x})$ to $\mathcal{E}_j$ at $\boldsymbol{x}$ bisects the angle formed by the vectors $\boldsymbol{f}_1 - \boldsymbol{x}$ and $\boldsymbol{f}_2 - \boldsymbol{x}$, where $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ are the focal points of $\mathcal{E}_i$, see Equation (2.19). This approach could be used by other iterative algorithms if they also used a mapping to reduce the problem to an ellipse and a circle.

We therefore introduce the unit vectors

$$\boldsymbol{v}_1 = (\boldsymbol{f}_1 - \boldsymbol{c}_i)/\|\boldsymbol{f}_1 - \boldsymbol{c}_i\|, \tag{5.9}$$

$$\boldsymbol{v}_2 = (\boldsymbol{f}_2 - \boldsymbol{c}_i)/\|\boldsymbol{f}_2 - \boldsymbol{c}_i\|, \tag{5.10}$$

and the unit vector $\boldsymbol{v}$ defined as

$$\boldsymbol{v} = \frac{\boldsymbol{v}_1 + \boldsymbol{v}_2}{\|\boldsymbol{v}_1 + \boldsymbol{v}_2\|}. \tag{5.11}$$

The vector $\boldsymbol{v}$ can thus be viewed as an approximation of $-\boldsymbol{n}_j(\boldsymbol{x}_j)$ at $\boldsymbol{x}_j \in \mathcal{E}_j$. The line spanned by $\boldsymbol{v}$ and passing through the center $\boldsymbol{c}_i$ of $\mathcal{C}_i$ necessarily intersects the ellipse $\mathcal{E}_j$ at two points. Parametrizing the line as the set of points $\boldsymbol{c}_i + r\boldsymbol{v}$, $r \in \mathbb{R}$, the points at the intersection of the line and $\mathcal{E}_j$ satisfy the quadratic equation in r:

$$f_j(\boldsymbol{c}_i + r\boldsymbol{v}) = (\boldsymbol{c}_i + r\boldsymbol{v})^T \mathcal{D}_j (\boldsymbol{c}_i + r\boldsymbol{v}) - 1 = (\boldsymbol{v}^T \mathcal{D}_j \boldsymbol{v}) r^2 + 2(\boldsymbol{v}^T \mathcal{D}_j \boldsymbol{c}_i) r + (\boldsymbol{c}_i^T \mathcal{D}_j \boldsymbol{c}_i) - 1 = 0.$$

Introducing the positive quantities $\alpha = (\boldsymbol{v}^T \mathcal{D}_j \boldsymbol{v})$, $\beta = -(\boldsymbol{v}^T \mathcal{D}_j \boldsymbol{c}_i)$, and $\gamma = (\boldsymbol{c}_i^T \mathcal{D}_j \boldsymbol{c}_i) - 1$, the

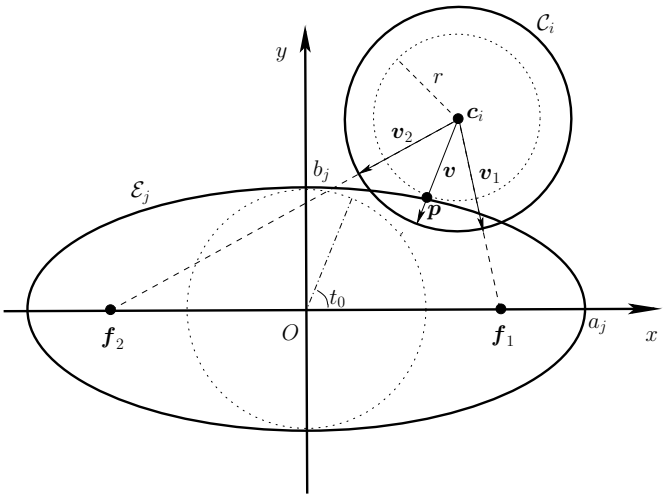


Figure 5.5 Location of initial point $\boldsymbol{p} = \boldsymbol{c}_i + r_{\min}\boldsymbol{v}$ on ellipse $\mathcal{E}_j$ where $\boldsymbol{c}_i$ is the center of the unit circle $\mathcal{C}_i$ and the unit vector $\boldsymbol{v}$ computed from vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. The point $\boldsymbol{p}$ is the closest point on $\mathcal{E}_j$ from $\boldsymbol{c}_i$ in the direction of $\boldsymbol{v}$. The vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are defined with respect to the focal points $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ of $\mathcal{E}_j$. The initial $t_0 \in \mathcal{S}$ is obtained such that $\boldsymbol{p} = (a_j \cos t_0, b_j \sin t_0)$.

solutions to the quadratic equation $\alpha r^2 - 2\beta r + \gamma = 0$ are thus given by

$$r = \frac{\beta \pm \sqrt{\beta^2 - \alpha\gamma}}{\alpha}. \tag{5.12}$$

We choose as an approximation of $\boldsymbol{x}_j$ the point $\boldsymbol{p}$, among the two intersection points, that is the closest to $\boldsymbol{c}_i$, that is, the point which is given by the smallest root of $f_j(\boldsymbol{c}_i + r\boldsymbol{v})$, i.e.

$$r_{\min} = \frac{\beta - \sqrt{\beta^2 - \alpha\gamma}}{\alpha}. \tag{5.13}$$

Therefore $\boldsymbol{p} = \boldsymbol{c}_i + r_{\min}\boldsymbol{v}$. This is illustrated in Figure 5.5. Finally, we compute the initial $t_0 \in [-\pi, \pi[$ from $\boldsymbol{p} = (p_x, p_y)$ as

$$t_0 = \arctan\left(\frac{a_j p_y}{b_j p_x}\right). \tag{5.14}$$

We note that by construction $t_0 \in \mathcal{S}$. Moreover, t_0 becomes a better approximation of t_j as the parameters a_j and b_j of $\mathcal{E}_j$ become larger.

5.4 Contact Point Algorithm

We now make some final overall comments on the contact detection algorithm, which is summarized below in pseudo-language in Algorithm 1. Using the same line of thought as other algorithms, the new algorithm is coined S-GPA, standing for Steered Geometric Potential Algorithm, since the iterative method is steered toward the solution to the minimization problem through the choice of the initial guess point and the use of the additional constraint.

First of all, we reiterate that the solution $\boldsymbol{x}_i$ to Problem (4.25) can be found in the same manner by transforming the ellipse $\mathcal{E}_i$ to the ellipse $\bar{\mathcal{E}}_i$ centered at the origin and the ellipse $\mathcal{E}_j$ to the unit circle $\bar{\mathcal{C}}_j$ centered at $\bar{\boldsymbol{c}}_j$.

Second of all, it is possible to avoid having to apply the second transformation, which is desirable because the two transformations usually dominate the total computational cost. Indeed, once the point $\bar{\boldsymbol{x}}_j$ is found in the configuration of $\bar{\mathcal{E}}_j$ and $\bar{\mathcal{C}}_i$, the point $\bar{\boldsymbol{x}}_i$ could be approximated by

$$\tilde{\boldsymbol{x}}_i = \bar{\boldsymbol{c}}_i - \bar{\boldsymbol{n}}_j(\bar{\boldsymbol{x}}_j).$$

Accuracy in $\tilde{\boldsymbol{x}}_i$ improves as the overlap gets smaller. Such an approximation could be used when high accuracy is not essential.

Algorithm 1 Algorithm to find a contact point $\boldsymbol{x}_c$ between two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$

```
Initialize ellipses E_i = {Q_i, c_i} and E_j = {Q_j, c_j}
for k = i and j do
  Transform the pair of ellipses into ellipse Ē_k and unit circle using the mapping of Section 3.7.2
  Compute the initial point t_0 using (5.14) and set t to t_0
  while |h(t)| > ε (given tolerance) do
    Set t = t − h(t)/h′(t) (one iteration of the Newton's method)
    if q(t) > 0 then
      Set t = t + h(t)/h′(t)
      Set t = t + αq̂′(t) (where α is obtained by the line search method)
    end if
  end while
  Set x̄_k = (ā_k cos t, b̄_k sin t)
  Map x̄_k to point x_k in original coordinate system using transformation (3.53)
end for
Compute contact point x_c using (4.1) and normal vector n_c(x_c) using (4.2)
```

5.5 Extension to Ellipsoids

The contact detection algorithm in the case of two ellipsoids is similar to that of two ellipses. Here we only concentrate on the distinctive features of the algorithm in 3D. We thus consider an ellipsoid $\mathcal{E}_j$ centered at the origin in its local coordinate system and a sphere $\mathcal{S}_i$ centered at $\boldsymbol{c}_i = (c_x, c_y, c_z)$, obtained after transforming two arbitrary ellipsoids in near perfect contact using the mapping of Section 3.7.2. We recall that an ellipsoid $\mathcal{E}_j$ in its local coordinate system, with semi-axes a_j, b_j, and c_j, can be parameterized in terms of the angles $\psi \in [0, 2\pi[$ and $\phi \in [0, \pi[$ such that

$$\boldsymbol{x}(\psi, \phi) = \begin{bmatrix} a_j \cos\psi \sin\phi \\ b_j \sin\psi \sin\phi \\ c_j \cos\phi \end{bmatrix}. \tag{5.15}$$

The co-gradient vector-valued function is given by:

$$\boldsymbol{H}(\boldsymbol{x}) = \nabla f_i(\boldsymbol{x}) \times \nabla f_j(\boldsymbol{x}),$$

whose components read:

$$\begin{aligned} H_x(\boldsymbol{x}) &= (z - c_z)y/b_j^2 - (y - c_y)z/c_j^2, \\ H_y(\boldsymbol{x}) &= (x - c_x)z/c_j^2 - (z - c_z)x/a_j^2, \\ H_z(\boldsymbol{x}) &= (y - c_y)x/a_j^2 - (x - c_x)y/b_j^2, \end{aligned}$$

from which we straightforwardly obtain that:

$$\frac{x}{a_j^2} H_x(\boldsymbol{x}) + \frac{y}{b_j^2} H_y(\boldsymbol{x}) + \frac{z}{c_j^2} H_z(\boldsymbol{x}) = 0.$$

Therefore, a point $\boldsymbol{x}$ belongs to the co-gradient locus $\mathcal{H}_{ij} = \{\boldsymbol{x} \in \mathbb{R}^3;\ \boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{0}\}$ if it satisfies e.g. $H_x(\boldsymbol{x}) = 0$ and $H_y(\boldsymbol{x}) = 0$, as the third equation $H_z(\boldsymbol{x}) = 0$ is necessarily satisfied if $H_x(\boldsymbol{x}) = 0$ and $H_y(\boldsymbol{x}) = 0$ (and $z \neq 0$). It follows that the points $\boldsymbol{x} = \boldsymbol{x}(\psi, \phi)$ in $\mathcal{E}_i \cap \mathcal{H}_{ij}$ are given by the roots of the vector-valued function $\boldsymbol{H} = (h_1, h_2)$ in $[0, 2\pi[\times [0, \pi[$:

$$\begin{aligned} h_1(\psi, \phi) &= (b_j^2 - c_j^2) \sin\psi \cos\phi \sin\phi - b_j c_y \cos\phi + c_j c_z \sin\psi \sin\phi, \\ h_2(\psi, \phi) &= (a_j^2 - c_j^2) \cos\psi \sin\psi \sin\phi - a_j c_x \cos\phi + b_j c_y \cos\psi \sin\phi. \end{aligned} \tag{5.16}$$

Moreover, we can introduce the scalar function on $[0, 2\pi[\times[0, \pi[$

$$\begin{aligned} q(\psi, \phi) &= \frac{1}{4}\nabla f_i(\boldsymbol{x}(\psi, \phi)) \cdot \nabla f_j(\boldsymbol{x}(\psi, \phi)) \\ &= \frac{x(\psi, \phi)}{a_j^2}(x(\psi, \phi) - c_x) + \frac{y(\psi, \phi)}{b_j^2}(y(\psi, \phi) - c_y) + \frac{z(\psi, \phi)}{c_j^2}(z(\psi, \phi) - c_z) \\ &= 1 - \frac{c_x}{a_j}\cos\psi\sin\phi - \frac{c_y}{b_j}\sin\psi\sin\phi - \frac{c_z}{c_j}\cos\psi. \end{aligned}$$

As in the 2D case, we know that there is only one root of $\boldsymbol{H}(\psi, \phi)$ that belongs to the constraint set

$$\mathcal{S} = \{(\psi, \phi) \in [0, 2\pi[\times[0, \pi[;\ q(\psi, \phi) < 0\}.$$

We also consider the function $\hat{q}(\psi, \phi)$ defined on $[0, 2\pi[\times[0, \pi[$ as

$$\hat{q}(\psi, \phi) = \frac{\nabla f_i(\boldsymbol{x}(\psi, \phi)) \cdot \nabla f_j(\boldsymbol{x}(\psi, \phi))}{\|\nabla f_i(\boldsymbol{x}(\psi, \phi))\|\ \|\nabla f_j(\boldsymbol{x}(\psi, \phi))\|},$$

which attains its minimum for the root (ψ_i, ϕ_i) of $\boldsymbol{H}(\psi, \phi)$ in $\mathcal{S}$, such that $\boldsymbol{x}_j = \boldsymbol{x}(\psi_j, \phi_j)$.

We now adapt the algorithm of Section 5.3 to find an initial guess in 3D. We note that the notion of focal points does not straightfowardly extend to the case of ellipses. However, one can still define the two sets of pairs, $(\boldsymbol{f}_{1x}, \boldsymbol{f}_{2x})$ and $(\boldsymbol{f}_{1y}, \boldsymbol{f}_{2y})$, such that

$$\begin{aligned} \boldsymbol{f}_{1x} &= (+d_x, 0, 0), \quad \boldsymbol{f}_{2x} = (-d_x, 0, 0), \\ \boldsymbol{f}_{1y} &= (0, +d_y, 0), \quad \boldsymbol{f}_{2y} = (0, -d_y, 0), \end{aligned}$$

where $d_x = \sqrt{a_j^2 - b_j^2}$ and $d_y = \sqrt{b_j^2 - c_j^2}$. Following the steps described in the equations (5.9), (5.10), and (5.11), one can compute the vector $\boldsymbol{v}_{ab}$ from the set $(\boldsymbol{f}_{1x}, \boldsymbol{f}_{2x})$ and the vector $\boldsymbol{v}_{bc}$ from the set $(\boldsymbol{f}_{1y}, \boldsymbol{f}_{2y})$. We then introduce the unit vector $\boldsymbol{v}$ as

$$\boldsymbol{v} = \frac{\boldsymbol{v}_{ab} + \boldsymbol{v}_{bc}}{\|\boldsymbol{v}_{ab} + \boldsymbol{v}_{bc}\|}.$$

The initial point $\boldsymbol{p} = (p_x, p_y, p_z)$ on the ellipsoid $\mathcal{E}_j$ is thus obtained as

$$\boldsymbol{p} = \boldsymbol{c}_i + r_{\min}\boldsymbol{v}, \tag{5.17}$$

where $r_{\min}$ is given by (5.13). We finally derive the initial angles ψ_0 and ϕ_0 to find the roots

of (5.16) by Newton's method as

$$\phi_0 = \arccos\left(\frac{p_z}{c_j}\right),$$
$$\psi_0 = \arctan\left(\frac{a_j p_y}{b_j p_x}\right).$$

In conclusion, the algorithm to identify the contact point between two ellipsoids follows the same steps as those described in Algorithm 1.

CHAPTER 6 NUMERICAL RESULTS

The objective of this chapter is to provide direct comparisons in accuracy and efficiency between the different contact detection algorithms presented in Chapters 4 and 5. In fact, the existing literature only provides a few comparisons between existing algorithms, mostly on a few pairs of ellipses, and so many of results presented below are new. Comparisons are difficult in practice because

(i) algorithms based on different minimization problems have fundamentally different solutions;

(ii) some algorithms may be based on root finding of scalar polynomial equations, others on coupled nonlinear systems, and others may have a key transformation to normalized coordinate systems. These choices dramatically affect accuracy and cost, how they are initialized, and even how accuracy is measured;

(iii) the wide variety of numerical techniques deployed imply that the computational efficiency is highly dependent on specific implementation details (code optimization, language strengths and platform choice);

(iv) some algorithms have weakness in robustness, or may not be able to exploit prior accurate estimates of contacts, which render them undesirable in certain applications.

These difficulties will be circumvented by focusing mostly on the underlying minimization problems driving the algorithms and then dedicating Sections 6.5, 6.6, and 6.7 to applying the algorithms on large number of pairs of ellipses in close/almost contact. The goal of the comparison is to demonstrate on several numerical examples the performance and accuracy of the new contact detection algorithm and compare its efficiency with the existing ones. For this purpose, To overcome these difficulties, we will compare different algorithms on large sets of pairs of ellipses and ellipsoids generated randomly by the algorithms described in Section C.1.1 and Section C.1.2, respectively. These algorithms produce a large number of pairs of ellipses or ellipsoids in near perfect contact for which we know the exact position of the point $\boldsymbol{x}_j$ from Equation 3.9. This will help us in particular to assess the accuracy in finding an approximation of $\boldsymbol{x}_j$ when using the Geometric Potential Algorithms (GPAs). We note however that the exact position of the associated $\boldsymbol{x}_i$ is not known for these pairs of particles.

