Journal of Computational Physics 228 (2009) 3288–3306

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Computational cost of the Fekete problem I: The Forces Method on the 2-sphere

E. Bendito [a], A. Carmona [a], A.M. Encinas [a], J.M. Gesto [a,*], A. Gómez [b], C. Mouriño [b], M.T. Sánchez [b]

[a] Departament de Matemàtica Aplicada III, Universitat Politècnica de Catalunya, Barcelona, Spain
[b] Fundación Centro Tecnológico de Supercomputación de Galicia, CESGA, Santiago de Compostela, Spain

ARTICLE INFO

Article history:
Received 9 June 2008
Received in revised form 13 January 2009
Accepted 16 January 2009
Available online 30 January 2009

Keywords:
Fekete points
Computational complexity
Nonlinear optimization

ABSTRACT

Here, we study the computational complexity of the Fekete point problem. Namely, we give an exhaustive description of the main properties of an algorithm for the minimization of the logarithmic potential energy on the 2-sphere, and we characterize the probability distribution of the cost of the different minima. In particular, we show that a local minimum can be found with an average cost of about $O(N^{2.8})$.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Many mathematical problems with applications in Physics, Numerical Methods and Complexity Theory are posed in terms of the minimization of a certain potential energy functional associated with a system of point particles under general constraints. Usually, the involved functionals depend on the relative distances between pairs of particles. For instance, a variation of the Thomson problem posed as part of the development of the plum pudding model of the atom consists of determining the minimum energy configuration of N classical electrons on the surface of the 2-sphere. The electrons repel one another with a force given by Coulomb's law. Another classical problem is that of maximizing the product of the relative distances between N points on the 2-sphere, which is equivalent to minimizing their logarithmic potential energy. Finding an efficient algorithm for the search of good estimations of such optimal points is the center of S. Smale's 7th "Mathematical problem for the 21st century". This problem is closely related to the efficient resolution of polynomial equation systems, and hence also with the resolution of general nonlinear equation systems. In Molecular Mechanics, each atom is regarded as a weighted point particle, and the stable molecular geometries are local minimizers of a potential energy functional depending on the relative distances between the atoms and other geometric parameters. In Dynamic Systems, the so-called planar central configurations are local minimizers of a generalized potential energy functional involving the Newtonian potential energy and the distances from each point mass to the mass center of the system. Different potential energy functionals have also been used in numerical integration, polynomial interpolation, mesh generation and also in computer graphics to distribute points on an implicit surface as an initial step for its visualization.

The literature on all these interesting topics is vast, and the authors use different notations and terminologies. Originally, the name Fekete points was related to the answer provided by M. Fekete to an algebraic problem posed by I. Schur in 1918, see [9]. Further works in Potential Theory showed deep relations of the Fekete points with the concept of capacity, and this

* Corresponding author. Tel.: +34 93 401 69 14; fax: +34 93 401 18 25.
E-mail address: jose.manuel.gesto@upc.edu (J.M. Gesto).

0021-9991/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2009.01.021

name became widely used in different contexts. We call the *Nth order Fekete points* the N-tuples $\omega_N = \{x_1, \ldots, x_N\}, x_i \in \mathbb{R}^d$, that minimize under general constraints any potential energy functional $\mathcal{I}_N$ involving the relative Euclidean distances between N points, and we call the *Fekete problem* that of determining these N-tuples. This framework includes potential energy functionals of the form

$$\mathcal{I}_N(x) = \sum_{1 \leqslant i < j \leqslant N} \mathcal{K}(x_i, x_j),$$

where $x = \{x_1, \ldots, x_N\}, x_i \in \mathbb{R}^d$, and the *kernel* $\mathcal{K}$ is a function of the Euclidean distance between x_i and x_j, $|x_i - x_j|$. Some interesting kernels are the logarithmic kernel, $-\log|x_i - x_j|$, and Riesz's kernels, defined by $|x_i - x_j|^{-s}$, with $s > 0$. The limit case $s \to 0$ recovers the logarithmic kernel, whereas the other limit case $s \to \infty$ leads to the so-called best-packing problem or Tammes problem. The particular case $s = d - 2$ plays a specially important role. The corresponding kernel is known as the Newtonian kernel and its potential energy functional is called the electrostatic potential energy.

Several results have been obtained regarding the asymptotic behavior of the Fekete points in different areas. In particular, recent works show that N point configurations that minimize Riesz's potential energy on a general rectifiable $(d-1)$-manifold K are asymptotically uniformly distributed (with $N \to \infty$) with respect to $(d-1)$-dimensional Hausdorff measure on K when $s \geqslant d - 1$, see [14,15]. The good separation properties of the Fekete points have been deeply analyzed in the context of numerical integration, where valuable results have been produced specially for the case of numerical integration on spheres, see, for instance, [4–7] and the references therein. The problem of the approximation of potentials and electrostatic fields generated by continuum distributions of charges on conductor bodies by means of Fekete points has also been studied, see [18].

Although the available theoretical results on the Fekete problem are of great interest and help us to understand some of the main features of the problem, often the information that they provide is rather vague from the point of view of the applications. Optimal point configurations in the interval and in the circle have been obtained, but only a handful of solutions for low values of N are known even in the simplest three-dimensional case of the 2-sphere. Thus, the use of numerical methods becomes necessary in practice.

The Fekete problem is a prototype of a highly nonlinear constrained optimization problem. One of its main characteristics is the massive multiextremality. In general, many local minimizers exist and some of them are fairly close (by their energy values) to the "true" global solutions. In fact, the evoked problems are well suited to test global optimization routines, see [17]. Thomson's problem requires a global solution and remains largely unsolved after almost 100 years. In [25], P.M. Pardalos stated the maximization of the product of relative distances on the 2-sphere as an open global optimization problem. Certificates of global optimality are not available; the usual first-order and second-order optimality conditions are easy to derive but of little value. Anyway, in many practical applications not only the global solutions but the local ones, approximated with a required accuracy degree depending on the context, are also demanded.

Smale's 7th problem gives a new dimension to the algorithmic and numerical treatment of the problem and focuses on its computational complexity, see [29,30]. A wide variety of optimization algorithms have been used to tackle the search of local and global minima for different potential energies. These algorithms can be divided into two categories: Classic Optimization Algorithms, basically Gradient and Conjugate Gradient methods, Relaxation methods and Newton and quasi-Newton methods; and Combinatorial Optimization Algorithms, for instance, Simulated Annealing, Tabu Search and Genetic algorithms. The essential difference is that the algorithms in the first family are fully deterministic, whereas randomness is a fundamental ingredient for the second group.

Classic optimization algorithms use analytic tools to construct a sequence of approximations tending to an optimal solution in a deterministic way. This iterative process starts in general from a given first approximation that leads univocally to a solution that in general is only a local minimum. So, global minimization strategies typically include multistart procedures where the random component can be introduced.

Combinatorial optimization methods were developed mainly for combinatorial or discrete optimization problems. Some of these problems are NP-complete, hence it is very unlikely that an efficient algorithm exists. However, these algorithms can be applied to get relatively good solutions in a reasonable time. Since the dawn of these algorithms they have been applied to continuous problems, too, with varying success.

For many versions of the Fekete problem classic algorithms or their combinations have been typically used. In the 1990s, intensive computations for the basic problems of the Newtonian and the logarithmic energy as well as the best-packing problem on the 2-sphere started to be performed. Many different optimization algorithms were proved by different authors. Classic algorithms usually require the elimination of the problem constraints by means of some parametrization, namely spherical or stereographic coordinates. Combinatorial algorithms can be directly applied by combining them with some projection technique to make the particles return to the surface.

K.J. Nurmela is particularly interested in the best-packing problem. This problem can be viewed as the limit of Riesz's energy problems, so this author also focuses on them. In [21], one can find a useful description of a great variety of different algorithms and combinations, with some comments on their practical implementation, their advantages and their limitations when they are applied to the Fekete problem. Not only the 2-sphere is considered by this author, but he also uses different optimization approaches to deal with the best-packing and covering problems in flat domains such as the circle, the triangle and the square; see, for instance, [22–24]. In [31], one can find tables in which different authors gather, among other

things, best energy values for the Lennard-Jones potential on the 2-sphere for N up to 75 and best sphere packings in 3, 4 and 5 dimensions for N up to 130.

We can also mention the Dissertation of Y. Zhou, [36], where in particular energy values are shown for the logarithmic and the Newtonian kernels in the 2-sphere for N up to 200. In general, all these results were obtained by proving and combining classical methods. T. Erber and G.M. Hockney used a relaxation algorithm for the Newtonian problem on the 2-sphere, and their computer trials indicated that in the range $70 \leqslant N \leqslant 112$ the number of distinct local minima associated with each value of N grows exponentially with N, see [8]. This exponential growth has been confirmed by several other authors. We must remark the work carried out by R. Womersley and by M.J. Bowick. In [34], Womersley gives energy values for the Newtonian kernel in the 2-sphere for $N = (m+1)^2$ with m up to 80, so the biggest case studied has $N = 6561$. He used a combination of local and global large-scale optimization techniques running on a cluster for the problem in the spherical coordinate parametric space. Using the same procedures, Womersley has also obtained configurations corresponding to different kernels on a torus, see [35]. In [3], Bowick describes many classical and combinatorial algorithms that he has used to tackle the Fekete problem. In particular, he shows energy values and configurations for the Newtonian energy for many N values up to 5000 on the 2-sphere. He also considers other energies and the case of the torus.

Genetic algorithms have rarely been applied to physical systems. Mating, mutation and crossover in these systems are often difficult to define, and determining an organism's fitness is often computationally intensive. These algorithms can be very effective as global optimization methods, but usually they are extremely slow. In [20], J.M. Morris et al. use a genetic algorithm for the Fekete problem. They combine this algorithm with a conjugate gradient technique. The authors provide values for the Newtonian energy on the 2-sphere up to $N = 200$. In the recent work [12], L. Giomi and Bowick apply a genetic algorithm to minimize the Newtonian potential energy of N points, N up to 200, on a paraboloidal surface with boundary. These authors emphasize the dramatic slowing down of the algorithm when applied to large-scale optimization.

An alternative formulation of the Fekete problem is based on the mechanical interpretation of the behavior of a system of particles that interact according to forces derived from a potential energy functional. If no dissipative forces are introduced, the system moves indefinitely while maintaining its total energy constant. However, if an adhesion force is implied on the particles, the total energy diminishes with time and finally a static stable position is found. The equivalence between these static stable equilibrium positions and minimum energy configurations constitutes the key of this approach. From this new perspective, minimum potential energy configurations are obtained by integrating the movement equations of the particle system with additional dissipative terms under the constraints of the corresponding optimization problem.

W.J.H. Stortelder et al., see [32], consider the complete Lagrangian movement equations of a system of point particles interacting according to the forces derived from the logarithmic potential energy and constrained to the 2-sphere. They include the adhesion forces necessary to reach a final equilibrium position. To solve the system, the RADAU5 code by E. Hairer and G. Eanner, see [13], was used. The work [32] contains the results of a comparison between this algorithm and other more classical optimization algorithms when they search for optimal configurations of N up to 150 particles. The authors show that for the studied cases the ODE approach is competitive, but they emphasize that memory limitations will become a serious drawback for this method in the case of increasing N due to the storage required by RADAU5.

