Maria Emelianenko

masha@cmu.edu

http://www.math.cmu.edu/~masha

Research Statement

My research is in the area of applied mathematics with main focus on scientific computing and development of efficient models and numerical algorithms. My work reflects an interplay of rigorous mathematical theory with the complexity of real life phenomena. I have been fortunate to be part of several interdisciplinary projects and gained valuable research experience working with colleagues from materials science, physics, chemistry, statistics, computer science and engineering.

1. Overview of research interests.

- The focus of my recent work has been on the development of novel **statistical theories for microstructure evolution**. By investigating the intimate relationship between grain growth dynamics and fractional continuous time random walk theory, I found myself involved in various aspects of **stochastic modeling** and uncovered the wealth of fascinating research topics associated with this field. I am actively involved with the Carnegie Mellon NSF Materials Science Research and Engineering Center (MRSEC) dedicated to the understanding, control and optimization of interface dominated materials properties. This work triggered my interest in biological, chemical and medical applications and jump started several fruitful collaborations in those areas.
- Some other directions of my current research include the design of fast new algorithms for **computational geometry** and **computational materials science**, with the use of the concepts like clustering, centroidal Voronoi tessellations and optimization methods for the determination of phase diagrams for multicomponent materials. In studying these topics, the ideas from the theories of multigrid methods and adaptive computations have been extended to solve wider classes of nonlinear problems. I have been an active member of the MatCASE project at Penn State which is funded by a major NSF-ITR grant to develop computational tools for multicomponent materials design in support of National Priorities.
- My earlier work included the analysis of multidimensional birth-death processes in **industrial applications** and development of efficient pricing schemes in next generation telecommunication networks, linear algebra with the focus on solving **ill-conditioned systems** and **mathematical models in biology**.

Close collaboration with colleagues in other fields of science, technology and engineering is a constant source of motivation for me. It poses new challenges that are rich in scale and complexity, some of which are suitable for undergraduate or graduate students, while others require a coordinated team effort of dozens of people. Below I provide a more detailed description of my ongoing research projects and some directions to be pursued in the near future.

2. Mathematical modeling of complex systems

2.1 Materials science applications.

Recent years have witnessed a changing paradigm in experimental science: automated data acquisition technologies, now practiced in disciplines as varied as materials science and molecular biology, allow vast interrogation at certain scales. Typically most interesting are those mesoscales rich in information. The yield has been huge amounts of data being collected in many scientific disciplines, demanding novel approaches for interpretation. These advances pose new challenges for our understanding of such systems through mathematical modeling, simulation, and analysis. Taking part in the Carnegie Mellon MRSEC initiative, I have been exposed to a wide range of such problems.

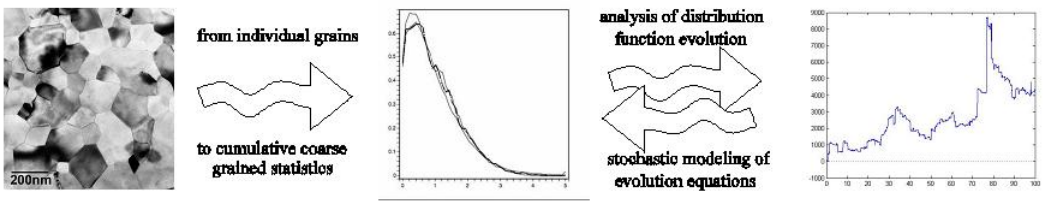

Figure 1: From individual grain statistics to stochastic analysis of distributions

In particular, many interesting questions remain about the evolution of materials microstructure. We know that most technologically useful materials are polycrystalline, comprised of many small grains separated by interfaces, called grain boundaries. The energetics and connectivity of this network of interfaces play an important

role in many material properties and across many scales. Preparing arrangements of grains and boundaries, a texture, suitable for a given purpose is a central problem in materials science. However, the mechanisms by which the distribution develops from an initial population are not yet understood. As a polycrystalline configuration coarsens, facets are interchanged and some grains grow larger and others disappear. We believe that these critical events determine the evolution of our distribution. This is likely because the system coarsens by the motion of the triple junctions, with low energy boundaries sweeping out those of higher energy. The triple junction population, in turn, is determined by the critical events. The regular evolution of the network is governed by the Mullins Equations of curvature driven growth [1] supplemented by the Herring Condition force balance at triple junctions, a system of parabolic equations with complementing boundary conditions, that has been successfully modeled via simulations [2]. The critical events, on the other hand, are stochastic. Figure 1 shows the three phases of the modeling process.

