ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

REDUCED BASIS METHOD FOR POISSON-BOLTZMANN EQUATION

Cleophas Kweyu[1], Martin Hess[2], Lihong Feng[3], Matthias Stein[4], and Peter Benner[5]

[1, 2, 3, 5] Computational Methods in Systems and Control Theory (CSC)
[4] Molecular Simulations and Design (MSD)
[1, 2, 3, 4, 5] Max Planck Institute of Dynamics of Complex Technical Systems
Sandtorstr. 1, 39106, Magdeburg (Germany)
e-mail: {kweyu, hessm, feng, matthias.stein, benner}@mpi-magdeburg.mpg.de

Keywords: Reduced basis method, Poisson-Boltzmann equation, ionic strength, finite differences scheme, preconditioned conjugate gradient, algebraic multigrid.

Abstract. *The Poisson-Boltzmann equation (PBE) is a nonlinear elliptic PDE that arises in biomolecular modeling and is a fundamental tool for structural biology. It is used to calculate electrostatic potentials of biomacromolecules in liquid solutions. To efficiently compute the electrostatic potential numerically, a very large domain is required to sufficiently accommodate both the biomacromolecule and the electrolyte. This yields high degrees of freedom in the resultant algebraic system of equations ranging from several hundred thousands to a few millions. This poses great computational challenges to conventional numerical techniques, especially when many simulations for varying parameters, for instance, the ionic strength, are to be run. The reduced basis method (RBM) greatly reduces this computational complexity by constructing a reduced order model of typically low dimension. We discretize the linearized PBE (LPBE) with a centered finite differences scheme and solve the resultant linear system by the preconditioned conjugate gradient (PCG) method with algebraic multigrid as the preconditioner. We then apply the RBM to the high-fidelity full order model (FOM) and present the numerical results. We notice that the RBM reduces the model order from* $\mathcal{N} = 1,614,177$ *to* $N = 6$ *at an accuracy of* 10^{-9} *and reduces computational time by a factor of approximately over* 1300.

1 INTRODUCTION

Electrostatics plays a crucial role in almost all processes that involve biomolecules in ionic solutions, for instance, in areas such as protein structural stability, biomolecular recognition, enzyme catalysis, among others [1]. Biomacromolecules such as nucleic acids and proteins have quite low internal dielectric constants ranging from 2 to 5 because of the fixed or frozen state of their constituent polar groups and side chains that cannot significantly reorient themselves in response to an applied external dielectric field [2, 3, 4]. On the other hand, water as the solvent has a high dielectric constant of around 80 because its dipoles are able to reorient freely. In principle, the Poisson-Boltzmann equation (PBE) is capable of modeling a system comprising of two dielectric materials with totally different dielectric properties. In this approach, the biomacromolecule is treated as a low dielectric cavity in which partial charges are discretely embedded at atomic positions, whereas the solvent is treated as a continuum through the Boltzmann distribution [1, 2, 3].

The PBE and its linear form have been widely applied in the field of structural biology. First, they are used to determine the electrostatic potential at the surface of a biomacromolecule. This is anticipated to provide information about the concentration of charged solute molecules in its vicinity which may, through inspection, indicate possible docking sites for biomolecules. Secondly, they determine the electrostatic potential outside the biomacromolecule that provides information on the free energy of interaction of micromolecules at different positions in the proximity of the biomacromolecule. Thirdly, they determine the free energy of a biomacromolecule or of its different states which provides information about its stability. Finally, they are used to determine the electrostatic field that is used to derive mean atomic forces which are added to the interatomic calculations in standard molecular dynamics [2, 3, 5].

In this study, we consider a charged protein molecule immersed in some electrolyte, for example, a salt solution at physiological concentration, and determine the electrostatic potential triggered by the interaction between the two particles. The kind of electrolyte considered is monovalent, in which the cations and anions are present in a ratio of $1:1$, which implies that the ionic strength equals the concentration of the ions. The ionic strength acts as a physical parameter of the LPBE, and we obtain the electrostatic potential under the variation of this parameter. The PBE is given as:

$$-\vec{\nabla}.(\epsilon(x)\vec{\nabla}u(x)) + \bar{k}^2(x)\sinh(u(x)) = (\frac{4\pi e_c^2}{k_B T})\sum_{i=1}^{N_m} z_i\delta(x-x_i), \quad u(\infty)=0 \tag{1}$$