Some other authors have also applied the mechanical analogy to deal with the Fekete problem. We can mention [19], where R. van Liere et al. use symplectic integrators for the problem of logarithmic energy on the 2-sphere (symplectic methods exploit the special structure of the problem and retain certain physical properties of the dynamic system, see [27]). In [16], J.S. Hesthaven uses a 7(6) embedded Nyström–Runge–Kutta scheme with error control to compute interpolation points in the triangle. In both cases, only small systems are considered. Several other authors have also used simplified approaches inspired in a mechanical conception of the problem to design mesh generators and to sample and render implicitly defined surfaces, see, for instance, [26,28,33] and the references therein. In general the rigorous definition of equilibrium is lost under these simplifications, but the developed algorithms are usually efficient and versatile from the point of view of their applications.

Here, we summarize the main conclusions that can be derived from the available information regarding the problem of the numerical minimization of the logarithmic and Riesz's energies under general constraints:

- Two approaches have been used to deal with the different variants of the Fekete problem: optimization algorithms and ODE integration algorithms. The latter approach has mainly been used in contexts such as mesh generation and computer aided design, where usually only reasonably well-distributed points are needed, and the accurate resolution of the optimization problem is not relevant.
- Few authors have considered the optimization problem out of the 2-sphere (apart from the Computer Graphics context).
- For the case of the 2-sphere, many authors have carried out rather exhaustive searches for global solutions for the logarithmic and Newtonian potential energies with $N \leqslant 200$. Their strategies cover practically the whole spectrum of available optimization methods, and in many cases combinations of them have been used. Moreover, some authors have applied ODE integration algorithms for the Fekete problem obtaining competitive results.
- For $N > 200$, only very partial results are available even in the case of the 2-sphere. Sometimes symmetries are used to reduce the number of variables and to obtain some information about the asymptotical behavior of the problem.
- No general results about convergence, stability, robustness and the computational cost of the employed algorithms have been published.

This paper is the first in a series of which we made a systematic study of the computational complexity of the Fekete problem. Here, we focus almost exclusively in the case of the logarithmic energy on the 2-sphere. In particular, we present a statistical analysis based on the extensive numerical experiments of the computational cost of an approximate local minimum for this case. The robustness, convergence and stability properties of the employed algorithm are identified and corroborated by the results of a massive computation program.

2. Background

Our first works on the numerical estimation of Fekete points focused on obtaining an efficient and robust algorithm for the identification of local minima of a potential energy restricted to a general regular surface. Then, we developed additional techniques that allowed us to apply the mentioned algorithm to a wide variety of objects, which we call W-compact sets while keeping their good properties. The W-compact sets are essentially the finite union of boundaries of open sets, surfaces with boundary and curves with boundary and they include, in particular, non-smooth surfaces. In [1], we gave a detailed description of the technical aspects of our approach. Nevertheless, we consider it convenient to briefly describe the algorithm, which we call *the Forces Method*, for the sake of completeness.

The basic structure of the Forces Method is classical. It can be viewed as a relaxation gradient-like descent algorithm in such a way that each step consists of the advance direction obtained and the step size in a deterministic way. This leads to the scheme

$$\hat{x}^{k+1} = x^k + \lambda_k w^k,$$

where x^k, w^k and λ_k are the position, the advance direction and the step size at the kth step, respectively, and $\hat{x}^{k+1}$ is the non-corrected position at the $(k+1)$th step. Position x^{k+1} is obtained by projecting each $\hat{x}_i^{k+1}$ to the surface S. With regard to the advance direction, let us start by observing that the potential energy of a system of N unitary particles $x_i \in \mathbb{R}^3, i = 1, \ldots, N$, is given by $\mathcal{I}_N = \frac{1}{2}\sum_{i=1}^N V_i$, where $V_i = \sum_{\substack{j=1 \\ j\neq i}}^N \mathcal{K}(x_i, x_j)$ is the potential created at x_i by all the other particles. If we fix the position of the $N-1$ particles $\{x_j \in S : j = 1, \ldots, N, j\neq i\}$, then V_i is a function of x_i and the opposite of its gradient, which we denote by $F_i = -\nabla V_i \in T_{x_i}(\mathbb{R}^3)$, represents the repulsive force acting on the ith particle due to the existence of the rest. If the particles lie on a regular surface $S \subset \mathbb{R}^3$, and F_i^T denotes the tangential component to S of the force F_i at x_i, then we choose $w = (w_1, \ldots, w_N)$ as the advance direction, where $w_i = \frac{F_i^T}{|F_i|}$. Moreover, we call the *disequilibrium degree* of the ith particle the scalar $|w_i|$. The magnitude of the step size is obtained from the expression

$$\lambda_k = a \min_{1\leqslant i<j\leqslant N} \{|x_i^k - x_j^k|\},$$

where the coefficient a is a positive scalar that we keep constant throughout the descent process (it does not depend on k), and the minimum distance between particles allows us to adapt the step size to the difficulty of the different configurations that appear during the calculation. From a mechanical point of view, we can see the descent scheme as the integration scheme of the autonomous system of ODE $x' = w(x)$, or, equivalently, the system $x' = \varphi(x)w(x)$, where $\varphi(x) = \min_{1\leqslant i<j\leqslant N}\{|x_i - x_j|\}$, by means of Euler's forward method. The variable factor $\varphi(x^k)$ reduces the step size when there exist very close particles to prevent them from "running helter-skelter", breaking the continuity of the movement, and it increases the step size when the relative distances grow in benefit of the convergence ratio. The magnitude $a|w_i|$ represents the fraction of the minimum distance between particles that the ith particle advances at each step.

For the study of the convergence of the algorithm, we use the maximum disequilibrium degree, $w_{\max} = \max_{1\leqslant i\leqslant N}|w_i|$, as a measure of the error at each step. We call the *convergence curve* the graph that displays the evolution of the error $w_{\max}$ with the step number, n_{step}. Fig. 1 corresponds to the application of the algorithm on the 2-sphere with the logarithmic kernel for $N = 1000$ and $a = 17.234$, and it shows its general behavior. The starting configuration (left) was generated according to a uniform probability density on the 2-sphere. In the center, we can see the configuration corresponding to step 8000 ($n_{\text{step}} = 8000$). The corresponding convergence curve (right) attains a final linear convergence ratio after a highly nonlinear phase. In this case, $x_i^{k+1} = \frac{\hat{x}_i^{k+1}}{|\hat{x}_i^{k+1}|}$, $i = 1, \ldots, N$.

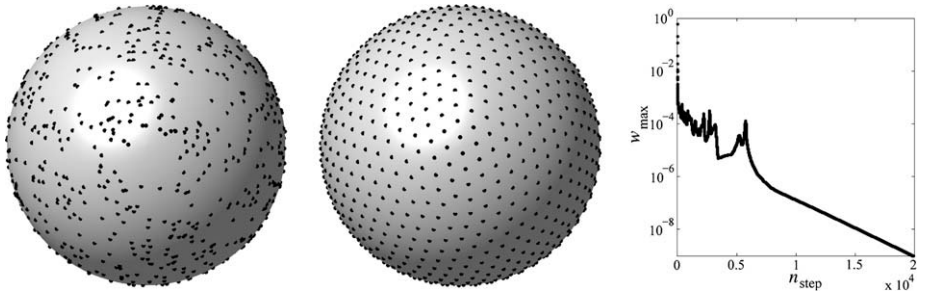


Fig. 1. Initial and final configurations of $N = 1000$ particles on the 2-sphere and the corresponding convergence curve.

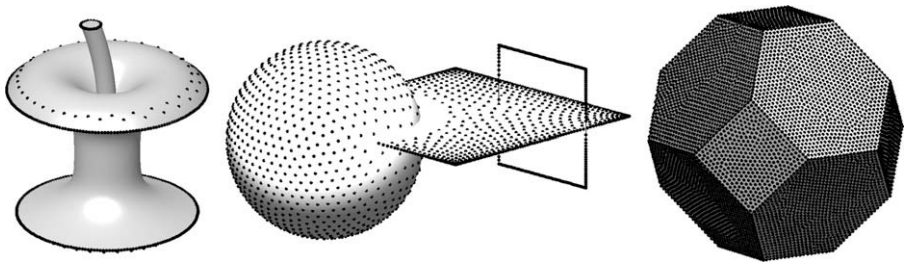


Fig. 2. $N = 500$ particles on an apple with the logarithmic kernel, $N = 2000$ particles on a W-compact set obtained by combining different dimension objects with the Newtonian kernel and $N = 10,000$ particles on the Kelvin Polyhedron with Riesz's kernel for $s = 2$.

To apply the above-mentioned algorithm in a non-regular surface $\mathcal{S}$, it is necessary to design a strategy for the generation of good starting configurations, because if the initial position is not good enough, the singularities of the surface could cause the particles to become trapped at unsatisfactory minima or stationary points. To solve this, we use a sequence of acceptable equilibrium configurations on a small number of approximating smooth surfaces. In [1], we described the W-compact sets that admit an approximation by regular surfaces, as well as a technique used in Computer Graphics for the construction of approximating regular surfaces by means of the composition of implicit equations. Moreover, we analyzed the potential energy and its gradient restricted to a W-compact set, which is essential for determining the disequilibrium degrees of particles at non-regular points. We also gave some additional details to carry out the implementation of our procedure, whose versatility and robustness we would like to mention here. Fig. 2 shows some equilibrium configurations for different kernels and W-compact sets.

3. The Forces Method

In this section, we focus on the case of the logarithmic kernel in the 2-sphere, which establishes an interesting framework to carry out a profuse study of the properties of our algorithm. So, throughout this section S stands for the 2-sphere and $\mathcal{I}_N(x)$ is the logarithmic potential energy of an N point set $x = \{x_1, \ldots, x_N\}$.

The rest of this section is organized as follows: after some notation, we analyze the convergence properties of our descent algorithm as well as the influence of the random starting configurations in the accessibility to the different local minima and in the computational cost. After that, we present the results of a statistical analysis carried out to characterize the computational cost of identifying a local minimum.

3.1. Terminology and notation

We have already introduced the forces F_i, their tangential components F_i^T, the advance direction $w = (w_1, \ldots, w_N)$, where $w_i = \frac{F_i^T}{|F_i|}$, the disequilibrium degree of a particle, $|w_i|$, and the error at a step, $w_{\max} = \max_{1 \leqslant i \leqslant N} |w_i|$. We have also defined the convergence curve as the graph of $w_{\max}$ as a function of the step number, n_{step}. A typical convergence curve contains a first highly nonlinear phase and a final linear phase or linear tendency; see Fig. 1.

After generating the initial position x^0 of N particles on S and fixing the magnitude of the coefficient a, the Forces Method can be applied recursively. The algorithm stops when $w_{\max}$ reaches a certain prescribed threshold value $\varepsilon > 0$. We call this ε-*convergence*. This definition of convergence is useful in practice, but it is not entirely satisfactory from a more theoretical point of view. We can introduce a new definition of convergence based on the concept of approximate local minimum, which is analogous to the concept of approximate zero of a system of polynomial equations proposed by M. Shub and Smale; see, for instance, [30]. Thus, we say that the algorithm has converged when it has found an approximate local minimum; that is, when the current position x^k is close to a local minimum in such a way that Newton's algorithm converges quadratically to this local minimum from x^k. This definition of convergence is associated to the size of the local influence zone of the minima and does not depend on ε.

