Computational Cost of the Fekete Problem

E. Bendito, A. Carmona, A. M. Encinas and J. M. Gesto

e-mail: jose.manuel.gesto@upc.edu

Departament de Matemàtica Aplicada III

Universitat Politècnica de Catalunya. España

Abstract

We present here strong numerical and statistical evidence of the fact that the Smale's 7th problem can be answered affirmatively. In particular, we show that a local minimum for the logarithmic potential energy in the 2-sphere satisfying the Smale's conditions can be identified with a computational cost of approximately $O(N^{10})$.

1 Introduction

In a wide sense, the Nth order Fekete points of a compact subset $S \subset \mathbb{R}^n$ can be defined as the N point sets ω_N that minimize in S a potential energy functional of the form

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

where $x = \{x_1, \ldots, x_N\}$, $x_i \in \mathbb{R}^n$, 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, $-\ln|x_i - x_j|$, and the 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 $s = n - 2$ plays a specially important role. The corresponding kernel is known as Newtonian kernel and its potential energy functional is called electrostatic potential energy.

The problem of the numerical estimation of the Fekete points of a given compact set S has become a paradigm of computational complexity and a model of non-linear optimization problem with non-linear constraints. In particular, the case of the distribution of points on the 2-sphere appears as the 7th problem of the list *Mathematical problems for the next century* that S. Smale published in 1998, see [7]. Specifically, the problem consists in finding an algorithm which on input N produces in halting time polynomial in N a N-tuple $x = \{x_1, \ldots, x_N\}$ of distinct points on the 2-sphere satisfying the Smale's condition

$$\mathcal{I}_N(x) - \mathcal{I}_N(\omega_N) \leq c \log N, \tag{1}$$

where ω_N are the Nth order Fekete points for the logarithmic kernel in the 2-sphere and c is a universal constant. So, the problem asks for an algorithm that finds in polynomial time a good local minimum for the Fekete problem, tacking into account that the number of local minima grows enormously with N.

Many optimization algorithms have been used to tackle the search of local minima for different potential energies. In [4], for instance, one can find a brief description of a variety of them, and also some comments about their advantages and limitations. Several authors have published energy tables that gather the best energies found for different kernels. We can mention [11], where in particular energy values are showed for the logarithmic and the Newtonian kernels in the 2-sphere for N up to 200. In [8], one can find tables in which different authors propose, 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 must also remark the works [9, 10], where the author gives energy values for the Newtonian kernel in the 2-sphere for $N = (m+1)^2$ with m up to 80 and shows configurations corresponding to different kernels for $N = 1000$ and $N = 4000$ particles on a torus. The author comments that these values are approximate local minimizers of the potential energy, and that he has made attempts to find global minima with decreasing reliability as N increases. According to the author, he used a combination of local and global large-scale optimization techniques running on a cluster.

As for the Fekete problem as it appears in [7], Smale cites [5, 6], where the generalized spiral points were presented and numerical evidence was provided that these configurations support Condition (1) for $N \leq 12000$ with $c = 114$. The spiral points are designed according to geometric and heuristic criteria, and they can be constructed explicitly with a computational cost of order N, but they do not correspond to the minimization of any functional. Moreover, as it is shown in [5] Fig. 2, the difference $\mathcal{I}_N(x) - \mathcal{I}_N(\omega_N)$ for the spiral points tends to grow linearly with N, so it seems that the generalized spiral points cannot support the Smale's condition with N growing indefinitely for any value of c. As far as we known, there have not been other attempts to propose neither a complete nor a partial solution for the Smale's 7th problem.

In [1] we presented an algorithm for the search of local minima of potential energy functionals in a wide variety of compact sets. Here we make an exhaustive analysis of the behavior of this descent algorithm in the context of the Smale's 7th problem, including the study of its convergence properties and its computational cost. As the main result of this work, we provide strong numerical and statistical evidence of the fact that this algorithm satisfies the conditions to be proposed as a solution for the Smale's 7th problem. The statistical analysis is based on the results provided by around $2 \cdot 10^6$ experiments. In particular, we show that a local minimum for the logarithmic kernel in the 2-sphere can be identified by our algorithm with an average computational cost of order less than N^3 and that the probability of obtaining a local minimum satisfying Condition (1) for a given c decreases as $\left(\frac{N}{c}\right)^{-p}$, where $p \simeq 7$. As a consequence, our algorithm is able to localize a minimum fulfilling the Smale's requirements with a total computational cost of order less than N^{10}.

The results presented in this paper and in [1] as well as some extensions constitute the Ph.D. of J.M. Gesto, entitled *Estimation of Fekete Points*.

2 Background

Our first works related to the numerical estimation of Fekete points focused in obtaining an efficient and robust algorithm for the localization 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, that 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 convenient to describe the algorithm briefly for the sake of completeness.

The basic structure of the algorithm for smooth surfaces is classical; each step consists in choosing an advance direction and applying a step size. With regards to the advance direction, let us start by observing that the potential energy of a system of N unitary particles $x_k \in \mathbb{R}^3$, $k = 1, \ldots, N$, is given by $\mathcal{I} = \frac{1}{2}\sum_{i=1}^{N} V_i$, 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, that we denote by $F_i = -\nabla V_i \in T_{x_i}(\mathbb{R}^3)$, represents the repulsive force that acts on the i-th 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 advance direction, where $w_i = \dfrac{F_i^T}{|F_i|}$. Moreover, we call disequilibrium degree of the ith particle the scalar $|w_i|$. As for the step size, it is determined according to a mechanic conception of the paths described by the particles towards the equilibrium. If $x(t) = (x_1(t), \ldots, x_N(t))$ represents the position of N particles as a function of a certain parameter t, then we consider the ODE $x' = w$ with the initial condition corresponding to a given starting position. Its numerical resolution is carried out by means of the explicit forward Euler scheme $x^{k+1} = x^k + a\,\varphi(x^k)\,w^k$, where the coefficient a is a positive scalar that depends on N, $\mathcal{K}$ and S, and φ depends on the position of all the particles of the system and it allows us to adapt the step size to the difficulties of the different configurations that appear throughout the descent process. In particular, the function $\varphi(x) = \min_{1\leq i<j\leq N}\{|x_i - x_j|\}$ has always showed to be an appropriated choice. The descent process also includes a return algorithm to make the particles come back to the surface S after each step.

For the study of the convergence of the algorithm we use the maximum disequilibrium degree, $w_{\max} = \max_{1\leq i\leq N} |w_i|$, as a measure of the error at each step. Figure 1 corresponds to the application of the algorithm in the 2-sphere with $N = 1000$ and the Newtonian kernel,

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 the step 8000 ($n_{\text{step}} = 8000$). The convergence curve (right), obtained by displaying $w_{\max}$ as a function of n_{step}, allows us to observe that a final linear convergence ratio is attained after a highly non-linear phase. When this final linear tendency is reached, it can be assumed that the points are close enough to a local minimum in such a way that the Newton's algorithm could be used with guarantee of convergence.

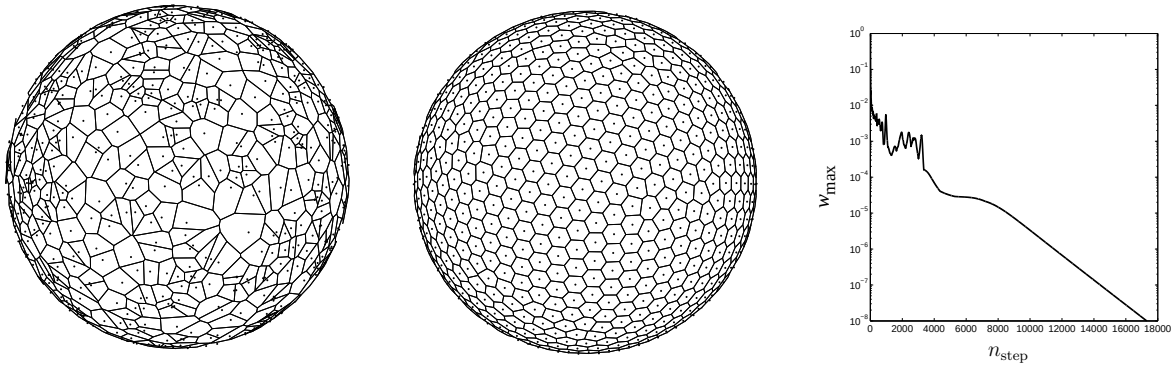


Figure 1: Initial and final configurations of $N = 1000$ particles on the 2-sphere with their Dirichlet cells and the correponding convergence curve.

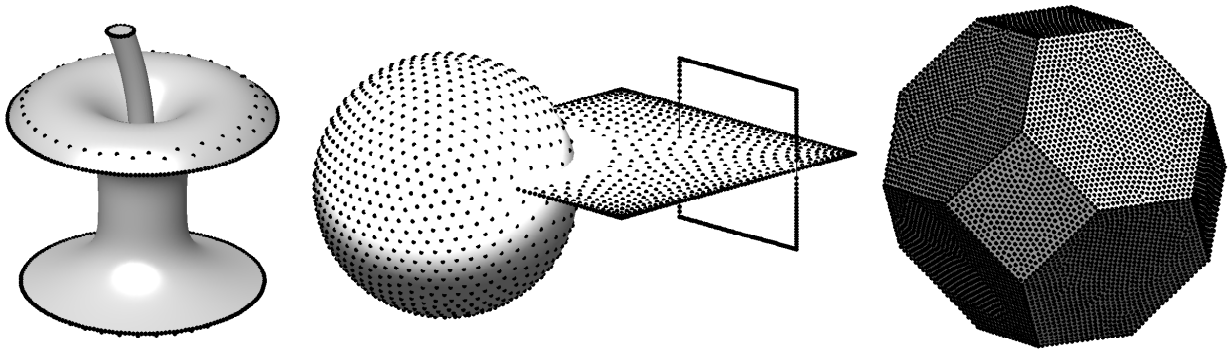


Figure 2: $N = 500$ particles on an apple with the logarithmic kernel, $N = 2000$ particles on a W-compact set obtained by combining different dimension objects with the Newtonian kernel and $N = 10000$ particles on the Kelvin Polyhedron with the Riesz's kernel for $s = 2$.