All algorithms are prototyped in MATLAB. The algorithms for L-GPA, M-GPA, C-GPA, and IA, require finding the roots of a polynomial function of degree four in 2D. For consistent comparison between the methods, we have chosen the same algorithm to find the roots in all cases, that consists in defining the companion matrix associated with the polynomial function and evaluating all eigenvalues using the Francis' algorithm [74]. For P-GPA, we combine the golden-section search method and Newton's method to search for the critical points associated with Problem (3.8). However, we note that in the absence of an additional constraint such as the one introduced in Section 5.2, the algorithm may converge for some pairs of ellipses to a local extremum instead of the global minimum. We nevertheless report the results obtained with this method in the examples below.

The first four examples, detailed in subsections 6.1-6.4, will present pairs of ellipses for which the differences are attributable solely to the minimization problem on which they are based. This implies that the computational cost will be ignored and all the algorithms will be run with a high tolerance in order to provide to machine precision the *exact* solution to the minimization problem. For these first four test problems, none of the contact detection algorithms failed, so we could focus only on the minimization problem.

In Section 6.5, our objective is to highlight the algorithmic aspects of the resolution of the different minimization problems. To accomplish this, ten thousand pairs of ellipses in close/almost contact are generated according to an algorithm in Section C.1. For each algorithm, the total computational time is found and then further subdivided into its different steps; see Tables 6.10, 6.11 and 6.12. This allows us to address issues (ii) and (iii). For those tests, uniformly low tolerances were used.

We compare the algorithms for $1,000$ pairs of ellipses in Section 6.6 with respect to their aspect ratio a/b and relative size $a_i b_i / a_j b_j$. In this section, we also study the influence of overlap size on performance and accuracy of the algorithms in GPA class. In Section 6.7, we compare performance of L-GPA and the S-GPA for $1,000$ pairs of ellipses and ellipsoids while keeping the same accuracy. Overall, the comparison indicates that the new contact detection algorithm is the fastest, but because of issues just mentioned at the beginning of this introduction, any comparison between algorithms comes with significant caveats.

6.1 Comparing the Intersection Set (IS), the MDP, and the MPP

The objective of this example is to demonstrate that with strict tolerances, the contact point obtained by a given algorithm will depend only on the minimization problem which the algorithm attempted to solve. Following this remark, the numerical results of Sections

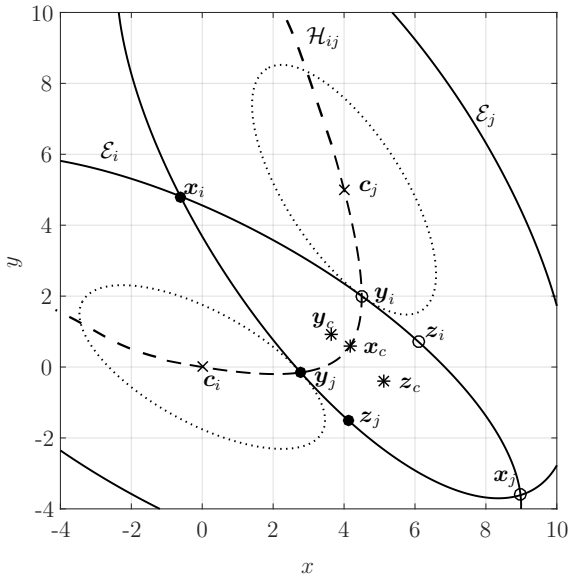


Figure 6.1 Locations of the contact points using different contact detection algorithms : the Intersection Set is the pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ with a contact $\boldsymbol{x}_c$, the MPP is $(\boldsymbol{y}_i, \boldsymbol{y}_j)$ with a contact point $\boldsymbol{y}_c$, and the MDP is $(\boldsymbol{z}_i, \boldsymbol{z}_j)$ with a contact point $\boldsymbol{z}_c$. The co-gradient locus $\mathcal{H}_{ij}$ is drawn as a dashed line and it traverses both centers $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$. The dotted line presents the scaled ellipses that are tangent at the MPP.

6.2 and 6.3 will focus only on the different solutions associated either to MDP and MPP. Nevertheless, the example in this section is also interesting in its own right since it presents a pair of overlapping ellipses, given in Table 6.1, for which the Intersection Set $(\boldsymbol{x}_i, \boldsymbol{x}_j)$, the Minimum Distance Pair $(\boldsymbol{z}_i, \boldsymbol{z}_j)$, and the Minimum Potential Pair $(\boldsymbol{y}_i, \boldsymbol{y}_j)$ are different. In particular, this example shows that the resulting contact points $\boldsymbol{x}_c, \boldsymbol{z}_c$ and $\boldsymbol{y}_c$ will be distinct, as seen in Figure 6.1. On the other hand, the normals at each of these three contact points are roughly the same, computed according to either (4.2) or (4.3). In general, we have found the normals to be less sensitive to the choice of the algorithm than the estimates of the contact point.

The numerical results in Tables 6.2 and 6.3 show that the MPP $(\boldsymbol{y}_i, \boldsymbol{y}_j)$ are the same whether they are computed by the P-GPA, L-GPA, M-GPA, C-GPA, or the S-GPA. Similarly, using sufficiently high tolerances we find the estimated MDP is the same whether computed by the CNA or the CCNA.

Examining Figure 6.1 and comparing it to the definition of the MPP and MDP, it is easy to explain the differences between the contact points $\boldsymbol{x}_c$, $\boldsymbol{y}_c$, and $\boldsymbol{z}_c$. For example, the MPP $(\boldsymbol{y}_i, \boldsymbol{y}_j)$ clearly belong to the co-gradient locus but their normals are different at each point.

On the other hand, it is clear in Figure 6.1 that the normals at the MDP $(\boldsymbol{z}_i, \boldsymbol{z}_j)$ have the same normals. The contact point $\boldsymbol{z}_c$ is far from the co-gradient locus because both ellipses have roughly parallel surfaces along a large portion of the intersection $E_i \cup E_j$.

Table 6.1 The two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ of Example 6.1.

	a	b	$\boldsymbol{c}$	θ	a/b
Ellipse $\mathcal{E}_i$	10	4.1	$(0, 0)$	-0.5	2.44
Ellipse $\mathcal{E}_j$	10	4.1	$(4, 5)$	-1	2.44

Table 6.2 The contact points $\boldsymbol{x}_c$ for the different contact detection algorithms in Example 6.1.

Algorithm	$\boldsymbol{x}_i$	$\boldsymbol{x}_j$	$\boldsymbol{x}_c$
IA	$(8.977, -3.611)$	$(-0.619, 4.816)$	$(4.179, 0.602)$
P-GPA	$(4.497, 2.002)$	$(2.778, -0.154)$	$(3.638, 0.923)$
L-GPA	$(4.497, 2.002)$	$(2.778, -0.154)$	$(3.638, 0.923)$
M-GPA	$(4.497, 2.002)$	$(2.778, -0.154)$	$(3.638, 0.923)$
C-GPA	$(4.497, 2.002)$	$(2.778, -0.154)$	$(3.638, 0.923)$
S-GPA	$(4.497, 2.002)$	$(2.778, -0.154)$	$(3.638, 0.923)$
CNA	$(6.108, 0.703)$	$(4.118, -1.504)$	$(5.113, -0.401)$
CCA	$(6.108, 0.703)$	$(4.118, -1.504)$	$(5.113, -0.401)$

Table 6.3 The normal vectors $\boldsymbol{n}_c$ for the different contact detection algorithms in Example 6.1.

Algorithm	$\boldsymbol{n}_i$	$\boldsymbol{n}_j$	$\boldsymbol{n}_c$
IA	$(0.994, 0.112)$	$(-0.890, -0.455)$	$(0.660, 0.751)$
P-GPA	$(0.587, 0.809)$	$(-0.744, -0.668)$	$(0.669, 0.743)$
L-GPA	$(0.587, 0.809)$	$(-0.744, -0.668)$	$(0.669, 0.743)$
M-GPA	$(0.587, 0.809)$	$(-0.744, -0.668)$	$(0.669, 0.743)$
C-GPA	$(0.587, 0.809)$	$(-0.744, -0.668)$	$(0.669, 0.743)$
S-GPA	$(0.587, 0.809)$	$(-0.744, -0.668)$	$(0.669, 0.743)$
CNA	$(0.670, 0.743)$	$(-0.670, -0.743)$	$(0.670, 0.743)$
CCA	$(0.670, 0.743)$	$(-0.670, -0.743)$	$(0.670, 0.743)$

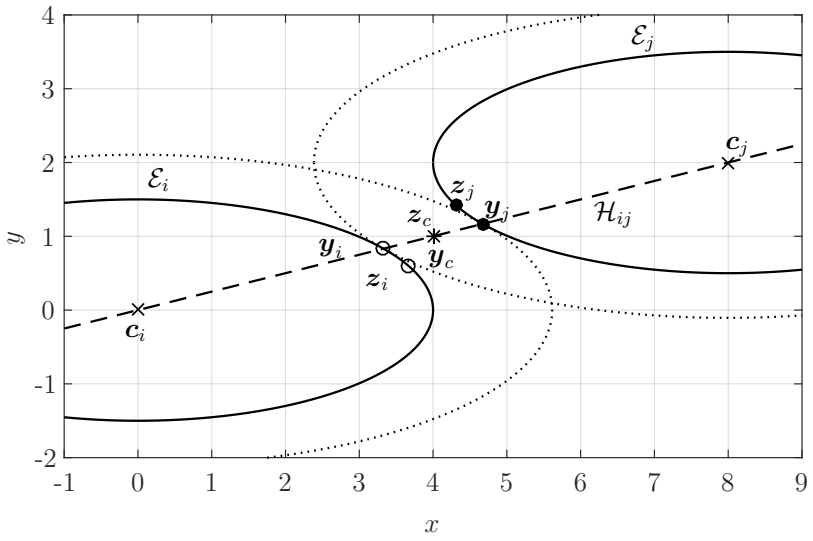


Figure 6.2 Two disjoint ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ with the same major and minor axis are aligned with x-axis. In this configuration, $\mathcal{H}_{ij}$ degenerates to a straight line passing through $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$. The MDP $(\boldsymbol{z}_i, \boldsymbol{z}_j)$ are not located on the co-gradient locus $\mathcal{H}_{ij}$ as the MPP $(\boldsymbol{y}_i, \boldsymbol{y}_j)$. However, the contact points $\boldsymbol{z}_c$ and $\boldsymbol{y}_c$ are identical and are both located on the co-gradient locus $\mathcal{H}_{ij}$.

6.2 Different MDP and MPP with the Same Contact Points

In this example, we present a pair of ellipses, described in Table 6.4, for which the contact point $\boldsymbol{z}_c$ of the MDP $(\boldsymbol{z}_i, \boldsymbol{z}_j)$ and the contact point $\boldsymbol{y}_c$ of the MPP $(\boldsymbol{y}_i, \boldsymbol{y}_j)$ are the same, but the associated normals are different. This expands on the previous example because the pairs are different, yet lead to equal contacts. The fact that the normals are different but the contact points are the same implies that the choice of using the MDP or the MPP could lead to significantly different forces between ellipses in a DEM model.

The example is quite simple because, as Table 6.4 shows, both ellipses have the same aspect ratio and the co-gradient locus degenerates to a straight line through both centers. The quartic equations (4.19) and (4.28) derived from M-GPA and C-GPA, respectively, degenerate in this case to quadratic equations, i.e. $a_4 = a_3 = a_1 = 0$. This implies that the mappings in Sections 3.7.1 and 3.7.2, applied respectively in M-GPA and C-GPA, send both ellipses to circles. In those new coordinates, the contact is easy to compute analytically. This shows that the coincidence $\boldsymbol{z}_c = \boldsymbol{y}_c$ is exact, and not simply an artifact of the numerical algorithms.

In Table 6.5, we present the MDP and the MPP obtained by respectively a GPA algorithm and the CCN with high tolerance. It is a coincidence in this example that the normals are

opposite, both for the MDP and the MPP.

Table 6.4 The two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ of Example 6.2.

	a	b	$\boldsymbol{c}$	θ	a/b
Ellipse $\mathcal{E}_i$	4	1.5	$(0, 0)$	0	2.67
Ellipse $\mathcal{E}_j$	4	1.5	$(8, 2)$	0	2.67

Table 6.5 The contact points $\boldsymbol{x}_c$ and their normal vectors $\boldsymbol{n}_c$ for the two classes of contact detection algorithms in Example 6.2.

	GPA	CCA
$\boldsymbol{x}_i$	$(3.328, 0.832)$	$(3.662, 0.604)$
$\boldsymbol{x}_j$	$(4.672, 1.168)$	$(4.338, 1.396)$
$\boldsymbol{x}_c$	$(4, 1)$	$(4, 1)$
$\boldsymbol{n}_i(\boldsymbol{x}_i)$	$(0.490, 0.872)$	$(0.649, 0.761)$
$\boldsymbol{n}_j(\boldsymbol{x}_j)$	$(-0.490, -0.872)$	$(-0.649, -0.761)$
$\boldsymbol{n}_c(\boldsymbol{x}_c)$	$(0.490, 0.872)$	$(0.649, 0.761)$

6.3 Ellipses in Perfect Contact

Following the work of Dziugys et al. [24, 72], we consider a pair ellipses of high-aspect ratio in perfect contact, according to Definition 4. This is a numerically challenging case, yet both the MDP and the MPP coincide mathematically in this case. As expected, every algorithm identified the same contact point. This perfect contact is illustrated in Figure 6.3. The ellipses are described in Table 6.6 and the resulting MDP and MPP pair are given in Table 6.7.

Table 6.6 The two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ of Example 6.3.

	a	b	$\boldsymbol{c}$	θ	a/b
Ellipse $\mathcal{E}_i$	1 0.025	$(-0.7073277, 0)$	0.753	40	
Ellipse $\mathcal{E}_j$	1	0.025	$(0.7073277, 0)$	-0.753	40

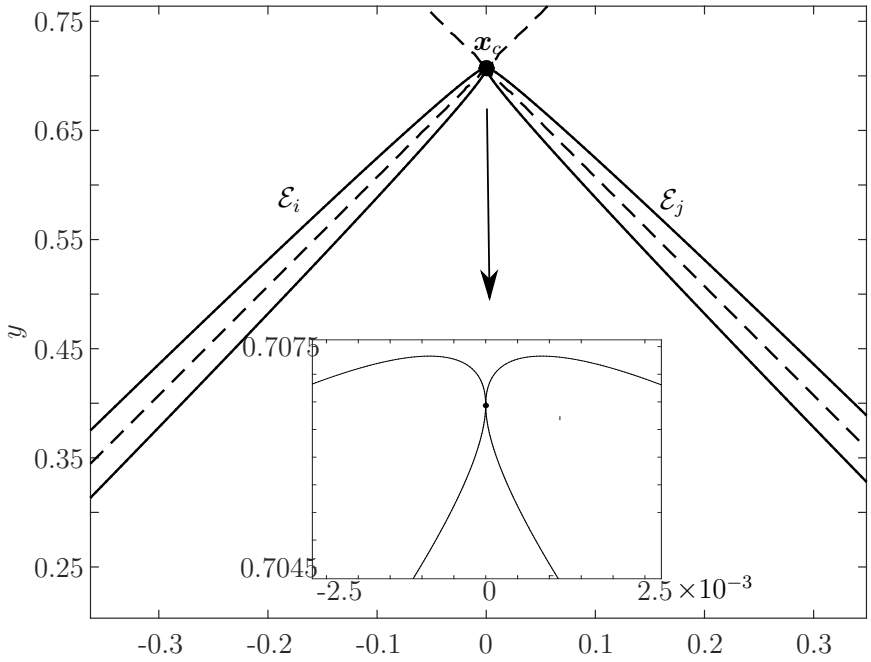


Figure 6.3 The location of the contact points $\boldsymbol{x}_c$ is the same for both GPA and CCA.

Table 6.7 Contact point $\boldsymbol{x}_c$ and normal vector $\boldsymbol{n}_c$ obtained by contact detection algorithms in Example 6.3.