The major difficulty in developing a theory of the grain boundary character distribution lies in the lack of understanding of these stochastic events. As a postdoc at the Center for Nonlinear Analysis working with Prof. Kinderlehrer and Prof. Ta'asan and closely collaborating with Prof. Golovaty from University of Akron, Prof. Rohrer and Prof. Rollett of the Materials Science Department at Carnegie Mellon, I concentrated my efforts on the mechanisms governing these processes and investigated a simplified 1-d model that served as a surrogate, exhibiting the main features of the interacting grain boundary network.

Motivated by the apparent presence of stationary distributions coming out of the statistical analysis of such a system, we considered theoretical frameworks capable of describing it. One possible such framework based on statistical mechanics is introduced in [3], where we constructed a Boltzmann-type equation modeling grain interaction, which can successfully reproduce simulation data in long time scale with the adequate choice of parameters and has a good potential for generalization to higher dimension. In fact, it appears that this approach can yield even better results in 2-d due to the the restrictions that 1-d problem poses on grain interactions, in contrast to its visual simplicity. In [4], we turned our attention to a stochastic description by means of jump processes and coupled it with the traditional continuous time random walk theory, which turned out to be the most flexible way to handle the complexity observed in this type of jump processes. We uncovered the fractional nature of the grain kinetics and proposed unified approach to model it in terms of generalized fractional master equations. In addition, we discovered that there exists a time variable scaling that puts our model into the scope of homogeneous jump processes with asymmetric jump size distributions and makes it amenable to the wide range of stochastic analysis tools. The generalized master equation approach developed through CTRW theory is uniform for any dimension of the state space and has been confirmed to produce statistics that matches that obtained by simulation.

This new view on the stochastic behaviour of grain boundaries opens new research directions and provides powerful mathematical tools for the analysis of grain growth phenomenon in general. In fact, the scope of this work goes far beyond the realm of materials science applications. Same tools of stochastic analysis are applicable to a large class of biological, medical and physical applications, that have recently started to play an increasing role in my research.

2.2 Medical and chemical applications.

Stochastic modeling can be successfully used to predict outbreaks in patients with various deseases. Working with Prof. Ta'asan and scientists from the University of Pittsburgh Medical Center, I have been involved in studying one such model based on the data from patients suffering from the Bipolar disorder, often called manic-depression. This desease is characterized by abnormal brain functioning that results in severe changes in mood, energy, and performance. The idea is to develop a quantitative understanding of the illness that addresses clinical questions and includes ways to predict patients reaction to medications and possible outcomes. Based on preliminary studies, the ups and downs of patients with bipolar disorder can be modeled via stochastic differential equations with Laplacian jump distributions, versus Gaussian fluctuations observed in normal patients behavior. If successful, similar techniques will be applicable to a wide range of medical problems, and seem have great potential for facilitating the process of diagnosis and treatment of a disease. In the future, I hope to have a chance to continue collaborating with scientists and physicians in these exciting areas of clinical research.

Recently I also started collaborating with Dr. Ryabov from the Chemistry Department of Carnegie Mellon on green oxidation chemistry problems. Achieving design control over activity and half-life of small-molecule synthetic homogeneous oxidation catalysts is very important for advancing the utility of oxidation catalysts and minimizing hazards associated with their use and disposal. We are working on a general approach to allow for the simultaneous determination of the rate parameters in this model. The type of ODE systems appearing in these

applications resembles the Lotka-Volterra equations which I studied before in the area of mathematical biology, but they often contain additional interaction terms that prevent simple solution. Many open questions remain pertaining to the type of equilibria to be found in these systems, as well as development of stable numerical solution techniques, which I am trying to explore.

3. Computational algorithms in geometry, materials science and engineering

One of the foci of my Ph.D research was related to the Centroidal Voronoi tessellations (CVTs) which are special Voronoi tessellations of a bounded geometric domain such that the generating points of the tessellations are also the centroids (mass centers) of the corresponding Voronoi regions with respect to a given density function [9]. Centroidal Voronoi Tessellations may also be defined in more abstract and more general settings. Figure 2 gives some examples of such tessel Due to the natural optimization properties enjoyed by CVTs, they have many applications in diverse fields such as data compression, optimal quadratures, vector quantization, image analysis, clustering, resource distribution, sensor networks, cellular biology, territorial behavior of animals, mesh generations and numerical solution of PDEs.