To apply the above algorithm in a non-regular surface $\mathcal{S}$ it is necessary to design an strategy for the generation of good starting configurations, because if the initial position is not good enough, the singularities of the surface could make the particles to became 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 results essential

for the determination of the disequilibrium degrees of particles at non-regular points, and we gave some additional details to carry out the implementation of our procedure, whose versatility and robustness we would like to mention here. Figure 2 shows some equilibrium configurations for different kernels and W-compact sets.

3 Distribution of points on the 2-sphere

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 a N point set $x = \{x_1, \ldots, x_N\}$. As a consequence, ω_N represents the Nth order Fekete points for the logarithmic kernel, that we simply call Fekete points.

A possible strategy for the estimation of ω_N consists in generating different random starting positions on S and applying a descent algorithm from each one of them to identify different local minima of $\mathcal{I}_N$. If we increase the number of random starting configurations, we also increase the probability of finding a good local minimum and, eventually, the Fekete points. The Smale's 7th problem asks for an algorithm that in polynomial calculation time produces for each N a N-tuple x of points on S satisfying Condition (1). Following the above strategy, it is clear that a descent algorithm satisfies the requirements of the problem in average if the following facts hold: first, the average calculation time to identify a local minimum from a given starting configuration is polynomial in N; second, the average number of starting positions needed to attain a local minimum x satisfying the Smale's condition grows polynomially with N. In fact, the second condition depends more on the intrinsic character of the Fekete problem itself than on the properties of the descent algorithm.

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 determine the average computational cost of identifying a local minimum and the average number of starting configurations needed to obtain a minimum satisfying Condition (1). Finally, we summarize the main conclusions of the study.

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 $w_i = \dfrac{F_i^T}{|F_i|}$, the disequilibrium degree of a particle, $|w_i|$, and the error at a step, $w_{\max} = \max\limits_{1 \le i \le N} |w_i|$. 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 non linear phase and a final linear phase or linear tendency, see Figure 1.

Our algorithm starts with the generation of an initial position for N particles on S. Once we fix the magnitude of the coefficient a, the step size $a\varphi$ can be applied in the direction w according to the forward Euler scheme. Then, the updated coordinates must be normalized to return the particles to S and the procedure is repeated obtaining a convergence curve. 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. Since the purpose of our algorithm is to identify local minima, a precise definition of convergence corresponds to be close to a local minima in such a way that the Newton's method converges, [3], and it is also associated to the final linear tendency in a convergence curve. It is not easy to determine neither the beginning of the linear tendency nor the step from which the Newton's algorithm converges. For this reason 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 precedes the linear tendency. We call non-return point the last maximum in a convergence curve. The non-return point indicates the entrance in the local influence zone of a minimum. We say that our algorithm 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 call c-minimum any local minimum that satisfies Condition (1) for a given c. In order to search a c-minimum we fix a number of random starting positions, n_{sp}, and we run the descent algorithm from each one of them up to attain the ε-convergence, with ε small enough. Each run produces sample values for different random output variables. Specifically, we define the random variables X, number of steps for nr-convergence in a run, and Y, number of steps for ε-convergence in a run. Moreover, taking into account that

$$\mathcal{I}_N(\omega_N) = -\frac{1}{4}\log\left(\frac{4}{e}\right)N^2 - \frac{1}{4}N\log N + O(N), \tag{2}$$

see [7], we define the random variable $U = I + \frac{1}{4}\log\left(\frac{4}{e}\right)N^2 + \frac{1}{4}N\log N$, where the random variable I is the energy of the local minimum identified in a run. The exact value of I remains unknown after the run, because the energy of the final configuration has an error depending essentially on ε. Taking this into account, we use the same symbol I both for the exact energy value and for the value associated to any ε small enough. For simplicity of notation we have not used indexes in the definitions of the variables X, Y, U, I. In each case it will be clear the values for N and ε. 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, $i = 1, \ldots, n_{\text{sp}}$, the sample data obtained in an experiment for Z, $m_z^k = \dfrac{1}{n_{\text{sp}}}\displaystyle\sum_{i=1}^{n_{\text{sp}}} z_i^k$ the sample moments of $\{z_i\}$, $\overline{z} = m_z^1$ the sample mean of $\{z_i\}$, $(m_z^k)' = \dfrac{1}{n_{\text{sp}}}\displaystyle\sum_{i=1}^{n_{\text{sp}}}(z_i - \overline{z})^k$ 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 resolution of an ODE system the coefficient a is bounded by a critical upper value related to the stability of the forward Euler's scheme. To guarantee this stability in a highly non-linear problem as the Fekete's one 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 choose the coefficient a according to the next criteria: its calculation must be as cheap as possible at each step and, under this condition, it must lead to convergence in a number of steps as small as possible. As for the cost at each step, it would be a favorable situation if the coefficient a did not depend on the initial configuration and if its value could remain constant throughout all the convergence process, since then its cost at each step would be null. In this subsection we study the existence of a coefficient a that satisfyes these properties.

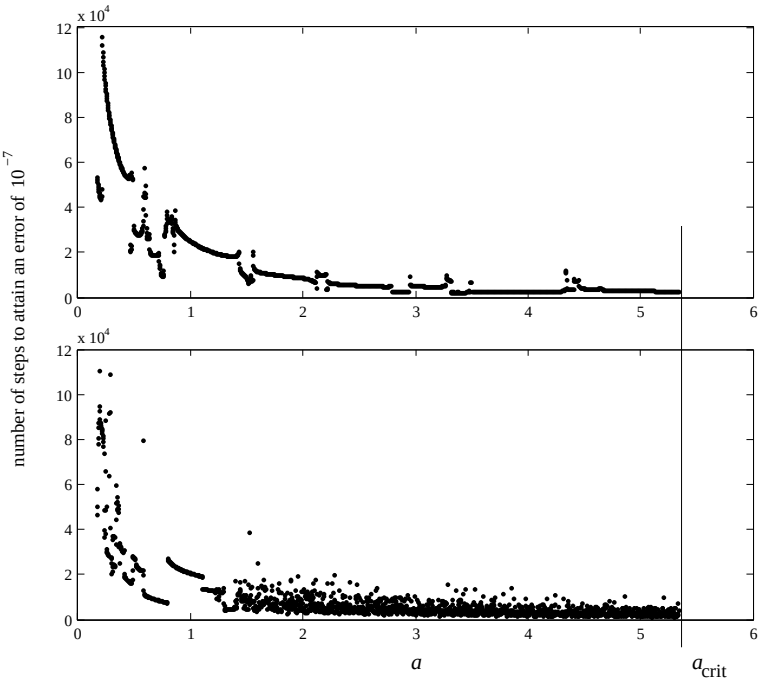


Figure 3: Evolution with a of the number of steps necessary to attain ε-convergence ($\varepsilon = 10^{-7}$) from a uniform starting configuration (up) and from a delta starting configuration (down).

Taking into account the enormous growing with N of the number of local minima of the Fekete problem, we start the study by considering the case $N = 87$, that still has a

relatively small amount of minima and that allows us to have a substantial information for each minimum. Figure 3 shows the evolution with a of the number of steps necessary to attain ε-convergence ($\varepsilon = 10^{-7}$) starting from two fixed initial configurations: one generated according to a uniform probability density on the whole 2-sphere (up) and the other generated according to a uniform probability density on a spherical cap of area $\pi \cdot 10^{-6}$ (down). 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 $\pi \cdot 10^{-6}$. In both cases we have considered 2000 different values for the coefficient a, that in each run remains constant from the starting configuration to the convergence position.

Figure 3 makes it clear that the loss of the stability in the forward Euler's scheme happens for extremely low values of a. Nevertheless, there exists an important range of values of a that guarantee the convergence to a minimum. Moreover, it results specially remarkable that the critical value of a from which the algorithm diverges, a_{crit}, is practically the same for both initial configurations. It can be observed that in the two cases the evolution with a of the number of steps necessary to converge is both qualitatively and quantitatively similar and that the smallest calculation times appear near a_{crit} from the left.

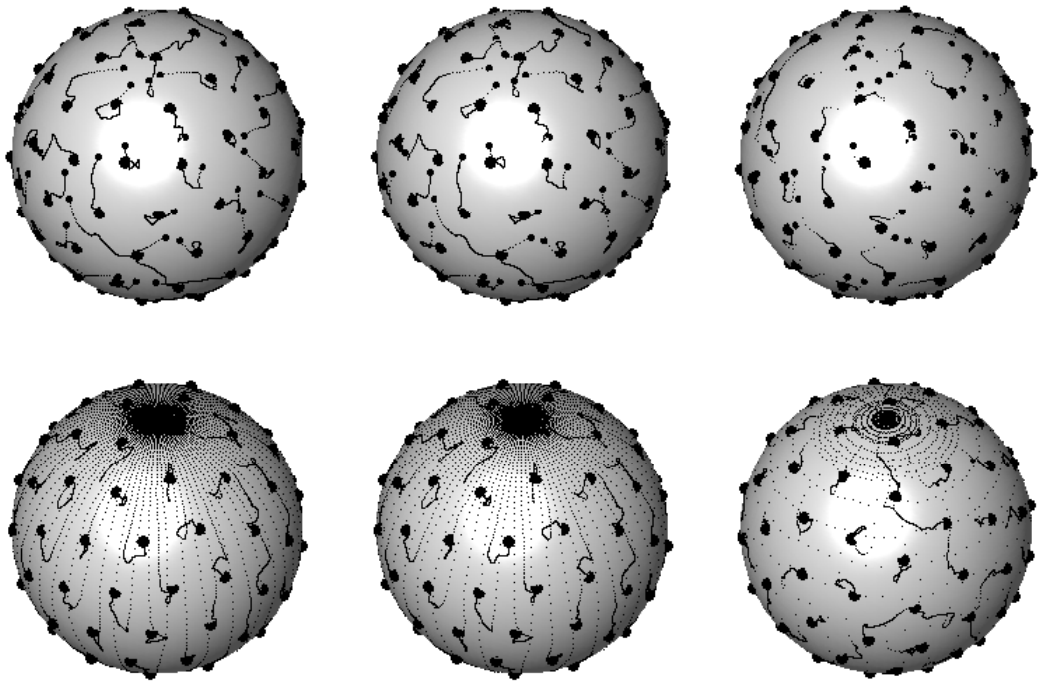


Figure 4: Paths described by the particles from a uniform starting configuration with $a = 0.9$, $a = 1.3$ and $a = 5.08$ (up from left to right) and from a delta starting configuration with $a = 0.85$, $a = 1.05$ and $a = 5.08$ (down from left to right).