In general, only when the final linear tendency of the convergence curve is reached can it be assumed that we have an approximate local minimum. Nevertheless, in practice it is not easy to accurately determine the beginning of the final linear tendency, and it is impractical to determine the precise step from which Newton's algorithm converges quadratically. To eliminate this ambiguity, we define a third kind of convergence: if we consider a long enough convergence curve; that is, with ε small enough, the step corresponding to its last maximum can be determined without ambiguity and it precedes the linear tendency. We call the last maximum in a convergence curve the *non-return point*. The non-return point anticipates the entrance to the attraction zone of a minimum. We say that the Forces Method has *nr-converged* when the non-return point has been attained. Obviously, the *nr*-convergence cannot be detected throughout the calculation and cannot be used in practice to stop the descent process. However, if we lead the algorithm to ε-convergence with ε small enough, then the non-return point is the last registered maximum in the convergence curve. We use the *nr*-convergence as an indirect way of measuring the computational cost of identifying an approximate local minimum.

The concept of *nr*-convergence characterizes different features of the Fekete problem and it is not exclusively related to the size of the influence area of the local minima. We consider that this notion of convergence is more appropriate and useful than that associated to the quadratic convergence of Newton's method in the Fekete point problem context. We use the term approximate local minimum both in the sense introduced by Shub and Smale and to denote the configurations x^k that the Forces Method provides after the non-return point is attained.

On the other hand, if the starting positions x^0 are randomly generated, then the number of steps needed to converge in a run of the Forces Method is a random variable. Specifically, we define the random variables X, the number of steps for *nr*-convergence in a run, and Y, the number of steps for ε-convergence in a run. For simplicity of notation, we have not used indices in the definitions of the variables X, Y. In each case, the values for N and ε will be clear. For a generic random variable Z with probability density function f_Z and probability distribution function F_Z, we call $M_Z^k = E[Z^k], k \in \mathbb{N}$, the kth order moment of Z, $\mu_Z = M_Z^1$ the mean of Z, $(M_Z^k)' = E[(Z-\mu_Z)^k]$ the kth order centered moment of Z and $\sigma_Z = \sqrt{(M_Z^2)'}$ the standard deviation of Z. Moreover, we call $z_i, i = 1, \ldots, n_{\mathrm{sp}}$, the sample data obtained in an experiment for Z, $m_z^k = \frac{1}{n_{\mathrm{sp}}}\sum_{i=1}^{n_{\mathrm{sp}}} z_i^k$ the sample moments of $\{z_i\}$, $\bar{z} = m_z^1$ the sample mean of $\{z_i\}$, $(m_z^k)' = \frac{1}{n_{\mathrm{sp}}}\sum_{i=1}^{n_{\mathrm{sp}}}(z_i-\bar{z})^k$ the sample centered moments of $\{z_i\}$ and $S_Z = \sqrt{(m_z^2)'}$ the sample standard deviation of $\{z_i\}$.

3.2. The coefficient a

All the components of our descent algorithm except the coefficient a can be determined at each step from the position of the particles. From the point of view of the cost of the algorithm at each step, it is clear that the optimal choice consists of keeping this coefficient constant throughout the descent process, since then its cost at each step is null. This reasoning could seem excessively simplistic. Certainly, we should consider some crucial issues such as the total computational cost of the process (that is, the number of steps needed to converge), the possibility of divergence or the influence of the starting configuration and the value of N. In this section, we present the results of a series of numerical experiments designed to determine the behavior of the algorithm when the coefficient a is kept constant.

3.2.1. Experiment 1

Taking into account the enormous growth with N of the number of local minima of the Fekete problem, we start the study by considering the case $N = 87$, which still has a relatively small amount of minima and provides us with substantial information for each minimum.

In the first experiment, we use two different starting configurations: one generated according to a uniform probability density on the whole 2-sphere, and the other generated according to a uniform probability density on a spherical cap of area $\pi \cdot 10^{-6}$. In the sequel, we call uniform starting configurations the ones randomly generated from a uniform probability density on the whole 2-sphere, and delta starting configurations the ones randomly generated according to a uniform probability density on a spherical cap of area $\pi \cdot 10^{-6}$.

The experiment consisted of running the algorithm until attaining ε-convergence ($\varepsilon = 10^{-7}$) from both starting positions taking 2000 different values for the coefficient a in each case. In each run, the value of a remained constant from the starting configuration to the ε-convergence position. Fig. 3 shows the evolution with a of the number of steps necessary to attain ε-convergence ($\varepsilon = 10^{-7}$) from the uniform starting configuration (up) and from the delta starting configuration (down).

Fig. 3 shows that there exist an important range of values of a that guarantee the convergence to a minimum. Moreover, it can be observed that the critical value of a from which the algorithm diverges, a_{crit}, is practically the same for both initial configurations, which is especially remarkable. Moreover, in both cases the evolution with a of the number of steps necessary to converge is qualitatively and quantitatively similar and the smallest calculation times appear when a tends to a_{crit} from the left.

Fig. 4 shows the paths described by the particles throughout the convergence process from the uniform starting configuration with $a = 0.9, a = 1.3$ and $a = 5.08$ (up from left to right), and from the delta starting configuration with $a = 0.85$, $a = 1.05$ and $a = 5.08$ (down from left to right). For each path the large and medium points correspond to the final and initial configurations, respectively, whereas the small points correspond to the intermediate steps (the delta starting position is confined within a small cap and it cannot be seen). Note that the paths corresponding to the first pair of figures of each group look alike. In fact, the values of a corresponding to each pair lead to the same minimum with different speed, and they correspond to "continuous" fragments of the curves in Fig. 3. There are few of such continuous fragments in the curve down in Fig. 3 because the starting configuration is very extreme. In any case, Fig. 3 makes it clear that small variations in the coefficient a can imply strong variations in the paths followed by the particles from both starting positions. So, in general, the loss of the stability of the descent scheme when it is seen as Euler's forward integration scheme occurs in general for extremely low values of a.

It is interesting to observe how the Forces Method works from a delta starting configuration. Fig. 5 shows, from left to right and from above to below, the configurations for $n_{\mathrm{step}} = 0, 3, 8, 19, 32, 37, 99, 8075$ corresponding to the convergence process associated to Fig. 4 (down to the right). The steps $n_{\mathrm{step}} = 0$ and $n_{\mathrm{step}} = 8075$ correspond to the starting configuration and to the final configuration after the ε-convergence ($\varepsilon = 10^{-7}$) has been attained, respectively. In each case, the scale has been conveniently adjusted to make the visualization of the process easier. As can be observed, in the first steps the algorithm constructs a small ring with all the particles. The diameter of that ring grows until it becomes comparable to the distance between two neighboring particles in the final configuration. From this moment, the ring leaves some particles in its

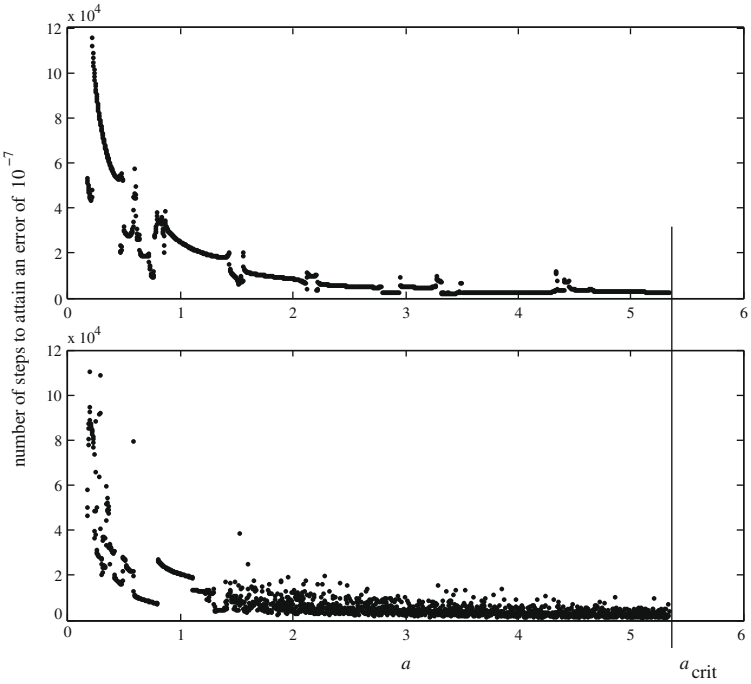


Fig. 3. Evolution with a of the number of steps necessary to attain ε-convergence ($\varepsilon = 10^{-7}$) from a uniform starting configuration (up) and from a delta starting configuration (down).

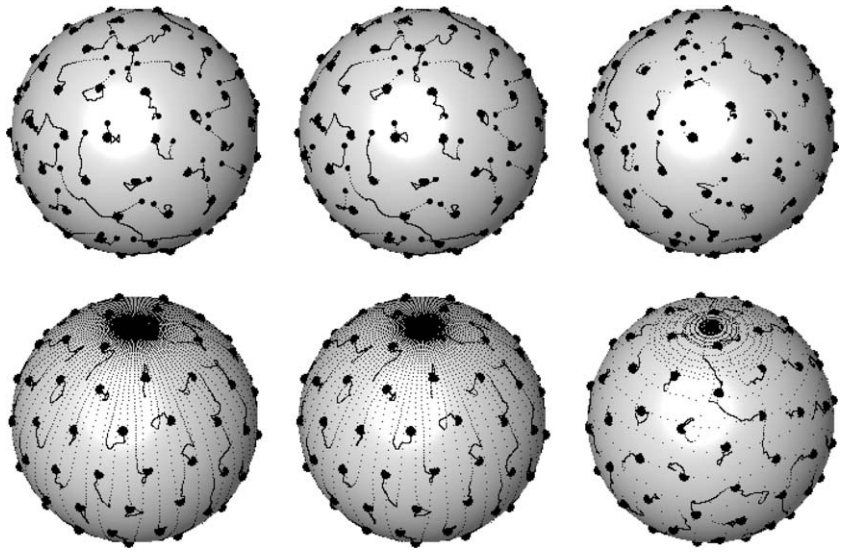


Fig. 4. Paths described by the particles from a uniform starting configuration with $a = 0.9, a = 1.3$ and $a = 5.08$ (top: from left to right) and from a delta starting configuration with $a = 0.85, a = 1.05$ and $a = 5.08$ (bottom: from left to right).

interior. At step $n_{\text{step}} = 32$, the ring has already left some particles, and a few steps later it has "evenly distributed" all the particles on the sphere's surface. Most of the remaining steps are spent in localizing a minimum accurately.

3.2.2. Experiment 2

This experiment complements the previous one and was designed to confirm that the value of a_{crit} (that is, the value of the coefficient a from which the descent scheme diverges) is strongly independent of the starting configuration. This has been corroborated by all the tests that we have carried out, and it is perhaps the most relevant result for the Fekete point problem that we have obtained.