where $u(x) = \frac{e_c\psi(x)}{k_B T}$ is the dimensionless potential and $\psi(x)$ the electrostatic potential in centimeter-gram-second (cgs) units at $x \in \mathbb{R}^3$. The dielectric coefficient $\epsilon(x)$ and the modified Debye-Hückel screening parameter $\bar{k}$, a function of the ionic strength of the solvent, are discontinuous functions at the interface between the charged biomacromolecule and the solvent, and at an ion exclusion region (Stern layer) surrounding the molecule, respectively. k_B is the Boltzmann constant and T the temperature. The term $k^2 = \frac{8\pi e_c^2 I}{1000\epsilon k_B T}$ describes ion accessibility and concentration, where I is the ionic strength given by $I = \frac{1}{2}\sum_{i=1}^{N} c_i z_i^2$. The right-hand side term represents the molecular charge density of the biomacromolecule consisting of N_m point partial charges $q_i = z_i e_c$ at atomic positions x_i and the distributions $\delta(.)$ reflect the point charge behaviour. Details on how to map the dielectric coefficient and the Debye-Hückel parameter to a computational grid can be found in [6].

The PBE (1) has some features which pose great challenges in both analytical and numerical

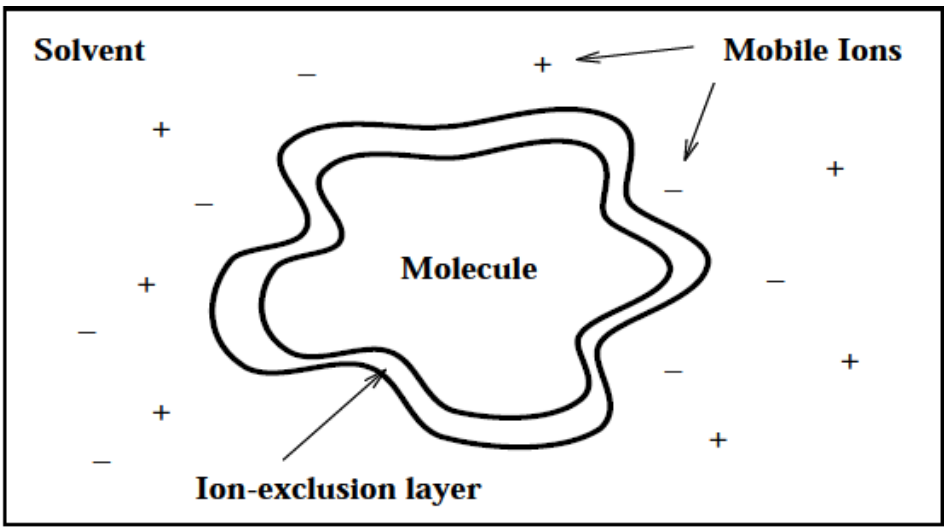

Figure 1: 2-D view of the 3-D Debye-Hückel model, a reproduction of Figure 1.1 in [7].

approaches. These include the infinite (unbounded) domain, delta functions, rapid nonlinearity, and discontinuous coefficients [7, 8]. Figure 1 shows the two dimensional view (cross-section) of the three dimensional Debye-Hückel model of electrostatic interactions between a biomolecule and a salt solution [7]. Proteins are not highly charged as compared to nucleic acids and, therefore, it suffices to consider the linearized PBE (LPBE) and still obtain accurate results. The PBE can be linearized under the assumption that the electrostatic potential is very small (relative to k_BT) or that the ions' electrostatic energy is much less than their thermal energy [1]. Therefore, the nonlinear function $\sinh(u(x))$ can be expanded by a Taylor series and only the first term is retained. We obtain the linearized PBE (LPBE) given by;

$$-\vec{\nabla}.(\epsilon(x)\vec{\nabla}u(x)) + \bar{k}^2(x)u(x) = (\frac{4\pi e_c^2}{k_BT})\sum_{i=1}^{N_m} z_i\delta(x-x_i). \tag{2}$$

However, we must note that the LPBE yields very inaccurate results for highly charged biomacromolecules such as nucleic acids (DNA and RNA) or polylysine, though it can be applied to low charged biomacromolecules such as proteins [3]. More information about the PBE, including its derivation from first principles can be found in [7].

