The Robin Hood method for electrostatic problems

P. Lazić, H. Štefančić & H. Abraham

Theoretical Physics Division, Rudjer Bošković Institute, Zagreb, Croatia

Abstract

We present a new, efficient and robust method for solving electrostatic problems. The basic idea of the method is rather simple, but has not been exploited so far. The essence of the method is achieving of the equipotentiality of the conducting surfaces by iterative nonlocal charge transfer. Besides the simple physical idea, the computational behavior of the method is very appealing. It scales linearly in memory with the number of elements and it converges geometrically without the occurrence of Critical Slowing Down. The presented method can be extended in application to other types of problems, electrostatics being a very specific example in which one can remain only on the boundaries of the objects involved in the calculation. Due to high efficiency and low resource demands, this method could prove useful in many areas that require electrostatic calculations of high precision and detail—medical applications, charged particle detector/accelerator construction, printed electronics being just some of them.

Keywords: boundary element, electrostatics, Robin Hood, nonlocal charge transfer, equipotentiality, Critical Slowing Down, real space DFT, numerical methods.

1 Introduction

The intention of this article is to give an overview of a new numerical method, the Robin Hood (RH) method for solving electrostatic problems. Implementation details, complexity analysis and other technical features of the method are only briefly sketched. A technically detailed account is available elsewhere [4]. We would like to show ideas collected along development of the RH method and our reflections upon obtained results. Possible future expansions of application are outlined.
1.1 A simple concept

It is surprising how a simple concept used as a main idea had led us towards a method with such appealing numerical properties as linear scaling in memory and geometric convergence. That sole concept, rather than its particular implementation, is where all the properties of the RH method come from and that is the reason why we describe it here using a simple example. From general physics one knows that ideal metal surfaces are equipotential in static situations. Argument for that is that the charge on the metal surface can redistribute freely and will do so until the surface is equipotential and there is no net force on the charges which could rearrange them any further. This equipotentiality principle has been used till now for example in solutions using the method of images [1], but those analytical solutions are possible only for limited number of problems, usually very symmetrical. To make use of it numerically for arbitrary metal shapes, consider a problem of an isolated neutral metal cube and a point charge positioned somewhere in the vicinity of the cube. What happens is that the charge on the surface of the cube redistributes (surface of the cube remains always globally charge neutral, but it can polarize thus having areas of positive and of negative surface charge density) in such a way that the surface of the cube is equipotential which is an equilibrium distribution and an electrostatic solution of this problem. To solve this problem numerically we suggest the following: divide the surface of the cube into finite elements (triangles for instance) and give them some initial surface charge densities (zero for example). Then, calculate potential due to finite elements of a cube and a point charge at each of triangles barycenters (barycenters of triangles being the points on the surface of the cube at which we require equal electric potential i.e. that is the condition of equipotentiality that the solution distribution should obey up to a certain precision).

To achieve the condition of equipotentiality at these points, iterate the following procedure: among the barycenter points find the one with the highest (maximum) and the one with the lowest (minimum) potential and transfer a charge from a triangle of maximum potential to the one of minimum potential in such a way that after the transfer those two triangles have exactly the same potential at their barycenters. That is the main idea of the method and such a procedure leads to the electrostatic solution even though we have not provided a strictly mathematical proof for this. The name of the method, is inspired by the spirit of the main idea, which is taking from the maximum (richest) and giving to the minimum (poorest). Electrostatic problems involving metals in vacuum is a very specific example where one can calculate potential in the whole space knowing only the surface charge distribution on metal surfaces—i.e. on boundaries of metals. Electrostatic problems also have a nice property of uniqueness of the solution which is not the case, for example in quantum problems described by Schrödinger equation where one can have more than one solution characterized by different eigenvalues. In some sense the value of the potential at which the cube from the previous example ends up is the single eigenvalue for that problem.
1.2 Implementation and convergence of the RH method