Figure 4 shows the paths described by the particles throughout the convergence process from the same uniform starting configuration with $a = 0.9$, $a = 1.3$ and $a = 5.08$ (up from left to right) and from the same delta starting configuration with $a = 0.85$, $a = 1.05$ and

$a = 5.08$ (down from left to right). For each path the big 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 in a small cap and it cannot be appreciated). 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 Figure 3. There are few of such continuous fragments in the curve down in Figure 3 because the starting configuration is very extreme.

It results interesting to observe how our algorithm works from a delta starting configuration. Figure 5 shows, from left to right and from above to below, the configurations for $n_{\text{step}} = 0, 3, 8, 19, 32, 37, 99, 8075$ corresponding to the convergence process associated to Figure 4 (down to the right). The steps $n_{\text{step}} = 0$ and $n_{\text{step}} = 8075$ correspond to the starting configuration and to the configuration after the ε-convergence ($\varepsilon = 10^{-7}$) has been attained, respectively. In each case, the scale has been conveniently adjusted to make easier the visualization of the process. As it 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 neighbor particles in the final configuration. From this moment the ring leaves some particles in its interior. At the step $n_{\text{step}} = 32$ the ring has already left some particles, and a few steps later it has "equidistributed" all the particles on the sphere's surface. Most of the rest of steps are wasted in localizing a minimum accurately.

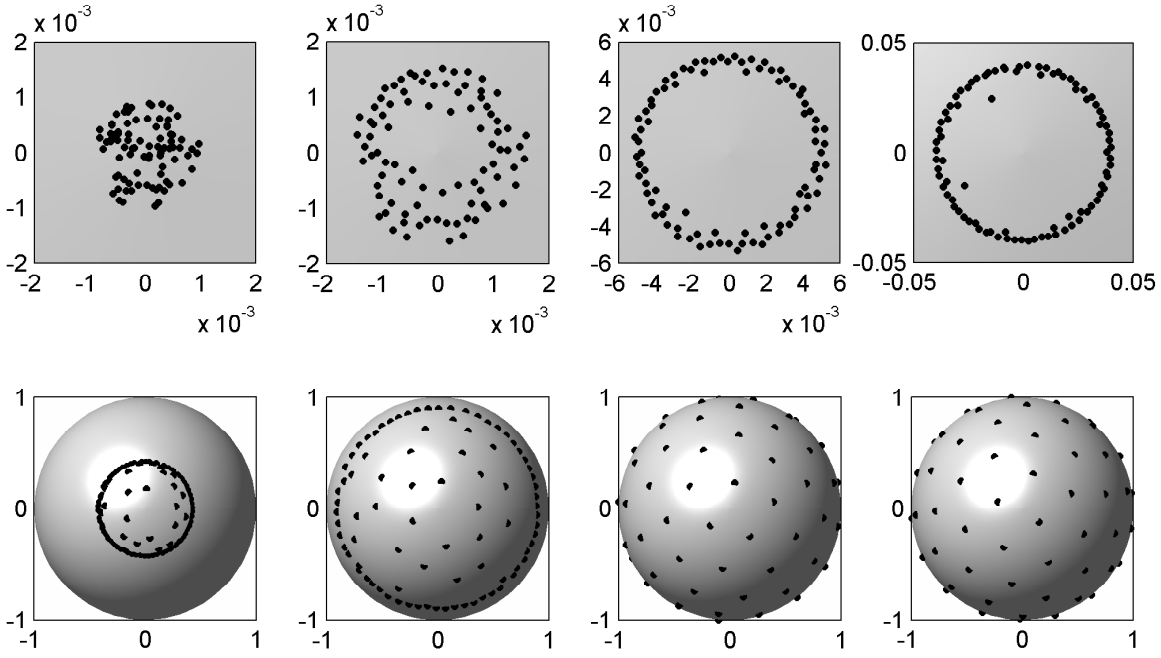
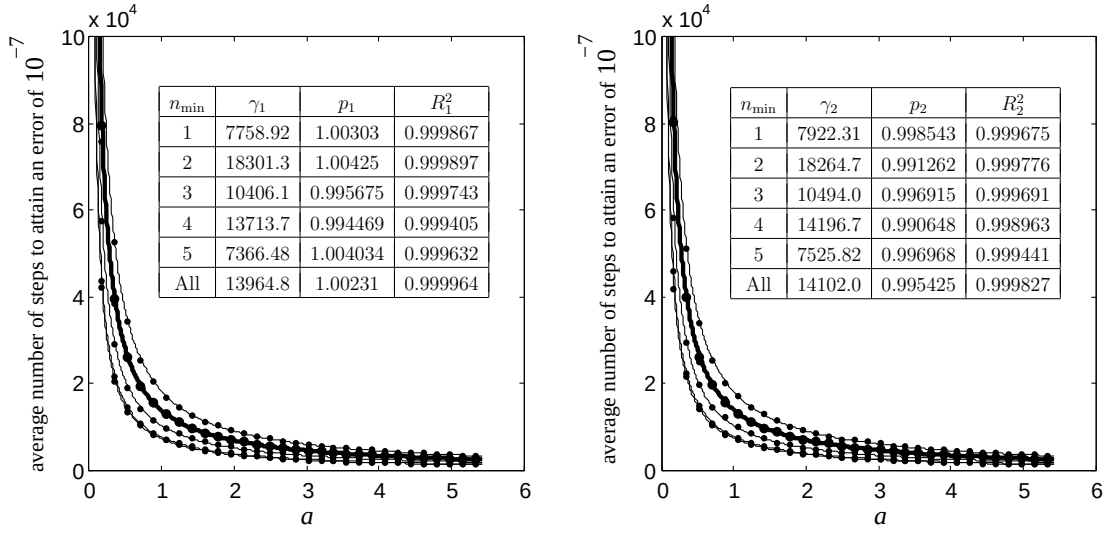


Figure 5: Convergence process associated to a delta starting configuration. From left to right and from above to below the steps $n_{\text{step}} = 0, 3, 8, 19, 32, 37, 99, 8075$ are displayed.

We have carried out an analysis in average for different random starting positions of the evolution with a of the number of steps necessary to converge. Figure 6 summarizes the

results corresponding to a test in which 120000 total runs of the algorithm with $N = 87$ have been performed. These runs correspond to 2000 uniform starting configurations (left) and 2000 delta starting configurations (right). For each one of these starting configurations we have considered 30 different values of a. The figure displays for each kind of starting configurations the average number of steps necessary for ε-convergence ($\varepsilon = 10^{-7}$) computed from the 2000 total data for each value of a (big 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 $\overline{n}_{\text{step}} = \dfrac{\gamma}{a^p}$ 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 of the coefficient R^2 for each one of the five analyzed minima, that are indexed by $n_{\min}$, and for all the data. The indexes 1 and 2 in γ, p, R^2 denote uniform and delta starting configurations, respectively. All this indicates 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 $p = 1$.



$n_{\min}$	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

Figure 6: Evolution with a of the average number of steps for ε-convergence ($\varepsilon = 10^{-7}$) for uniform starting configurations (left) and delta starting configurations (right). The average data corresponding to the best five minima and their interpolations are also included.

Figure 7 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, that has been determined from the 60000 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 results 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 with the procedure used for generating the random initial positions and also with 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 appears only once among the 120000 total runs. The worst minima have also very small probabilities.

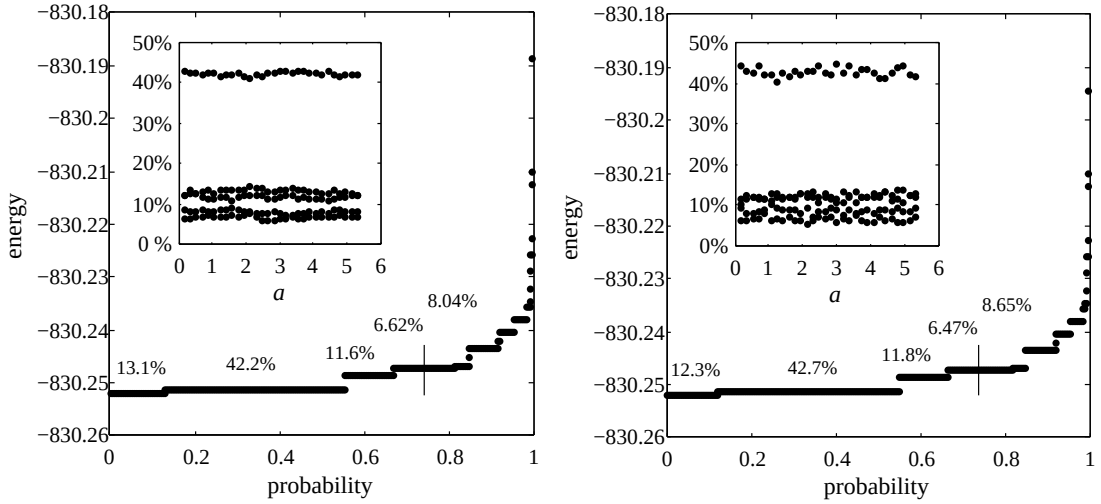


Figure 7: Sample distribution of the probability of obtaining the different minima of the case $N = 87$ from uniform starting configurations (left) and from delta starting configurations (right).

Maybe the most relevant result for the Fekete point problem we can extract from all the tests that we have carried out is that the value of a_{crit} is practically independent of the random starting configuration. Figure 8 shows this independence as well as the process of loss of convergence when the coefficient a is greater than a_{crit}. The two diagrams on the left show the behavior of the algorithm corresponding to $a = 5.4$, $a = 5.4275$, $a = 5.43$, $a = 5.433$, $a = 5.44$, $a = 5.7$ and $a = 8$ from the same uniform starting configuration (up) and from the same delta starting configuration (down). For the first value $a = 5.4$ the process converges in linear tendency until the machine error is reached (we work in double precision). For the next a values it is still possible to reach the linear tendency, but this tendency is left in a certain moment 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 happens for 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 little variations of a can lead to different minima. In any case, there exists 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 same figure we have included the paths described by the particles in the cases $a = 5.7$ (center) and $a = 8$ (right) from a uniform starting configuration (up) and from a delta starting configuration (down). The big points correspond to the starting position. We have not remarked a final position because the

process diverges. As it can be observed, the divergence process corresponds to a bifurcation both of the trajectories and of the convergence curves. For higher values of a the process seems to lead to a sort of "rotation" of two similar alternating configurations with respect to the same axis.

