*AMS IP STUDIES IN ADVANCED MATHEMATICS,*
42, no. 2 (2008): 731-752

# INTELLIGENT AND INFORMATIVE SCIENTIFIC COMPUTING, TRENDS AND EXAMPLES

QIANG DU

Abstract. In this exposition, we consider some emerging issues in intelligent and informative scientific computation. We present some examples to illustrate the needs of such a computation paradigm, and discuss a few possible potential applications with background ranging from microstructure evolution in alloys, deformation of cell membranes, to multi-scale modeling of complex fluids.

## 1. Introduction

In recent years, we have been involved in a number of interdisciplinary research projects on large scale scientific computation problems in diverse applications [15, 63, 29]. A recurring theme in these different problems is the need of intelligent and informative scientific computation ($I^2SC$) which is the focus of this short exposition.

Starting from the later half of the last century, scientific computation is becoming one of the most significant development in the practice of scientific research. Indeed, we are witnessing the sweeping changes brought to the world by the powerful modern computers including both gigantic machines and diminutive devices. As often quoted, computing is redefining the framework of information flow and human interactions, and it is accelerating the pace of scientific discovery by enhancing our ability to predict and understand the behavior of complex systems. The latter is made possible by computational algorithms and softwares, which translate the ever increasing computing power into sophisticated scientific tools which compliment and bridge, and sometimes even surpass, the capabilities of physical experiments and analytic studies.

With the advent of computing power, the landscape of scientific computation changes over time. Here, we begin in Section 2 with some terse examination on a couple of recent trends. We then discuss the needs of identifying useful information and finding new scientific knowledge through simulation results in the subsequent

*Key words and phrases.* Scientific computation, knowledge discovery, intelligence and information, cognitive computing, model reduction, centroidal Voronoi orthogonal decomposition, complex fluids, vesicle membrane, topological information.

Department of Mathematics, Pennsylvania State University, University Park, PA 16802 (qdu@math.psu.edu).

sections. These are illustrated through several examples that include retrieving useful statistics such as topological information on micro-structures, identifying important modes and building reduced models. The applications range from membrane deformation, micro-structure evolution and complex fluid simulations.

The study made in this paper is not intended to be a comprehensive review but rather a brief account of several examples and issues confined to the particular problems we have worked on. They are nevertheless presented with a greater perspective to reflect some emerging trends in the development of intelligent and informative scientific computation. We have chosen a generic and less descriptive title for the paper so that the issues discussed in the paper may attract the interests from readers of this proceeding volume who are outside the area of numerical analysis. As most of our discussions are fairly rudimentary and limited in scope, making connections to broader issues may seem excessive. Still. they serve our purpose of provoking further discussions and debates on this important subject in the community.

## 2. Background: the coming of I$^2$SC

In more traditional applications of scientific computation, the O&O concept, referring to the fact that the computational tools and softwares must be *operational* and *optimized*, has played a dominant role. A piece of computational software is often compartmentalized and the pressing issue is to get the code operational and then to improve the performance via optimization. Integrating existing software components together seamlessly to form more complex problem solving environment remains important [14, 23], yet, for many applications, as scientific computation is the only viable approach and the expectation of making computation based virtual laboratory is mounting, there are also many new trends emerging from recent studies. For example, multi-scale modeling and simulation approaches are attracting much attention in the community [39, 48, 50]. From a practical point of view, there is also an increasing call for the expansion of the predictive power of the computational tools, thus demanding a shift of emphasis from the traditional O&O mode to the V&V paradigm, where the latter refers to the important *validation* and *verification* processes that provide both quantitative confidence and reliability assessment for computational simulations [2, 70, 73]. While *validation* assures that the computer implementation represents the model and the solution from the developer's perspective, *verification* determines the degree to which a model is an accurate description of the real world from the user's perspective [71]. Naturally,

the V&V paradigm relies even more on the interpretation of computational findings in meaningful ways which is at the heart of the intelligent and informative scientific computation ($I^2SC$). Recall the famous quotation from Henri Poincare: *just as houses are made of bricks, so is science made of facts; but a pile of bricks is not a house and a collection of facts is not necessarily science*, an analogy can be made: the objective of the O&O mode is to mostly provide factual data, but it is through the predictive power gained from the V&V process that scientific computing is ultimately making its impact in the advance of science.