Many different kinds of numerical methods have been used to solve the PBE and LPBE [9]. The finite difference method (FDM) is the most widely used method and has been implemented in many programs, for example, Amber, Delphi, and adaptive Poisson-Boltzmann solver (APBS) [10]. Other methods such as the finite element method (FEM), the boundary element method, the interface method, and an adaptive method have also been developed and a thorough review can be found in [9]. APBS, for instance, has versions for both FDM and FEM. As stated earlier, an infinite domain, $u(\infty) = 0$, coupled with the fact that molecules of interest such as enzymes and proteins have a spatial scale of 10 to over 100 Ångströms, provides a great challenge to numerical techniques. Therefore, it is customary to choose a truncated domain of at least threefold the size of the biomacromolecule to accurately approximate the boundary conditions [7]. Nevertheless, this leads to a very large algebraic system which becomes computationally expensive if many simulations for different ionic strengths have to be run. The reduced basis method may largely reduce the computational complexity due to its capability of constructing a reduced order model of typically low dimension, which can replace

the high-fidelity system in the multi-query tasks, and can be solved at a low computational cost [11]. In this paper, we apply the reduced basis method to the system of PBE, and have achieved significant acceleration of orders of magnitude.

The outline of this paper is as follows: In Section 2, we provide the basics of RBM which include the problem formulation, the solution manifold, the greedy algorithm, and the *a posteriori* error estimation. In Section 3 we provide numerical results of the FOM (via the FDM) and those of the ROM. Conclusions and future work are given in the end.

2 ESSENTIALS OF THE REDUCED BASIS METHOD

RBM is a projection-based parametrized model order reduction (PMOR) technique that aims at computing numerical solutions to parametrized PDEs in real-time and many-query scenarios by exploiting an offline/online procedure which ensures accurate approximation of the truth solution in a rapid and inexpensive manner. For a thorough review, see [12]. We consider a physical domain $\Omega \subset \mathbb{R}^3$ with boundary $\partial\Omega$, and a parameter domain $D \subset \mathbb{R}$. Since the exact solution is seldom available, we discretize the PDE (LPBE) with a centered finite differences (FD) scheme. The discrete problem of the LPBE now becomes, for any $\mu \in D$, find $u^{\mathcal{N}}(\mu)$ that satisfies the linear system

$$A(\mu)u^{\mathcal{N}}(\mu) = f, \quad \mu \in D, \tag{3}$$

where $A(\mu) \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ and $f \in \mathbb{R}^{\mathcal{N}}$. We notice that the $\mathcal{N}\times\mathcal{N}$ system is indeed computationally expensive to solve for an accurate approximation of $u(\mu)$ and therefore, we apply model order reduction (in this case, RBM) to save computational costs by providing an accurate approximation of $u^{\mathcal{N}}(\mu)$ at a greatly reduced dimension $N \ll \mathcal{N}$.

2.1 The solution manifold and the greedy algorithm

A very important assumption in RBM is the assumption that a typically very low dimensional solution manifold already covers all the high-fidelity solutions of (3) under variation of parameters [11]:

$$\mathcal{M}^{\mathcal{N}} = \{u^{\mathcal{N}}(\mu) : \mu \in D\}. \tag{4}$$

The RB approximation space is then built upon this solution manifold and is given by the subspace spanned by the snapshots, i.e. the subspace spanned by the high-fidelity $u^{\mathcal{N}}(\mu)$ solutions corresponding to a number of samples of the parameters

$$\text{range}(V) = \text{span}\{u^{\mathcal{N}}(\mu_1), ..., u^{\mathcal{N}}(\mu_N)\}, \quad \forall \mu_1, ..., \mu_N \in D. \tag{5}$$

The greedy algorithm is used to generate the reduced basis space (5) through an iterative procedure where a new basis is computed at each iteration. The procedure, as given by [13] is shown in Algorithm 1. The RB approximation is then formulated as, for any given $\mu \in D$, find $u_N(\mu) \in X_N$ which satisfies

$$A_N(\mu)u_N(\mu) = f_N, \tag{6}$$

where $A_N = V^T AV$ and $f_N = V^T f$. V is the orthonormal matrix computed from the greedy algorithm. From the fact that $N \ll \mathcal{N}$, solving the small dimensional ROM is much cheaper than solving the high-fidelity model (3) [11]. Efficient implementation of Algorithm 1 depends on an efficient error estimation $\Delta_N(\mu)$ of the ROM. In the next subsection, we introduce an *a posteriori* error estimation derived from the residual of the approximate RB solution.