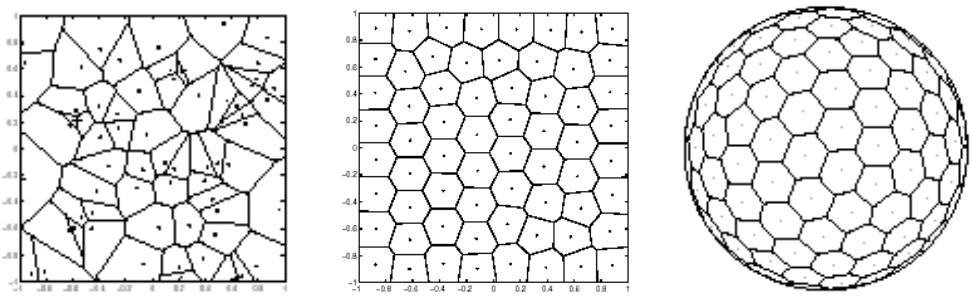

Figure 2: Voronoi tessellation on a square (left), CVT on a square (middle) and on a sphere (right) for constant density

3.1. Theoretical analysis of CVTs.

For a given domain, the Lloyd algorithm originally proposed for the optimal vector quantizer design [10] is one of the most popular iterative schemes for computing the CVTs. It is a simple iteration between generators and mass centroids of Voronoi cells, but its theoretical analysis is far from complete. In our recent work [8], several new convergence theorems were established, including the global convergence of subsequences for any density functions, the global convergence of the whole sequence in one space dimension, and the global convergence under some non-degeneracy conditions. We also performed some theoretical studies on the local convergence properties of the Lloyd algorithm including estimates on the convergence rates. The geometric rate of convergence was shown to be achieved under strong type of log-concavity conditions.

For practical applications of CVTs, such as in quantization and clustering analysis, the Lloyd method slows down for large values of k, the total number of generators. Even in the 1-d case, both our theoretical estimates and the experiments indicate that the convergence of Lloyd iterations is at most linear with the rate on the order of $1 - c/k^2$. Our efforts have been directed at the design of new computational schemes with superior convergence properties. The research in this direction resulted in two sets of algorithms. One of them explores the coupling of Lloyd scheme with Newton like methods [5]. We have both numerical and analytical results justifying super-linear convergence of the 1-d and multi-dimensional algorithms inside the convergence region, which can be reached after some initial Lloyd iterations.

Another work of ours is the development of multigrid methods in a nonlinear energy-based optimization setting [6]. The problem of constructing a CVT is nonlinear in nature and hence cannot be analyzed using standard linear multigrid approach. Some recent attempts by other groups to construct a multigrid method for quantization problems via conventional methods have resulted only in limited success for some 1-d problems. We hope to overcome the difficulties of this approach by essentially relying on the energy minimization. Since the energy functional is in general non-convex, a dynamic nonlinear preconditioner is proposed to relate our problem to a convex optimization problem. In the case of 1-d problems, we have shown that the nonlinear multigrid algorithms enjoy uniform convergence properties independent of k, the problem size, thus a significant speedup comparing to the Lloyd iteration is achieved.

Most recently, collaborating with Prof. Ludmil Zikatanov at Penn State, we investigated the question of extending our theoretical framework to higher space dimension. The general construction remains unchanged, since we have established proper relationship between CVTs, optimization problems and dynamic nonlinear preconditioning. However, the problem of minimizing the energy functional with multiple local minima is far more complex in 2-d and requires an efficient coarsening procedure in order to preserve the efficiency of the method. We have established that changes in the incidence matrix may have an adverse effect on the convergence rate of the iteration. However, when used as a local accelerator, the method exhibits all the superior features attributed to the multigrid theory, including uniform convergence with respect to the problem size. In [7], we demonstrate the successful use of this multigrid local acceleration routine for both linear and nonlinear densities in the square domain.

3.2. Discrete CVT applications

Recently I started a collaboration with physicists from a group at Rutherford Appleton Laboratory led by Dr. Ibon Bustinduy. The group's main interest concerns neutron spectroscopy data where signals span a 4-d space comprised by three spatial plus the time dimension. A priori binning schemes are inadequate in this setting since physically relevant information is usually concentrated within rather small regions of such a space. The group aims on generating optimally-binned 2D surfaces to enhance the experimenters ability to carry out searches within a 4D space. My role in the project consists in providing mathematical expertise in the area of CVT-based methods and investigating the possibility of accelerating the binning process by using a discrete analogue of the multilevel quantization scheme.