Ever since the invention of computers, artificial intelligence has been a subject of popular pursuit after the work of Alan Turing. Ironically, once a culture icon, it also had risked to lose credibility in the scientific community near the end of the last century. We do not intend to lead our discussion into this direction. For the more solemn subject of intelligent scientific computation, there have also been collections of works in the direction [1, 49, 89]. Intelligence at large is understood as the ability to take actions that reflect the knowledge of the situation [69]. Intelligence in the context of scientific computation may include the ability for model selection and reduction, algorithm adoption, data retrieving and manipulation, as well as reasoning and self-evolution, to say the least. Moreover, intelligence is also embedded in the effectiveness of knowledge and data processing in the information age, and being informative (or cognitive) goes side by side with being intelligent in scientific computation. Thus, we have the term $I^2SC$ (intelligent and informative scientific computation). As much of the larger issues are beyond the scope of this paper, we instead only focus on some aspects related to our earlier works in this area. The examples we discuss include the ability of the numerical codes to automatically retrieve useful statistics from the simulation results, a framework for systematic closure approximations of complex systems and a simple tool for identifying important modes and building reduced models.

## 3. Retrieving statistics

We now illustrate an example of collecting useful statistics within a computational framework where the importance of such information has often been overlooked in the earlier works. This example came from our recent study in [30] which was motivated from a number of problems in different applications.

### 3.1. **Phase field modeling.**

In the multis-cale modeling and simulation of multi-component alloys, large scale numerical simulation of the micro-structures in the

material plays an important step [63]. A recent development is the successful application of spectral methods to solve the phase field models for microstructure evolution [92]. Phase field methods employ diffusive interface representations of the micro-structures. In contrast to the sharp interface approach, they are very effective in handling complex evolutions in the morphology and are insensitive to topological events. While working on the integration of the phase field modelling with other simulation tools ranging from first principle calculations, phase diagram computation and the computation of micro-mechanical effect, it becomes important to be able to efficiently collect useful statistics on the underlying microstructure such as particle numbers, average grain sizes, etc, from the computed data or images (see figure.1 for an illustration of both 2D and 3D micro-structures computed in [63, 92]).

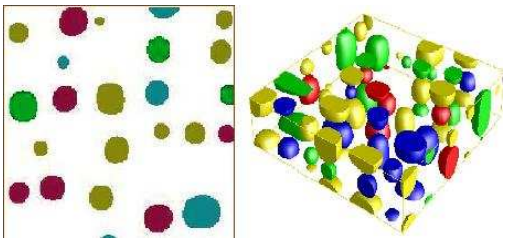

FIGURE 1. Two and three dimensional microstructure simulations for the precipitate growth in different alloys: colored domains represent various precipitates.

Though the insensitivity to the appearance of topological event is regarded as an advantage of the phase field simulation, it becomes very interesting to see if such events can in fact be effectively detected and possibly even controlled. This has many obvious applications such as collecting statistics during the microstructure evolution. While we search for numerical schemes achieving such goals, we also become strongly motivated from our recent study on the modelling of cell membrane and lipid bilayer deformations. The detection and control of topological events in the membrane deformation have significant implications in the design of bilayer liposome as drug delivery vehicles. A simple continuum mechanical model for bilayer membranes studied by Helfrich [44] and others is based on minimizing the bending elasticity energy given by

$$E = \int_\Gamma (H - c_0)^2 \, ds \,, \tag{3.1}$$

subject to given enclosed volume and surface area. Here $H$ is the mean curvature of the membrane surface $\Gamma$ and $c_0$ denotes the spontaneous curvature. Earlier studies

using the Helfrich energy include [76, 77, 81]. In [29] and subsequent works [30, 33, 27, 31], a phase field bending elasticity model via diffuse interface approximations has been systematically developed based on the bending elasticity energy model, including extensions to membrane/fluid interactions and multicomponent and open membranes [28, 88].