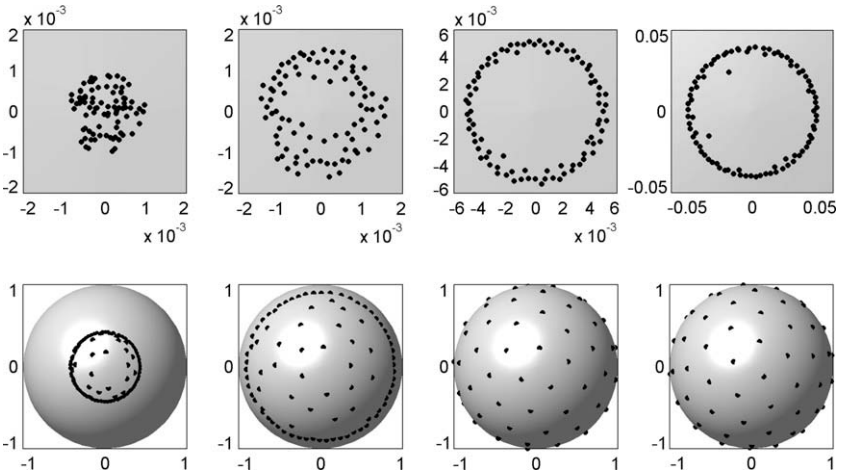


Fig. 5. Convergence process associated to a delta starting configuration. The steps $n_{\text{step}} = 0, 3, 8, 19, 32, 37, 99, 8075$ are displayed from left to right and from top to bottom.

Here, the algorithm runs again from the uniform and delta starting configurations used in Experiment 1, and different values of a were considered in each case. Specifically, the chosen values were $a = 5.4, a = 5.4275, a = 5.43$, $a = 5.433, a = 5.44, a = 5.7$ and $a = 8$. The two diagrams in Fig. 6 (left) show the convergence curves corresponding to these values of a from the uniform configuration (up) and from the delta configuration (down). For the first value $a = 5.4$, the convergence curve goes down in linear tendency until the precision limits are attained (we work in double precision). For the next a values, it is still possible to reach the linear tendency, but this tendency is left at a certain step and finally the process diverges. For higher values of a, the divergence process starts even before the linear tendency is reached. In the case corresponding to the uniform starting configuration, the loss of convergence occurs for the values of $w_{\max}$ growing with a, whereas with the delta starting configuration and in this particular case, this monotonicity is lost for $a = 5.4275, a = 5.43$ and $a = 5.433$, which can be explained by the fact that with a delta starting configuration even these small variations of a can lead to different minima. In any case, there exist a short range of values of a for which the process goes from ε-convergence for any ε to divergence, independently of the random starting position. On the spheres displayed in the figure, we have

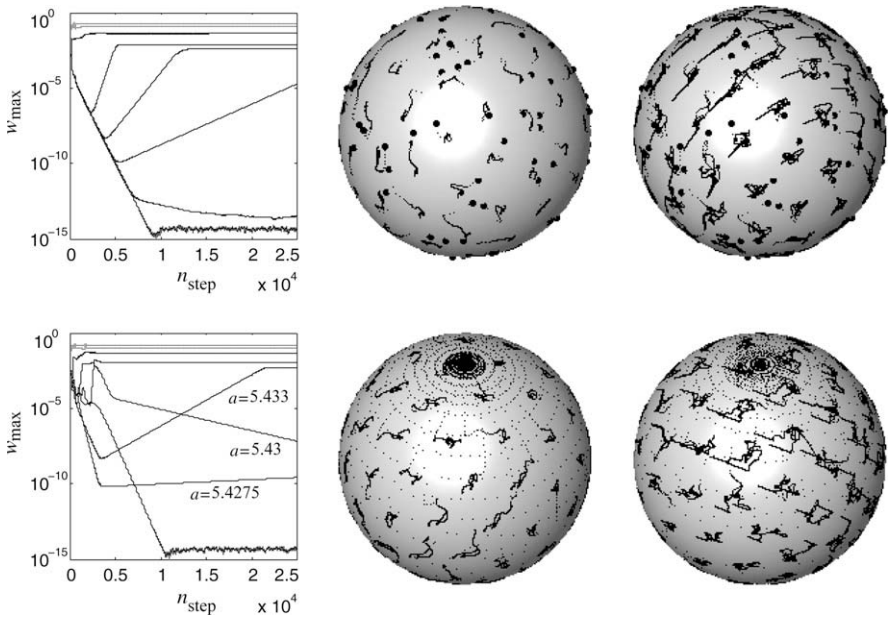


Fig. 6. The divergence process from uniform starting configurations (up) and from delta starting configurations (down). In the diagrams on the left, the values of a for each curve are, from bottom to up (the exceptions are indicated), $a = 5.4, a = 5.4275, a = 5.43, a = 5.433, a = 5.44, a = 5.7$ and $a = 8$, respectively. The spheres correspond to the cases $a = 5.7$ (center) and $a = 8$ (right).

included the paths described by the particles in the cases $a = 5.7$ (center) and $a = 8$ (right) from the uniform starting configuration (up) and from the delta starting configuration (down). The large points correspond to the starting position. As can be observed, the divergence process corresponds to a bifurcation of both the trajectories and the convergence curves. For high values of a, the process seems to lead to a sort of "rotation" of two similar alternating configurations with respect to the same axis.

3.2.3. Experiment 3

In this experiment, an analysis was carried out on average for different random starting positions with $N = 87$. The objective was to obtain information about the evolution with a of the average number of steps necessary to converge.

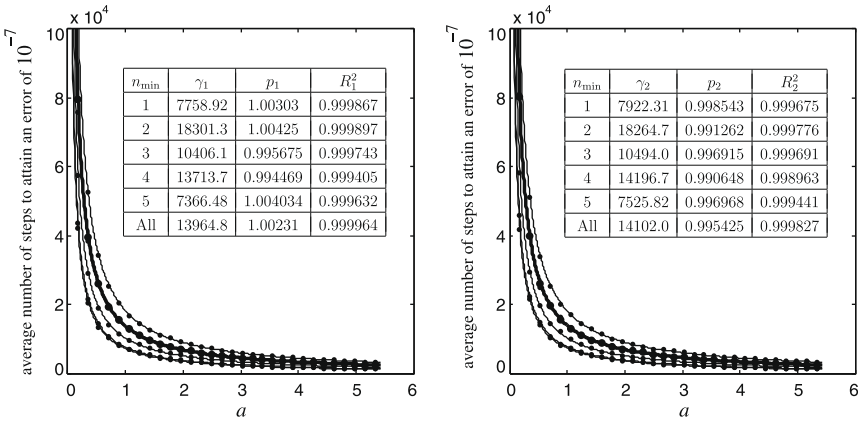
Fig. 7 summarizes the results of this test, in which 120,000 total runs of the algorithm were performed. These runs corresponded to 2000 uniform starting configurations (left) and 2000 delta starting configurations (right). For each one of these starting configurations, we considered 30 different values of a. For each kind of starting configurations, the figure displays the average number of steps necessary for ε-convergence ($\varepsilon = 10^{-7}$) computed from the 2000 total data for each value of a (large points), and the average number of steps for ε-convergence ($\varepsilon = 10^{-7}$) computed only from the data associated to the same minimum energy value for each value of a (small points). We include the information corresponding to the best five minima that we have found, whose energy values are -830.25191515, -830.25122722, -830.24870458, -830.24727726 and -830.24727422, respectively. Moreover, we display the regression curves of the form $\bar{n}_{\text{step}} = \frac{\gamma}{a^p}$ obtained for all the data and only for the data associated to each one of the five best minima. The included tables show the values of the regression parameters, γ, p, and the coefficient R^2 for each one of the five analyzed minima, which are indexed by $n_{\min}$, and for all the data. The indices 1 and 2 in γ, p, R^2 denote uniform and delta starting configurations, respectively. The results obtained indicate that there exists a strong independence between the average behavior of the algorithm and the method used to generate the random initial configurations. In addition, the values obtained for the regression parameters clearly confirm the intuitive result $p = 1$.

Fig. 8 shows the sample distribution of the probability of obtaining the different minima of the problem from random starting configurations. We specify the probability of obtaining the five best minima, which were determined from the 60,000 data corresponding to each kind of starting configurations (left and right, respectively). The separation between the data corresponding to the fourth and fifth minima has been marked with a vertical segment because the difference of their energies is invaluable. The figure also shows (in the small boxes) the probability of obtaining the five best minima, taking into account only the results corresponding to each a. These results confirm the independence of the probability distribution of the minima, the procedure used for generating the random initial positions and also the value of a. Note that there exist equilibrium configurations with extremely low probabilities. For instance, the seventh best minima obtained from uniform starting configurations appeared only once among the 120,000 total runs. The worst minima also have very small probabilities.

3.2.4. Experiment 4

In this experiment, 60,000 total runs were performed for $N = 200$. These runs corresponded to 2000 uniform starting positions and 30 different values of a (from now on all the starting configurations are uniform except for the robustness experiment performed for $N = 10^6$ – see Section 3.4).

Fig. 9 (left) shows the evolution with a of the average number of steps necessary for ε-convergence ($\varepsilon = 10^{-7}$) for this case. The corresponding interpolation curve is also included. The same figure (center) shows the sample probability distribution of the minima that were obtained from the 60,000 total runs. Note that the appearance of the probability distribution



$n_{\min}$	γ_1	p_1	R_1^2
1	7758.92	1.00303	0.999867
2	18301.3	1.00425	0.999897
3	10406.1	0.995675	0.999743
4	13713.7	0.994469	0.999405
5	7366.48	1.004034	0.999632
All	13964.8	1.00231	0.999964

$n_{\min}$	γ_2	p_2	R_2^2
1	7922.31	0.998543	0.999675
2	18264.7	0.991262	0.999776
3	10494.0	0.996915	0.999691
4	14196.7	0.990648	0.998963
5	7525.82	0.996968	0.999441
All	14102.0	0.995425	0.999827

Fig. 7. Evolution with a of the average number of steps for ε-convergence ($\varepsilon = 10^{-7}$) from uniform starting configurations (left) and delta starting configurations (right). The average data corresponding to the best five minima and their interpolation curves are also included.

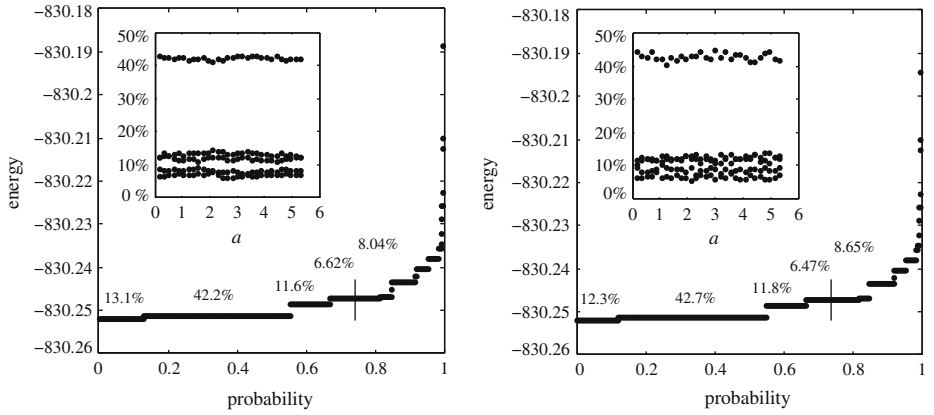


Fig. 8. Sample distribution of the probability of obtaining the different minima of the case $N = 87$ from uniform starting configurations (left) and from delta starting configurations (right).