In a working paper, we suggested and implemented a multilevel strategy for accelerating and improving the CVT solution obtained by the usual Lloyd scheme, along the lines of the method proposed in [7]. We have demonstrated that this approach carries computational advantages, at the same time improving the quality of the solution. This work is the first in a series of investigations I plan to undertake in the future to test the effectiveness of our novel numerical scheme in various discrete and continuous contexts.

The methods and tools used in our analysis of CVT related algorithms are applicable in a wide range of scientific areas, from clustering, data mining and image analysis to mesh generation and domain decomposition and can be adjusted to various optimization contexts. The research in this area triggered my interest in the interplay of these methods with large scale problems, parallelization techniques and GRID computing, and motivated me to pursue the Graduate minor in High Performance Computing. During the study in the minor, I completed several major computational projects, which included the parallel implementation of Voronoi diagram construction in MPI using divide-and-conquer and sweepline methods along with several 3d Monte Carlo molecular simulations.

3.3. Algorithms for computational materials science

I have been a part of the Penn State Computational Methods for Materials Design team working on the NSF supported ITR project on multicomponent materials design and simulation. My primary role in the project involving researchers from Penn State, Ford and NIST, was to help bring algorithmic innovations and advances to material science applications. Collaborating with Prof. Zi-Kui Liu from the Materials Science and Engineering Department, I have recently made progress on developing robust and reliable numerical methods for phase diagram computation. Phase diagrams are visual representations of the phase equilibrium in a material as a function of temperature, pressure and concentrations of the constituent components and are frequently used as basic blueprints for materials research and development. Phase equilibria are obtained via minimization of the total Gibbs energy of a system by adjusting the compositions and amounts of all individual phases in the system. As in any minimization procedure, the starting values play an important role due to the existence of many possible metastable states.

Many existing software packages lack the ability to automatically determine system properties from initial data and can produce metastable equilibria instead of stable ones or simply diverge if initial guess is not good enough. Such situations are especially common to occur when the Gibbs energy exhibits multiple minima, the situation referred to as a miscibility gap. Figure 3 shows how much the results of calculations made by a widely used package Thermocalc ([13]) can differ if the existence of a miscibility gap is not known a priori. Several algorithms were proposed to automate the process of finding suitable starting positions (e.g. [14], all of which carry an increased computational cost. We made an attempt to improve on the existing strategies for automating phase diagram calculations by introducing a novel reduced complexity algorithm based on adaptive critical point detection approach. The main advantage of the new scheme lies in its ability to effectively reduce

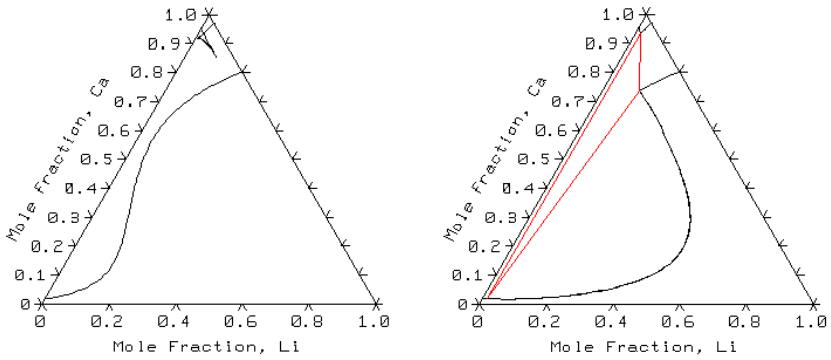

Figure 3: Incorrect (left) phase diagram of a Ca-Li-Na system produced by Thermocalc when miscibility gap is not identified vs. correct diagram (right)

the total number of trial calculations by recognizing the importance of geometry specific properties of the Gibbs energies. A set of numerical routines and a suite of Matlab programs was developed implementing the proposed algorithm for binary and ternary systems. The description of the algorithm together with the analysis of its performance in comparison with widely spread commercial packages Thermocalc and Pandat were presented in [11]. Generalization of the algorithm to higher dimensions without loss of efficiency requires the use of more advanced sampling schemes. In an upcoming publication [12], we incorporate advanced multidimensional sampling schemes to make phase diagram calculations more computationally effective.

My work in the field of numerical methods for materials science gave me a chance to apply and enhance my computational, optimization and statistical analysis skills, at the same time making me broaden my knowledge of thermodynamic properties of multicomponent systems and opening a completely new and exciting field of mathematical problems in the area of multiscale analysis of material properties.