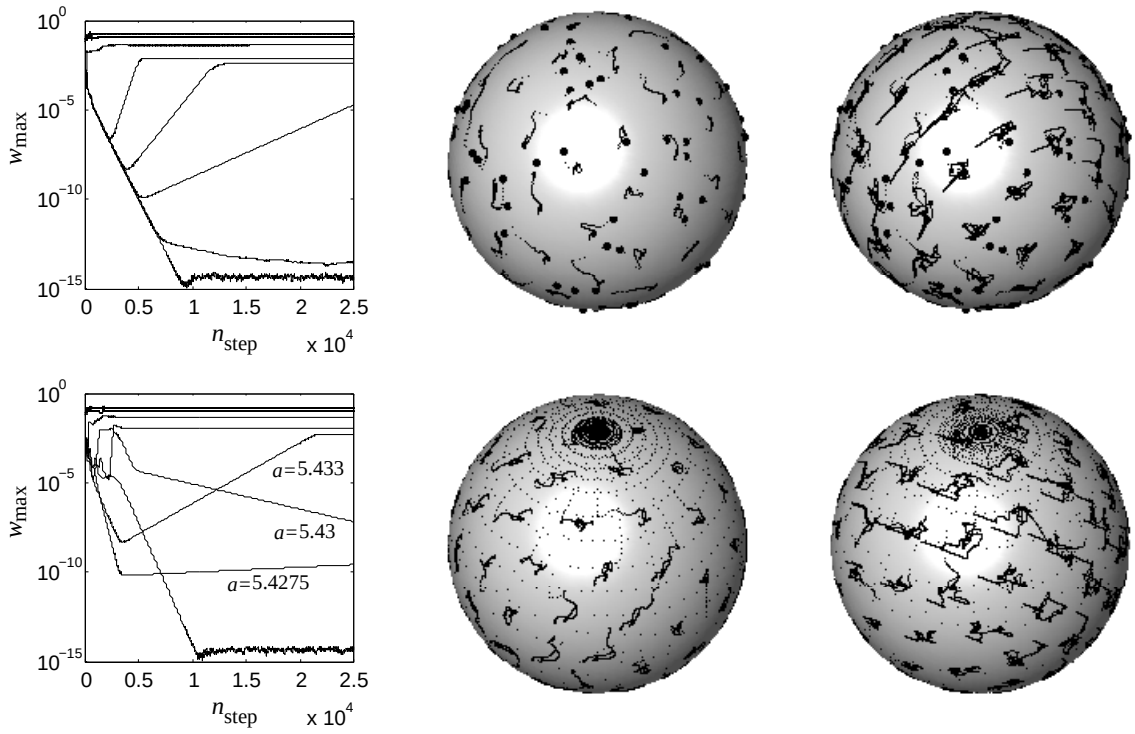


Figure 8: The divergence process from uniform starting configurations (up) and from delta starting configurations (down).

The results obtained for the case $N = 87$ suggest that if we impose a constant value of a throughout all the convergence process then the evolution with a of the average cost and the value of a_{crit} depend only on N. We have confirmed this fact in all the studied cases. Figure 9 (left), for instance, shows the evolution with a of the average number of steps necessary for ε-convergence ($\varepsilon = 10^{-7}$) for the case $N = 200$. There are data from 60000 total runs corresponding to 2000 uniform starting positions and 30 different values of a (from now on all the starting configurations are uniform). The corresponding interpolation curve is also included. The same figure (center) shows the sample probability distribution of the minima that have been obtained from the 60000 total runs. Note that the appearance of the probability distribution corresponds to a continuous random variable rather than to a discrete one, whereas the probability distribution for the case $N = 87$ has the typical appearance of a discrete random variable distribution. Figure 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 $a = 12$.

So, from all the numerical evidence we have accumulated, we can conclude that if the scheme $x^{k+1} = x^k + a \min_{1\leq i<j\leq N}\{|x_i^k - x_j^k|\}w^k$ (combined with the projection to the sphere

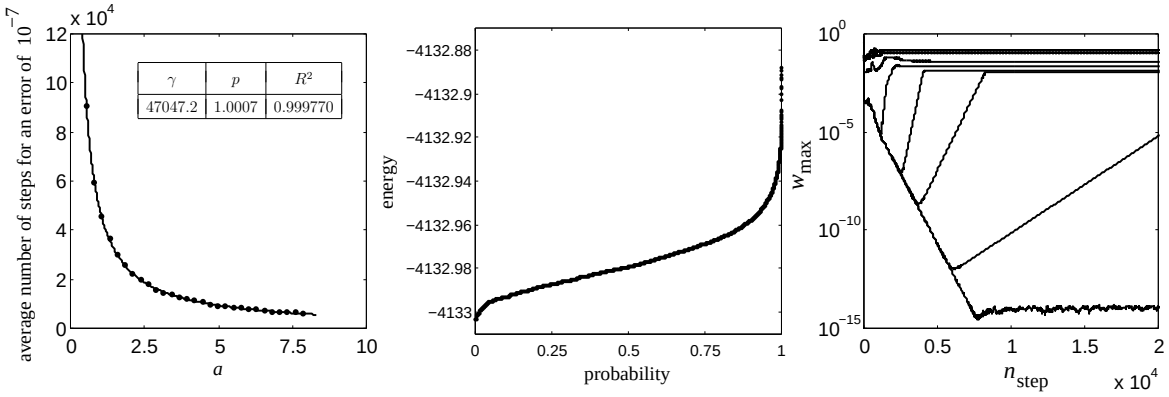


Figure 9: Evolution with a of the average cost, sample probability distribution of the minima and divergence for the case $N = 200$.

at each step) is adopted as the descent algorithm and the coefficient a is kept constant throughout all 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 is independent of the initial configuration. Taking also into account that the probability distribution of the minima obtained from random initial configurations is independent of a, it is clear that the optimum efficiency of the algorithm is obtained when for each N a value of a, a^*, is chosen close to a_{crit} from the left.

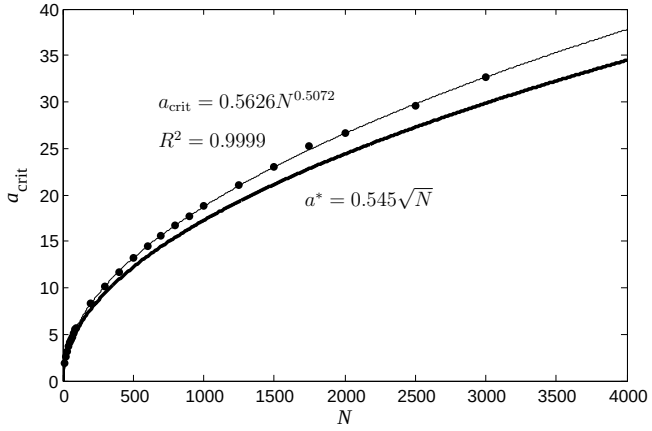


Figure 10: Evolution with N of a_{crit} and of a^*.

With regards to the computation of a_{crit}, an accurate estimation can be obtained for each N by the bisection method after determining by inspection a value of a that leads to convergence and another that leads to divergence from a given starting configuration. Figure 10 shows the results of a test in which this procedure is carried out for $N = 10, 20, ..., 100, 200, ..., 1000, 1250, ..., 2000, 2500, 3000$ and their interpolation. Taking into account this information, we establish the formula $a^* = 0.545\sqrt{N}$. All the results presented in the rest of this section correspond to computations with $a = a^*$.

3.3 The average cost of a local minimum

We can now carry out a study about the computational cost of the process of identifying a local minimum. It is clear that the cost at each step of our algorithm is always the same and it essentially corresponds to the computation of the N forces $F_i = \sum_{j \neq i} \frac{x_i - x_j}{|x_i - x_j|^2}$. More specifically, each one of the $\binom{N}{2}$ vectors $\frac{x_i - x_j}{|x_i - x_j|^2}$, $1 \leq i < j \leq N$, requires in cartesian coordinates five sums, a division and six products, to which we must add the six sums necessary to update the forces F_i and F_j. Hence, the computational cost of the identification of a local minimum depends only on the number of steps necessary to converge. In this subsection we analyze for each N the random variables X, number of steps for nr-convergence, and Y, number of steps for ε-convergence.

For the statistical analysis of the random variables X and Y we have used the sample data provided by $n_{\text{sp}} = 5000$ runs of the algorithm for each $N = 500, 1000, ..., 3000$ and by $n_{\text{sp}} = 1000$ runs for each $N = 4000, 5000$. The results corresponding to $N = 4000, 5000$ are 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 fundamental parameters for the analysis of the cost. In Figure 11 it can be observed the way the sample means $\overline{x}$, $\overline{y}$ stabilize with n_{sp}. 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 for $\varepsilon = 10^{-8}$ (right) for all considered N.