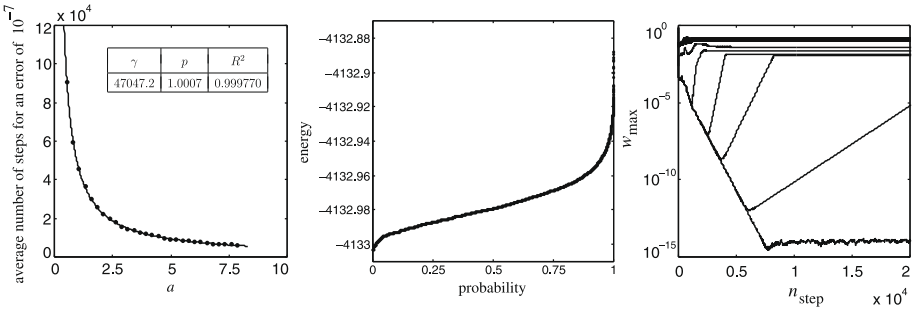


Fig. 9. Evolution with a of the average cost, sample probability distribution of the minima and divergence for the case $N = 200$. In the diagram on the right, the values of a for each curve are, from bottom to up, $a = 8.25, a = 8.27, a = 8.28, a = 8.3, a = 8.35, a = 8.8$ and $a = 12$, respectively.

corresponds to a continuous random variable rather than a discrete one, whereas the probability distribution for the case $N = 87$ has the typical aspect of a discrete random variable distribution. Fig. 9 (right) also shows the divergence process and includes the curves corresponding to $a = 8.25, a = 8.27, a = 8.28, a = 8.3, a = 8.35, a = 8.8$ and $a = 12$.

3.2.5. Experiment 5

The previous experiments suggest that if the scheme $\hat{x}^{k+1} = x^k + a \min_{1\leqslant i<j\leqslant N}\{|x_i^k - x_j^k|\}w^k, x_i^{k+1} = \frac{\hat{x}_i^{k+1}}{|\hat{x}_i^{k+1}|}, i = 1,\ldots,N$, is adopted as the descent algorithm and the coefficient a is kept constant throughout the optimization process, then the average number of steps necessary for ε-convergence is inversely proportional to a until a certain value a_{crit}, beyond which the scheme diverges. Moreover, all this process is practically independent of the initial configuration. Also taking into account that the probability distribution of the minima obtained from random starting configurations is independent of a, it is clear that the optimum efficiency of the algorithm is obtained when for each N a value a^* close to a_{crit} from the left is chosen for a.

The objective of this experiment was to determine the way the coefficient a_{crit} varies with N. We obtained an accurate estimation of a_{crit} for each $N = 10, 20, \ldots, 100, 200, \ldots, 1000, 1250, \ldots, 2000, 2500, 3000$ just by applying the bisection method after determining by inspection a value of a that leads to convergence, and another that leads to divergence from a given (any) starting configuration. Fig. 10 shows the results of the experiment and the interpolation of the obtained data. Taking into account this information, we establish the formula $a^* = 0.545\sqrt{N}$ for the logarithmic energy on the 2-sphere.

In a small complementary experiment, we computed the value of a_{crit} corresponding to 10 different starting positions for each $N = 50, 100, 200$. In all the cases we used the bisection method until attaining an error smaller than $10^{-4}a_{\text{crit}}$. For $N = 50$, the mean of the 10 obtained values was 4.084 and the corresponding standard deviation was 0.0079, which gives a quotient of about 0.002. For $N = 100, 200$, the mean and the standard deviation were 5.757, 0.065 and 8.269, 0.075, respectively. In both cases, the quotient is around 0.01. According to the curves displayed in Fig. 10, $a^* \simeq 0.9a_{\text{crit}}$ for $N \leqslant 10^6$. Our experience indicates that this criterion is sufficiently conservative. In fact, the formula $a^* = 0.545\sqrt{N}$ has been used in about $6 \cdot 10^7$ runs of the algorithm for different N from $N = 87$ to $N = 10^6$ in the case of the logarithmic energy on the 2-sphere (see Section 3.4 and [2]). In particular, more than $3 \cdot 10^7$ runs were performed for $N = 500$ and more than 10^7 runs were carried out for $N = 600$. The Forces Method never diverged, which can be understood as a sort of "convergence proof" for our algorithm.

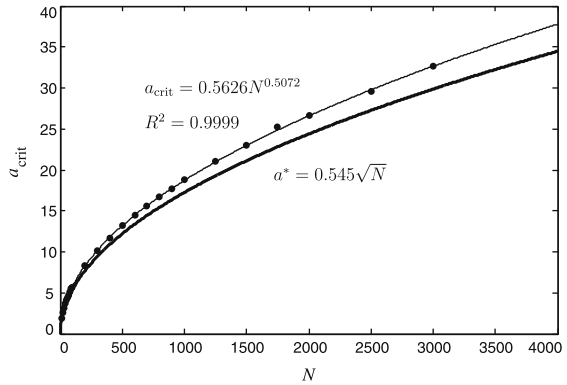


Fig. 10. Evolution with N of a_{crit} and a^*.

3.3. The cost of a local minimum

It is clear that the cost at each step of the Forces Method is always the same, and when the logarithmic energy is considered it essentially corresponds to the computation of the $\binom{N}{2}$ vectors $\frac{x_i - x_j}{|x_i - x_j|^2}, 1 \leqslant i < j \leqslant N$, which in cartesian coordinates requires only elementary operations. Hence, the computational cost of the identification of a local minimum depends only on the number of steps necessary to converge.

We introduced in Section 3.1 the random variables X, the number of steps for *nr*-convergence in a run, and Y, the number of steps for ε-convergence in a run. For the statistical analysis of these random variables, we used the sample data provided by $n_{\text{sp}} = 5000$ runs of the Forces Method for each $N = 500, 1000, \ldots, 3000$ and $n_{\text{sp}} = 1000$ runs for each $N = 4000, 5000$. The results corresponding to $N = 4000, 5000$ were used to confirm the tendencies given by the rest of the data. In all the runs the ε-convergence with $\varepsilon = 10^{-8}$ was attained.

The average values μ_X and μ_Y are the main parameters for the analysis of the cost. In Fig. 11, the way the sample means $\bar{x}, \bar{y}$ stabilize with n_{sp} can be observed. Specifically, we show the evolution with n_{sp} of the sample means associated to *nr*-convergence (left) and to ε-convergence for $\varepsilon = 10^{-6}$ (center), and $\varepsilon = 10^{-8}$ (right) for all the considered N.

Fig. 12 (left) shows the average number of steps necessary to attain *nr*-convergence and ε-convergence for $\varepsilon = 5 \cdot 10^{-5}, 2 \cdot 10^{-5}, 10^{-5}, \ldots, 10^{-8}$ obtained from the above-described sample data. We also display the regression curves of the form $\mu = \gamma N^p$ obtained from the 5000 data corresponding to each $N = 500, 1000, \ldots, 3000$ (big points). The table on the right contains the regression data, γ, p, R^2, of these curves and the intersection point $N_\perp$ of the non-return average cost curve (thick curve) and each ε-convergence average cost curve. In any case, it can be assumed that the choice $\varepsilon = 10^{-8}$ guarantees that the information contained in Fig. 12 about the *nr*-convergence is reliable (in [2] we show that only a small fraction of the runs are not in linear tendency when the ε-convergence with $\varepsilon = 10^{-8}$ is attained).

The curves of ε-convergence in Fig. 12 are useful in practice to estimate calculation times, but from a more conceptual point of view the information about the *nr*-convergence is crucial; the intrinsic average complexity of obtaining an approximate local minimum, which is independent of ε, is approximately $O(N^{2.77})$. It must be taken into account that entering into the neighborhood of a local minimum has been identified with the bottleneck of the Fekete problem; see, for instance, [3,10]. We must also consider that the algorithms with a high convergence ratio, whose prototype is Newton's algorithm, diverge in general if the initial estimation is not good enough. The presented results imply that the average computational cost of the identification of an approximate local minimum is asymptotically negligible in comparison with the application of a qua-

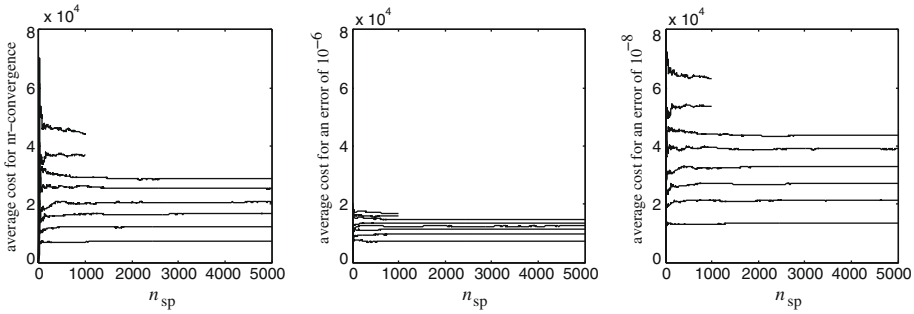
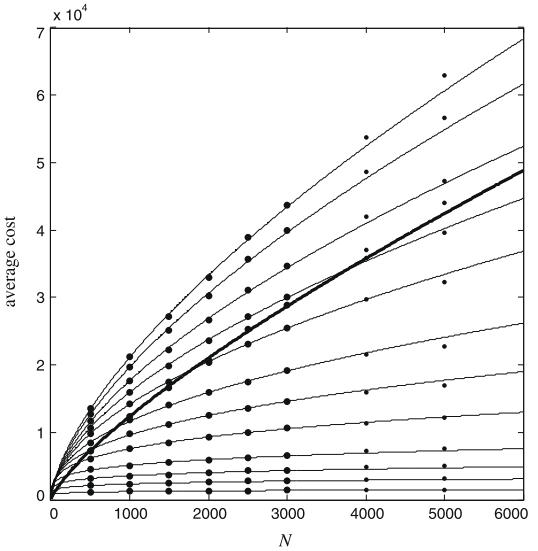


Fig. 11. Evolution with n_{sp} of the average cost for *nr*-convergence (left) and for ε-convergence for $\varepsilon = 10^{-6}$ (center) and $\varepsilon = 10^{-8}$ (right).



ε	γ	p	R^2	$N_\perp$
$5\cdot10^{-5}$	543.64	0.1161	0.9892	28
$2\cdot10^{-5}$	818.92	0.1521	0.966	67
$1\cdot10^{-5}$	1043.4	0.1774	0.9732	121
$5\cdot10^{-6}$	1119.0	0.2197	0.9936	200
$2\cdot10^{-6}$	932.67	0.3026	0.9979	347
$1\cdot10^{-6}$	684.91	0.3819	0.9987	520
$5\cdot10^{-7}$	505.6	0.4536	0.9992	824
$2\cdot10^{-7}$	351.02	0.5349	0.9997	1794
$1\cdot10^{-7}$	297.33	0.5761	0.9996	3776
$5\cdot10^{-8}$	264.28	0.608	0.9996	9353
$2\cdot10^{-8}$	240.28	0.6376	0.9994	35989
$1\cdot10^{-8}$	229.7	0.6546	0.9994	116855
nr	61.306	0.7678	0.9995	

Fig. 12. Average cost of the process of identifying a local minimum.

dratic method, because the Hessian matrix of the potential energy is full and each step of Newton's method has, apart from eventual conditioning problems, a cost $O(N^3)$.