	GPA	CCA
$\boldsymbol{x}_i$	$(1.73\text{e}{-8}, 0.706)$	$(1.73\text{e}{-8}, 0.706)$
$\boldsymbol{x}_j$	$(-1.73\text{e}{-8}, 0.706)$	$(-1.73\text{e}{-8}, 0.706)$
$\boldsymbol{x}_c$	$(0, 0.706)$	$(0, 0.706)$
$\boldsymbol{n}_i(\boldsymbol{x}_i)$	$(1, 0)$	$(1, 0)$
$\boldsymbol{n}_j(\boldsymbol{x}_j)$	$(-1, 0)$	$(-1, 0)$
$\boldsymbol{n}_c(\boldsymbol{x}_c)$	$(1, 0)$	$(1, 0)$

6.4 Ellipses with Small Overlap

In this challenging test, the two ellipses of high-aspect ratio have a relatively innocuous overlap which, with respect to their size, one would not expect to cause trouble. Yet, as Figure 6.4 shows, the contact point $\boldsymbol{x}_c$ from intersection set is located close to the boundary of ellipse $\mathcal{E}_i$. The point $\boldsymbol{z}_c$ associated to the MDP does not belong to the intersection of both ellipses, i.e. $E_i \cap E_j$. However, the MPP produces a reasonable contact point $\boldsymbol{y}_c$ inside the intersection of both ellipses. We insist here that these ellipses are not in **near perfect contact**, that is according to the Definition 6, because the two disks of radius $\underline{\rho}_i$ and $\underline{\rho}_j$,

tangent to respectively $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$, are disjoint. The description of the ellipses is given in Table 6.8. The estimates of the contact points for both the IS, the MDP, and the MPP are given in Table 6.9. We note that the estimated normals are very close to each other.

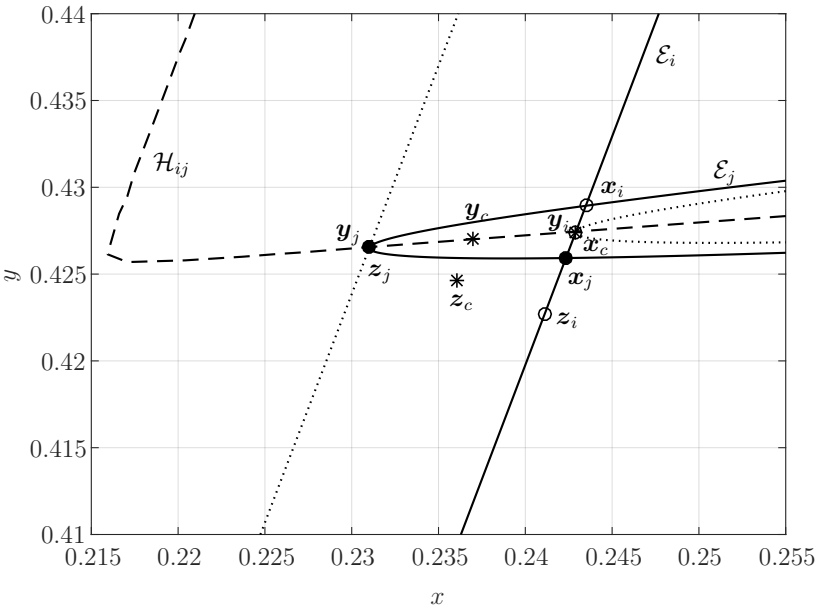


Figure 6.4 Locations of contact points using different contact detection algorithms, $(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_c)$ are obtained by using IA, $(\boldsymbol{y}_i, \boldsymbol{y}_j, \boldsymbol{y}_c)$ are obtained by using GPA, and $(\boldsymbol{z}_i, \boldsymbol{z}_j, \boldsymbol{z}_c)$ are obtained by using CCA.

Table 6.8 The ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ of Example 6.1.

	a	b	$\boldsymbol{c}$	θ	a/b
Ellipse $\mathcal{E}_i$	7	0.025	$(0.235, 0.477)$	1.2077	280
Ellipse $\mathcal{E}_j$	7	0.025	$(7.2113, 0.9515)$	0.0750	280

Table 6.9 Contact point $\boldsymbol{x}_c$ and normal vector $\boldsymbol{n}_c$ obtained by different algorithms in Example 6.4.

	IA	GPA	CCA
$\boldsymbol{x}_i$	$(0.24348, 0.42893)$	$(0.24291, 0.42743)$	$(0.24111, 0.42270)$
$\boldsymbol{x}_j$	$(0.24234, 0.42592)$	$(0.23102, 0.42653)$	$(0.23102, 0.42653)$
$\boldsymbol{x}_c$	$(0.24291, 0.42743)$	$(0.23696, 0.42698)$	$(0.23606, 0.42462)$
$\boldsymbol{n}_i(\boldsymbol{x}_i)$	$(0.934815, -0.355135)$	$(0.934815, -0.35513)$	$(0.934814, -0.35514)$
$\boldsymbol{n}_j(\boldsymbol{x}_j)$	$(0.012235, -0.99992)$	$(-0.934807, 0.35515)$	$(-0.934814, 0.35514)$
$\boldsymbol{n}_c(\boldsymbol{x}_c)$	$(0.934815, -0.35513)$	$(0.934810, -0.35514)$	$(0.934814, -0.35514)$

6.5 Statistical Comparison of the Algorithms for Ellipses

In contrast to the previous four sections, we attempt to analyze and compare the accuracy and efficiency of the algorithms presented in Chapters 4 and 5. More precisely, we study the individual numerical approximations implemented in the algorithms used to solve the two main contact detection problems we identified, that is the MDP and the MPP, independent of the characteristics of the problems themselves studied in the previous four sections. As we argued in the introduction of this chapter, there are many reasons why comparisons between contact detection algorithms could be difficult. Nevertheless, as we will explain below, we can come to a limited number of conclusions by studying the behavior of these algorithms on a large sample of pairs of ellipses in close contact, and then separately considering the contributions to accuracy and efficiency of the individual components of the algorithm. The tests in this section are new to the literature and should help to establish benchmarks for comparisons between such algorithms for contact detection.

The first step is to generate a random set of $10,000$ pairs of ellipses in almost/close contact, according to an algorithm described in Appendix C.1.1, and to apply to each pair of ellipses one of the seven algorithms of Chapter 4 and the S-GPA in Chapter 5. In order to reproduce pairs of ellipses one might encounter in the DEM, the generating algorithm provided some control on the aspect ratio of each ellipse, on their relative orientation, on their relative closeness, and on the location of the contact point along the boundary of each ellipse. First of all, the first ellipse $\mathcal{E}_i$ was permitted to have maximum aspect ratio of $a/b = 5$ while ellipse $\mathcal{E}_j$ was permitted to have an aspect ratio as large as $a/b = 20$. Figure 6.5 presents the actual distribution of the aspect ratio a/b for $\mathcal{E}_i$ and $\mathcal{E}_j$. The generating algorithm assumes that $a_j b_j = 1$ for $\mathcal{E}_j$, but the distribution of the area πab for $\mathcal{E}_i$ is randomly determined and shown in Figure 6.6a.

The algorithm generating these pairs was able to provide the exact MPP, thus allowing us to estimate the relative error of the algorithms in the GPA family, i.e the P-GPA, L-GPA, M-GPA, C-GPA, and the S-GPA. On the other hand, for the IS algorithm and the MDP algorithms we were not able to provide estimates of the error. In any case, this implies that we are able to provide in Figure 6.6b the distribution of the penetration/separation distance, measured according to $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|/\min\{\|\boldsymbol{x}_i - \boldsymbol{c}_i\|, \|\boldsymbol{x}_j - \boldsymbol{c}_j\|\}$ supporting our claim that the pairs of ellipses in our tests were relatively close. Furthermore, the generation of the MPP for the pair of ellipses also allowed for the pair to be selected uniformly along the boundary of the first ellipse, thereby ensuring that we tested MPP occurring both in the flatter or more curved regions of the boundary of the ellipses.

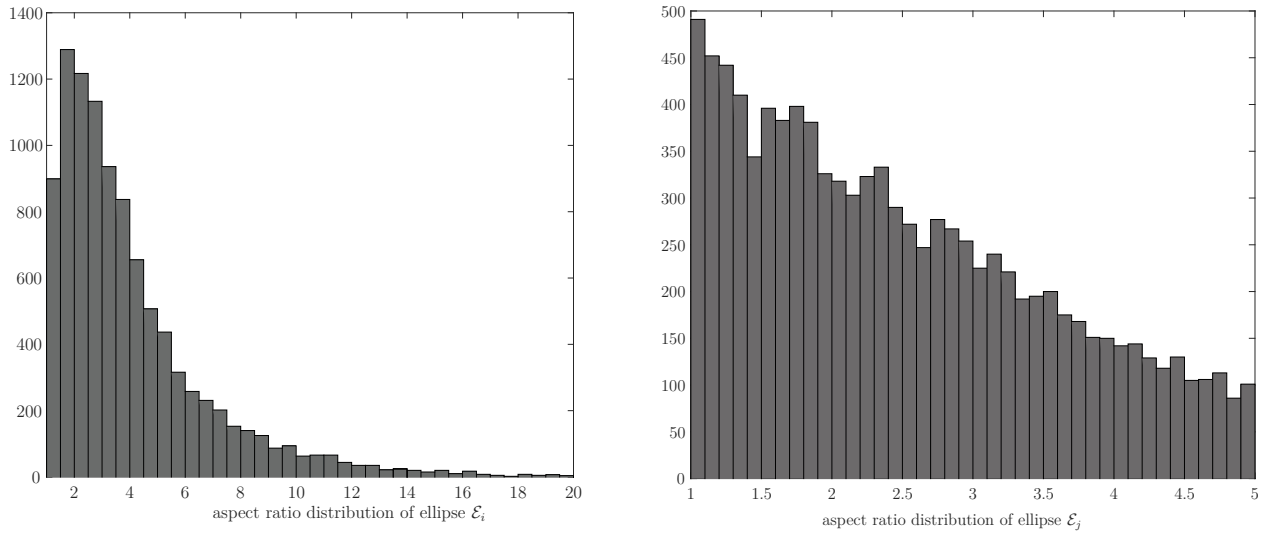
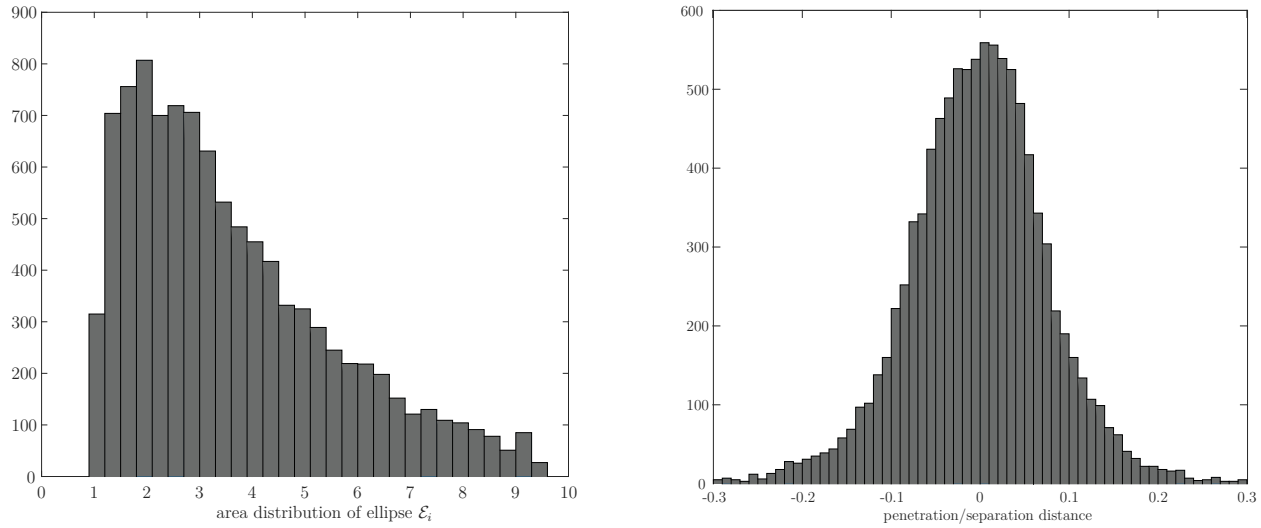


Figure 6.5 Distribution of the aspect ratio of the ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$.

The numerical results of the experiments are summarized in Tables 6.10 and 6.11, where a relative tolerance of 10^{-5} was used, and in Table 6.12 where a stricter relative tolerance of 10^{-9} was used. We will discuss the results of Table 6.12 later, since it mostly concerns the observed convergence and how it depends on the underlying numerical approximations. Tables 6.10 and 6.11 present for each of the seven algorithms of Chapter 4 and the S-GPA, the total computational time required for the resolution of the 10^4 pairs of ellipses, the statistics of the number of iterations required for the resolution, and the statistics of the error in those approximate solutions. First of all, the error could only be measured for the algorithms estimating the MPP, since the algorithm generating the pairs only provided the exact MPP. This explains why the error for the CNA and the CCA was not tabulated. The tables also include the number iterations that were required to attain the desired relative tolerance, but one



(a) (b)

Figure 6.6 (a) Distribution of the area of the ellipses $\mathcal{E}_i$. (b) Distribution of the penetration/separation distance for the pairs of ellipses, where a positive (negative) distance represents a separation (penetration) distance.

must be careful when comparing these values because the nature of the iterations in, say, the golden search algorithm, Newton's method, or in Francis' algorithm are completely different. Finally, we have included the total computational time required to solve the 10^4 pairs of ellipses, as measured by the profiler in MATLAB. Although estimates of computational time in MATLAB are known to be somewhat variable, we have performed many such studies and found the estimates of computational time to be consistently reproducible to within 10%.

We now proceed to analyze the results of the experiments in Tables 6.10 and 6.11, going from the most general to the most specific conclusions. First of all, we remark that the median error was roughly 10^{-11}, that is several orders of magnitude lower than the chosen tolerance, showing that for most pairs of ellipses, the algorithms converged quickly to the MPP. The relatively low standard deviation further shows that the error was close to the relative tolerance of 10^{-5} for only a small subset of the pairs of ellipses. In Tables 6.10 and 6.11, we immediately remark that the total computational time is only roughly equal to the sum of the time required for the different components because we omitted the computationally insignificant but necessary step of computing the coefficients in the systems of equations we needed to solve. We observe that in the L-GPA, M-GPA and C-GPA algorithms, a significant fraction of the computational effort is spent on finding the roots of a polynomial. However, in the S-GPA, mapping costs more than finding a root by Newton's method. The data also

shows that the most expensive algorithm was CNA because it required a system of equations to be solved rather than simply finding the roots of a polynomial of degree four as in the case of the algorithms L-GPA, M-GPA, and C-GPA.

Tables 6.10 and 6.11 contain some interesting observations about the efficiency of the iterative solvers underlying some of these algorithms. First of all, it appears that the Lagrange approach leads to the system requiring the smallest number of iterations. On the other hand, the P-GPA uses a golden search algorithm to compute initial estimates of the minima and maxima before invoking Newton's method to converge rapidly to the minima and maxima. Unfortunately, our tests indicate that it is difficult to reduce the number of iterations in the golden search without obtaining initial estimates for which Newton's method will not converge. As a matter of fact, even using the golden search algorithm with the recommended tolerances, roughly 1% of the pairs of ellipses did not converge to a contact point for the P-GPA. In order to maintain the consistency of our tests, the pairs of ellipses for which P-GPA failed to converge were also excluded from our tests with the other methods. We remark that the standard deviation of number of iterations for M-GPA and C-GPA are higher than the other algorithms, which leads us to conclude the number of iterations of M-GPA and C-GPA depend on the relative configuration of the two ellipses. In contrast, the convergence of the S-GPA appears to be independent of the relative geometry between the ellipses.

Later in this section, we will examine the pairs of ellipses for which different algorithms attained the maximum number of iterations; see Figures 6.7, 6.8, and 6.9. Although the number of iterations required by different algorithms are not necessarily correlated, the pairs of ellipses for which a specific algorithm had more difficulty could indicate that certain geometrical properties reduce robustness.

Overall, the data indicates that the new algorithm was the most efficient. Our hypothesis is this algorithm combines a good initial estimate of the MPP (for an ellipse at origin and a unit circle) with Newton's method quadratic convergence. In practice, we found that the S-GPA converged in a single iteration. Finally, the numerical experiments indicate that the CNA and CCA algorithms were by far the most costly alternatives. We will therefore refrain from discussing them any further.

Table 6.10 Computational cost, number of iterations, and logarithmic error for MPP algorithms using a relative tolerance of 10^{-5}.