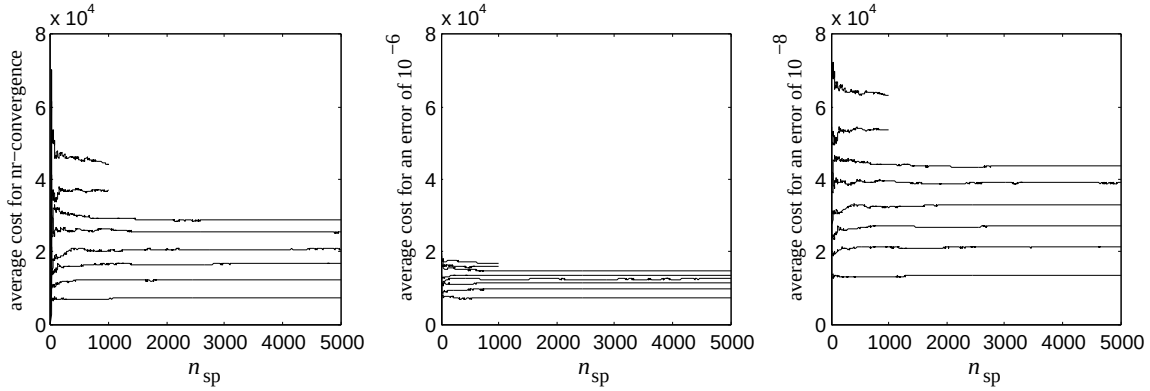
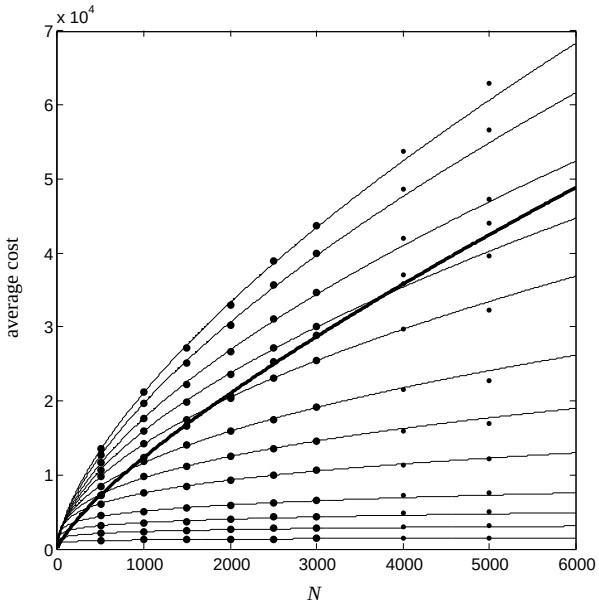


Figure 11: Evolution with n_{sp} of the average cost for nr-convergence (left) and for ε-convergence for $\varepsilon = 10^{-6}$ (center) and for $\varepsilon = 10^{-8}$ (right).

Figure 12 (left) shows the average number of steps necessary to attain nr-convergence and ε-convergence for $\varepsilon = 5 \cdot 10^{-5}, 2 \cdot 10^{-5}, 10^{-5}, ..., 10^{-8}$ 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 $N = 500, 1000, ..., 3000$ (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 last maximum in the convergence curve is the true non-return point for all the sample data.



ε	γ	p	R^2	$N_\perp$
$5 \cdot 10^{-5}$	543.64	0.1161	0.9892	28
$2 \cdot 10^{-5}$	818.92	0.1521	0.966	67
$1 \cdot 10^{-5}$	1043.4	0.1774	0.9732	121
$5 \cdot 10^{-6}$	1119.0	0.2197	0.9936	200
$2 \cdot 10^{-6}$	932.67	0.3026	0.9979	347
$1 \cdot 10^{-6}$	684.91	0.3819	0.9987	520
$5 \cdot 10^{-7}$	505.6	0.4536	0.9992	824
$2 \cdot 10^{-7}$	351.02	0.5349	0.9997	1794
$1 \cdot 10^{-7}$	297.33	0.5761	0.9996	3776
$5 \cdot 10^{-8}$	264.28	0.608	0.9996	9353
$2 \cdot 10^{-8}$	240.28	0.6376	0.9994	35989
$1 \cdot 10^{-8}$	229.7	0.6546	0.9994	116855
nr	61.306	0.7678	0.9995	

Figure 12: Average cost of the process of identifying a local minimum.

3.4 The probability of finding a c-minimum

The next essential point is to determine if the average number $\overline{n}_{\text{sp}}$ of random starting configurations needed to find a minimum with an energy value satisfying Condition (1) for a certain positive constant c is also polynomial in N. In this context, the energy of a local minimum can be seen as a random variable whose support is an interval with the unknown $\mathcal{I}_N(\omega_N)$ as lower limit.

We work here with the random variable U previously defined. To extract reliable conclusions about this random variable we have performed the following statistical experiment: $n_{\text{sp}} = 10^5$ runs of the algorithm from different uniform starting positions for each $N = 300, 400, 600, 700, ..., 1000$ and $n_{\text{sp}} = 10^6$ runs for $N = 500$. All these runs arrive to ε-convergence for $\varepsilon = 2 \cdot 10^{-7}$. Moreover, we have the sample information cited above; that is, $n_{\text{sp}} = 5000$ runs of the algorithm for each $N = 1500, 2000, 2500, 3000$ and $n_{\text{sp}} = 1000$ runs for each $N = 4000, 5000$ up to ε-convergence for $\varepsilon = 10^{-8}$.

Let us start the analysis of all this sample information with Figure 13, that shows the stabilization with n_{sp} of the kth roots of the sample centered moments, $\pm\sqrt[k]{|(m_u^k)'|}$, $k = 2, ..., 10$, where the sign is given by $(m_u^k)'$, for the cases $N = 300$ (left), $N = 500$ (center), $N = 1000$ (right). For instance, we note that the value of the 10th root of the 10th centered moment for $N = 1000$ is 0.11457357 when $n_{\text{sp}} = 5 \cdot 10^4$ and it is 0.11499766 when $n_{\text{sp}} = 10^5$. We do not include the evolution of the sample mean $\overline{u}$, which stabilizes earlier than the higher order moments.

Figure 14 (left) shows the evolution with N of the final values obtained for the sample mean $\overline{u}$ and for $\pm\sqrt[k]{|(m_u^k)'|}$, $k = 2, ..., 10$. Note the different scaling in the vertical axis for

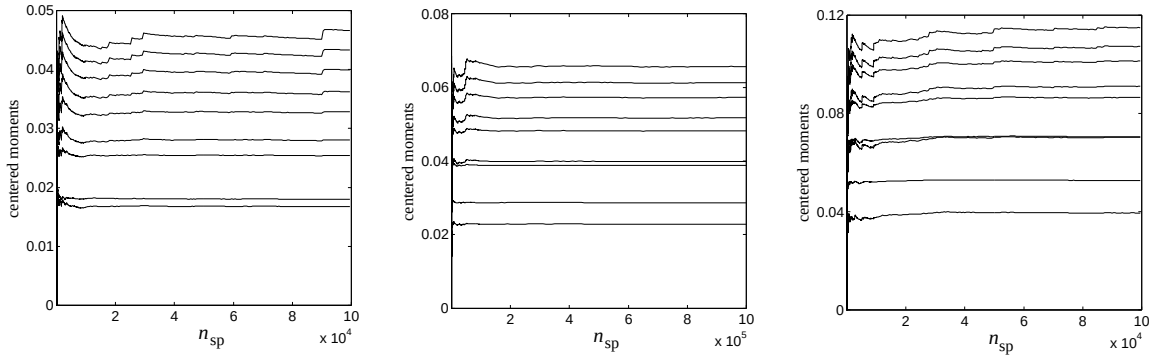
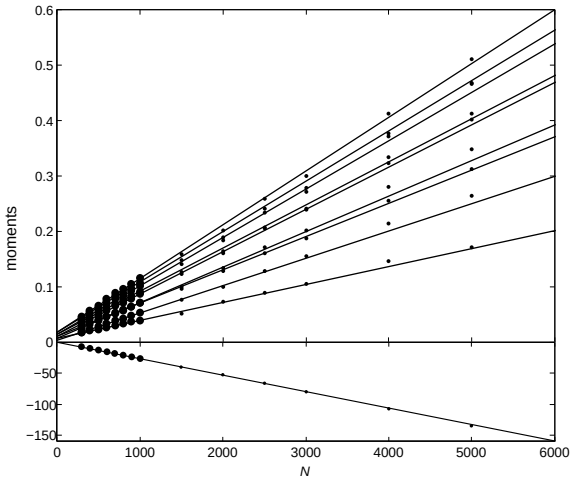


Figure 13: Evolution with $n_{\rm sp}$ of the kth roots of the sample centered moments, $\pm\sqrt[k]{|(m_u^k)'|}$, with $k = 2, ..., 10$, for $N = 300$ (left), $N = 500$ (center) and $N = 1000$ (right).

positive and negative values. The figure also includes the straight lines obtained by linear regression from the data corresponding to $N = 300, 400, ..., 1000$ (big points). The rest of the data (small points) come from a smaller sample, and they are displayed only to confirm the tendency given by the first points. The table on the right shows the linear regression parameters, A_k, B_k, R^2, for each straight line. The index 1 is for the mean and the rest of values are for centered moments. It seems clear that the expressions $\mu_U \simeq A_1 N + B_1$, $(M_U^k)' \simeq (A_k N + B_k)^k$, $k = 2, ..., 10$, can be used as good approximations for the moments.



order	A_k	B_k	R^2
1	-0.026656	0.26840	0.99999
2	$4.9296 \cdot 10^{-5}$	0.0035699	0.99925
3	$3.2442 \cdot 10^{-5}$	0.0069032	0.99795
4	$6.4333 \cdot 10^{-5}$	0.0063944	0.99944
5	$6.0243 \cdot 10^{-5}$	0.010139	0.99930
6	$7.6582 \cdot 10^{-5}$	0.010175	0.99916
7	$7.8273 \cdot 10^{-5}$	0.013186	0.99850
8	$8.7542 \cdot 10^{-5}$	0.014208	0.99803
9	$9.1198 \cdot 10^{-5}$	0.016632	0.99695
10	$9.7274 \cdot 10^{-5}$	0.018202	0.99603

Figure 14: Evolution with N of the sample mean $\overline{u}$ and of the kth roots of the sample centered moments, $\pm\sqrt[k]{|(m_u^k)'|}$, with $k = 2, ..., 10$.

Let us consider now the standardized variable $V = \dfrac{U - \mu_U}{\sigma_U}$. If we assume the above approximating expressions for the moments, we can conclude that $M_V^k \simeq \left(\dfrac{A_k N + B_k}{A_2 N + B_2}\right)^k$, $k = 2, ..., 10$. The sample information related to higher order moments, that we do not show here, clearly suggests that this tendency is general. As a consequence, we also assume the

existence of an asymptotic standardized distribution whose moments are given by $\left(\frac{A_k}{A_2}\right)^k$, to which the distribution V tends when $N \to \infty$. This implies in particular that there exists $\alpha_{\lim} < 0$ such that for $N \to \infty$

$$\mathcal{I}_N(\omega_N) \simeq -\frac{1}{4}\log\left(\frac{4}{e}\right)N^2 - \frac{1}{4}N\log N + \mu_U + \alpha_{\lim}\sigma_U.$$

Note that this expression agrees with Equation (2) taking $O(N) = \mu_U + \alpha_{\lim}\sigma_U = (-0.026656 + 0.000049296\alpha_{\lim})N + (0.26840 + 0.0035699\alpha_{\lim})$. In [11] it was proposed the expression $O(N) = -0.026422N + 0.13822$, obtained from numerical results with N up to 200.