Variables X and Y have a considerable dispersion. Fig. 13, for instance, shows the sample probability distribution of variable Y for $N = 500$ and $\varepsilon = 10^{-5}$ (left) and that of variable X for $N = 3000$ (right). As can be seen, some starting configurations lead to convergence curves much larger than the average ones.

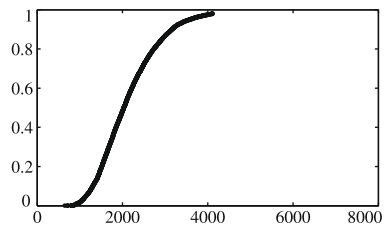
Fig. 14 shows the evolution with N of the sample deviation of the number of steps necessary to attain nr-convergence and ε-convergence for $\varepsilon = 2\cdot10^{-6}, 10^{-6}, 5\cdot10^{-7}, \ldots, 10^{-8}$. As for the study of the means, we also display regression curves of the form $\mu = \gamma N^p$. The table on the right contains the regression data, γ, p, R^2, of these curves. Naturally, the standard deviations of X, Y grow with N, and, moreover, that of Y grows with ε.

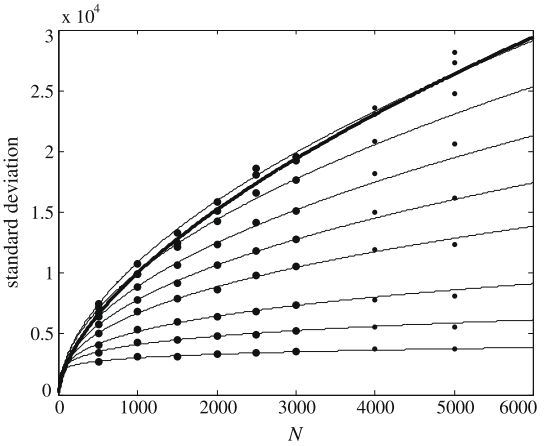
Let us consider now the random variables $A = \log X$ and $B = \log Y$, and their corresponding standardized variables $\widetilde{A} = \frac{A-\mu_A}{\sigma_A}, \widetilde{B} = \frac{B-\mu_B}{\sigma_B}$. Fig. 15 displays the sample probability distributions of $\widetilde{A}$ for $N = 1000, 5000$, and that of $\widetilde{B}$ for each pair $(N,\varepsilon) = (500, 10^{-7}), (1500, 10^{-8}), (3000, 10^{-5})$. The figure also includes the probability distribution of a standard gaussian variable (thick line).

It seems clear that for the purposes of this subsection, variables X, Y can be considered to be sufficiently well approximated by log-normal variables. This allows us to give a complete characterization of the cost of a local minimum by means of simple tools. So, for instance, $P[Y \leqslant y] \simeq \frac{1}{2\pi}\int_{-\infty}^{\frac{\log y-\mu_B}{\sigma_B}} e^{-\frac{s^2}{2}} ds$, and analogously for X. For each pair (N,ε) good approximations for $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ can be obtained from Figs. 12 and 14, and then the well-known expressions for the mean and the standard deviation of a log-normal variable, $\mu_B = \log\frac{\mu_Y^2}{\sqrt{\mu_Y^2+\sigma_Y^2}}, \sigma_B = \sqrt{\log\frac{\mu_Y^2+\sigma_Y^2}{\mu_Y^2}}$, can be used in the above-mentioned expression for $P[Y \leqslant y]$.

3.4. Large-scale experiments: The FinisTerrae challenge

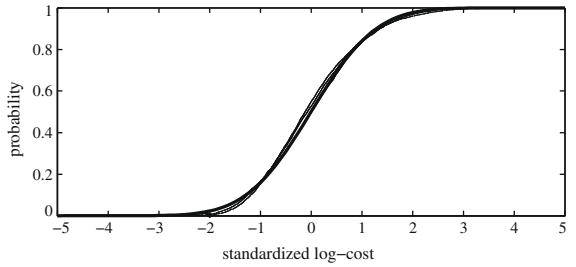
We used CESGA's supercomputer FinisTerrae to carry out some large-scale experiments on the Fekete point problem. Around 350,000 h of computation were spent in this computational challenge. The main objectives were to confirm the pre-





ε	γ	p	R^2
$2 \cdot 10^{-6}$	1174.5	0.1372	0.9892
$1 \cdot 10^{-6}$	882.93	0.2231	0.9811
$5 \cdot 10^{-7}$	592.05	03144	0.9912
$2 \cdot 10^{-7}$	407.73	0.4052	0.9975
$1 \cdot 10^{-7}$	354.15	0.4477	0.9991
$5 \cdot 10^{-8}$	308.74	0.4866	0.9991
$2 \cdot 10^{-8}$	287.64	0.5147	0.9972
$1 \cdot 10^{-8}$	245.59	0.5489	0.9971
nr	158.05	0.6007	0.9968

Fig. 14. Standard deviation of variables X, Y.



sented results about convergence, robustness and computational cost of the Forces Method for large N (190,000 h), and to obtain sample information about Smale's 7th problem (160,000 h). In [2], we give more information about this challenge and describe the data obtained for Smale's 7th problem. Regarding the behavior of the Forces Method for large N, we performed the following experiments:

- For $N = 10,000$, a total of 1000 runs attaining ε-convergence with $\varepsilon = 10^{-9}$ from uniform starting configurations.
- For $N = 20,000$, a total of 100 runs attaining ε-convergence with $\varepsilon = 5 \cdot 10^{-10}$ from uniform starting configurations.
- For $N = 50,000$, a total of 10 runs attaining ε-convergence with $\varepsilon = 10^{-10}$ from uniform starting configurations.
- For $N = 10^6$, a total of 3000 steps from a delta starting configuration.

In all the cases, we considered $a = a^*$ and all the runs converged. The runs for N = 10,000, 20,000, 50,000 allow us, in particular, to check the validity of the interpolation curves for the average cost presented in Fig. 12. These curves were used to predict calculation times and to design the assignation of the available resources during the execution of the challenge. Fig. 16 is an enlargement of Fig. 12 and includes the sample information obtained for N = 10,000, 20,000, 50,000. For the case N = 10,000, we have 1000 data and the sample average costs are already close to the corresponding interpolation curves. For N = 20,000, 50,000, there are still little data (remember the high dispersion of the cost shown in Section 3.3), but their sample information is in reasonable agreement with the interpolation curves, especially that corresponding to the nr-convergence. These experiments confirm that for a moderately high amount of runs (say n_{sp} in the hundreds), the interpolation curves presented in Fig. 12 can be used to obtain accurate estimations of the total computation time.

Figs. 17 and 18 show two convergence curves (left) and the triangulation of the corresponding final positions (right) for runs with N = 20,000 and N = 50,000, respectively. The nonlinear phase and the final linear tendency can be clearly appreciated in both cases. When N grows the probability of that the maximum disequilibrium degree reaches really low values before the non-return point is attained increases. The configurations corresponding to these low disequilibrium degree values are not approximate local minima (they are probably close to saddle points). This must be taken into account in practice for designing the stopping criteria of the descent algorithms in order to obtain reliable results.

The experiment for $N = 10^6$ was designed as a *tour de force* both for the Forces Method and for FinisTerrae. It consisted of reproducing with a million points the results shown in Fig. 5 for the case $N = 87$. So, the position of 10^6 points was randomly generated according to a uniform probability density on a spherical cap of area $\pi \cdot 10^{-6}$ (a disk of radius about 10^{-3} with center at the North pole of the unit 2-sphere), and the Forces Method was applied from this starting position. The value

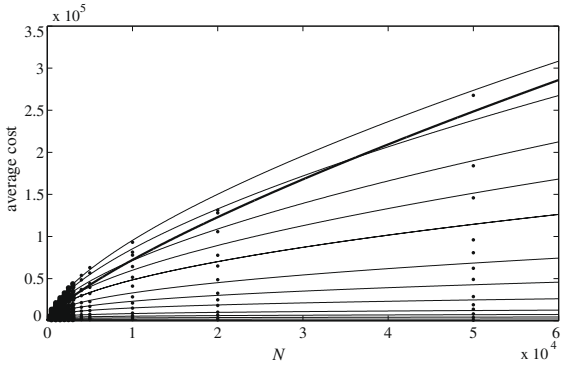


Fig. 16. Average cost of the process of identifying a local minimum.

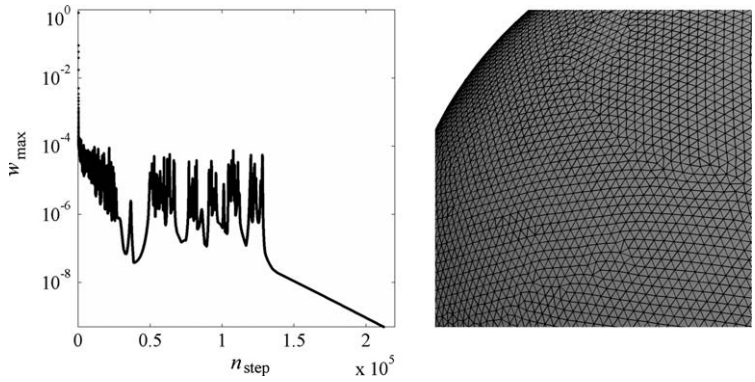


Fig. 17. Convergence curve and final position for a run with $N = 20,000$.