In this subsection we sketch briefly the implementation of the RH method that we used. With code based on that implementation we solved a number of problems which could not be solved until now, at least with such a precision, and show the solution for one such problem along with some values one can calculate once the surface charge distributions are found. The convergence of the method is also discussed. The most important part in the scheme of the RH method for electrostatic problems is the calculation of the electric potential due to surface charges. But before the calculation of the potential we have to divide the objects into finite elements. We choose to represent every object (surface) as a set of triangles (except for point charges, of course). One reason for choosing triangles is that the 3D graphics today is done with triangles, so it is trivial to get such representations without the need to write additional code for division of surfaces into triangles. Moreover, we chose triangles to be right-angled triangles. The reason for this is the way we calculate the potential. Namely, for the calculation of the potential, the multipole expansion for the right angled triangle is used. To guarantee a certain precision of the solution, which means the precision of the potential calculated in every point used, we would need to include the contributions of higher order multipoles. Instead of including large number of multipole terms, we decided to employ a trick which enables us to have the same precision using the multipole expansion up to quadrupole term only. The precision of the potential calculated in such a way can be compared with the analytical results, obtained by the direct integration, expressed as a function of the distance of the point in which the potential is calculated and the catheti ratio of the triangle. If the point is too close to the triangle, which means that the precision of the calculated potential is insufficient, the triangle is divided into four new triangles by the bisection of its sides. In that subdivision, the surface charge density remains the same. If any of these four new triangles still does not meet the criteria, it is further divided in a same way. The procedure is repeated until the precision criteria for the potential is met for all triangles. One such division for the calculation of the potential at point T is shown in fig. 1.

During the calculation, only the coordinates of the vertices of the triangles and the values of the potential at their barycenters are kept in the memory, so the memory requirement grows linearly with the number of triangles in calculation. Initial calculation of the potential at all barycenters due to all elements (triangles and point charges, if any) in the problem unavoidably requires a number of calculations of order of \( N^2 \). But these \( N^2 \) calculations have to be done only once, and after that the new potential values, after the charge transfer is done, are obtained only by updating current potential values due to charge transfer. This requires only a number of calculations of order of N for each charge transfer.

For detailed analysis of the complexity of the RH method and more implementation details see [4].
1.3 Solutions and convergence

To show what the RH method can do, we choose a very difficult problem by today’s methods standards. A problem solved is shown in fig. 2.
Figure 3: The speed of convergence for the system of a point charge and two insulated conducting spheres (one curve for each sphere). Similar calculation was done with point charge and only one conducting sphere. In both cases, the error decreases exponentially with the number of iterations which shows that the addition of one more separate conducting surface does not deteriorate the speed of convergence.

Convergence of the method is tested on various examples and always shows the same behavior as the one shown in the plot in fig. 3.

2 Absence of Critical Slowing Down

There are two main classes of methods for numerical solving of electrostatic problems. The first one is the class of Boundary Element Methods which as a main disadvantage has a very unpleasant $N^2$ memory requirement scaling. The other are Finite Differences Methods which are based on the relaxation of the solution on a grid at which the values of the potential are given. The main problem of that approach is that Critical Slowing Down occurs. That problem is solved by the employment of the multigrid techniques [3]. Finite Differences Methods also suffer from some technical problems: the grid has to end somewhere in space, some parts of space are more interesting than the others so the grid has to be finer there, etc. Similarity of the RH method with BEM is that it uses boundaries only and similarity with the FDM is that it relaxes the solution in a certain way. A great advantage of RH method over the BEM is the linear scaling in memory requirement and advantage over FDM is in the absence of Critical Slowing Down. To explain our understanding of the absence of CSD in RH method we sketch briefly
the basics of FDM. Consider 1D problem with grid points \( x_i \) and potential values at these points \( \phi(x_i) \). Relaxation procedure is derived using the action functional minimization theory which gives the relaxation formula for Poisson equation:

\[
\phi(x_i)^{t+1} = (1 - \omega)\phi(x_i)^t + \frac{\omega}{2} \left[ \phi(x_{i-1})^t + \phi(x_{i+1})^t + 4\pi \rho h^2 \right].
\]  