The next natural step is to adjust the probability density function f_V of the standardized variable V. For this we use the following simple model, that is based on the composition of two queues. Let us consider the random variable Z, $Z \in (0,1)$, whose probability density function has the form $f_Z(z) = A^{-1}c_1(z)c_2(1-z)$, where $A = \int_0^1 c_1(s)c_2(1-s)\,ds$ and a generic expression for the queues c_1, c_2 is given by $c_i(z) = \dfrac{1}{z g_i\left(h^{-1}\left(\frac{z}{20e}\right)\right)}$, $z \in (0,1)$, where h is defined by $h(x) = \dfrac{\log x}{x}$ with $x > e$ and the function g_i is chosen to define different approximating models. The coefficient $20e$ is only a scaling parameter whose value has been fixed by the authors. We present here the results given by two different models obtained by taking $g_1(t) = t^{p_1}$, $g_2(t) = t^{p_2}$ (Model 1) and $g_1(t) = e^{p_1 t}$, $g_2(t) = e^{p_2 t}$ (Model 2). These models correspond to the hypotheses of polynomial and exponential cost for the Fekete problem, respectively. For both models, the probability density function of the standardized random variable $W = \dfrac{Z - \mu_Z}{\sigma_Z}$, f_W, and its support, $(\alpha, \omega) = \left(\dfrac{-\mu_Z}{\sigma_Z}, \dfrac{1-\mu_Z}{\sigma_Z}\right)$, depend only on two parameters p_1, p_2.

Both models have been used to adjust the first ten moments M_V^k for each N. In Fig. 15 we show the kth roots of the moments M_V^k obtained from the linear expressions showed in Fig. 14 for $N = 200, 400, 600, 1000, 3000$ and for the limit case $N \to \infty$. These moments are indicated by means of empty circles. The figure also displays the approximations given by models 1 and 2, that are marked with points and crosses, respectively, for an appropriated choice of the parameters a, b for each N. For the determination of these parameters we have used the following procedure: for a given value of α we minimize the maximum relative error in the moments with $k = 3, \ldots, 10$ along the curve in the space p_1, p_2 whose points satisfy the condition $-\dfrac{\mu_Z(p_1,p_2)}{\sigma_Z(p_1,p_2)} = \alpha$. Then, we vary α and chose the parameters corresponding to the value of α that produces the minimum maximum relative error in the moments. As it can be seen, this procedure allows us to finely adjust the moments both for the polynomial and exponential hypotheses, even when only two parameters control the model. In the worst cases the maximum relative errors are around 1%, but they often are smaller than 0.5%.

Fig. 15 makes clear that the analysis of the moments of the random variable V is not enough to determine if the probability of finding a c-minimum decreases polynomially or exponentially. However, the analysis of the support (α, ω) provided by models 1 and 2 results

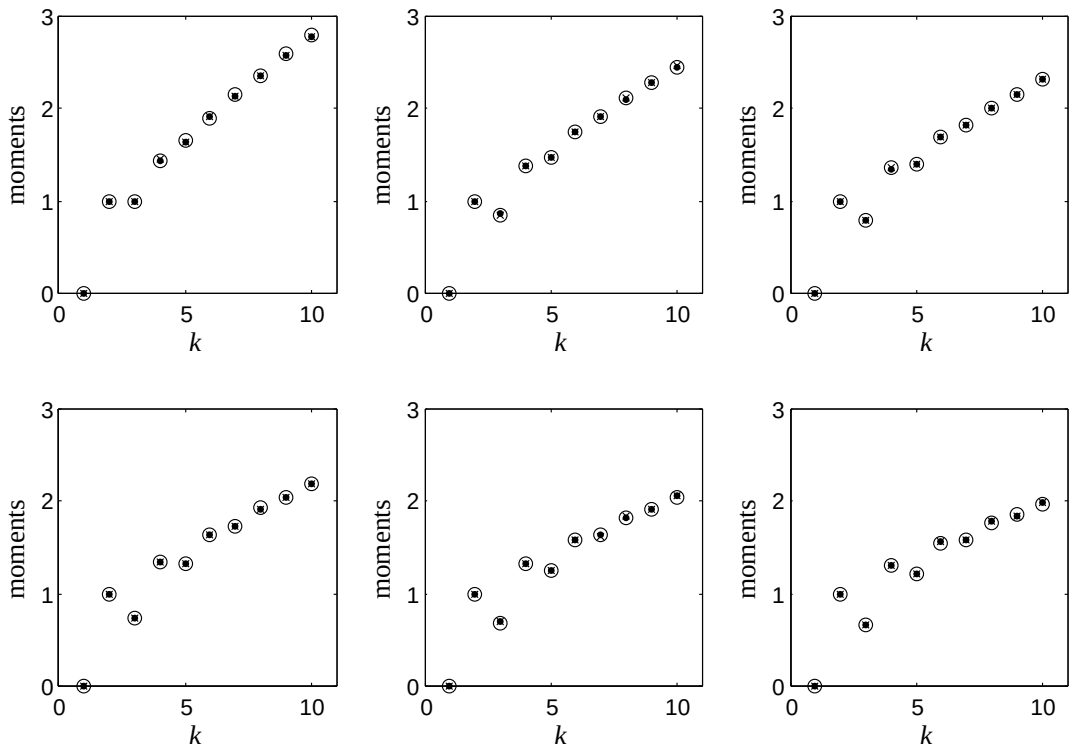


Figure 15: Evolution with N of the kth order moments of V, $k = 1, ..., 10$, given by the linear approximations $A_k N + B_k$ ($\circ$) and their approximations by means of models 1 ($\bullet$) and 2 ($\times$). From left to right and from above to below, we show the results corresponding to $N = 200, 400, 600, 1000, 3000$ and $N \to \infty$.

crucial to make a decision. Fig. 16 (up) shows the evolution with N of the parameters p_1, p_2 obtained as it has been indicated for Model 1 (left) and Model 2 (right). We have used the logarithmic scale in the vertical axis. The same figure (down) displays the evolution with N of the support (α, ω) corresponding to the values p_1, p_2 obtained in each case.

Let us start by observing the high values for ω given by both models for the first values of N, that considerably overestimate the sample data. To understand this, it is convenient to look again at Fig. 7, that corresponds to the case $N = 87$. In this case there is only a little amount of minima, one of them with a probability of around 42% and some others with probabilities of around 10%. Nevertheless, even in this case there exist some bad minima far from the average energy value and with extremely low (of order 10^{-5}) probabilities. From the sample data corresponding to this case we have obtained a value of around 14 for the standardized energy value of the worst minimum. The values provided for ω by the continuous models for the first values of N correspond to these extremely low probabilities of bad isolated minima in the upper queue. Obviously, the continuous approximating models can not be used for very small values of N, whose probability distributions are clearly discrete. For this reason, we only use the models for $N \geq 200$. The biggest value for ω obtained from the sample data corresponding to $N \geq 200$ is around 7 and it has been obtained for $N = 300$. Nevertheless, taking into account the results corresponding to $N = 87$

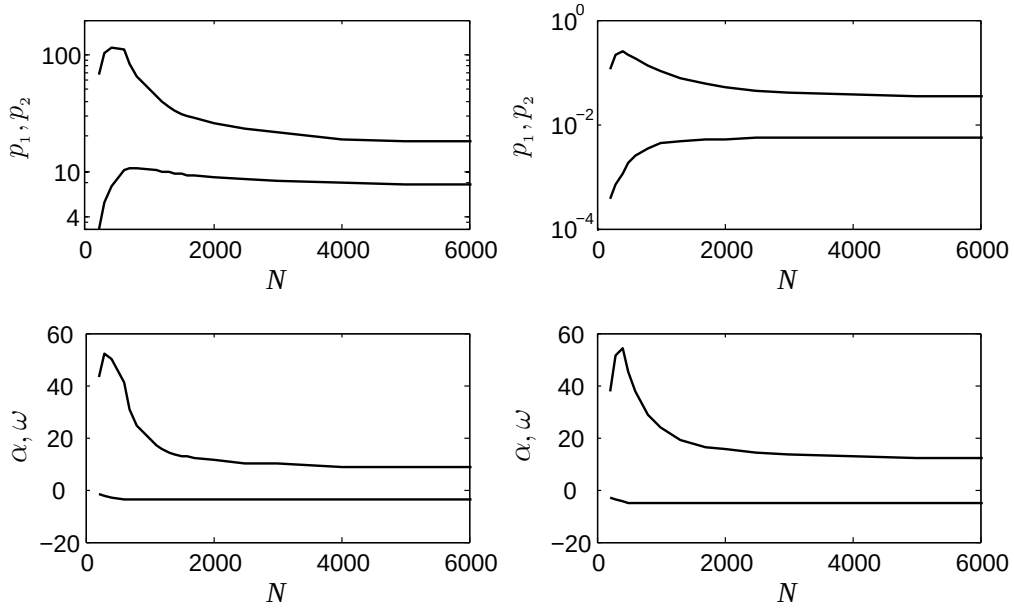


Figure 16: Evolution with N of the parameters p_1, p_2, α, ω for the Models 1 (left) and 2 (right).

and the huge decrease with N of the probability of finding any local minimum, it is reasonable to assume that there exists a relatively little amount of bad isolated minima with probabilities of order 10^{-8} or even less for the first values of N. However, as N increases and the problem tends to be continuous this isolated bad minima tend to disappear, which both models reflect with a strong decrease of ω.