Algorithm 1 Greedy algorithm

Input: A training set Ξ including all the samples of μ, i.e., $\Xi := \{\mu_1, \ldots, \mu_m\}$.
Output: RB basis represented by the projection matrix V.

Choose $\mu^* \in \Xi \subset D$ arbitrarily
Solve (3) for $u^{\mathcal{N}}(\mu^*)$
$S_1 = \{\mu^*\}$
$V_1 = [u^{\mathcal{N}}(\mu^*)]$
$N = 1$
while $\max_{\mu \in \Xi} \Delta_N(\mu) \geq \epsilon$ **do**
    $\mu^* = \arg\max_{\mu \in \Xi} \Delta_N(\mu)$
    Solve (3) for $u^{\mathcal{N}}(\mu^*)$
    $S_{N+1} = S_N \cup \mu^*$
    $V_{N+1} = [X_N \quad u^{\mathcal{N}}(\mu^*)]$
    Orthonormalize the columns of V_{N+1}
    $N = N + 1$
end while

2.2 *A posteriori* error estimation

A posteriori error estimators are computable indicators that estimate the actual solution error by utilizing the residual of the approximate RB solution. An error estimator is required to fulfill three major properties: it is supposed to be sharp (close to the unknown actual error), asymptotically correct (tend to zero with increasing RB space dimension N, at a similar rate as the actual error), and computationally cheap. Therefore, they guarantee both reliability and efficiency of the reduction process [14].

We first compute the residual;

$$\begin{aligned} r_N(u_N;\mu) &= f - A(\mu)u_N(\mu) \\ &= A(\mu)u^{\mathcal{N}}(\mu) - A(\mu)u_N(\mu) \\ &= A(\mu)e(\mu) \end{aligned} \tag{7}$$

where the error $e(\mu) := u^{\mathcal{N}}(\mu) - u_N(\mu)$ is given by

$$e(\mu) = (A^{\mathcal{N}}(\mu))^{-1} r_N(u_N;\mu) \tag{8}$$

We obtain an upper bound for the 2-norm of the error by taking the 2-norm on both sides of equation (8), that is;

$$\|e(\mu)\|_2 \leq \|(A^{\mathcal{N}})^{-1}(\mu)\|_2 \|r_N(u_N;\mu)\|_2 = \frac{1}{\sigma_{min}(A^{\mathcal{N}}(\mu))} \|r_N(u_N;\mu)\|_2 =: \Delta_N(\mu), \tag{9}$$

where $\sigma_{min}(A^{\mathcal{N}}(\mu))$ is the smallest singular value of $A^{\mathcal{N}}(\mu)$ [14]. The quantity $\Delta_N(\mu)$ is a rigorous error bound, and can be used to select snapshots within the greedy algorithm in the offline stage and consequently to measure the accuracy of the RB approximation in the online stage [13]. For efficient computation of the norm of the residual and error bounds, see [15, 14]. It is computationally expensive, especially in the offline-online context, to compute $\sigma_{min}(A^{\mathcal{N}}(\mu))$ as it entails solutions of large-scale eigenvalue problems [13]. Therefore, in our computations, we use the norm of the residual as our error estimator, an upper bound which satisfies the inequality

(9) and provides an estimation of the true error that works well for our problem. It also provides rapid convergence as depicted in the numerical results in Figures 6 and 7. It is given by

$$\|e(\mu)\|_2 \approx \|r_N(u_N;\mu)\|_2 = \Delta_N(\mu). \tag{10}$$

3 NUMERICAL RESULTS

3.1 Finite differences results

We consider the LPBE in (2), a parameter domain $\mu \in D = [0.05, 0.15]$, and a physical domain $\Omega = [7.4415, 64.5655]\text{Å} \times [-1.5120, 44.2180]\text{Å} \times [-20.2755, 35.8775]\text{Å}$. The parameter domain is chosen for a feasible physiological process and μ resides in the second term in the kappa function. Information about the molecular charge density (right-hand side of the LPBE) was obtained from a PQR file provided by the MSD group at our institute. The file contains 1228 protein atoms (constituents of amino acid residues) at various atomic positions (P), with respective charges (Q), and radii (R). The PQR file is generated from the PDB (Protein Data Bank) file by the PDB2PQR software which allows users to add charges and assign atomic radii as additional parameters. The PDB file provides a standard representation for macromolecular structure data (in terms of atomic coordinates), derived from X-ray crystallography and nuclear magnetic resonance (NMR) studies for electrostatic potential analysis [16].