In the phase field formulation, a phase function $\phi = \phi(x)$, defined on the physical (computational) domain $\Omega$, is used to label the inside and the outside of the vesicle $\Gamma$. We visualize that the level set $\{x : \phi(x) = 0\}$ gives the membrane, while $\{x : \phi(x) > 0\}$ represents the inside of the membrane and $\{x : \phi(x) \leq 0\}$ the outside. As $\phi$ is defined on the Eulerian reference coordinates, there is no need to follow the motion of the specific level set, a feature that is similar to the level set methods and volume of fluids (VOF) methods [7, 74, 82]. The modified elastic energy used in the phase field bending elasticity model is given by:

$$W(\phi) = \int_\Omega \frac{\epsilon}{2}|\Delta\phi - \frac{1}{\epsilon^2}(\phi^2 - 1)(\phi + C\epsilon)|^2 \, dx, \tag{3.2}$$

where $\epsilon$ is a *transition* parameter that is taken to be very small and $C$ is $\sqrt{2}$ times of the spontaneous curvature. For the volume constraint, we use the functional

$$A(\phi) = \int_\Omega \phi(x) \, dx \ , \tag{3.3}$$

and for the surface area constraint, we define the functional

$$B(\phi) = \int_\Omega [\frac{\epsilon}{2}|\nabla\phi|^2 + \frac{1}{4\epsilon}(\phi^2 - 1)^2] \, dx \ . \tag{3.4}$$

The problem of minimizing the bending surface energy with the prescribed surface area and bulk volume constraints is now effectively transformed into the problem of finding, for small $\epsilon$, the function $\phi = \phi(x)$ on the whole domain that minimizes the energy $W = W(\phi)$ with prescribed values for the constrain functionals $A$ and $B$. Rigorous theoretical analysis have been made to explore the sharp interface limit as $\epsilon \to 0$ [33], and extensive simulations have also been carried out. For different ratios of the volume and surface area, we effectively simulated the transformation of the membrane shapes with the minimum bending elasticity energy. Some examples are shown in Figure 2 which includes the well known discocytes and stomatocytes shapes as well as confirms the existence of a nonsymmetric energy minimizing torus previously studied in [52, 61]. For each shape, the 3d view and the cutting or cross-section view are presented.

3.2. **Euler-Poincare index in phase field framework.** As the phase field model is based on the diffuse interface approximation of the surface curvature, it becomes then natural for us to employ the Gauss-Bonnet formula, so that useful topological

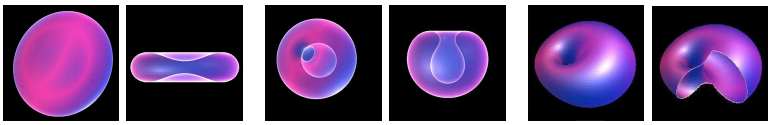

FIGURE 2. Discocyte, Stomatocyte and Dupin Cyclide.

information, in particular, the Euler-Poincare index, can be determined through the computation of the Gaussian curvature. This in turn leads to statistics of practical interests. Briefly, given the phase field function $\phi$, we define the $3 \times 3$ matrix

$$M_{ij}(\phi) = \frac{1}{2\sqrt{\pi(a-b)|\nabla\phi|}}(\nabla^2\phi - \frac{\nabla|\nabla\phi|^2 \cdot \nabla\phi}{2|\nabla\phi|^4}\nabla_i\phi\nabla_j\phi) \ , \tag{3.5}$$

where the constant parameters maybe taken to be $a = 0.5$ and $b = -0.5$; and let $F$ denote the coefficient of the linear term of the characteristic polynomial of $M = (M_{ij})$, i.e.