	P-GPA	L-GPA	M-GPA	C-GPA	S-GPA
total computational cost (s)	8.24	10.79	15.22	13.76	1.96
mapping	0	0	1.31	1.07	1.31
initialization (focal points)	0	0	0	0	0.21
initialization (golden search)	5.55	0	0	0	0
root finding (Francis' algorithm)	0	9.90	12.79	11.65	0
root finding (Newton's method)	0.09	0	0	0	0.09
number of iterations					
maximum	199	41	56	50	4
minimum	15	11	5	5	1
median	18	18	23	22	3
standard deviation	7.31	3.27	8.31	8.11	0.68
logarithmic error					
maximum	−5.18	−5.17	−5.05	−5.12	−5.12
median	−10.84	−11.97	−12.00	−12.46	−10.22
standard deviation	1.65	2.14	2.4261	2.386	2.61

Table 6.11 Computational cost, number of iterations, and logarithmic error for MDP algorithms using a relative tolerance of 10^{-5}.

	CNA	CCA
total computational cost (s)	107.58	22.03
root finding (MATLAB functions)	101.22	17.47
number of iterations		
maximum	109	13
minimum	4	1
median	10	4
standard deviation	17.90	2.07

We now turn to the data in Table 6.12 obtained using the same set of pairs of ellipses but with a stricter tolerance of 10^{-9}. A priori, we expect the results to indicate the same overall trends but the stricter tolerance should help to identify any robustness issues. It is clear that the stricter tolerance produces more accurate estimates and requires a larger number of iterations. Yet, it is noticeable that the median number of iterations is almost identical while the median error is roughly 10^{-4} smaller. This indicates that for the majority of the test

cases we considered, the algorithm was already within the asymptotic regime of convergence and in many cases one or no iterations were required to satisfy the desired tolerances.

Among the L-GPA, M-GPA, and C-GPA, the L-GPA is again the most efficient and accurate, but it is still an order of magnitude slower than the S-GPA. The low standard deviation of the number of iterations suggests that the S-GPA was also the most robust algorithm. It appears that the M-GPA and C-GPA algorithms both required more iterations of their root-finding algorithm, Francis' method, in order to obtain the MPP, particularly when contrasted with L-GPA. We hypothesize that the mapping step, present in M-GPA and C-GPA but not in L-GPA, might make the root-finding problem harder, although further tests would be required to confirm this.

Table 6.12 Computational cost, number of iterations, and logarithmic error for different algorithms using a relative tolerance of 10^{-9}.

	P-GPA	L-GPA	M-GPA	C-GPA	S-GPA
total computational cost (s)	8.26	12.12	16.57	15.20	1.98
mapping	0	0	1.31	1.07	1.31
initialization (focal points)	0	0	0	0	0.21
initialization (golden search)	5.55	0	0	0	0
root finding (Francis' algorithm)	0	11.25	14.14	13.09	0
root finding (Newton's method)	0.11	0	0	0	0.11
root finding (Matlab functions)	0	0	0	0	0
number of iterations					
maximum	201	46	59	55	6
minimum	15	14	5	5	2
median	19	21	26	24	5
standard deviation	7.24	3.37	8.66	8.53	0.80
logarithmic error					
maximum	−9.08	−9.14	−9.16	−9.05	−9.16
median	−15.14	−14.61	−14.75	−15.41	−15.67
standard deviation	2.01	1.47	1.01	0.88	0.84

Finally, we conclude this section by examining pairs of ellipses for which certain algorithms required the largest number of iterations from among our sample set. Figure 6.7 presents the pair of ellipse that required the largest number of iterations in Francis' method. In this case, the two ellipses appear to have their principal axis roughly aligned and to be in contact near the regions of highest curvature. We also note that in this configuration, the points on the segment formed by $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ cross the segment formed of $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$. It is somewhat surprising

that this configuration might be difficult for the M-GPA to handle. Figure 6.8 presents the pair of ellipses that required the most iterations of the L-GPA. This configuration appears very similar to the one associated to the M-GPA. Finally, Figure 6.9 presents the worst case scenario for the C-GPA and again, the configuration seems unexceptional although the principal axis are angled by roughly $\pi/4$ and the contact occurs at points where the curvature is intermediate. In other words, the pair of ellipse in Figure 6.9 is completely the opposite than what we found in the previous two configurations. We did not present the pair of ellipse associated to the worst case scenario for the new algorithm because there was very little variation on the number of iterations. Furthermore, the new algorithm profited from a unique initialization, hence a comparison with the other GPA-type algorithm would be biased.

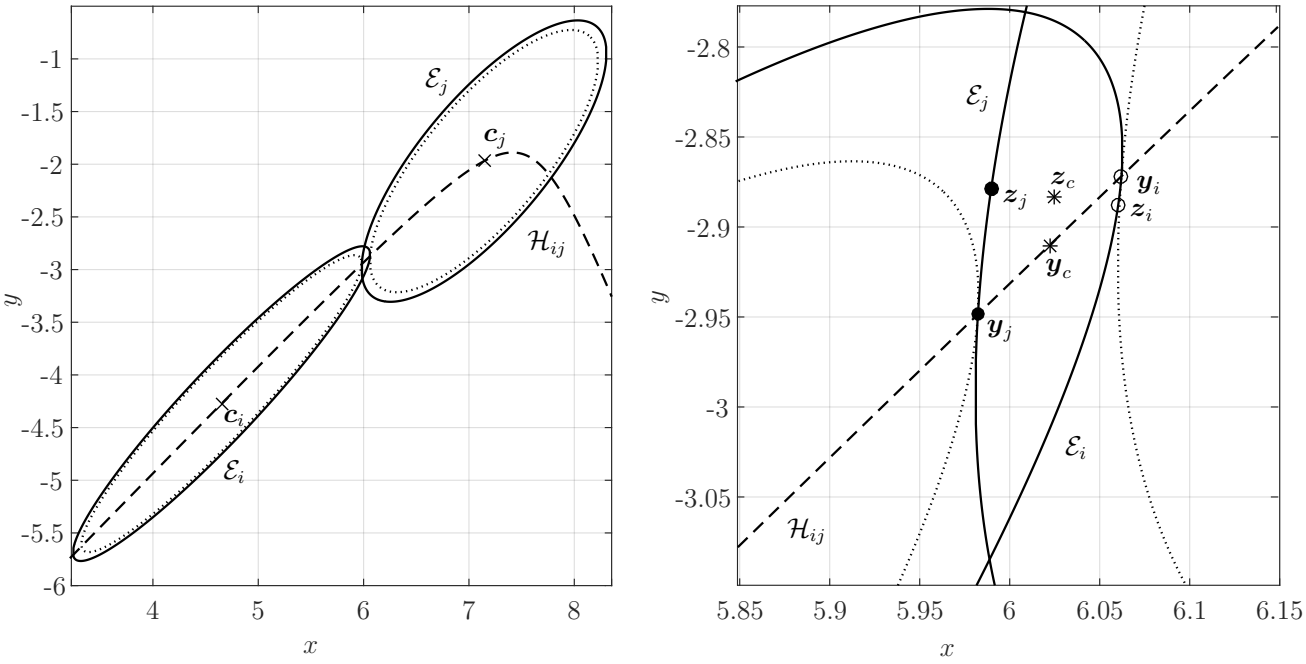


Figure 6.7 In this configuration of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$, M-GPA requires the largest number of iterations to find $\boldsymbol{y}_j$. The points $(\boldsymbol{y}_i, \boldsymbol{y}_j, \boldsymbol{y}_c)$ are the MPP and the contact point while $(\boldsymbol{z}_i, \boldsymbol{z}_j, \boldsymbol{z}_c)$ is the MDP and its contact point.

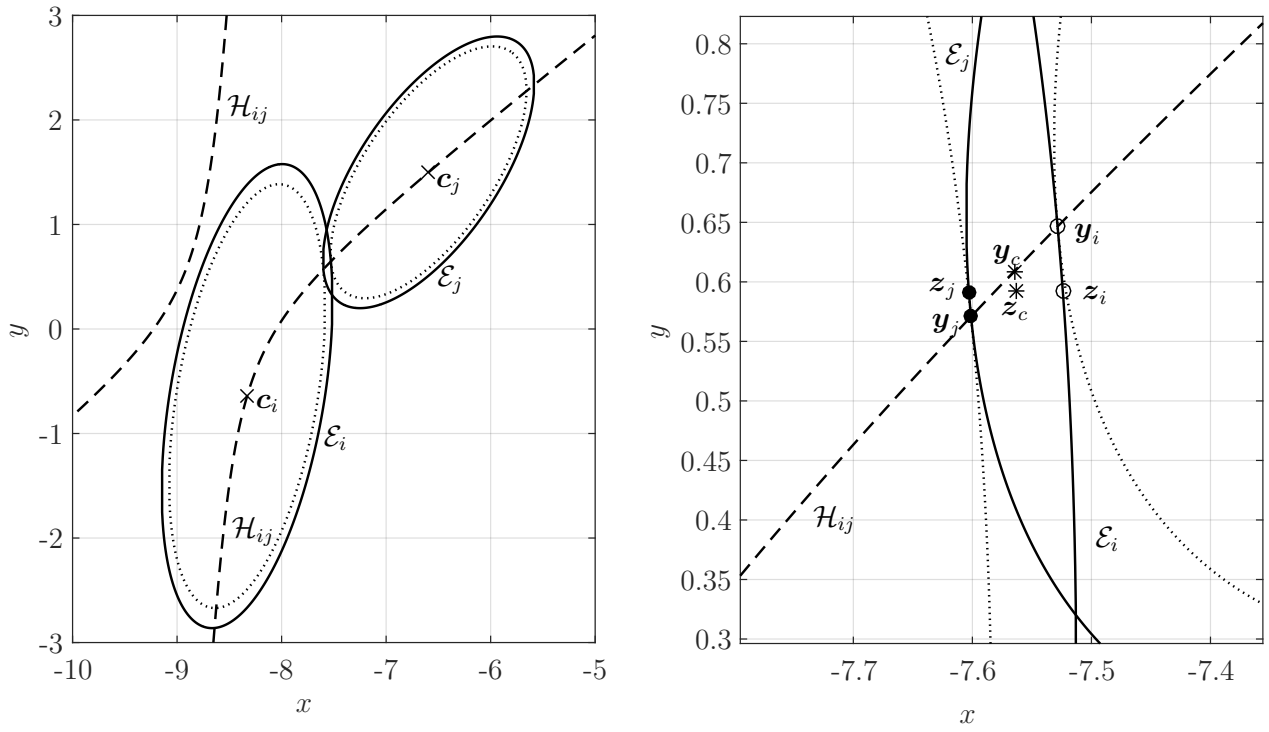


Figure 6.8 In this configuration of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$, C-GPA requires its largest number of iterations to converge to $\boldsymbol{y}_i$.

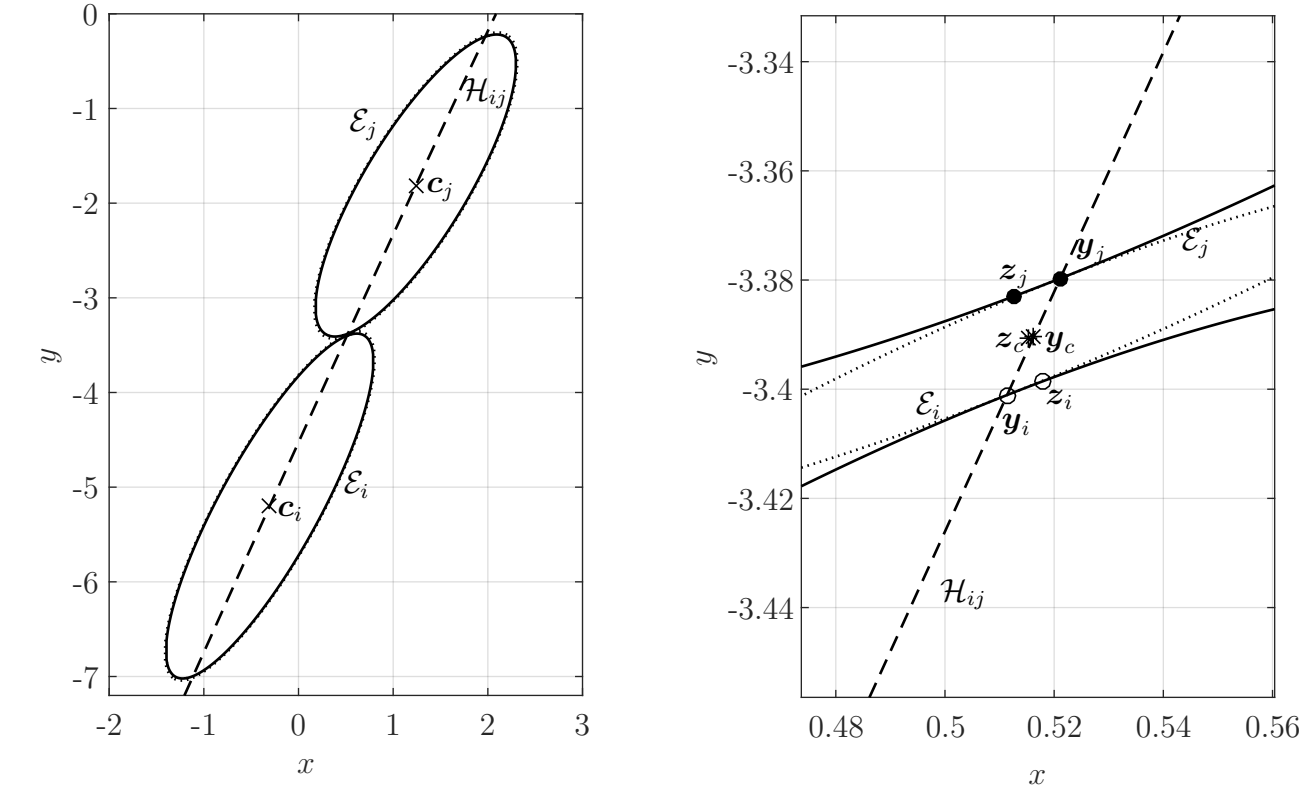


Figure 6.9 In this configuration of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$, L-GPA required its largest number of iterations to find $\boldsymbol{y}_i$.

6.6 Performance and Accuracy of the Algorithms for Ellipses with Respect to their Aspect Ratio, Size and Overlap

The main objective of this section is to compare the performance of existing algorithms for contact detection with the S-GPA. We shall compare in particular the computational time needed to compute contact points as we are interested in fast algorithms. In order to do so, we generate sample sets made of $1,000$ pairs of ellipses randomly generated using the algorithm described in Section C.1.1 or a variant of it. We also study the influence of certain parameters such as the aspect ratio a/b of the ellipses, their size ab, or the penetration length or separation distance ε between the ellipses. We note that the generator can either create ellipses that overlap ($\varepsilon < 0$) or are separated ($\varepsilon > 0$), with $\varepsilon = 0$ being the case of perfect overlap. We will not consider the Intersection Method in this study as the intersection set is empty when the two ellipses do not overlap.

Effect of the aspect ratio: In this test, the size ab of ellipse $\mathcal{E}_i$ and the overlap ε are uniformly distributed in the ranges $[1, 100]$ and $[-0.2, 0.2]$, respectively. However, we control the aspect ratio a/b of the two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ following a uniform distribution in the specific ranges listed in Table 6.13. We report in the table the total computational time in seconds to compute the contact point for the 1,000 pairs of ellipses for each interval of the aspect ratio. On the one hand, we observe that the computational time in the case of the S-GPA and P-GPA is independent of the aspect ratio. On the other hand, we see some significant variations in the computational times for L-GPA, M-GPA, and C-GPA. More specifically, by increasing the aspect ratio of the ellipses, the efficiency of L-GPA and M-GPA tends to improve while it deteriorates for C-GPA. The trend is less clear in the case of CCA but the algorithm is consistently slower than all the others.

Table 6.13 Computational time (in seconds) for 1,000 pairs of ellipses with respect to their aspect ratio.

Aspect ratio	L-GPA	P-GPA	M-GPA	C-GPA	CCA	S-GPA
1–5	1.16	0.93	1.66	1.79	2.21	0.31
5–10	0.91	0.95	1.53	1.95	2.19	0.32
10–20	0.76	0.95	1.45	1.99	2.23	0.31
20–40	0.65	0.93	1.32	2.05	2.16	0.32
40–80	0.56	0.93	1.29	2.08	2.13	0.31
80–160	0.42	0.92	1.22	2.13	2.10	0.32

Effect of the relative size: We consider here an experiment similar to the one previously described in order to study the effect of the relative size of the ellipses on the computational time. In this test, the aspect ratio a/b of the ellipses and the overlap ε are uniformly distributed in the ranges $[1, 10]$ and $[-0.2, 0.2]$, respectively. However, the size of $\mathcal{E}_j$ is kept equal to one while the size of $\mathcal{E}_i$ is uniformly drawn within the intervals given in Table 6.14. It is clear from the results that the relative size of the ellipses has little effect or even no effect on the performance of the algorithms.

Table 6.14 Computational time (in seconds) for 1,000 pairs of ellipses with respect to their relative size.