We discretize the LPBE with a centered finite differences scheme and the resulting parametrized linear system (3) is of more than 1.6×10^6 degrees of freedom. We solve this full order model (FOM) by the preconditioned conjugate gradient (PCG) method with algebraic multigrid as the preconditioner at different samples of ionic strength. We use a tolerance of 10^{-10} for the PCG and the solution converges within 50 iterations. The choice of this tolerance directly affects that of the greedy algorithm proportionally and therefore, it is prudent to ensure that the truth solution is highly accurate. Zero Dirichlet boundary conditions are used to avoid the dependence of the boundary data on the parameter as well. Figures 2-5 below plot cross-sections of the $z-$ axis and show the exponential decay of the high-fidelity solutions $(u^{\mathcal{N}}(\mu), \mathcal{N} = 1, 614, 177)$ with variation of the parameter μ. This is attributed to the large force constant (332 kcal/mol) of electrostatic interactions. In the absence of ions (that is, at $\mu = I = 0$), these interactions are long ranged, but in the presence of ions (that is, $\mu > 0$), they are damped and gradually decay to zero [1]. The computational time taken to obtain the high-fidelity solution $u^{\mathcal{N}}(\mu)$ is on average approximately 90 seconds, and varies depending on the value of the ionic strength used.

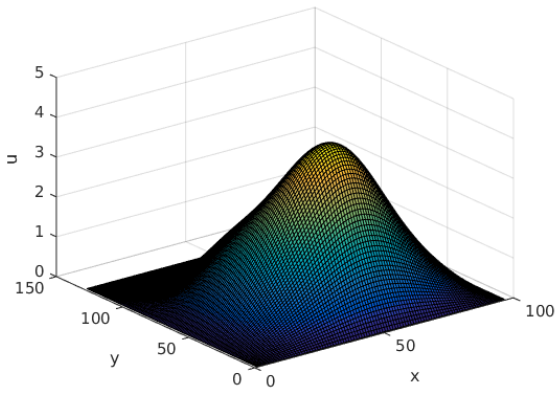

Figure 2: $u^{\mathcal{N}}(\mu)$ at μ = 0.05

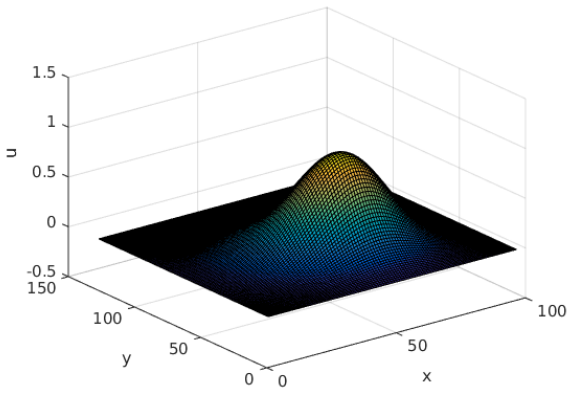

Figure 3: $u^{\mathcal{N}}(\mu)$ at μ = 0.5

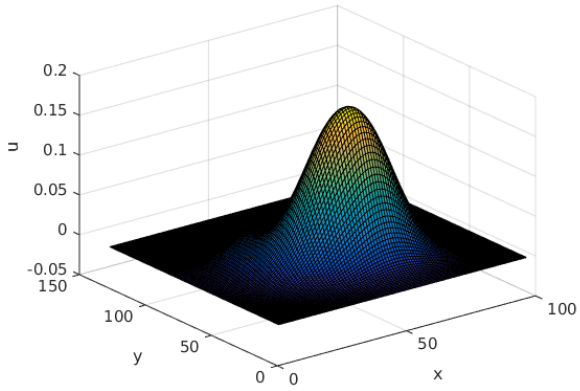

Figure 4: $u^{\mathcal{N}}(\mu)$ at $\mu = 1.5$

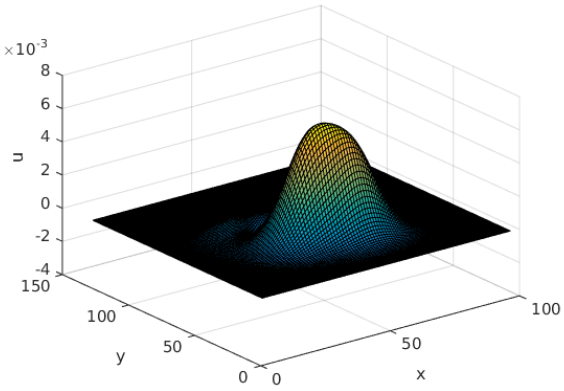

Figure 5: $u^{\mathcal{N}}(\mu)$ at $\mu = 5$

3.2 Reduced basis results