$$F(M) = M_{11}M_{22} + M_{11}M_{33} + M_{22}M_{33} - M_{12}^2 - M_{13}^2 - M_{23}^2 \ .$$

The Euler number can then be calculated by

$$\frac{\chi}{2} \int_{b\le\phi(x)\le a} F(x)\, dx. \tag{3.6}$$

In [30], simplifications based on the asymptotic profiles of the phase function $\phi$ for the energy minimizing membrane have also been considered, for instance, the modified matrix with entries

$$\tilde{M}_{ij}(\phi) = \sqrt{\frac{35\epsilon}{64\sqrt{2}\pi}}\left((1-\phi^2)\nabla^2\phi + 2\phi\nabla_i\phi\nabla_j\phi\right) \ , \tag{3.7}$$

can be used to give the Euler number [30]:

**Theorem 3.1.** *The Euler number $\chi$ of the zero level set of $\phi$ satisfies:*

$$\frac{\chi}{2} \ = \int_\Omega F(\tilde{M}(x))\, dx \ + o(1) \ , \qquad as \quad \epsilon \to 0 \tag{3.8}$$

*for the $\tilde{M}$ and $F$ given above. In the 2d case, it simplifies to*

$$\chi = \lim_{\epsilon\to 0}\frac{1}{4\pi}\int_\Omega \left(-\triangle\phi + \frac{1}{\varepsilon^2}(\phi^2-1)\phi\right) dx \ + o(1) \ , \qquad as \quad \epsilon \to 0 \ . \tag{3.9}$$

Here we give an example on how the formulae (3.6) and (3.5) can be used to detect topological change. In the left group of pictures of the Figure 3, the merging of several smaller vesicles has been simulated. The change of the Euler number in this experiment is shown in Figure 3 (right), illustrating the near stair-case jumps of the Euler number from 5 to 0, demonstrating the effectiveness of our formulae. For more detailed discussions, we refer to [30] and [31].

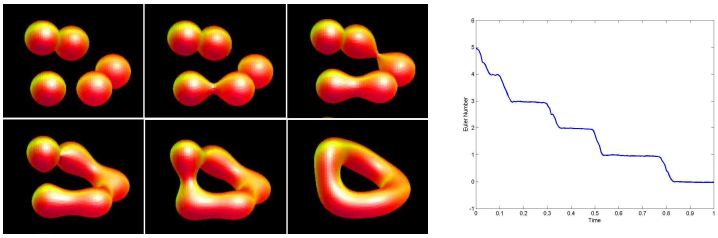

FIGURE 3. The Euler number comes from 6 to 0 after the 6 spheres merges together.

The extraction of Euler number from the phase field models is only one of the examples where useful statistics may be computed during the simulation runs and such information can help users to gauge the development of important events and thus make more intelligent decisions.

## 4. Model reduction

As the multi-scale and multi-physics simulations become a popular approach in modelling complex systems, intelligent schemes that help us understand the results of such simulations and extract new physical laws are also becoming increasingly important. We now discuss a few examples in our earlier works that illustrate several techniques for developing new and reduced models for complex systems.

4.1. **Closure approximations.** With a large amount of cells present in a background fluid, the effective properties of the fluid become significantly different. Such a situation also occurs in polymeric fluids where there is strong coupling between the flow-induced evolution of polymer chains and macroscopic rheological response. This coupling calls naturally for the development of multi-scale methods to analyze the flow of rheologically complex fluids.

For example, using the incompressible Navier-Stokes equation as the background fluid model, the dilute solutions of linear polymers can be described in some detail by modelling the polymer chains as a single dumbbell, namely two beads connected by an elastic spring. A configuration variable $\vec{Q}$ is used to simply represent the vector connecting the two beads of the dumbbell. Thus, the multi-scale approach to modelling dilute solutions of flexible polymers (treated as dumbbells) involves solving the coupled partial differential equations consisting of the macroscopic momentum equation, the incompressible Navier-Stokes equation, and a Fokker-Planck equation describing the probability distribution function (PDF) $f = f(\vec{x}, \vec{Q}, t)$ of the dumbbell orientation $\vec{Q}$ on the microscopic level. One of such coupled system