Relative size	L-GPA	P-GPA	M-GPA	C-GPA	CCA	S-GPA
1–5	0.93	0.97	1.63	2.30	2.42	0.33
5–10	0.89	0.99	1.70	2.29	2.39	0.34
10–20	0.91	0.96	1.67	2.27	2.33	0.33
20–40	0.89	0.99	1.69	2.24	2.35	0.33
40–60	0.85	0.99	1.68	2.30	2.43	0.34
60–180	0.87	0.98	1.67	2.29	2.36	0.34

Effect of the overlap size: In this example, we only consider the geometric potential methods, the reason being that the generator of Section C.1.1 provides pairs of ellipses for which we know the exact solution $\boldsymbol{x}_i$ to Problem (3.8). In that case, we can compute the relative error in the approximations of $\boldsymbol{x}_i$ by the various algorithms for each pair of ellipses and use as a metric of accuracy the maximum relative error over the 1,000 pairs of ellipses. The objective here is to study the influence of the overlap size ε on the computational time while controlling the overall accuracy of the algorithms. We emphasize that the GPAs involve iterative methods with distinct criteria for convergence. In order to provide meaningful comparisons between the algorithms, we therefore aim at calculating solutions with a similar accuracy. The sample sets consist here of 1,000 pairs of ellipses whose aspect ratio varies in the range $[1, 20]$ and relative size varies in the range $[1, 10]$. Each sample set involves pairs of ellipses whose overlap size ε is either 10^{-1}, 10^{-5}, 10^{-10}, or 10^{-15}. The convergence criterion of each algorithm is tuned so that the maximum relative error lies between 10^{-9} and 10^{-10}, as shown in Table 6.15. The computational times are reported in Table 6.16. We observe that the overlap size has in fact limited effect on the performance of the algorithms.

The main conclusions from this series of experiments are as follows: 1) The performance of the S-GPA is insensitive to variations in the aspect ratio, the relative size, or the overlap size

Table 6.15 Logarithmic maximum error for 1,000 pairs of ellipses with respect to the overlap size.

Overlap size	L-GPA	P-GPA	M-GPA	C-GPA	S-GPA
10^{-1}	-9.34	-9.34	-9.38	-9.38	-9.45
10^{-5}	-9.17	-9.23	-9.55	-9.48	-9.45
10^{-10}	-9.17	-9.31	-9.56	-9.48	-9.45
10^{-15}	-9.17	-9.31	-9.56	-9.48	-9.45

Table 6.16 Computational time (in seconds) for 1,000 pairs of ellipses with respect to the overlap size.

Overlap size	L-GPA	P-GPA	M-GPA	C-GPA	S-GPA
10^{-1}	0.83	0.99	1.61	2.22	0.33
10^{-5}	0.85	1.01	1.61	2.05	0.33
10^{-10}	0.87	1.00	1.58	1.99	0.33
10^{-15}	0.83	1.01	1.58	1.84	0.33

of the pair of ellipses; 2) Only the aspect ratio of the ellipses may have a significant effect on the performance of the other algorithms; 3) All Geometric Potential Algorithms (GPAs) outperform the Closest Normal Algorithm (CCA); 4) L-GPA seems to be the most efficient algorithm among the existing GPAs; 5) The S-GPA is consistently more efficient than the other algorithms by a factor between two and seven when considering ellipses.

6.7 Performance of L-GPA and S-GPA for Ellipses and Ellipsoids

In this section, we refine our analysis of the performance of the S-GPA, by limiting the comparison to only the L-GPA, which we demonstrated was consistently the second fastest. As before, we generate sample sets of 1,000 pairs of ellipses and ellipsoids using the algorithms of Sections C.1.1 and C.1.2. The parameters for the generation of the ellipses are $\gamma_{\max} = 20$, $\omega_{\max} = 50$, $\varepsilon_{\max} = 0.2$, $N_{\min} = 0$, $N_{\max} = 15$, $r_{\max} = 10$. The parameters for the generation of the ellipsoids: are $\gamma_{1,\max} = 20$, $\gamma_{2,\max} = 20$, $\omega_{\max} = 50$, $\varepsilon_{\max} = 0.2$, $N_{\min} = 0$, $N_{\max} = 15$, and $r_{\max} = 10$. In other words, the ellipsoids have aspect ratios a/b and b/c in the range $[1, 60]$ for $\mathcal{E}_i$ and in the range $[1, 10]$ for $\mathcal{E}_j$. The size abc of $\mathcal{E}_i$ lies in the range $[1, 20]$ and that of $\mathcal{E}_j$ is 1.

We report in Table 6.17 the computational time, the number of iterations in the root finding algorithms, and some statistics for the relative error. We first observe that the two algorithms

provide similar accuracy whether in 2D or in 3D. We note however that while the S-GPA is three times faster than L-GPA for ellipses, the speed-up reaches a factor of five for ellipsoids. In the case of the S-GPA, more than half of the time is spent on the mapping and the calculation of the guess point for the Newton's method. On the other hand, the iterative solver requires only a few iterations (the median number is 4 iterations in 2D and 7 iterations in 3D) to converge to the solution of the problem. By contrast, the median number of iterations in L-GPA almost triples in 3D when compared to 2D. Moreover, each iteration is more costly as it requires finding the six roots of a polynomial of degree six in 3D compared to the four roots of a polynomial of degree four in 2D. We also remark that if these algorithms were used in DEM, where initial estimates of the contact point could be obtained by using the estimates at the previous time step, then the new iterative algorithm would be vastly superior to the L-GPA.

Finally, it is worth mentioning that, for the S-GPA, the standard deviation in the number of iterations remains close to unity in 2D and 3D and that the maximum number of iterations never exceed 14 for these two sets of ellipses and ellipsoids. Moreover, the Newton's method accounts for a fraction of the total computational cost of the algorithm. These features thus makes it a suitable candidate for the development of a parallel version as the loads would reasonably be well distributed among processors.

Table 6.17 Computational time, number of iterations, and maximum relative error for L-GPA and the S-GPA for 1,000 pairs of ellipses and ellipsoids.

	2D		3D	
	L-GPA	S-GPA	L-GPA	S-GPA
Computational time (in seconds)				
total	1.05	0.34	2.57	0.47
guess point	–	0.05	–	0.06
mapping	–	0.19	–	0.21
root finding	0.89	0.02	2.34	0.16
Number of iterations				
maximum	37	7	71	14
minimum	11	2	28	3
median	14	4	40	7
standard deviation	4.14	0.82	6.26	1.34
Log of relative error				
maximum	−7.80	−7.81	−7.88	−7.94
minimum	−16.63	−17.00	−16.31	−17.00
median	−12.88	−13.01	−12.21	−12.24
standard deviation	1.71	1.77	1.77	1.79

CHAPTER 7 CONCLUSION AND RECOMMENDATIONS

7.1 Summary of Works

This research has been concerned with the development of contact detection methods for ellipses and ellipsoids in near perfect contact. We hope that we were able to convey the idea that the problem may seem deceptively simple. The main reasons are essentially twofold: 1) the definition of contact point and normal is not unique for ellipsoidal particles; 2) there is no analytical solution that allows one to solve the problem in a finite number of operations and one has then to resort to numerical methods to identify such points and normals. The major challenge is therefore to develop very efficient algorithms in order to be able to consider very large systems of ellipsoidal particles in Discrete Element Simulations. An ideal solution method should be fast, robust, and precise to handle all possible configuration pairs and particle sizes.

A major objective was to define a rigorous mathematical framework to study the problem of contact detection for elliptical and ellipsoidal particles and formalize key concepts in contact detection. Starting with disjoint particles, we have identified two minimization problems to compute their separation distance. The solutions of these problems are given in terms of two points, one on each ellipse, that we have referred here to as the Minimum Distance Pair (MDP) and the Minimum Potential Pair (MPP). The notion of the MDP and MPP happens to coincide in the case of two ellipses or ellipsoids in perfect contact. In that particular case, one can also find the contact point by looking at the Intersection Set (IS), which obviously consists here of a single point. Our contribution was then to extend those three notions to the case of overlapping particles and show that the corresponding minimization problems were still well-posed. We have in particular conceptualized the notion of near-perfect contact and proposed a formal definition of what is meant by small overlap between particles. The latter relies on the introduction of the so-called co-gradient locus associated with a pair of ellipses, which was previously identified as the line of common slope. We have shown that the co-gradient locus is in fact a hyperbola in 2D. Using these concepts and definitions, we were able to extend the minimization problems to the case of overlapping ellipses and ellipsoids and show existence and uniqueness of their solutions, which has been rarely, if ever, addressed in the literature. We have also highlighted the role played by non-binding constraints involving the normals in the solution of the minimization problems.

The mathematical analysis of contact detection for ellipses and ellipsoids has led us to conclude that there are in fact only three broad classes of methods, each associated with a

specific definition of the contact point: the intersection methods, which search for the intersection set between particles, 2) the geometric potential methods, which provide the MPP, and 3) the common normal methods, which look for the MDP. Geometric potential methods can also be distinguished by the fact that the minimization problem imposes, or not, the non-binding constraints that the solution be on the co-gradient locus or that the normals at each point of the MPP be opposite. We have also identified variants within the common normal methods. It is then interesting to realize that the known algorithms all falls within one of these three classes of methods, sometimes unbeknown to the authors. We note that the contact points obtained by these methods get closer to each other as the overlap between particles become smaller and eventually coincide when the particles are in perfect contact. However, the difference can be noticeable if the pair of ellipses does not satisfy our definition of small overlap. The intersection methods are conceptually simple and straightforward to implement in two dimensions, but lacks stability and accuracy when the two ellipses approaches the configuration of perfect contact. Moreover, the method does not provide any information about the separation distance when the particles are fully disjoint. The common normal methods assume that one simultaneously find the contact pair of points by solving one minimization problem, which results in a higher computational cost when compared to the other methods. Moreover, we have shown that for some configurations of ellipses and ellipsoids, the MPP may lie outside of the overlap region, which may be counterintuitive. In our opinion, the geometric potential methods seem to provide the method of choice in terms of computational cost and physical interpretation. We note that several algorithms within the geometric potential methods have been developed to date. The algorithms vary in essence from each other depending on the approach chosen to solve the constrained minimization problems, on the use of normalization mappings of the particle pair or not, on the parametrization of one of the ellipses or ellipsoids or not, on the choice of the root finding or optimization approach and the choice of the initial guess point, and whether the non-binding constraint on the gradients or normals is used or not. Finally, we emphasize that the solution process for several of the geometric algorithms leads to finding the roots of polynomials of degree four for ellipses and of degree six for ellipsoids. In this case, one usually computes all roots and select among those the one that provides the global minimizer of the minimization functional. It is in our opinion counterproductive and one should design an algorithm that produces the global minimizer only without the need to evaluate the other minimizers, or maximizers, associated with the objective function.

The mathematical analysis of the contact detection problem for ellipsoidal particles in near perfect contact has thus led us to develop a novel algorithm that falls within the geometric potential methods. The proposed algorithm make use of the normalization transformation that

maps one of the two ellipses (ellipsoids) into an ellipse (ellipsoid) centered at the origin and aligned with the coordinate axes and the other into a unit circle (sphere). Each minimization problem to determine the contact pair is then recast into one of finding the roots of a simple trigonometric scalar-valued function and vector-valued function after parametrization of the ellipse and the ellipsoid, respectively. A byproduct in using such a transformation is the possibility to efficiently construct an accurate initial guess point for the root finding problem. The problem is solved using a few iterations of the Newton's method and convergence to the desired root is guaranteed by enforcing an additional constraint that only the solution to the minimization problem satisfies. The combination of all these ingredients has allowed us to design an algorithm that is fast, robust, and suitable for any pair of ellipses and ellipsoids in near perfect contact. The performance of the algorithm was assessed on large sample sets of pairs of particles for which we have shown that it was several times faster than existing algorithms for similar accuracy. Moreover, we have run experiments to demonstrate that the computational time of the novel algorithm was independent of the aspect ratio, relative size, and overlap size of the ellipses and ellipsoids and that its computational cost did not significantly increase when passing from the 2D case to the 3D case, in contrast with the existing algorithms. For verification purposes, we have actually developed a general algorithm to generate random pairs of ellipses and ellipsoids in near perfect contact. The originality of these algorithms lies in the fact that one constructs pairs of ellipses or ellipsoids for which the solution to one of the minimization problems in the geometric potential method is exactly known, which allows one to precisely verify the accuracy of the contact detection algorithm.

7.2 Future Research

The mathematical analysis presented in this thesis has brought to light several new concepts and notions for a better understanding of the contact detection problem for ellipses and ellipsoids. We nevertheless recognize that this original work, while bringing some mathematical rigor to the problem, also opens up new opportunities for future research. We list some examples below:

1. The co-gradient locus has been fully characterized in the 2D case as it was rigorously proved to be a hyperbola. However, the co-gradient locus in 3D still needs to be properly characterized apart from the fact it is the intersection of two surfaces, each defined by a quadratic equation. In the same spirit, one should also extend the definition of small overlap to overlapping ellipsoids and complete the proof of some theorems for the 3D case.

2. We have proposed a novel algorithm based on our mathematical analysis of the contact detection problem and showed that it is more efficient than existing algorithms. However, we do not rule out the possibility that one could design an even faster algorithm. In order to do so, one could explore different combinations of transformations, parameterizations, root finding algorithms, methods to find initial guesses, constraints, etc. It is possible that one could find an optimal combination that lower the number of operations need to identify the contact point and contact normal for ellipses and ellipsoids. Moreover, the algorithm should eventually be implemented on a parallel machine in order to study very large assemblies of particles by the Discrete Element Method and verify that it properly scales up for high-performance computing.

3. There is a growing interest in the use of super-quadrics to represent particles as they cover a wide variety of shape geometries that resemble cubes, octahedra, cylinders, lozenges, or spindles, with rounded or sharp corners. It would then be interesting to extend the current mathematical framework to the contact detection problems involving convex super-quadrics. For example, an intriguing problem would be to characterize the co-gradient locus for a pair of two super-quadratics.

4. In DEM simulations, it is customary to decompose the contact detection problem into a narrow phase search and a broad phase search. The latter is concerned with identifying the list of candidate neighboring particles associated with each particle of an assembly that should be considered in the former. Only in the narrow phase search does one actually employ the accurate contact detection algorithm to determine the contact point and contact normal. It is clear that the fewer neighboring particles that one marks during the broad phase, the more efficient the overall algorithm for contact detection would be. We believe that one could improve on the broad phase algorithms by using tools from machine learning. One could for example design a deep neural network to estimate the distance, with some level of confidence, between pairs of ellipsoids. Training and validation of the model could then be performed using both the algorithm to generate random pairs of ellipsoids and the contact detection algorithm to determine the separation distance that were developed in this work.

REFERENCES

[1] T. Sitharam and S. Dinesh, "Numerical simulation of liquefaction behaviour of granular materials using discrete element method," *Journal of Earth System Science*, vol. 112, no. 3, p. 479, 2003.

[2] A. Manne and N. Satyam, "A review on the discrete element modeling of dynamic laboratory tests for liquefaction assessment," *Electronic Journal of Geotechnical Engineering*, vol. 20, no. 1, pp. 21–46, 2015.

[3] B. J. Alder and T. E. Wainwright, "Studies in molecular dynamics. i. general method," *The Journal of Chemical Physics*, vol. 31, no. 2, pp. 459–466, 1959.

[4] P. Cundall, "A computer model for simulating systems," in *Proc. Symp. on Rock Fracture (ISRM)*, vol. 29, 1971.

[5] P. A. Cundall and O. D. Strack, "A discrete numerical model for granular assemblies," *Geotechnique*, vol. 29, no. 1, pp. 47–65, 1979.

[6] X. Zhuang, Q. Wang, and H. Zhu, "Effective properties of composites with periodic random packing of ellipsoids," vol. 10, no. 2, p. 112. [Online]. Available: http://www.mdpi.com/1996-1944/10/2/112

[7] P. Stroeven and H. He, "Packing of non-spherical aggregate particles by DEM," 2013.

[8] J. Zhou, Y. Zhang, and J. Chen, "Numerical simulation of random packing of spherical particles for powder-based additive manufacturing," *Journal of manufacturing science and engineering*, vol. 131, no. 3, p. 031004, 2009.

[9] R. P. Jensen *et al.*, "Effect of particle shape on interface behavior of dem-simulated granular materials," *International Journal of Geomechanics*, vol. 1, pp. 1–19, 01 2001.

[10] J. Wang *et al.*, "Particle shape effects in discrete element modelling of cohesive angular particles," *Granular Matter*, vol. 13, no. 1, pp. 1–12, 2011.

[11] J. M. Ting *et al.*, "An ellipse-based discrete element model for granular materials," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 17, no. 9, pp. 603–623, 1993.

[12] B. Yan, R. A. Regueiro, and S. Sture, "Three-dimensional ellipsoidal discrete element modeling of granular materials and its coupling with finite element facets," *Engineering Computations*, vol. 27, no. 4, pp. 519–550, 2010.