Indices \( t \) and \( t + 1 \) represent relaxation steps, for details of eqn. (1) see [3]. What is important to notice in this equation if that the value of the potential at point \( x_i \) in a next relaxation step is determined only by the current values of the potential in its nearest neighbors on the grid i.e. points \( x_{i\pm 1} \). Such an usage of local information only in relaxing the solution is what causes CSD [2]. The fact that the RH method uses global information on the system and changes the system according to it is the reason that it never slows down. To illustrate relaxation, CSD, the general multigrid idea for dealing with the CSD and the RH method approach we give a simple example. Consider pits of different depths and some amount of water in them. Water flows from one pit to the neighboring ones proportionally to the difference of water levels in them. To have the parallel with electrostatics, we suppose that the geometry of the pits and amount of water is such that only one solution of the problem exists. Also due to rules which govern the water flow, we conclude that solution must have the property that water levels in all pits are the same because only then the flow will stop. With such property we can readily employ RH approach on this problem. Let the initial distribution of water be as in fig. 4.

Water will flow rather fast at first due to large level differences. But eventually differences will become smaller and the flow will become slower. If at that point
system has excess of water in pit 1 and absence of it in pit 8, a rather large time (i.e. number of relaxation steps) will be used until the excess water form pit 1 "relaxes" into pit 8. That is due to the fact that system has certain dynamics. That dynamics is introduced in FDM by the action functional approach and it causes Critical Slowing Down and gives no additional information on the system, i.e. that dynamics is not physical, it is an artificial numerical feature. To handle the CSD in this example one can use the multigrid technique. The very general idea of it would be to represent the system at a coarser grid, for example making one averaged pit out of pits $1 - 2 - 3 - 4 = 1^*$ and the other out of pits $5 - 6 - 7 - 8 = 2^*$. Then one makes water transfer in the coarser grid from pit $1^*$ to $2^*$ and projects that transfer back to the finer grid of 8 pits. In such a way a water transfer at large distances is achieved and errors introduced by that are of short wavelengths and are therefore efficiently handled at finer grid representation. The RH method deals with this problem quite differently. It finds the minimum and the maximum of water levels (minimum level can be in a pit with no water in it yet) and makes a transfer such that levels are equal at those two pits after the transfer. A very good solution is achieved after only a few steps but RH still improves the solution with equal effectiveness. If we zoom out around the level of water as shown in fig. 5, we can see that differences of water levels still exist but are of different orders of magnitude than at the beginning of the relaxation. But the RH method still does the same, finds the minimum and the maximum and makes them equal, of course at this point much smaller quantities of water are being transferred. For the RH method the system looks the same no matter how big or small the differences in water levels are. The system far away from the solution and the one very close to it are equally well relaxed (improved) in the RH method steps, which is the reason of the geometric convergence of the RH method at all levels of error as shown in fig. 3, and that’s the reason why the RH method does not run into the CSD effect.

3 Conclusions

The RH method represents a new way of solving electrostatic problems with high precision and efficiency. When we started developing it from the simple idea of equipotentiality we hoped at best that it would work, but we were astonished when we ended up with the method that possesses such features as the absence of Critical Slowing Down. Nevertheless once we grasped the possibilities of the RH approach we could, owing to simplicity of the basic principle of the method, clearly see where all the nice properties of the method come from and why it works as well as it does. Extension to solving other types of problems requires further effort [5], but one can write virtually any differential equation in terms of generalized potential and generalized charge density. For example, the single particle Schrödinger equation can be cast in such a form with energy corresponding to the potential and the wavefunction to charge density. So far we can say that the existence of multiple solutions of a problem introduces some new interesting moments in the RH approach. We hope that the RH method will be further developed and used
Figure 5: Water in the pits after 8 steps of RH method starting from the initial
distribution shown in fig. 4. Area around average water level is zoomed in.

because the areas of application are immense and are of great practical and funda-
mental importance.

References

[2] T. Torsti, M. Heiskanen, M.J. Puska et al., MIKA: Multigrid-based pro-
gram package for electronic structure calculations, Int. J. Quantum Chem. 91
numerical method for electrostatic problems based on a nonlocal charge