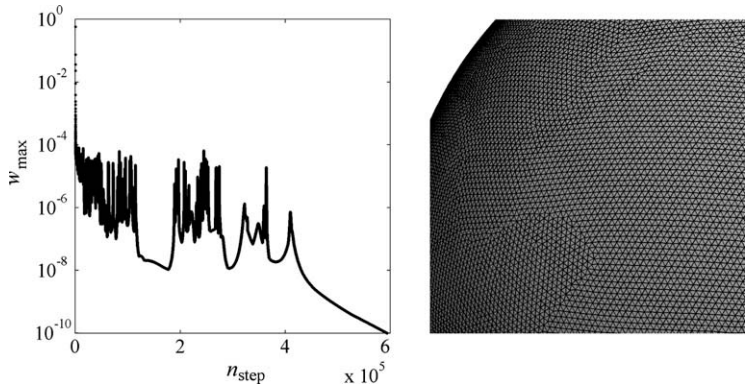


Fig. 18. Convergence curve and final position for a run with $N = 50,000$.

of the coefficient a was computed again with the formula $a = a^* = 0.545\sqrt{10^6} = 545$. Note that with the chosen starting position, $w_{\max} \simeq 1$ during the first steps. This implies that some particles jump around 545 times the minimum distance between particles at each step when the convergence process starts. From Fig. 12, it is found that the average number of steps necessary to attain nr-convergence for $N = 10^6$ is about $2.5 \cdot 10^6$, which represents years of real computation time even with all the about 2500 CPUs of FinisTerrae working in parallel. The objective of this experiment was not to obtain an approximate local minimum for the case $N = 10^6$, but to check the robustness of the Forces Method and the validity of the formula established for the determination of the coefficient a under really hard conditions. The iterative process was stopped at $n_{\text{step}} = 3000$. This was enough to observe the way the Forces Method was able to deal with the extreme starting position described above. For this computation, 1024 CPUs of FinisTerrae worked in parallel during around one day and a half

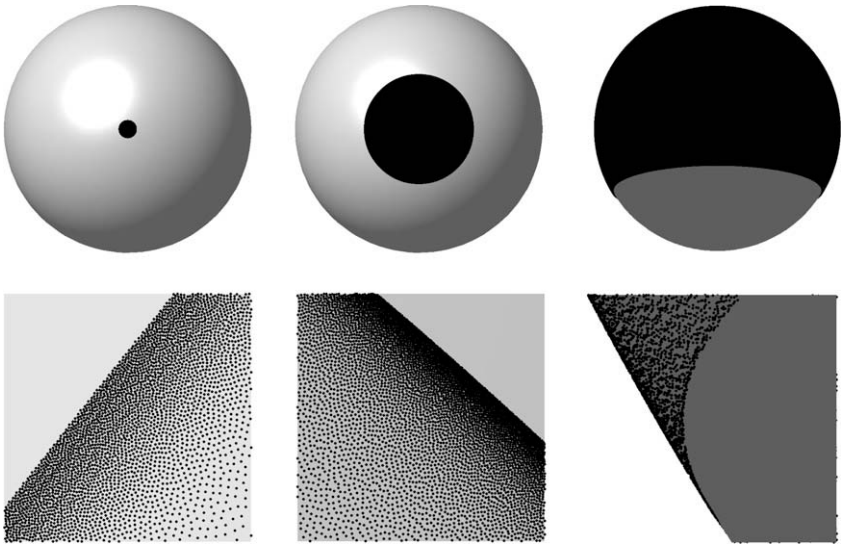


Fig. 19. The steps $n_{\text{step}} = 100$ (left), $n_{\text{step}} = 200$ (center) and $n_{\text{step}} = 220$ (right) for the experiment with $N = 10^6$.

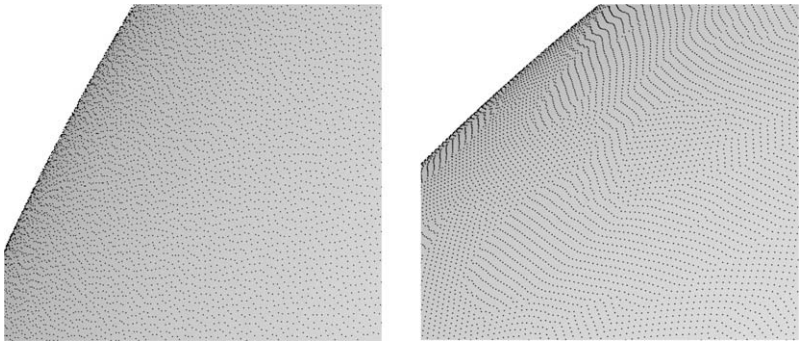


Fig. 20. The steps $n_{\text{step}} = 230$ (left) and $n_{\text{step}} = 3000$ (right) for the experiment with $N = 10^6$.

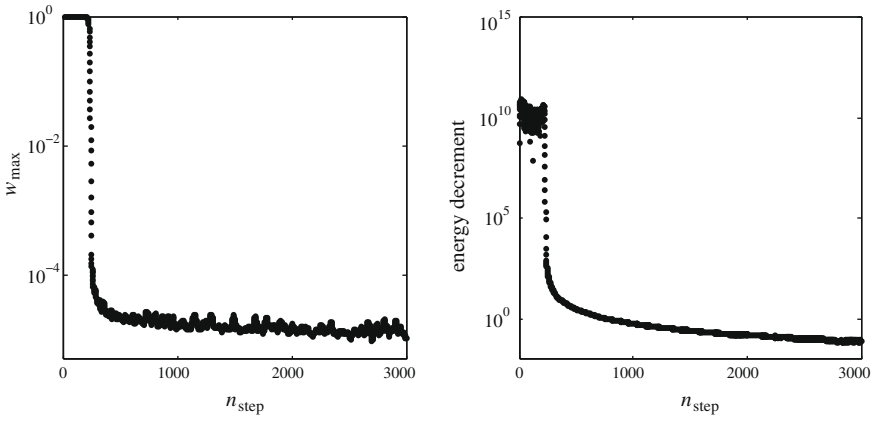


Fig. 21. Convergence curve and the energy decrements at each step for the experiment with $N = 10^6$.

of real time, which represents approximately 40,000 h of computation. Figs. 19 and 20 show the positions of the points at steps $n_{\text{step}} = 100, 200, 220, 230, 3000$, respectively. The behavior of the algorithm is entirely analogous to that observed

for $N = 87$. In particular, Fig. 19 (bottom left) shows that only 100 steps were sufficient to "put in order" the particles. Between the steps $n_{step} = 200$ and $n_{step} = 230$, the algorithm expands the spherical cap containing all the particles until all the points are "evenly distributed" on the surface of the whole sphere. At the step $n_{step} = 3000$, the achieved configuration is far from being an approximate local minimum. However, its regularity is really notable. Fig. 21 shows the convergence curve (left) and the evolution with n_{step} of the energy decrement at each step. As can be seen, the energy systematically descended.

4. The energy

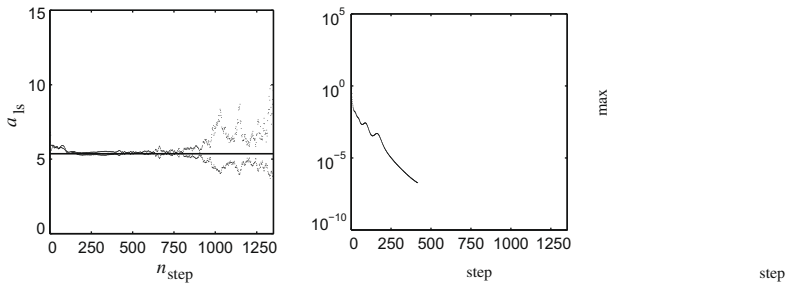
The Forces Method, which has been designed to minimize a potential energy, can be run with guarantee of convergence without the necessity of evaluating this energy throughout the optimization process. The energy value of an approximate local minimum can be obtained *a posteriori* as a byproduct of its calculation. Nevertheless, it is possible to check that the energy descents at each step of the convergence process. In this section we focus on the analysis of the behavior of the energy throughout the convergence process.

We have compared our criteria for the choice of the coefficient a with the classical line search procedure, which determines the step size from the minimization of the objective function in the advance direction. For this experiment we have carried out some calculations by means of the line search procedure following the advance direction given by the disequilibrium degree (in the case of the sphere this direction practically equals the gradient direction at most of the steps). The computation of the coefficient a that minimizes the energy at each step has been made with great accuracy in order to evaluate to what extent the line search procedure can improve the optimization process (as for the number of steps). To be exact, if we call a_{ls} the magnitude of the coefficient a that minimizes the energy in the direction φw, we have required that at each step the absolute error in a_{ls} is smaller than $10^{-16} a_{crit}$. By using this procedure in the case $N = 87$ ($a_{crit} \simeq 5.37$), we obtained that after 500 runs corresponding to different starting configurations the ε-convergence ($\varepsilon = 10^{-6}$) was attained after 1884 steps on average, whereas the Forces Method reached the same accuracy level after 2009 steps on average, which represents approximately only a 6% increment in the number of steps. However, it is clear that the cost at each step is much higher with the line search procedure, which requires several energy evaluations per step, than with the Forces Method, in which the step size is analytically obtained and the evaluation of the energy is not necessary.

It is even more interesting to study the values that a_{ls} takes throughout the descent process. Fig. 22 shows the evolution with n_{step} of a_{ls} (left), the energy decrement at each step (center), and w_{max} (right) for a calculation with line search for $N = 87$ from a random starting configuration. In the diagram on the left, we also display the horizontal straight line corresponding to the value 5.37. We do not include in this figure the value $a_{ls} = 33.6570189174919$, which corresponds to $n_{step} = 1$. The value corresponding to $n_{step} = 2$ is $a_{ls} = 11.1625964050175$, but it cannot be distinguished in the figure because it is an isolated point placed practically on the vertical axis. As can be observed, the value of a_{ls} oscillates around a_{crit} throughout all the convergence process, and the amplitude of the oscillation grows with the proximity to the minimum. This oscillation can also be observed in the disequilibrium degree, but not in the energy decrements, which describe a smooth curve. So the coefficient a_{crit} seems to be a sort of "average" of a_{ls}. Moreover, regarding the average number of steps necessary to convergence, the choice $a \simeq a_{crit}$ is practically equivalent to take $a = a_{ls}$, but a_{crit} is much cheaper at each step than a_{ls} and, in addition, $a \simeq a_{crit}$ guarantees a smooth variation of all the convergence control parameters, w_{max} in particular.

The behavior indicated in Fig. 22 is also observed when working with other kernels. Fig. 23 shows the evolution with n_{step} of a_{ls} for the cases $s = 1$ with $N = 100$, $s = 2$ with $N = 200$ and $s = 3$ with $N = 300$, where s denotes the power of Riesz's kernel. We also include the horizontal straight lines associated to the values that a_{crit} takes in these three cases ($a_{crit} \simeq 1.17$, $a_{crit} \simeq 0.298$ and $a_{crit} \simeq 0.098$, respectively). These estimations of a_{crit} have been obtained by following the same procedure as that used for the preparation of Fig. 10; that is, by using the bisection method from a value of a that converges and another that diverges for each kernel and for each N.

Fig. 24 shows the evolution with N of a_{crit} for Riesz's kernels with $s = 1, 2, 3$. For the case of the Newtonian kernel, the coefficient a_{crit} remains practically constant except for the first values of N, whereas for $s = 2, 3$ the evolution of a_{crit} fits the curve given by a negative power of N. The figure includes interpolation curves with their R^2 values for both cases.



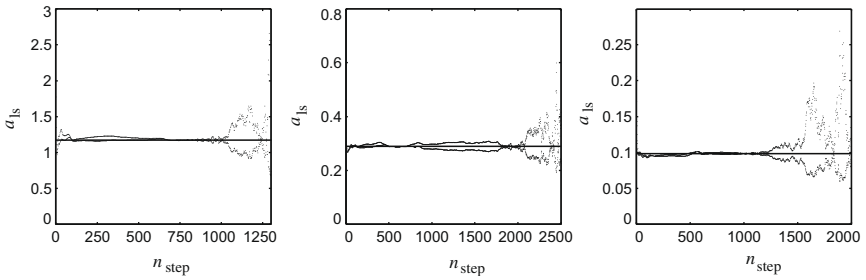


Fig. 23. Evolution with n_{step} of a_{ls} for the cases $s = 1$ with $N = 100, s = 2$ with $N = 200$ and $s = 3$ with $N = 300$.