[13] T.-T. Ng and W. Zhou, "DEM simulations of bi-disperse ellipsoids of different particle sizes," *Comptes Rendus Mécanique*, vol. 342, no. 3, pp. 141–150, 2014.

[14] J. Gan, A. Yu, and Z. Zhou, "DEM simulation on the packing of fine ellipsoids," *Chemical Engineering Science*, vol. 156, pp. 64–76, 2016.

[15] J. R. Williams and A. P. Pentland, "Superquadrics and modal dynamics for discrete elements in interactive design," *Engineering Computations*, vol. 9, no. 2, pp. 115–127, 1992.

[16] A. Podlozhnyuk, S. Pirker, and C. Kloss, "Efficient implementation of superquadric particles in discrete element method within an open-source framework," *Computational Particle Mechanics*, vol. 4, no. 1, pp. 101–118, 2017.

[17] D. Zhao *et al.*, "Three-dimensional discrete element simulation for granular materials," *Engineering Computations*, vol. 23, no. 7, pp. 749–770, 2006.

[18] B. Smeets *et al.*, "Polygon-based contact description for modeling arbitrary polyhedra in the discrete element method," *Computer Methods in Applied Mechanics and Engineering*, vol. 290, pp. 277–289, 2015.

[19] X. Garcia *et al.*, "A clustered overlapping sphere algorithm to represent real particles in discrete element modelling," *Geotechnique*, vol. 59, no. 9, pp. 779–784, 2009.

[20] Q. Zhang *et al.*, "A novel non-overlapping approach to accurately represent 2D arbitrary particles for DEM modelling," *Journal of Central South University*, vol. 24, no. 1, pp. 190–202, 2017.

[21] D. Ilin and M. Bernacki, "A new algorithm for dense ellipse packing and polygonal structures generation in context of FEM or DEM," *MATEC Web of Conferences*, vol. 80, p. 02004, 01 2016.

[22] B. Yan and R. A. Regueiro, "A comprehensive study of MPI parallelism in three-dimensional discrete element method (DEM) simulation of complex-shaped granular particles," *Computational Particle Mechanics*, vol. 5, no. 4, pp. 553–577, 2018.

[23] Z. Zhou *et al.*, "Discrete particle simulation of gas fluidization of ellipsoidal particles," *Chemical Engineering Science*, vol. 66, no. 23, pp. 6128–6145, 2011.

[24] A. Džiugys and B. Peters, "A new approach to detect the contact of two-dimensional elliptical particles," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 25, no. 15, pp. 1487–1500, 2001.

[25] J. W. Perram and M. Wertheim, "Statistical mechanics of hard ellipsoids. I. overlap algorithm and the contact function," *Journal of Computational Physics*, vol. 58, no. 3, pp. 409–416, 1985.

[26] W. Wang, J. Wang, and M.-S. Kim, "An algebraic condition for the separation of two ellipsoids," *Computer Aided Geometric Design*, vol. 18, no. 6, pp. 531–539, 2001.

[27] X. Jia *et al.*, "An algebraic approach to continuous collision detection for ellipsoids," *Computer Aided Geometric Design*, vol. 28, no. 3, pp. 164–176, 2011.

[28] G. Delaney *et al.*, "Random packing of elliptical disks," *Philosophical Magazine Letters*, vol. 85, no. 2, pp. 89–96, 2005.

[29] A. Donev *et al.*, "Improving the density of jammed disordered packings using ellipsoids," *Science*, vol. 303, no. 5660, pp. 990–993, 2004.

[30] A. Pankratov, T. Romanova, and I. Subota, "An efficient algorithm for the ellipse packing problem," *Electronics and Informatics*, no. 1 (60), 2013.

[31] L. Wang, A. D. Ames, and M. Egerstedt, "Multi-objective Compositions for Collision-Free Connectivity Maintenance in Teams of Mobile Robots," *2016 Decisions and Control Conference*, pp. 2659–2664, 2016.

[32] E. Rimon and S. P. Boyd, "Obstacle collision detection using best ellipsoid fit," *Journal of Intelligent and Robotic Systems*, vol. 18, no. 2, pp. 105–126, 1997.

[33] M.-Y. Ju *et al.*, "A novel collision detection method based on enclosed ellipsoid," in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*, vol. 3. IEEE, 2001, pp. 2897–2902.

[34] G. Nolan and P. Kavanagh, "Random packing of nonspherical particles," *Powder technology*, vol. 84, no. 3, pp. 199–205, 1995.

[35] C.-Y. Wang, C.-F. Wang, and J. Sheng, "A packing generation scheme for the granular assemblies with 3D ellipsoidal particles," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 23, no. 8, pp. 815–828, 1999.

[36] S. Johnson, J. R. Williams, and B. Cook, "Contact resolution algorithm for an ellipsoid approximation for discrete element modeling," *Engineering Computations*, vol. 21, no. 2/3/4, pp. 215–234, 2004.

[37] C. Hogue, "Shape representation and contact detection for discrete element simulations of arbitrary geometries," *Engineering Computations*, vol. 15, no. 3, pp. 374–390, 1998.

[38] E. G. Gilbert, D. W. Johnson, and S. S. Keerthi, "A fast procedure for computing the distance between complex objects in three-dimensional space," *IEEE Journal on Robotics and Automation*, vol. 4, no. 2, pp. 193–203, April 1988.

[39] K.-W. Lim, K. Krabbenhoft, and J. E. Andrade, "On the contact treatment of non-convex particles in the granular element method," *Computational Particle Mechanics*, vol. 1, no. 3, pp. 257–275, 2014.

[40] I. Vlahinić *et al.*, "From computed tomography to mechanics of granular materials via level set bridge," *Acta Geotechnica*, vol. 12, pp. 85–95, 2017.

[41] L. Rothenburg and R. J. Bathurst, "Numerical simulation of idealized granular assemblies with plane elliptical particles," *Computers and Geotechnics*, vol. 11, no. 4, pp. 315–329, 1991.

[42] H. Ouadfel and L. Rothenburg, "An algorithm for detecting inter-ellipsoid contacts," *Computers and Geotechnics*, vol. 24, no. 4, pp. 245–263, 1999.

[43] X. Lin and T.-T. Ng, "Contact detection algorithms for three-dimensional ellipsoids in discrete element modelling," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 19, no. 9, pp. 653–659, 1995.

[44] C. Wellmann, C. Lillie, and P. Wriggers, "A contact detection algorithm for superellipsoids based on the common-normal concept," *Engineering Computations*, vol. 25, no. 5, pp. 432–442, 2008.

[45] G. Lu, J. Third, and C. Müller, "Discrete element models for non-spherical particle systems: From theoretical developments to applications," *Chemical Engineering Science*, vol. 127, pp. 425–465, 2015.

[46] J. M. Ting *et al.*, "An ellipse-based discrete element model for granular materials," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 17, no. 9, pp. 603–623, 1993.

[47] G. Mustoe and M. Miyata, "Material flow analyses of noncircular-shaped granular media using discrete element methods," *Journal of Engineering Mechanics*, vol. 127, no. 10, pp. 1017–1026, 2001.

[48] Cramer and Gabriel, *Introduction a l'analyse des lignes courbes algebriques.* chez les freres Cramer & Cl. Philibert, 1750.

[49] A. A. Kosinski, "Cramer's rule is due to cramer," *Mathematics Magazine*, vol. 74, no. 4, pp. 310–312, 2001.

[50] J. S. Marshall and S. Li, *Adhesive particle flow.* Cambridge University Press, 2014.

[51] W. Xu, H. Chen, and Z. Lv, "An overlapping detection algorithm for random sequential packing of elliptical particles," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 13, pp. 2452–2467, 2011.

[52] M. D. Carmo, *Differential Geometry of Curves and Surface.* Prentice Hall, 1976.

[53] S. Bektas, "Curvature of the Ellipsoid with Cartesian Coordinates," *Landscape Architecture and Regional Planning*, vol. 2, no. 2, pp. 61–66, 2017.

[54] G. Glaeser, H. Stachel, and B. Odehnal, *The Universe of Conics : From the ancient Greeks to the 21st century developments.* Springer Spektrum, 2016.

[55] J. W. Downs, *Practical Conic Sections : the geometric properties of ellipses, parabolas and hyperbolas.* Dover, 2003.

[56] I. Artobolevskii, *Mechanisms for the Generation of Plane Curves.* Pergamon Press, 1964.

[57] H. Goldstein, *Classical Mechanics.* Addison-Wesley, 1980.

[58] W. R. Hamilton, "Xi. on quaternions; or on a new system of imaginaries in algebra," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 33, no. 219, pp. 58–60, 1848.

[59] J. Perram *et al.*, "Ellipsoids contact potential: theory and relation to overlap potentials," *Physics Review E*, vol. 54, no. 6, pp. 6565–6572, 1996.

[60] J. Vieillard-Baron, "Phase Transitions of the Classical Hard-Ellipse System," *The Journal of Chemical Physics*, vol. 56, no. 10, pp. 4729–4744, 1972.

[61] A. Donev, "Jammed Packings of Hard Particles," Ph.D. dissertation, Princeton University, 2006.

[62] W. Wang *et al.*, "Efficient collision detection for moving ellipsoids using separating planes," *Computing*, vol. 72, no. 1-2, pp. 235–246, 2004.

[63] Y.-K. Choi *et al.*, "Continuous Collision Detection for Ellipsoids," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 2, pp. 311–325, 2008.

[64] H. Hilton, *Plane Algebraic Curves*. Clarendon Press, 1920.

[65] W. Fulton, *Algebraic Curves*. Addison-Wesley, 1989.

[66] W. Rudin, *Functional Analysis*. McGraw-Hill, 1991.

[67] J. M. Ting, "A robust algorithm for ellipse-based discrete element modelling of granular materials," *Computers and Geotechnics*, vol. 13, pp. 175–186, 1992.

[68] A. Sard, "The measure of the critical values of differentiable maps," *Bulletin of the American Mathematical Society*, vol. 48, no. 12, pp. 883–890, 1942.

[69] J. Milnor and D. W. Weaver, *Topology from the differentiable viewpoint*. Princeton university press, 1997.

[70] P. Griffiths and M. Harris, *Principles of Algebraic Geometry*, ser. Pure and Applied Mathematics : A Wiley-Interscience Series. New York: John Wiley & Sons, 1994.

[71] T.-T. Ng, "Numerical simulations of granular soil using elliptical particles," *Computers and Geotechnics*, vol. 16, no. 2, pp. 153–169, 1994.

[72] A. Džiugys and B. Peters, "An approach to simulate the motion of spherical and non-spherical fuel particles in combustion chambers," *Granular matter*, vol. 3, no. 4, pp. 231–266, 2001.

[73] J. M. Ting, "An ellipse-based micromechanical model for angular granular materials," in *Proceedings of the ASCE Engineering Mechanics Conference on Mechanics Computing in 1990+ and Beyond*, 1991.

[74] D. S. Watkins, "Francis's Algorithm," *The American Mathematical Monthly*, vol. 118, no. 5, pp. 387–403, 2011.

APPENDIX A INTERSECTION ALGORITHM

A.1 Intersection Algorithm

The coefficients of the Polynomial 4.5 corresponding to Intersection Algorithm from Section 4.1 are defined as following:

$$
\begin{aligned}
a_4 &= A_i P^2 + B_i Q^2 + 2C_i PQ, \\
a_3 &= 2A_i SP + 2B_i QU + 2C_i(PU + SQ) + 2D_i P^2 + 2E_i PQ, \\
a_2 &= A_i S^2 + B_i U^2 + 2B_i QV + 2C_i(PV + SU) + 4D_i SP + 2E_i(PU + SQ) + F_i P^2, \\
a_1 &= 2B_i UV + 2C_i SV + 2D_i S^2 + 2E_i(PV + SU) + 2F_i SP, \\
a_0 &= B_i V^2 + 2E_i SV + F_i S^2,
\end{aligned}
$$

where we have introduced the following parameters

$$
\begin{aligned}
P &= 2B_i C_j - B_j C_i \\
Q &= A_i B_j - A_j B_i \\
S &= 2B_i E_j - 2B_j E_i \\
U &= 2D_i B_j - 2D_j B_i \\
V &= F_i B_j - F_j B_i.
\end{aligned}
$$

APPENDIX B LAGRANGIAN GPA

B.1 Lagrangian GPA

B.1.1 Algorithm in 2-D

The two Equations (4.8) and (4.9) can be rewritten in the matrix form as

$$\begin{bmatrix} A_i + \lambda A_j & C_i + \lambda C_j \\ C_i + \lambda C_j & B_i + \lambda B_j \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -D_i - \lambda D_j \\ -E_i - \lambda E_j \end{bmatrix}. \tag{B.1}$$

with constraint:

$$A_j x^2 + B_j y^2 + 2C_j xy + 2D_j x + 2E_j y + F_j = 0.$$

By using Cramer's rule, we can write (x, y) in terms of λ

$$\begin{aligned} x &= \frac{D_x}{D^*}, \\ y &= \frac{D_y}{D^*}, \end{aligned} \tag{B.2}$$

where D_x, D_y, and D^* denote the following determinants

$$D_x = \begin{vmatrix} -D_i - \lambda D_j & C_i + \lambda C_j \\ -E_i - \lambda E_j & B_i + \lambda B_j \end{vmatrix}, \qquad D_y = \begin{vmatrix} A_i + \lambda A_j & -D_i - \lambda D_j \\ C_i + \lambda C_j & -E_i - \lambda E_j \end{vmatrix},$$

$$D^* = \begin{vmatrix} A_i + \lambda A_j & C_i + \lambda C_j \\ C_i + \lambda C_j & B_i + \lambda B_j \end{vmatrix}.$$

Developing above equations yields

$$\begin{aligned} x &= \frac{a_x \lambda^2 + b_x \lambda + c_x}{a^* \lambda^2 + b^* \lambda + c^*}, \\ y &= \frac{a_y \lambda^2 + b_y \lambda + c_y}{a^* \lambda^2 + b^* \lambda + c^*}, \end{aligned}$$

with

$$\begin{aligned} a_x &= -B_j D_j + C_j E_j, \\ b_x &= -D_j B_i - D_i B_j + E_i C_j + E_j C_i, \\ c_x &= -D_i B_i + C_i E_i, \end{aligned}$$

$$
\begin{aligned}
a_y &= -A_j E_j + C_j D_j, \\
b_y &= -A_j E_i - A_i E_j + D_i C_j + D_j C_i, \\
c_y &= -A_i E_i + C_i D_i, \\
a^* &= A_j B_j - C_j^2, \\
b^* &= A_j B_i + A_i B_j - 2C_i C_j, \\
c^* &= A_i B_i - C_i^2.
\end{aligned}
$$

By substituting the new expressions for x and y in $f_j(x, y) = 0$, we obtain Polynomial (4.11) respect to λ where its coefficients are defined as

$$
\begin{aligned}
a_4 &= A_j a_x^2 + B_j a_y^2 + 2C_j a_x a_y + 2D_j a_x a^* + 2E_j a_y a^* + F_j a^{*2}, \\
a_3 &= 2A_j a_x b_x + 2B_j a_y b_y + 2C_j (a_x b_y + b_x a_y) + 2D_j (a_x b^* + b_x a^*), \\
&\quad + 2E_j (a_y b^* + b_y a^*) + 2F_j a^* b^*, \\
a_2 &= A_j (b_x^2 + 2a_x c_x) + B_j (b_y^2 + 2a_y c_y) + 2C_j (b_x b_y + c_x a_y + c_y a_x), \\
&\quad + 2D_j (b_x b^* + c_x a^* + a_x c^*) + 2E_j (b_y b^* + c_y a^* + a_y c^*) + F_j (b^{*2} + 2a^* c^*), \\
a_1 &= 2A_j b_x c_x + 2B_j b_y c_y + 2C_j (b_x c_y + c_x b_y) + 2D_j (b_x c^* + c_x b^*), \\
&\quad + 2E_j (b_y c^* + c_y b^*) + 2F_j b^* c^*, \\
a_0 &= A_i c_x^2 + B_j c_y^2 + 2C_j c_x c_y + 2D_j c_x c^* + 2E_j c^* c_y + F_j c^{*2}.
\end{aligned}
$$

B.1.2 Algorithm in 3-D

Equation (B.1) in 3-D can be recast as

$$
\begin{bmatrix}
A_i + \lambda A_j & F_i + \lambda F_j & E_i + \lambda E_j \\
F_i + \lambda F_j & B_i + \lambda B_j & D_i + \lambda D_j \\
E_i + \lambda E_j & D_i + \lambda D_j & C_i + \lambda C_j
\end{bmatrix}
\begin{bmatrix} x \\ y \\ z \end{bmatrix}
=
\begin{bmatrix}
-G_i - \lambda G_j \\
-H_i - \lambda H_j \\
-I_i - \lambda I_j
\end{bmatrix}
$$

with constraint

$$
A_j x^2 + B_j y^2 + C_j z^2 + 2D_j yz + 2E_j zx + 2F_j xy + 2G_j x + 2H_j y + 2I_j z + J_j = 0
$$