In any case, the key parameter for the Fekete problem is α; that is, the limit of the lower queue, since there exists an excellent agreement between the moments, and hence also between the shapes of the probability distribution functions. Fig. 17 shows the evolution with N of the parameter α according to Model 1 (solid line) and 2 (dotted line). At this point it is important to observe that this figure shows a strong change in the evolution of the lower support between $N \simeq 500$ and $N \simeq 1000$. Both models exhibit the same qualitative behavior. In fact, the quotient between the parameters α corresponding to models 1 and 2 remains practically constant and equal to $0.69 \simeq \log 2$ for all N. After a rapid increase of the lower support up to $N \simeq 500$, this growing stops and the support stabilizes with N tending to the limit values $\alpha_{\text{lim}} \simeq -3.47$ (Model 1) and $\alpha_{\text{lim}} \simeq -4.97$ (Model 2). The upper support behaves similarly after attaining its maximum value, tending to $\omega_{\text{lim}} \simeq 7.3$ and $\omega_{\text{lim}} \simeq 10.3$ for models 1 and 2, respectively. Note that the quotient of both values of ω_{lim} is practically the same that the one obtained for the lower support.

In these conditions, we can say that it is the way the moments M_V^k change with N what leads to a specific shape of the lower support's evolution, and that different approximating models give essentially the same tendency, but affected by a scale factor. The objective of the intensive calculations we have performed was not only to extract reliable sample information about the moments, but also about the lower support. The points in Fig. 17

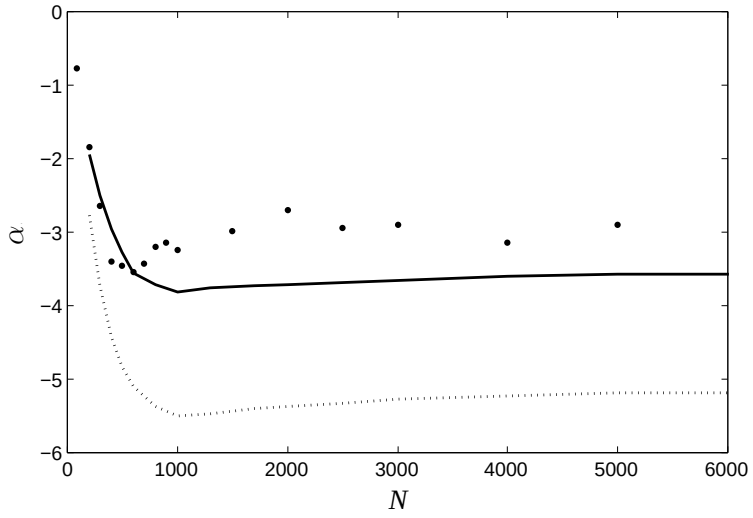


Figure 17: Evolution with N of the parameter α for the models 1 (solid line) and 2 (dotted line) and sample data.

correspond to the best standardized energies found for all the studied N; that is, $N = 87, 200, 300, \ldots, 1000, 1500, \ldots, 3000, 4000, 5000$. The best found energies for the cases $N = 87$ and $N = 200$ are -830.251915153 and -4133.00307952 respectively, and they equal the ones available in the literature, see for instance [11]. These values are assumed to be global minima. On the other hand, it results very meaningful that the point corresponding to $N = 500$, obtained from $n_{\text{sp}} = 10^6$ runs, is between the points corresponding to $N = 400$ and $N = 600$, obtained from $n_{\text{sp}} = 10^5$ runs. We consider that the points associated to $N = 300, 400, 500$, and likely also the one associated to $N = 600$, correspond to true global minima. The position of the rest of points clearly indicates that the global minima has not been attained yet, but they confirm the stabilization with N of V.

The sample data show that the growing of the lower support begins to stop at $N \simeq 500$, where the sample support arrives to -3.463. Model 2 gives $\alpha \simeq -4.8$ for $N = 500$ and $\alpha \simeq -2.8$ for $N = 200$, where the sample value is $\alpha = -1.844$. Therefore, Model 2 considerably overestimates the lower support of V, whereas Model 1 adjusts well the sample lower support as well as the moments. So, if W is the standardized random variable corresponding to Model 1, whose parameters p_1, p_2, α, ω change with N as indicated in Fig. 16 (left), then p_1 tends to approximately 7 when N grows and the probability of finding a c-minimum is given by $P[I - \mathcal{I}_N(\omega_N) \leq c \log N] = P[U - \mu_U - \alpha\sigma_U \leq c \log N] = F_V\left(\dfrac{c \log N}{A_2 N + B_2} + \alpha\right) \approx F_W\left(\dfrac{c \log N}{A_2 N + B_2} + \alpha\right)$. If we denote by $\delta_N = \dfrac{c \log N}{(\omega - \alpha)(A_2 N + B_2)}$, then the last expression equals $\dfrac{1}{A}\displaystyle\int_0^{\delta_N} c_1(s) c_2(1-s)\, ds$, that when $N \to \infty$ has order $O(\delta_N c_1(\delta_N)) \sim O\left(\left[h^{-1}\left(\lambda \dfrac{\log N}{N}\right)\right]^{-p_1}\right)$, where $\lambda = \dfrac{c}{20e(\omega - \alpha)A_2}$. Taking now into ac-

count that if $N \to \infty$ then $h\left(\frac{N}{\lambda}\right) \simeq \lambda h(N)$, we obtain

$$P[I - \mathcal{I}_N(\omega_N) \leq c \log N] \sim O\left(\left(\frac{N}{c}\right)^{-p_1}\right), \ p_1 \simeq 7.$$

3.5 Conclusions

We have studied the properties of our algorithm for the numerical estimation of Fekete points in the framework of the Smale's 7th problem; that is, for the case of the logarithmic potential energy in the 2-sphere. As a result of the performed tests, we can conclude that the algorithm converges for any constant value of the coefficient a lower than a certain value a_{crit} that depends only on N. The formula $a^* = 0.545\sqrt{N}$ has been used in practice, taking into account that the average number of steps to converge is inversely proportional to a. All the runs (about $2 \cdot 10^6$) carried out according to this criterion converged, which can be understood as a "convergence proof" for our algorithm. Moreover, the algorithm is extremely robust, since it converges even for really hard initial positions like the delta starting configurations. On the other hand, the probability distribution of the minima obtained from random starting positions results independent of the procedure used to generate them as well as of the coefficient a. This justifies the choice of a value for a close to a_{crit} in order to maximize the efficiency while keeping the convergence.

It has been shown that the average computational cost of identifying a local minimum for the Fekete problem is bounded by N^3, since the cost at each step is $O(N^2)$ and the average number of steps necessary to reach the ε-convergence and the nr-convergence grows sub-linearly with N. Specifically, the average number of steps needed to attain the nr-convergence grows approximately as $N^{0.77}$, and when ε decreases the exponent in N for the ε-convergence increases tending to a limit value of around 0.77. In particular, the intersection of the ε-convergence ($\varepsilon = 10^{-8}$) average cost curve with the nr-convergence average cost curve is located at $N > 10^5$.

An exhaustive statistical analysis has been performed to characterize the probability distribution function of the energy of a local minimum obtained with our algorithm from a random starting position. In particular, good linear approximations for the mean and for the kth roots of the order k centered moments up to order 10 of U, the linear part of I, have been obtained, and reliable estimations of the global minima corresponding to $N = 300, 400, 500, 600$ have been found. Two theoretical models based in the hypotheses of polynomial and exponential cost, models 1 and 2 respectively, have been used to adjust all the available sample information. Both models produce fine adjustments for the moments up to order 10 of the variable $V = \dfrac{U - \mu_U}{\sigma_U}$ for all N and give the same qualitative evolution with N of the lower support of this standardized variable, but Model 2 clearly overestimates this support. Model 1 reproduces well all the information provided by the sample data. According to this approximating model, the probability of obtaining a c-minimum tends to decrease approximately as N^{-7}.

If Model 1 is taken as a valid approximating model, we can conclude that the total average computational cost of the process of finding a c-minimum is polynomially bounded by N^{10}. In addition, the number of starting positions needed to attain a c-minimum with a given probability r is also polynomial in N. Effectively, if q is the probability of finding a c-minimum from a random starting position, then the number of starting positions needed to attain a c-minimum with probability r is $\dfrac{\log(1-r)}{\log(1-q)}$, and if $q \sim N^{-p}$ then the last expression has order N^p. On the other hand, taking into account that in general a big amount of starting positions are needed, the average computational cost of a minimum can be used to predict calculation times with high accuracy. Regarding the constant c, we can observe that taking $c = 0.043$ the 10^6 energy values of I obtained for $N = 500$ satisfy the condition $I - I_{\min} \leq c \log 500$, where $I_{\min}$ is the best found energy. For N big enough arbitrarily small values for $c > 0$ can be considered. Moreover, we have obtained the expression

$$\mathcal{I}_N(\omega_N) \simeq -\frac{1}{4}\log\left(\frac{4}{e}\right)N^2 - \frac{1}{4}N\log N - 0.02683N + 0.2560$$

for the asymptotic minimum logarithmic energy on the 2-sphere.

Finally, we must observe that the previous reasonings are based on numerical and statistical approaches and, as a consequence, the presented results are affected by a certain error. The error comes from the difference between exact and sample moments, linear regression of points non exactly aligned, and the choice of an approximating model and of a criterion to adjust the moments. All this could lead in particular to a variation in the exponent in N for the probability of finding a c-minimum. However, all the performed tests indicate that the hypothesis of exponential cost does not allow us to adjust both the moments and the lower support, whereas the parameters of the polynomial model can slightly vary maintaining the agreement between moments and support.

4 Further comments

We would like to start by observing that our algorithm, that has been designed for the minimization of a potential energy, does not need to evaluate this energy throughout the optimization process. The energy value of an equilibrium configuration can be obtained *a posteriori* as a byproduct of the calculation of this configuration. Nevertheless, it can be checked that the energy descents at each step of the convergence process. These considerations lead us to compare our advance criterion with the classical line search procedure, according to which the step size is determined from the minimization of the objective function in the advance direction. There exist remarkable similarities between both advance criteria, which are being studied by the authors. In any case, we want to show here some preliminary results. 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 high accuracy in

order to evaluate how much the line search procedure can improve the optimization process. To be exact, if we call $a_{\rm ls}$ the magnitude of the coefficient a that minimizes the energy in the direction w (we include the minimum distance between particles in the step size), we have demanded at each step that the absolute error in $a_{\rm ls}$ is smaller than $10^{-16}a_{\rm crit}$. Using this procedure in the case $N = 87$ ($a_{\rm crit} \simeq 5.37$) we obtain that after 500 runs corresponding to different starting configurations the ε-convergence ($\varepsilon = 10^{-6}$) is attained after 1884 steps in the average, whereas our algorithm reaches the same accuracy level after 2009 steps in the average, which represents approximately only a 6% of increment in the number of steps. However, it is clear that the cost at each step is much higher with the line search procedure than with our algorithm, in which the step size is analytically obtained.