The *a posteriori* error estimates the true error between the high-fidelity solution and the reduced basis solution. It also provides the exponential decay of the error in the greedy algorithm until a reasonable tolerance is achieved. Figures 6 and 7 show the decays of the error estimator and the true error during the greedy algorithm at the current RB dimension $i = 1, \ldots, N$. They satisfy the asymptotic correctness property stated in subsection 2.2, and it is evident that the error estimator is an upper bound to the true error. We also observe a high convergence rate of the error estimator with up to two orders of magnitudes. The error estimators in Figures 6 and 7 are the maximal error and relative maximal error, respectively, and are defined as, maximal error: $\Delta_N^{\max} = \max_{\mu \in \Xi} \|r_N(u_N;\mu)\|_2$, relative maximal error: $\frac{\Delta_N^{\max}}{\|u_N(\mu^*)\|_2}$, $\mu^* = \arg\max_{\mu \in \Xi} \|r_N(u_N;\mu)\|_2$.

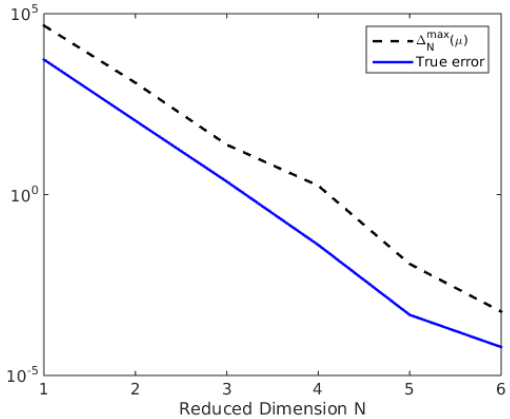

Figure 6: Maximal versus true error

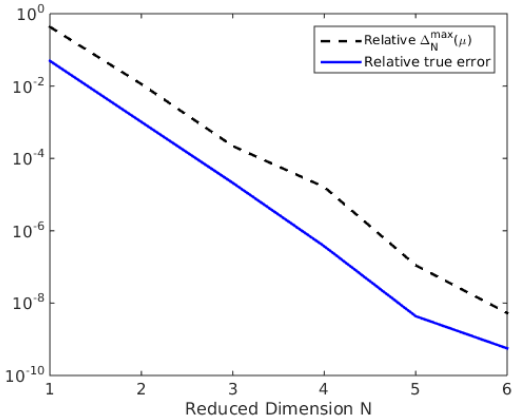

Figure 7: Relative maximal versus true error

Figure 8 shows the error estimator and the true error of the finally constructed reduced model over the whole parameter domain $\mu_i \in \Xi$ for $i = 1, ..., 100$. It is evident that the error estimate for the final RB approximation at dimension $N = 6$ is indeed an upper bound of the true error and a trend that both quantities behave similarly is clearly visible from the graphs. The

computational time taken to obtain the approximate solution $u_N(\mu)$ is on average approximately 0.065 seconds. Figure 9 validates the true error of the ROM of dimension $N = 6$, i.e., the final ROM constructed by the greedy algorithm, at some 100 randomly selected parameter values $\mu \in D$ which are different from those in Ξ, used in Algorithm 1. It is evident that the true error is still smaller than the error estimate.