By using Cramer's rule, we can solve for (x, y, z) in terms of λ such that

$$\begin{aligned} x &= \frac{D_x}{D^*} \\ y &= \frac{D_y}{D^*} \\ z &= \frac{D_z}{D^*} \end{aligned} \tag{B.3}$$

where

$$D_x = \begin{vmatrix} -G_i - \lambda G_j & F_i + \lambda F_j & E_i + \lambda E_j \\ -H_i - \lambda H_j & B_i + \lambda B_j & D_i + \lambda D_j \\ -I_i - \lambda I_j & D_i + \lambda D_j & C_i + \lambda C_j \end{vmatrix}, \qquad D_y = \begin{vmatrix} A_i + \lambda A_j & -G_i - \lambda G_j & E_i + \lambda E_j \\ F_i + \lambda F_j & -H_i - \lambda H_j & D_i + \lambda D_j \\ E_i + \lambda E_j & -I_i - \lambda I_j & C_i + \lambda C_j \end{vmatrix}$$

$$D_z = \begin{vmatrix} A_i + \lambda A_j & F_i + \lambda F_j & -G_i - \lambda G_j \\ F_i + \lambda F_j & B_i + \lambda B_j & -H_i - \lambda H_j \\ E_i + \lambda E_j & D_i + \lambda D_j & -I_i - \lambda I_j \end{vmatrix}, \qquad D^* = \begin{vmatrix} A_i + \lambda A_j & F_i + \lambda F_j & E_i + \lambda E_j \\ F_i + \lambda F_j & B_i + \lambda B_j & D_i + \lambda D_j \\ E_i + \lambda E_j & D_i + \lambda D_j & C_i + \lambda C_j \end{vmatrix}$$

Developing above equations yields

$$\begin{aligned} x &= \frac{a_x\lambda^3 + b_x\lambda^2 + c_x\lambda + d_x}{a^*\lambda^3 + b^*\lambda^2 + c^*\lambda + d^*} \\ y &= \frac{a_y\lambda^3 + b_y\lambda^2 + c_y\lambda + d_y}{a^*\lambda^3 + b^*\lambda^2 + c^*\lambda + d^*} \\ z &= \frac{a_z\lambda^3 + b_z\lambda^2 + c_z\lambda + d_z}{a^*\lambda^3 + b^*\lambda^2 + c^*\lambda + d^*} \end{aligned} \tag{B.4}$$

where

$$\begin{aligned} a_x &= -G_j(C_jB_j - D_j^2) - F_j(D_jI_j - H_jC_j) + E_j(I_jB_j - H_jD_j) \\ b_x &= -G_j(C_iB_j + B_iC_j - 2D_iD_j) - G_i(C_jB_j - D_j^2) \\ &\quad - F_j(-C_iH_j - C_jH_i + D_iI_j + D_jI_i) - F_i(D_jI_j - H_jC_j) \\ &\quad + E_j(-H_jD_i - H_iD_j + B_jI_i + B_iI_j) + E_i(I_jB_j - H_jD_j) \\ c_x &= -G_i(C_iB_j + B_iC_j - 2D_iD_j) - G_j(B_iC_i - D_i^2) \\ &\quad - F_i(-C_iH_j - C_jH_i + D_iI_j + D_jI_i) - F_j(D_iI_i - H_iC_i) \\ &\quad + E_i(-H_jD_i - H_iD_j + B_jI_i + B_iI_j) + E_j(I_iB_i - H_iD_i) \\ d_x &= -G_i(B_iC_i - D_i^2) - F_i(D_iI_i - H_iC_i) + E_i(I_iB_i - H_iD_i) \end{aligned}$$

$$
\begin{aligned}
a_y &= A_j(-C_jH_j + D_jI_j) + G_j(C_jF_j - D_jE_j) + E_j(E_jH_j - F_jI_j) \\
b_y &= A_j(-C_iH_j - C_jH_i + D_jI_i + D_iI_j) + A_i(-C_jH_j + D_jI_j) \\
&\quad + G_j(F_iC_j + F_jC_i - D_iE_j - E_iD_j) + G_i(C_jF_j - D_jE_j) \\
&\quad + E_j(-F_jI_i - F_iI_j + E_iH_j + E_jH_i) + E_i(E_jH_j - F_jI_j) \\
c_y &= A_i(-C_iH_j - C_jH_i + D_jI_i + D_iI_j) + A_j(-C_iH_i + D_iI_i) \\
&\quad + G_i(F_iC_j + F_jC_i - D_iE_j - E_iD_j) + G_j(C_iF_i - D_iE_i) \\
&\quad + E_i(-F_jI_i - F_iI_j + E_iH_j + E_jH_i) + E_j(E_iH_i - F_iI_i) \\
d_y &= A_i(-C_iH_i + D_iI_i) + G_i(C_iF_i - D_iE_i) + E_i(E_iH_i - F_iI_i)
\end{aligned}
$$

$$
\begin{aligned}
a_z &= A_j(-B_jI_j + H_jD_j) + F_j(F_jI_j - H_jE_j) - G_j(F_jD_j - B_jE_j) \\
b_z &= A_j(H_jD_i + H_iD_j - I_iB_j - I_jB_i) + A_i(H_jD_j - I_jB_j) \\
&\quad + F_j(F_jI_i + F_iI_j - H_jE_i - H_iE_j) + F_i(F_jI_j - H_jE_j) \\
&\quad - G_j(F_jD_i + F_iD_j - B_jE_i - B_iE_j) - G_i(F_jD_j - B_jE_j) \\
c_z &= A_i(H_jD_i + H_iD_j - I_iB_j - I_jB_i) + A_j(H_iD_i - I_iB_i) \\
&\quad + F_i(F_jI_i + F_iI_j - H_jE_i - H_iE_j) + F_j(F_iI_i - H_iE_i) \\
&\quad - G_i(F_jD_i + F_iD_j - B_jE_i - B_iE_j) - G_j(F_iD_i - B_iE_i) \\
d_z &= A_i(-B_iI_i + H_iD_i) + F_i(F_iI_i - H_iE_i) - G_i(F_iD_i - B_iE_i) \\
a^* &= A_j(B_jC_j - D_j^2) - F_j(F_jC_j - D_jE_j) + E_j(F_jD_j - B_jE_j) \\
b^* &= A_j(B_jC_i + B_iC_j - 2D_iD_j) + A_i(B_jC_j - D_j^2) \\
&\quad - F_j(F_jC_i + F_iC_j - D_jE_i - D_iE_j) - F_i(F_jC_j - D_jE_j) \\
&\quad + E_j(F_jD_i + F_iD_j - B_jE_i - B_iE_j) + E_i(F_jD_j - B_jE_j) \\
c^* &= A_i(B_jC_i + B_iC_j - 2D_iD_j) + A_j(B_iC_i - D_i^2) \\
&\quad - F_i(F_jC_i + F_iC_j - D_jE_i - D_iE_j) - F_j(F_iC_i - D_iE_i) \\
&\quad + E_i(F_jD_i + F_iD_j - B_jE_i - B_iE_j) + E_j(F_iD_i - B_iE_i) \\
d^* &= A_i(B_iC_i - D_i^2) - F_i(F_iC_i - D_iE_i) + E_i(F_iD_i - B_iE_i)
\end{aligned}
$$

Substituting the new expressions (B.4) for x, y, and z in $f_j(x, y, z) = 0$, we obtain a polynomial degree sixth with respect to λ

$$
a_6\lambda^6 + a_5\lambda^5 + a_4\lambda^4 + a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 = 0 \tag{B.5}
$$

where

$$
\begin{aligned}
a_6 &= A_j a_x^2 + B_j a_y^2 + C_j a_z^2 + 2D_j a_y a_z + 2E_j a_z a_x + 2F_j a_x a_y \\
&\quad + 2G_j a^* a_x + 2H_j a_y a^* + 2I_j a_z a^* + J_j {a^*}^2 \\
a_5 &= 2A_j a_x b_x + 2B_j a_y b_y + 2C_j a_z b_z + 2D_j (a_z b_y + a_y b_z) \\
&\quad + 2E_j (b_z a_x + b_x a_z) + 2F_j (b_x a_y + b_y a_x) + 2G_j (a_x b^* + b_x a^*) \\
&\quad + 2H_j (a_y b^* + b_y a^*) + 2I_j (a_z b^* + b_z a^*) + J_j (2a^* b^*) \\
a_4 &= A_j (b_x^2 + 2a_x c_x) + B_j (b_y^2 + 2c_y a_y) + C_j (b_z^2 + 2c_z a_z) \\
&\quad + 2D_j (a_y c_z + c_y a_z + b_y b_z) + 2E_j (a_z c_x + a_x c_z + b_x b_z) \\
&\quad + 2F_j (a_x c_y + c_x a_y + b_x b_y) + 2G_j (a_x c^* + a^* c_x + b_x b^*) \\
&\quad + 2H_j (a_y c^* + c_y a^* + b_y b^*) + 2I_j (c_z a^* + a_z c^* + b_z b^*) \\
&\quad + J_j ({b^*}^2 + 2a^* c^*) \\
a_3 &= A_j (2a_x d_x + 2c_x b_x) + B_j (2a_y d_y + 2b_y c_y) \\
&\quad + C_j (2a_z d_z + 2b_z c_z) + 2D_j (a_y d_z + d_y a_z + b_y c_z + c_y b_z) \\
&\quad + 2E_j (a_z d_x + d_z a_x + b_z c_x + c_z b_x) + 2F_j (a_x d_y + d_x a_y + b_x c_y + b_y c_x) \\
&\quad + 2G_j (a_x d^* + d_x a^* + c_x b^* + b_x c^*) + 2H_j (a_y d^* + d_y a^* + b_y c^* + c_y b^*) \\
&\quad + 2I_j (a_z d^* + d_z a^* + b_z c^* + c_z b^*) + J_j (2a^* d^* + 2c^* b^*) \\
a_2 &= A_j (c_x^2 + 2b_x d_x) + B_j (c_y^2 + 2d_y b_y) + C_j (c_z^2 + 2d_z b_z) \\
&\quad + 2D_j (b_y d_z + d_y b_z + c_y c_z) + 2E_j (b_z d_x + b_x d_z + c_x c_z) \\
&\quad + 2F_j (b_x d_y + b_y d_x + c_x c_y) + 2G_j (b_x d^* + b^* d_x + c_x c^*) \\
&\quad + 2H_j (b_y d^* + b^* d_y + c_y c^*) + 2I_j (b_z d^* + b^* d_z + c_z c^*) \\
&\quad + J_j ({c^*}^2 + 2b^* d^*) \\
a_1 &= A_j (2d_x c_x) + B_j (2d_y c_y) + C_j (2c_z d_z) \\
&\quad + 2D_j (c_y d_z + d_y c_z) + 2E_j (c_z d_x + c_x d_z) + 2F_j (c_x d_y + c_y d_x) \\
&\quad + 2G_j (c_x d^* + c^* d_x) + 2H_j (c_y d^* + d_y c^*) + 2I_j (c_z d^* + c^* d_z) + J_j (2d^* c^*) \\
a_0 &= A_j d_x^2 + B_j d_y^2 + C_j d_z^2 + 2D_j d_y d_z + 2E_j d_z d_x + 2F_j d_x d_y + 2G_j d_x d^* + 2H_j d_y d^* \\
&\quad + 2I_j d_z d^* + J_j {d^*}^2
\end{aligned}
$$

APPENDIX C ALGORITHM FOR THE GENERATION OF PAIRS OF RANDOM ELLIPSES AND ELLIPSOIDS

C.1 Algorithm for the Generation of Pairs of Random Ellipses and Ellipsoids

C.1.1 Algorithm for Ellipses

We present in this section an algorithm to generate arbitrary pairs of ellipses that may overlap or not for which we know the exact position of the solution $\boldsymbol{x}_j$ to the minimization problem (3.9) The idea is to start by defining an ellipse $\bar{\mathcal{E}}_j$, centered at the origin and aligned with the coordinate system $(\bar{O}, \bar{x}, \bar{y})$, and by constructing a unit circle $\bar{\mathcal{C}}_i$. The final ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ are obtained in the coordinate system (O, x, y) using the transformations of Section 3.7.2

Ellipse $\bar{\mathcal{E}}_j$: We propose to characterize the semi-axes $(\bar{a}_j, \bar{b}_j)$ of $\bar{\mathcal{E}}_j$ in terms of two dimensionless parameters, namely the aspect ratio γ_j of $\bar{\mathcal{E}}_j$ and the ratio ω_j between the volumes of the ellipse, $V_{\bar{\mathcal{E}}_j}$, and of the unit circle, $V_{\bar{\mathcal{C}}_i}$:

$$\gamma_j = \frac{\bar{a}_j}{\bar{b}_j}, \qquad \omega_j = \frac{V_{\bar{\mathcal{E}}_j}}{V_{\bar{\mathcal{C}}_i}} = \frac{\pi \bar{a}_j \bar{b}_j}{\pi 1^2} = \bar{a}_j \bar{b}_j.$$

Given γ_j and ω_j, the semi-axes are then obtained as:

$$\bar{a}_j = \sqrt{\omega_j \gamma_j},$$
$$\bar{b}_j = \sqrt{\omega_j / \gamma_j}.$$

Constraints on the ratios are as follows:

$$1 \leq \gamma_j \leq \gamma_{\max},$$
$$1 \leq \omega_j \leq \omega_{\max}.$$

with $\gamma_{\max}$ and $\omega_{\max}$ possibly large. The constraints on γ_j and ω_j ensure that the semi-axes are finite, such that $0 < \bar{b}_j \leq \bar{a}_j$, and that the volume of the ellipse $\bar{\mathcal{E}}_j$ is greater than that of the circle $\bar{\mathcal{C}}_i$. The values of γ_j and ω_j can be drawn using uniform distributions or lognormal distributions.

Circle $\bar{\mathcal{C}}_i$: The unit circle $\bar{\mathcal{C}}_i$ is constructed with respect to a point $\boldsymbol{x}_j$ chosen on ellipse $\bar{\mathcal{E}}_j$ with a small "overlap" ε. If $\varepsilon > 0$, the circle is fully outside of the ellipse. If $\varepsilon \leq 0$, $|\varepsilon|$ measures the penetration length of $\bar{\mathcal{C}}_i$ into $\mathcal{E}_j$. The center of the unit circle is determined in terms of $\boldsymbol{x}_j$ and ε as follows:

1. Using the parametric representation of $\bar{\mathcal{E}}_j$, one can take $t_j \in [0, 2\pi[$ so that

$$\bar{\boldsymbol{x}}_j = \begin{bmatrix} \bar{a}_j \cos t_j \\ \bar{b}_j \sin t_j \end{bmatrix}.$$

   If one chooses for example that t_j should follow a uniform distribution, one can draw $\rho \sim \mathrm{U}(0,1)$ and compute $t_j = 2\pi\rho$.

2. We suppose that, given $\varepsilon_{\max} \leq 1$, the "overlap" satisfies $-\varepsilon_{\max} \leq \varepsilon \leq \varepsilon_{\max}$. For instance, one can choose $\rho \sim \mathrm{U}(0,1)$, compute $N(\rho) = (1-\rho)N_{\min} + \rho N_{\max}$ and set $\varepsilon = \pm\varepsilon_{\max} 10^{-N(\rho)}$, with $N_{\min}$ and $N_{\max}$ given. In doing so, one can restrict the overlap to certain subintervals of $[-\varepsilon_{\max}, \varepsilon_{\max}]$ or consider all separation lengths up to about machine precision by taking e.g. $N_{\min} = 0$ and $N_{\max} = 15$.

Therefore, the unit circle $\bar{\mathcal{C}}_i$ is fully defined by providing the center $\bar{\boldsymbol{c}}_i$ as

$$\bar{\boldsymbol{c}}_i = \bar{\boldsymbol{x}}_j + (1+\varepsilon)\bar{\boldsymbol{n}}_j(\bar{\boldsymbol{x}}_j),$$

where $\bar{\boldsymbol{n}}_j(\bar{\boldsymbol{x}}_j)$ is the unit outward normal to $\bar{\mathcal{E}}_j$ at $\bar{\boldsymbol{x}}_j$ computed from (2.16).

Transformation: The objective here is to transform the circle $\bar{\mathcal{C}}_i$ and the ellipse $\bar{\mathcal{E}}_j$ into the ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ with an arbitrary orientation in the coordinate system (O, x, y) using the transformation of Section 3.7.2, that we recall here for convenience

$$\bar{\boldsymbol{x}} = \bar{\mathcal{R}}_j^T \mathcal{D}_i^{1/2} \mathcal{R}_i^T (\boldsymbol{x} - \boldsymbol{c}_j).$$

In other words, one needs to specify the two rotations $\mathcal{R}_i$ and $\bar{\mathcal{R}}_j$, the diagonal matrix $\mathcal{D}_i$ associated with $\mathcal{E}_i$, and the center $\boldsymbol{c}_j$ of $\mathcal{E}_j$.