It is even more interesting to study the values that $a_{\rm ls}$ takes throughout the descent process. Figure 18 shows the evolution with $n_{\rm step}$ of $a_{\rm ls}$ (left), of the energy decrement at each step (center), and of $w_{\rm max}$ (right) for a calculation with line search for $N = 87$ from a random starting configuration. In the left diagram we have also displayed the horizontal straight line corresponding to the value 5.37. We do not include in the figure the value $a_{\rm ls} = 33.6570189174919$, which corresponds to $n_{\rm step} = 1$, because of the scaling of the axis. The value corresponding to $n_{\rm step} = 2$ is $a_{\rm ls} = 11.1625964050175$, but it cannot be distinguished in the figure because it is an isolated point placed practically on the vertical axis. As it can be observed, the value of $a_{\rm ls}$ oscillates around $a_{\rm crit}$ 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, that describe a smooth curve. So the coefficient $a_{\rm crit}$ seems to be a sort of "average" of $a_{\rm ls}$. Moreover, regarding to the average number of steps necessary to convergence, the choice $a \simeq a_{\rm crit}$ is practically equivalent to take $a = a_{\rm ls}$, but $a_{\rm crit}$ results much cheaper at each step that $a_{\rm ls}$ and, in addition, $a \simeq a_{\rm crit}$ guarantees a smooth variation of all the convergence control parameters, $w_{\rm max}$ in particular.

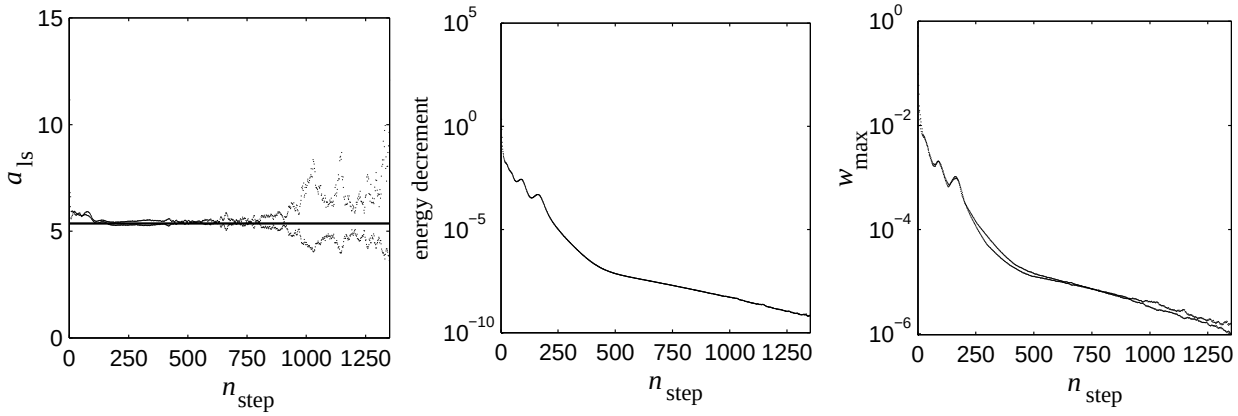


Figure 18: Evolution with $n_{\rm step}$ of $a_{\rm ls}$, of the energy decrement and of the maximum disequilibrium degree in a calculation with line search.

The behavior that Figure 18 indicates is also observed working with other kernels. Figure 19 shows the evolution with $n_{\rm step}$ of $a_{\rm ls}$ for the cases $s = 1$ with $N = 100$, $s = 2$ with $N = 200$ y $s = 3$ with $N = 300$, where s denotes the power of the Riesz's kernel. We have

also included the horizontal straight lines associated to the values that a_{crit} takes in these three cases ($a_{\text{crit}} \simeq 1.17$, $a_{\text{crit}} \simeq 0.298$ and $a_{\text{crit}} \simeq 0.098$, respectively). These estimations of a_{crit} have been obtained by following the same procedure that we used for the performance of Figure 10, that is, by using the bisection method from a value of a that converges and other that diverges for each kernel and for each N.

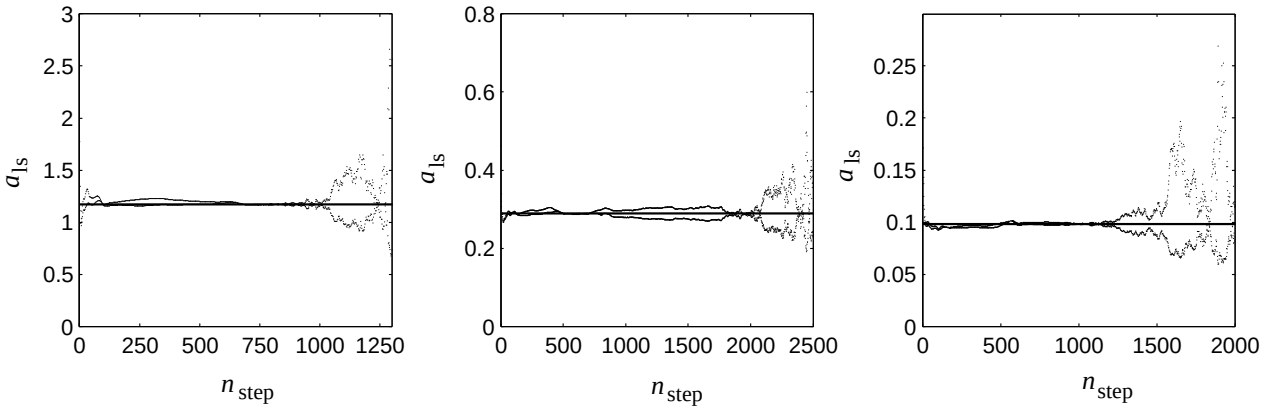


Figure 19: Evolution with n_{step} of a_{ls} for the cases $s = 1$ with $N = 100$, $s = 2$ with $N = 200$ and $s = 3$ with $N = 300$.

Figure 20 shows the evolution with N of a_{crit} for the 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 to the curve given by a negative power of N. The figure includes interpolation curves with their R^2 values for both cases.

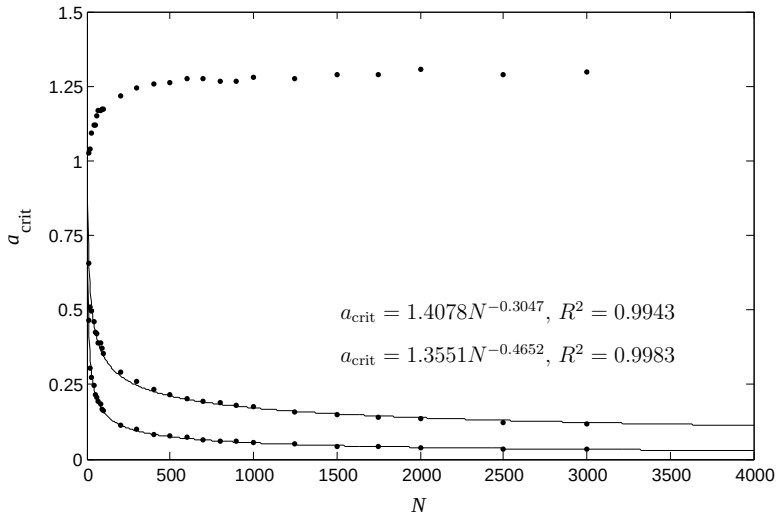


Figure 20: Evolution with N of a_{crit} for the Riesz's kernels with $s = 1, 2, 3$.

Although the relationships between a_{crit} and a_{ls} must be still more deeply analyzed, the results that we have showed here seem to indicate the "optimality" of our criterion for the

choice of the step size. In [1] we made some comments about the "optimality" of the advance direction w in comparison with the gradient direction.

Finally, we would like to comment that the structure of our algorithm can be adapted to the requirements of other classical problems. An interesting application that is being explored by the authors falls into the dynamic systems framework. In [2] we show some preliminary results of the application of our algorithm, that we call Forces' method, to the problem of the computation of planar central configurations and to the problem of the calculation of trajectories by means of action minimization. The problem of the central configurations is very similar to the Fekete point one. As for the minimization of the action, which is the integral of the difference between the kinetic and potential energies of a mechanic system, it is necessary to start by discretizing the problem by means of the numerical quadrature of that action integral, which leads to the minimization of a functional depending on a finite amount of relative distances.

Acknowledgements

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 Lacan from UPC, in whose cluster "Clonetroop" most of the intensive calculations performed for this work have been carried out.

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 and J.M. Gesto, Estimation of Fekete points, *J. Comput. Phys.*, **225** (2007), 2354-2376.

[2] E. Bendito, A. Carmona, A.M. Encinas and J.M. Gesto, Application of the Forces' method in dynamic systems, preprint, accessible in `http://www-ma3.upc.edu/users/bencar/papers`.

[3] D.P. Hardin and E.B. Saff, Discretizing manifolds via minimum energy points, *Notices Amer. Math. Soc.*, **51** (2004), 1186-1194.

[4] K.J. Nurmela, Constructing spherical codes by global optimization methods, *Helsinki University of Technology, Series A: Research Reports*, **32** (1995).

[5] E.A. Rakhmanov, E.B. Saff and Y. Zhou, Minimal discrete energy on the sphere, *Math. Res. Lett.*, **1** (1994), 647-662.

[6] E.B. Saff and A.B.J. Kuijlaars, Distributing many points on a sphere, *Math. Intelligencer*, **19** (1997), 5-11.

[7] S. Smale, Mathematical problems for the next century, *Math. Intelligencer*, **20** (1998), 7-15.

[8] N. Sloane, `http://www.research.att.com/ njas/`

[9] R. Womersley, `http://web.maths.unsw.edu.au/ rsw/Sphere/Energy/index.html`.

[10] R. Womersley, `http://web.maths.unsw.edu.au/ rsw/Torus/`.

[11] Y. Zhou, *Arrangements of points on the sphere*, Ph.D. Department of Mathematics, University of South Florida, (1995).