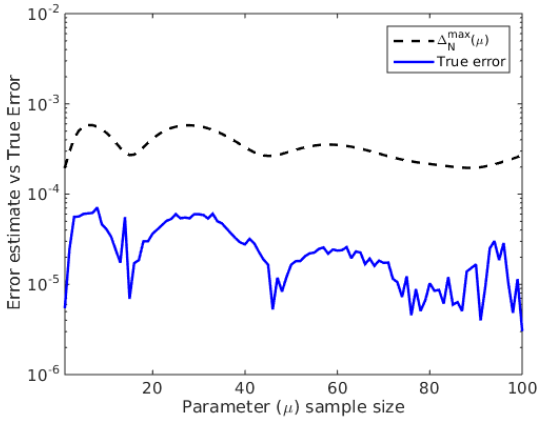

Figure 8: Error estimate versus true error

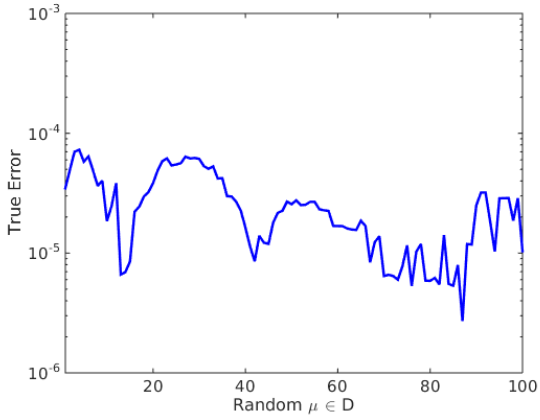

Figure 9: True error

4 CONCLUSIONS

In this paper, we have presented the RBM and its application to the LPBE with ionic strength as the meaningful parameter. The RBM reduces the high dimensional full order model by a factor of $269,029.5$ and the computational time by a factor of approximately over 1300. The error estimator provided fast convergence to the reduced basis approximation. Our future research is on employing non-zero Dirichlet boundary conditions from Debye-Hückel model for a spherical non-polarizable ion with uniform charge. These boundary conditions also vary with the parameter, which is non-affinely inherent in them and therefore we need to employ the discrete empirical interpolation method (DEIM) for a speed-up in the online stage.

5 ACKNOWLEDGEMENTS

We would like to thank the following organizations for financial and material support on this project: International Max Planck Research School (IMPRS) and Max Planck Society (MPI).

REFERENCES

[1] F. Fogolari, A. Brigo, H. Molinari, The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition*, **15**, 377–392, 2002.

[2] F. Fogolari, P. Zuccato, G. Esposito, P. Viglino, Biomolecular electrostatics with the linearized Poisson-Boltzmann equation. *Biophysical Journal*, **76**, 1–16, 1999.

[3] B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry. *Science, New Series*, **268**, 1144–1149, 1995.

[4] http://www.edusoft-lc.com/zap/manuals/100/chaptwo.html. Last visited 2016-02-07.

[5] K.A. Sharp, B. Honig, Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Chem.*, **19**, 301–332, 1990.

[6] N.A. Baker, M.J. Holst, F. Wang, The adaptive multilevel finite element solution of the Poisson-Boltzmann equation on massively parallel computers. *IBM J. Res. Devel.*, **45**, 427–438, 2001.

[7] M.J. Holst, *Multilevel methods for the Poisson-Boltzmann equation, PhD Thesis*. Numerical Computing group, University of Illinois at Urbana-Champaign, 1993.

[8] M. Holst, N. Baker, F. Wang, Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I: algorithms and examples. *Journal of Computational Chemistry*, **21**, 1319–1342, 2000.

[9] B.Z. Lu, Y.C. Zhou, M.J. Holst, J.A. McCammon, Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. *Commun. Comput. Phys.*, **3**, 973–1009, 2008.

[10] J. Wang, R. Luo, Assessment of linear finite difference Poisson-Boltzmann solvers. *J. Comp. Chem.*, **31**, 1689–1698, 2010.

[11] J.L. Eftang, *Reduced basis methods for parametrized partial differential equations, PhD Thesis*. Norwegian University of Science and Technology, Trondheim, 2011.

[12] P. Benner, S. Gugercin, K. Willcox, A survey of projection-based model order reduction methods for parametrized dynamical systems. *SIAM Review*, **57**, 483–531, 2015.

[13] M.W. Hess, P. Benner, Fast evaluation of time-harmonic Maxwell's equations using the reduced basis method. *IEEE Transactions on Microwave Theory and Techniques*, **61**, 2265–2274, 2013.

[14] A. Quarteroni, A. Manzoni, F. Negri, *Reduced basis methods for parametrized partial differential equations: an introduction*. Springer International Publishing, 2015.

[15] J.S. Hesthaven, G. Rozza, B. Stamm, *Certified reduced basis methods for parametrized partial differential equations*. Springer International Publishing, 2015.

[16] http://www.poissonboltzmann.org/docs/file-format-info/. Last visited 2016-02-25.