The center is conveniently generated in polar coordinates, i.e. $\boldsymbol{c}_j = (r\cos\theta, r\sin\theta)$. We choose $\rho \sim \mathrm{U}(0,1)$ and compute $\theta = 2\pi\rho$. For the length r, we provide a maximal value $r_{\max}$, choose $\rho \sim \mathrm{U}(0,1)$, and compute $r = r_{\max}\rho$.

The two rotations $\mathcal{R}_i$ and $\bar{\mathcal{R}}_j$ are defined in 2D in terms of the two angles θ_i and $\bar{\theta}_j \in [0, 2\pi]$ according to Eq. (2.2). We suppose that the angles follow a uniform distribution such that

$\theta_i = 2\pi\rho$ and $\bar{\theta}_j = 2\pi\rho$ with $\rho \sim \mathrm{U}(0,1)$ for each rotation.

In order to transform the circle into an ellipse, we introduce the diagonal matrix $\mathcal{D}_i$:

$$\mathcal{D}_i = \begin{bmatrix} 1/a_i^2 & 0 \\ 0 & 1/b_i^2 \end{bmatrix}$$

where the semi-axes a_i and b_i are selected in a fashion similar to that for $\bar{\mathcal{E}}_j$. However, we impose here that the volume ratio ω_i between $\bar{\mathcal{C}}_i$ and $\mathcal{E}_i$ be equal to unity, i.e. $\omega_i = 1$, which will be motivated below. In other words, we just need to draw a value for the aspect ratio γ_i, so that we have:

$$\begin{aligned} a_i &= \sqrt{\gamma_i}, \\ b_i &= 1/\sqrt{\gamma_i} = a_i^{-1}. \end{aligned}$$

Equations of ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ in (O,x,y): The equation of ellipse $\mathcal{E}_i$ is given by

$$f_i(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{c}_i)^T \mathcal{Q}_i (\boldsymbol{x} - \boldsymbol{c}_i) - 1 = 0,$$

where

$$\begin{aligned} \mathcal{Q}_i &= \mathcal{R}_i \mathcal{D}_i \mathcal{R}_i^T, \\ \boldsymbol{c}_i &= \boldsymbol{c}_j + \mathcal{R}_i \mathcal{D}_i^{-1/2} \bar{\mathcal{R}}_j \bar{\boldsymbol{c}}_i. \end{aligned}$$

The equation of $\mathcal{E}_j$ in (O,x,y) reads:

$$f_j(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{c}_j)^T \mathcal{Q}_j (\boldsymbol{x} - \boldsymbol{c}_j) - 1 = 0,$$

where

$$\mathcal{Q}_j = \mathcal{R}_i \mathcal{D}_i^{1/2} \bar{\mathcal{R}}_j \bar{\mathcal{D}}_j \bar{\mathcal{R}}_j^T \mathcal{D}_i^{1/2} \mathcal{R}_i^T.$$

The semi-axes a_i and b_i of $\mathcal{E}_i$ and angle of rotation θ_i can be obtained from the eigenvalues and eigenvectors of $\mathcal{Q}_j = \mathcal{R}_j \mathcal{D}_j \mathcal{R}_j^T$. The fact that we chose $\omega_i = 1$ implies that we also control the volume ratio ω_{ij} between the two ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ as it is the same as the volume ratio $\omega_j = \bar{a}_j \bar{b}_j$. Indeed, on the one hand, $\omega_i = 1$ implies that $a_i b_i = 1$. On the other hand, by definition of Q_j, we have:

$$\det Q_j = \frac{1}{a_j^2}\frac{1}{b_j^2} = \det \bar{\mathcal{D}}_j \det \mathcal{D}_i = \frac{1}{\bar{a}_j^2}\frac{1}{\bar{b}_j^2}\frac{1}{a_i^2}\frac{1}{b_i^2}$$

so that $a_j b_j = \bar{a}_j \bar{b}_j$. It follows that

$$\omega_{ij} = \frac{\pi a_j b_j}{\pi a_i b_i} = a_j b_j = \bar{a}_j \bar{b}_j = \omega_j.$$

Coordinates of point $\boldsymbol{x}_j$: The point $\boldsymbol{x}_j$ on $\mathcal{E}_j$ is obtained from $\bar{\boldsymbol{x}}_j$ through the transformation as

$$\boldsymbol{x}_j = \boldsymbol{c}_j + \mathcal{R}_i \mathcal{D}_i^{-1/2} \bar{\mathcal{R}}_j \bar{\boldsymbol{x}}_j.$$

This is the unique solution to Problem (3.9) for ellipses $\mathcal{E}_i$ and $\mathcal{E}_j$ in the coordinate system (O, x, y).

Input and output: Apart from the choice of the distributions, the main input data to generate a pair of ellipses are $\gamma_{\max}$, $\omega_{\max}$, $\varepsilon_{\max}$, $N_{\min}$, $N_{\max}$, $r_{\max}$. However, one can imagine variants of above algorithm, which could introduce additional input parameters. The output data consists mainly of the data for the two ellipses, namely $\mathcal{Q}_i$ and $\boldsymbol{c}_i$ for $\mathcal{E}_i$ and $\mathcal{Q}_j$ and $\boldsymbol{c}_j$ for $\mathcal{E}_j$, and the point $\boldsymbol{x}_j$ on $\mathcal{E}_j$. The ellipses can also be described in terms of their semi-axes and angle of rotation, i.e. $\{a_i, b_i, \theta_i, \boldsymbol{c}_i\}$ and $\{a_j, b_j, \theta_j, \boldsymbol{c}_j\}$.

C.1.2 Algorithm for Ellipsoids

We describe in this section the algorithm to generate pairs of ellipsoids. The approach is similar to the 2D case: it starts by constructing an ellipsoid $\bar{\mathcal{E}}_j$ centered at the origin, whose semi-axes are aligned with the coordinate system $(\bar{O}, \bar{x}, \bar{y}, \bar{z})$, and a sphere $\bar{\mathcal{S}}_i$ of unit radius overlapping, or not, with $\bar{\mathcal{E}}_j$. The final pair of ellipsoids $\mathcal{E}_i$ and $\mathcal{E}_j$ are obtained by extending to the 3D case the transformations proposed in [24].

Ellipsoid $\bar{\mathcal{E}}_j$: The semi-axes $(\bar{a}_j, \bar{b}_j, \bar{c}_j)$ of ellipsoid $\bar{\mathcal{E}}_j$ are characterized in terms of three dimensionless parameters: 1) the aspect ratio $\gamma_{j,1}$ between $\bar{a}_j$ and $\bar{b}_j$; 2) the aspect ratio $\gamma_{j,2}$ between $\bar{b}_j$ and $\bar{c}_j$; 3) and the ratio ω between the volumes of the ellipsoid, $V_{\bar{\mathcal{E}}_j}$, and of the sphere, $V_{\bar{\mathcal{S}}_i}$:

$$\gamma_{j,1} = \frac{\bar{a}_j}{\bar{b}_j}, \qquad \gamma_{j,2} = \frac{\bar{b}_j}{\bar{c}_j}, \qquad \omega_j = \frac{V_{\bar{\mathcal{E}}_j}}{V_{\bar{\mathcal{S}}_i}} = \bar{a}_j \bar{b}_j \bar{c}_j.$$

Given $\gamma_{j,1}$, $\gamma_{j,2}$, and ω_j, the semi-axes are then obtained as:

$$\bar{a}_j = \sqrt[3]{\omega_j \gamma_{j,1}^2 \gamma_{j,2}}, \qquad \bar{b}_j = \sqrt[3]{\omega_j \frac{\gamma_{j,2}}{\gamma_{j,1}}}, \qquad \bar{c}_j = \sqrt[3]{\frac{\omega_j}{\gamma_{j,1} \gamma_{j,2}^2}}.$$

Constraints on the ratios are as follows:

$$1 \leq \gamma_{j,1} \leq \gamma_{1,\max},$$
$$1 \leq \gamma_{j,2} \leq \gamma_{2,\max},$$
$$1 \leq \omega_j \leq \omega_{\max},$$

with $\gamma_{1,\max}$, $\gamma_{2,\max}$, $\omega_{\max}$ possibly large. The values of the three parameters can be drawn from uniform or lognormal distributions.

Sphere $\bar{\mathcal{S}}_i$: The construction of the sphere is similar to that of the circle in the 2D case. The unit sphere $\bar{\mathcal{S}}_i$ is constructed with a point $\boldsymbol{x}_j$ on ellipsoid $\bar{\mathcal{E}}_j$ with a small "overlap" ε. The center $\bar{\boldsymbol{c}}_i$ of $\bar{\mathcal{S}}_i$ is determined in terms of $\boldsymbol{x}_j$ and ε as follows:

1. Using the parametric representation of the ellipsoid $\bar{\mathcal{E}}_j$:

$$\bar{\boldsymbol{x}}(u,v) = \begin{bmatrix} \bar{a}_j \cos u \sin v \\ \bar{b}_j \sin u \sin v \\ \bar{c}_j \cos v \end{bmatrix}, \qquad u \in [-\pi, \pi],\ v \in [0, \pi],$$

   one can draw two numbers, e.g. $\rho_1 \sim \mathrm{U}(0,1)$ and $\rho_2 \sim \mathrm{U}(0,1)$, and compute $u_j = 2\pi\rho_1$ and $v_j = \pi\rho_2$. We then define the point $\bar{\boldsymbol{x}}_j$ on $\bar{\mathcal{E}}_j$ such that:

$$\bar{\boldsymbol{x}}_j = \begin{bmatrix} \bar{a}_j \cos u_j \sin v_j \\ \bar{b}_j \sin u_j \sin v_j \\ \bar{c}_j \cos v_j \end{bmatrix}.$$

2. We choose a value of ε in the same manner as in the 2D case, e.g. $-\varepsilon_{\max} \leq \varepsilon \leq \varepsilon_{\max}$.

The center $\bar{\boldsymbol{c}}_i$ of the sphere $\bar{\mathcal{S}}_i$ is thus defined as:

$$\bar{\boldsymbol{c}}_i = \bar{\boldsymbol{x}}_j + (1+\varepsilon)\bar{\boldsymbol{n}}_j(\bar{\boldsymbol{x}}_j)$$

where $\bar{\boldsymbol{n}}_j(\bar{\boldsymbol{x}}_j)$ is the unit outward normal to $\bar{\mathcal{E}}_j$ at $\bar{\boldsymbol{x}}_j$.

Transformation: We now transform the ellipsoid $\bar{\mathcal{E}}_i$ and the sphere $\bar{\mathcal{S}}_j$ into two ellipsoids $\mathcal{E}_i$ and $\mathcal{E}_j$ in the coordinate system (O, x, y, z) using the transformation

$$\bar{\boldsymbol{x}} = \bar{\mathcal{R}}_j^T \mathcal{D}_i^{1/2} \mathcal{R}_i^T (\boldsymbol{x} - \boldsymbol{c}_j),$$

where one needs to specify the two rotations $\mathcal{R}_i$ and $\bar{\mathcal{R}}_j$, the diagonal matrix $\mathcal{D}_i$ associated with $\mathcal{E}_i$, and the center $\boldsymbol{c}_j$ of $\mathcal{E}_j$.

The center is constructed using the spherical coordinates, i.e. $\boldsymbol{c}_j = (r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta)$. We compute $\theta = \pi\rho$ and $\varphi = 2\pi\rho$, with $\rho \sim \mathrm{U}(0,1)$ for each angle. For the length r, we provide a maximal value $r_{\max}$, choose $\rho \sim \mathrm{U}(0,1)$, and compute $r = r_{\max}\rho$.

Rotations in 3D can be represented in several ways. A rotation $\mathcal{R}$ will be expressed here in terms of three elementary rotations as follows:

$$\mathcal{R}(\alpha_1,\alpha_2,\alpha_3) = \mathcal{R}_{\alpha_3}\mathcal{R}_{\alpha_2}\mathcal{R}_{\alpha_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_3 & -\sin\alpha_3 \\ 0 & \sin\alpha_3 & \cos\alpha_3 \end{bmatrix} \begin{bmatrix} \cos\alpha_2 & 0 & \sin\alpha_2 \\ 0 & 1 & 0 \\ -\sin\alpha_2 & 0 & \cos\alpha_2 \end{bmatrix} \begin{bmatrix} \cos\alpha_1 & -\sin\alpha_1 & 0 \\ \sin\alpha_1 & \cos\alpha_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where $\alpha_1,\ \alpha_3 \in [0, 2\pi]$ and $\alpha_2 \in [0, \pi]$. We will consider the two rotations, $\mathcal{R}_i$ and $\bar{\mathcal{R}}_j$, with parameters chosen as $\alpha_1,\ \alpha_3 \sim 2\pi\mathrm{U}(0,1)$ and $\alpha_2 \sim \pi\mathrm{U}(0,1)$.

The semi-axes (a_i, b_i, c_i) of ellipsoid $\mathcal{E}_i$, which form the diagonal matrix $\mathcal{D}_i$:

$$\mathcal{D}_i = \begin{bmatrix} 1/a_i^2 & 0 & 0 \\ 0 & 1/b_i^2 & 0 \\ 0 & 0 & 1/c_i^2 \end{bmatrix},$$

are selected in the same fashion as that for $\bar{\mathcal{E}}_j$. However, we impose here that the volume ratio ω_i between $\mathcal{E}_i$ and $\bar{\mathcal{S}}_i$ be equal to unity, i.e. $\omega_i = 1$. In other words, we just need to draw values for the aspect ratios $\gamma_{i,1}$ and $\gamma_{i,2}$, so that:

$$a_i = \sqrt[3]{\gamma_{i,1}^2\gamma_{i,2}}, \qquad b_i = \sqrt[3]{\frac{\gamma_{i,2}}{\gamma_{i,1}}}, \qquad c_i = \sqrt[3]{\frac{1}{\gamma_{i,1}\gamma_{i,2}^2}}.$$

Equations of ellipsoids $\mathcal{E}_i$ and $\mathcal{E}_j$ in (O, x, y, z): As in 2D, the equation of ellipsoid $\mathcal{E}_i$ in (O, x, y, z) can be written as:

$$f_i(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{c}_i)^T \mathcal{Q}_i (\boldsymbol{x} - \boldsymbol{c}_i) - 1 = 0$$

where:

$$\begin{aligned} \mathcal{Q}_i &= \mathcal{R}_i \mathcal{D}_i \mathcal{R}_i^T, \\ \boldsymbol{c}_i &= \boldsymbol{c}_j + \mathcal{R}_i \mathcal{D}_i^{-1/2} \bar{\mathcal{R}}_j \bar{\boldsymbol{c}}_i. \end{aligned}$$

The equation of $\mathcal{E}_j$ in (O, x, y) reads:

$$f_j(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{c}_j)^T \mathcal{Q}_j (\boldsymbol{x} - \boldsymbol{c}_j) - 1 = 0,$$

where

$$\mathcal{Q}_j = \mathcal{R}_i \mathcal{D}_i^{1/2} \bar{\mathcal{R}}_j \bar{\mathcal{D}}_j \bar{\mathcal{R}}_j^T \mathcal{D}_i^{1/2} \mathcal{R}_i^T.$$

The semi-axes (a_j, b_j, c_j) are obtained from the eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ of $\mathcal{Q}_i$ as:

$$a_j = \frac{1}{\sqrt{\lambda_1}}, \qquad b_j = \frac{1}{\sqrt{\lambda_2}}, \qquad c_j = \frac{1}{\sqrt{\lambda_3}}.$$

Coordinates of point $\boldsymbol{x}_j$: The point $\boldsymbol{x}_j$ is obtained from $\bar{\boldsymbol{x}}_j$ through the transformation

$$\boldsymbol{x}_j = \boldsymbol{c}_j + \mathcal{R}_i \mathcal{D}_i^{-1/2} \bar{\mathcal{R}}_j \bar{\boldsymbol{x}}_j.$$

Output: The main input data to generate a pair of ellipses are $\gamma_{1,\max}$, $\gamma_{2,\max}$, $\omega_{\max}$, $\varepsilon_{\max}$, $N_{\min}$, $N_{\max}$, $r_{\max}$. The output data consists of the data for the two ellipsoids, namely $\mathcal{Q}_i$ and $\boldsymbol{c}_i$ for $\mathcal{E}_i$ and $\mathcal{Q}_j$ and $\boldsymbol{c}_j$ for $\mathcal{E}_j$, and the point $\boldsymbol{x}_j$ on $\mathcal{E}_j$.