3.4. Industrial applications of queuing theory

For a couple of years, I have worked with Prof. N. Gautam and his colleagues from the Department of Industrial and Manufacturing Engineering at Penn State on the development of an effective pricing scheme for the Next Generation Networks. We modeled connections to Next Generation Networks using a bandwidth sharing mechanism. With the help of queuing theory, we were able to define an appropriate pricing scheme to charge the users for their use of the network, depending on the class of traffic and required Quality of Service (minimum bandwidth). An optimization problem was formulated to determine the optimal resource allocation in terms of minimum bandwidth, subject to a call-blocking probability QoS constraint. We solved the problem for a general N-class system by modeling the system as an N-dimensional birth and death process. In our paper [15], we showed that in order to avoid degeneracy issues, one needs to make a novel modification to the Continuous Time Markov Chain by suitably inserting dummy nodes and perturbing the corresponding rate matrix. This approach has never been used before and it has generated huge interest in the community due to its robustness and computational efficiency. We also provided a mathematical framework for describing the interaction between user behavior and pricing structures. This paper has been rated as one of the Top 10 downloads from the Performance Evaluation journal website.

The mathematical analysis of industrial engineering applications is far from being trivial, with many challenging questions still left to answer. On the other hand, due to its high practical importance, it can serve as a great motivator for mathematics students starting their career in applied mathematical field.

Overall, my experience of working with scientists from many different fields helped me develop enormous interest in mathematical methods for physical sciences and engineering applications and greatly improved my research and interdisciplinary communication skills. I believe that these invaluable skills together with my diverse background and broad interests will help me maintain a successful research program in the future.

References

[1] Mullins, W.W. "Two dimensional motion of idealized grain boundaries", J. Appl. Phys., 27, 900 - 904, 1956

[2] Kinderlehrer, D., Livshits, I., Rohrer, G.S., Ta'asan, S., and Yu, P. "Mesoscale evolution of the grain boundary character distribution, Recrystallization and Grain Growth", Materials Science Forum vols 467-470, 1063-1068, 2004

[3] M. Emelianenko, D. Golovaty, D. Kinderlehrer, S. Ta'asan, "Boltzmann-type kinetic approach to grain growth dynamics", in preparation for submission, 2006

[4] M. Emelianenko, D. Golovaty, D. Kinderlehrer, S. Ta'asan, "Toward a statistical theory of texture", in preparation for submission, 2006

[5] Q. Du, M. Emelianenko *"Acceleration schemes for computing the centroidal Voronoi tessellations"*, Numer. Linear Algebra Appl.,13, Issue 2-3 (Special Issue on Multigrid Methods) (2006), 173-192

[6] Q. Du, M. Emelianenko *"Uniform convergence of a nonlinear energy-based multilevel quantization scheme via centroidal Voronoi tessellations"*, under revision for SIAM J. Numer. Anal., 2006

[7] Q. Du, M. Emelianenko, L. Zikatanov *"Uniformly convergent nonlinear quantization scheme in multidimension"*, in preparation for submission, 2006

[8] Q. Du, M. Emelianenko and L. Ju *"Convergence properties of the Lloyd algorithm for computing the centroidal Voronoi tessellations"*, SIAM J. Numer. Anal., 44, Issue 1 (2006), 102-119

[9] Q. Du, V. Faber and M. Gunzburger, "Centroidal Voronoi Tessellations, theory, algorithms and applications", SIAM Review, 41, pp637-676, 1999

[10] S. Lloyd, "Least square quantization in PCM", IEEE Trans. Infor. theory, 28 (1982), pp.129-137.

[11] M. Emelianenko, Z.-K. Liu and Q. Du *"A New Algorithm for the Automation of Phase Diagram Calculation"*, Comp. Mater. Sci., 35, issue 1 (2006), 61-74 (In ScienceDirect Top 25 Hottest Articles)

[12] M. Emelianenko, Z.-K. Liu and Q. Du, *"Automation of High-dimensional Phase Diagram Calculation"* , in preparation for submission, 2006

[13] Bo Jansson, A General Method for Calculating Phase Equilibria under Different Types of Conditions, TRITA-MAC-0233, 1984.

[14] S.-L. Chen,S. Daniel, F. Zhang, Y. A. Chang, X.-Y. Yan, F.-Y. Xie, R. Schmid-Fetzer, W. A. Oates, The Pandat Software Package and its Applications, CALPHAD, 26 (2002) 175-188

[15] Mohamed Yacoubi, Maria Emelianenko, Natarajan Gautam, "Pricing in next generation networks: a queuing model to guarantee QoS", Performance Evaluation, v. 52, issue 1, p. 59-84, 2003