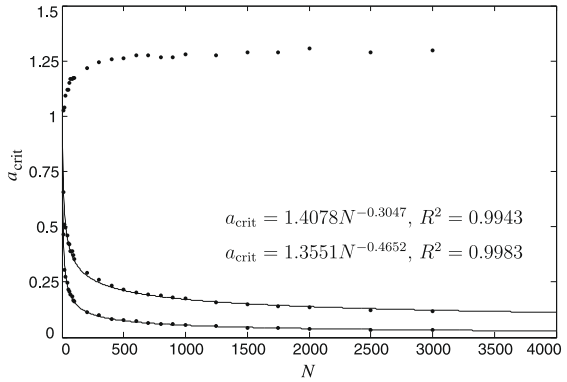


Fig. 24. Evolution with N of a_{crit} for Riesz's kernels with $s = 1, 2, 3$.

5. Interpretation of the method

The ideas behind the first versions of the Forces Method came from Physics. In a first interpretation, the Forces Method searches for a minimum of the energy by numerically integrating the quasi-static trajectories described by the particles in their movement towards the equilibrium, see [1]. Other authors have applied this mechanical analogy to identify local minima for the Fekete problem; see, for instance, [32]. In this case the integrated trajectories are not quasi-static and the authors must choose a value for the adhesion constant. This implies in particular that the potential energy does not necessarily descend at each step.

From the point of view of the resolution of an ODE system, the coefficient a is bounded by a critical upper value related to the stability of Euler's forward scheme. Guaranteeing this stability in a highly nonlinear problem such as the Fekete problem requires extremely low values of a, which implies prohibitive calculation times. However, for the optimization problem the scheme stability is not really important if the algorithm leads to a minimum of the potential energy, even when this minimum is not the one associated to the starting configuration in an initial value problem. Taking this into account, we have chosen the coefficient a according to the following criteria: its calculation must be as cheap as possible at each step (constant) and, under this condition, it must lead to convergence (that is, to an approximate local minimum) in the fewer number of steps possible.

This can be regarded as a change in priorities. We are forgetting the accurate integration of the trajectories in benefit of the efficiency of the optimization process. Nevertheless, we want to keep the versatility and the simplicity of Euler's forward method as well as to preserve some features of the mechanical conception of the problem. In particular, the determination of the advance direction and the step size for the Forces Method is completely independent of the parametrization chosen for the surface S, which is not guaranteed by classical optimization methods such as Gradient, Conjugate Gradient, Newton or quasi-Newton, and the advance direction w guarantees that each particle contributes to the energy descent. Moreover, the factor φ in the step size reminds one clearly the mechanical analogy.

Experiments 1–5 were designed according to those principles, and with the objective of determining the values of a that maximize the efficiency while keeping the convergence to a local minimum. A surprising result was the practically total independence between the limit value a_{crit} and the position of the particles, even when really hard positions such as the delta starting configurations were considered. One can expect that for each local minimum there exists an upper bound for the value of a that guarantees the convergence of the Forces Method's scheme when it has entered into the influence zone of this minimum. This upper value should be closely related with the size of the attraction area of the minimum, and hence also with the convergence radius for Newton's method. However, there are no *a priori* reasons for expecting this upper bound

to be almost the same for all the minima corresponding to each N, nor that it leads to convergence from any starting position and not only from a configuration close enough to a local minimum.

After performing these experiments, we saw that a step size guaranteeing the convergence could be obtained without the necessity of evaluating the potential energy, but evaluating only its derivatives (the forces). Many authors use relaxation algorithms in which the step size is fixed tentatively, but the proposed procedures are always associated to the evaluation of the objective function with a certain frequency throughout the optimization process. The most expensive of these procedures is the so-called line-search technique, which guarantees that the step size leads to the maximum descent of the objective function in the advance direction at each step. This prompted us to perform the experiment described in Section 4, which gave another surprising result: a_{ls} oscillates around a_{crit} throughout all the descent process.

Thus, although the conception of the Forces Method as the integration scheme of an EDO system corresponds to a mechanical analogy for the Fekete problem, it cannot be said by any means that this method integrates an initial value problem, because the work values of the coefficient a, a^*, are far higher than the upper bound of Euler's forward scheme's stability. On the other hand, the definition of a_{crit} is exclusively associated to the energy gradient and not to the energy itself, in such a way that the Forces Method can be run without evaluating the objective function throughout the descent process. This behavior is curious for an optimization algorithm with the classical advance direction–step size structure applied to a highly nonlinear problem. The particular relation between a_{ls} and a_{crit} places the Forces Method on the boundary between the two usual conceptions of the Fekete problem as an optimization problem and a mechanical problem.

For more details about the results presented here, see the Dissertation of J.M. Gesto [11].

Acknowledgments

The authors express their sincere gratitude to Professor Antonio Gens for his unconditional support. We are also pleased to thank Professors Agustín Medina and Juan Sánchez for their kind comments and the research group La CáN from UPC, in whose cluster "Clonetroop" some of the intensive calculations performed for this work have been carried out.

In February, 2008, we had the opportunity of using one of the biggest supercomputers in the world, the FinisTerrae. For this, and for the excellent relation that we have always had with CESGA's team, we are grateful to all of them.

This work has been partially supported by the CICYT under project MTM2007-62551, by the i-MATH project and by CESGA.

References

[1] E. Bendito, A. Carmona, A.M. Encinas, J.M. Gesto, Estimation of Fekete points, J. Comput. Phys. 225 (2007) 2354–2376.
[2] E. Bendito, A. Carmona, A.M. Encinas, J.M. Gesto, Computational cost of the Fekete problem II: on Smale's 7th problem, preprint, accessible in <http://www-ma3.upc.edu/users/bencar/papers.html>.
[3] M.J. Bowick, <http://phy.syr.edu/condensedmatter/thomson/thomsonapplet.htm>.
[4] S.B. Damelin, A walk through energy, discrepancy, numerical integration and group invariant measures on measurable sets of Euclidean space, Numer. Algorithms 48 (1–3) (2008) 213–235.
[5] S.B. Damelin, P. Grabner, Energy functionals numerical integration and asymptotic equidistribution on the sphere, J. Complex. 19 (3) (2003) 231–246 (Postscript) Corrigendum, J. Complex. 20(6) (2004) 883–884.
[6] S.B. Damelin, V. Maymeskul, On point energies separation radius and mesh norm for s-extremal configurations on compact sets in $\mathbb{R}^n$, J. Complex. 21 (6) (2006) 863–884.
[7] S.B. Damelin, V. Maymeskul, Minimal discrete energy problems and numerical integration on compact sets in Euclidean spaces, in: A. Iske, J. Levesley (Eds.), Algorithms for Approximation, Springer-Verlag, Heidelberg, 2007, pp. 369–378.
[8] T. Erber, G.M. Hockney, Comment on "Method of constrained global optimization", Phys. Rev. Lett. 74 (8) (1995) 1482.
[9] M. Fekete, Über die Verteilung der Wurzeln bei gewissen algebraischen Gleischungen mit ganzzahligen Koeffizienten, Math. Zeitschr. 17 (1923) 228–249.
[10] J. Fliege, U. Maier, The distribution of points on the sphere and corresponding cubature formulae, IMA J. Numer. Anal. 19 (1999) 317–334.
[11] J.M. Gesto, Estimation of Fekete points, Dissertation, Departament de Matemàtica Aplicada III, Universitat Politècnica de Catalunya, 2008, accessible in <http://www-ma3.upc.edu/users/bencar/personnel.html>.
[12] L. Giomi, M. Bowick, Crystalline order on Riemannian manifolds with variable Gaussian curvature and boundary, Phys. Rev. B 76 (2007) 054106.
[13] E. Hairer, G. Wanner, RADAU5, version July 1996, available at URL <ftp://ftp.unige.ch/pub/doc/math/stiff/radau5>.
[14] D.P. Hardin, E.B. Saff, Discretizing manifolds via minimum energy points, Notices Am. Math. Soc. 51 (2004) 1186–1194.
[15] D.P. Hardin, E.B. Saff, Minimal Riesz energy point configurations for rectificable d-dimensional manifolds, Adv. Math. 193 (2005) 174–204.
[16] J.S. Hesthaven, From electrostatics to almost optimal nodes sets for polynomial interpolation in a simplex, SIAM J. Numer. Anal. 35 (1998) 655–676.
[17] J.B. Hiriart-Urruty, A new series of conjectures and open questions in optimization and matrix analysis, doi:10.1051/cocv:2008040, in: ESAIM: Control, Optimisation and Calculus of Variations, in press, accessible in <http://www.mip.ups-tlse.fr/publis/files/07.21.pdf>.
[18] J. Korevaar, M.A. Monterie, Approximation of the equilibrium distribution by distributions of equal point charges with minimal energy, Trans. Am. Math. Soc. 350 (1998) 2329–2348.
[19] R. van Liere, J. Mulder, J. Frank, J. de Swart, Virtual Fekete point configurations: a case study in perturbing complex systems, 1999, accessible in <http://ieeexplore.ieee.org/iel5/6781/18163/00840498.pdf>.
[20] J.M. Morris, D.M. Deaven, K.M. Ho, Genetic-algorithm energy minimization for point charges on the sphere, Commun. Pure Appl. Math. 39 (1986) 149–186.
[21] K.J. Nurmela, Constructing spherical codes by global optimization methods, Helsinki University of Technology Series A: Research Reports, 32, 1995.
[22] K.J. Nurmela, P.R.J. Östergard, Packing up to 50 equal circles in a square, Discrete Comput. Geom. 18 (1997) 111–120.
[23] K.J. Nurmela, P.R.J. Östergard, Covering t-sets with $(t+2)$-sets, Discrete Appl. Math. 95 (1999) 425–437.
[24] K.J. Nurmela, Upper bounds for covering arrays by tabu search, Discrete Appl. Math. 138 (2004) 143–152.
[25] P.M. Pardalos, An open global optimization problem on the unit sphere, J. Global Optim. 6 (1995) 213.
[26] P.O. Persson, G. Strang, A simple mesh generator in MATLAB, SIAM Rev. 46 (2) (2004) 329–345.
[27] J.M. Sanz-Serna, M.P. Calvo, Numerical Hamiltonian Problems, Chapman & Hall, 1994.

[28] K. Shimada, D.C. Gossard, Automatic triangular mesh generation of trimmed parametric surfaces for finite element analysis, Computer Aided Geom. Design 15 (1998) 199–222.
[29] S. Smale, Mathematical problems for the next century, Math. Intell. 20 (1998) 7–15.
[30] S. Smale, Complexity theory and numerical analysis, Acta Numer. 6 (1997) 523–551.
[31] N. Sloane, <http://www.research.att.com/viewPage.cfm?pageID=171>.
[32] W.J.H. Stortelder, J.J.B. de Swart, J.D. Pintér, Finding elliptic Fekete points sets: two numerical approaches, J. Comput. Appl. Math. 130 (2001) 205–216.
[33] A. Witkin, P. Heckbert, Using particles to sample and control implicit surfaces, in: Computer Graphics Proc. SIGGRAPH94, 1994, pp. 269–278.
[34] R. Womersley, <http://web.maths.unsw.edu.au/~rsw/Sphere/Energy/index.html>.
[35] R. Womersley, <http://web.maths.unsw.edu.au/~rsw/Torus/>.
[36] Y. Zhou, Arrangements of points on the sphere, Dissertation. Department of Mathematics, University of South Florida, 1995.