[5] reads

$$\begin{aligned} \frac{\partial \vec{u}}{\partial t} + (\vec{u}\cdot\nabla)\vec{u} + \nabla p &= \nabla\cdot\tau_p + \nu\triangle\vec{u}\,, \\ \nabla\cdot\vec{u} &= 0\,, \\ \frac{\partial f}{\partial t} + (\vec{u}\cdot\nabla)f + \nabla_{\vec{Q}}\cdot(\nabla\vec{u}\vec{Q}f) &= \frac{2}{\zeta}\nabla_{\vec{Q}}\cdot(\nabla_{\vec{Q}}\Psi(\vec{Q})f) + \frac{2kT}{\zeta}\triangle_{\vec{Q}}f\,. \end{aligned}$$

In the momentum equation, $p$ is the hydrostatic pressure, $\nu$ is the fluid viscosity, and $\tau_p$ is a tensor representing the polymer contribution to stress,

$$\tau_p = \lambda\int(\nabla_{\vec{Q}}\Psi(\vec{Q})\otimes\vec{Q})f(\vec{x},\vec{Q},t)d\vec{Q} \tag{4.10}$$

where $\Psi(\vec{Q})$ is the elastic spring potential to be specified later, and $\lambda$ is the polymer density constant. This induced elastic stress $\tau_p$ in (4.10) can be derived from the least action principle [13]. In the Fokker-Planck equation, $\zeta$ is the friction coefficient of the dumbbell beads, $T$ is the temperature, and $k$ is the Boltzmann's constant. One may associate with the F-P equation the following stochastic differential equation [51],

$$d\vec{Q} + \vec{u}\cdot\nabla\vec{Q}dt = (\nabla\vec{u}\vec{Q} - \frac{2}{\zeta}\nabla_{\vec{Q}}\Psi)dt + \sqrt{\frac{4kT}{\zeta}}d\vec{W}_t, \tag{4.11}$$

where $\vec{W}_t$ is the standard Brownian motion. This stochastic differential equation defines Brownian dynamics of the connector vector $\vec{Q}$ and the probability distribution of dumbbell polymers evolves according to the Fokker-Planck equation [90].

The simplest spring potential is given by the Hookean law $\Psi(\vec{Q}) = HQ^2/2$, where $Q = |\vec{Q}|$ and $H$ is the elasticity constant. It leads to a macroscopic differential constitutive equation for the polymeric stress $\tau_p$ from the Fokker-Planck equation, and gives rise to the so-called Oldroyd-B fluids. However, for a more practical FENE (Finite Extendible Nonlinear Elasticity) potential that takes into account a finite-extensibility constraint, $\Psi(\vec{Q}) = -(HQ_0^2/2)\log(1-(Q/Q_0)^2)$, there does not exist an exact macroscopic constitutive equation for $\tau_p$, and thus the FENE model represents a truly multi-scale model. Here, $Q_0$ is the maximum dumbbell extension. The FENE spring force law reads $\nabla_{\vec{Q}}\Psi = H\vec{Q}/(1-(Q/Q_0)^2)$. We refer to [40, 59, 60, 62] for more recent mathematical works on the dumbbell models.

Simulating the coupled micro-macro system numerically is a demanding task, in that at each spatial point $\vec{x}$, the configuration space of $\vec{Q}$ needs to be resolved or at least sampled. A widely used method to solve the coupled system is the so-called CONNFFESSIT approach [57, 75] that simulates the stochastic differential equation corresponding to the Fokker-Planck equation to achieve a Monte-Carlo sampling

of the configuration space. The latter approach has the advantage that the complexity increases only linearly with the number of beads of the multi-bead polymer chains. In our recent work [32], a finite-difference scheme has been developed that shares several attractive features such as simplicity and positivity-preserving. A key scientific computation issue, however, is on the so-called moment-closure approach.

To illustrate the idea of moment-closure approximations, we multiply the Fokker-Planck equation in 2D by $\vec{Q} \otimes \vec{Q}$, and perform integration by parts on the configuration space $|\vec{Q}| < Q_0$. Denoting by $A$ the conformation tensor $\langle \vec{Q} \otimes \vec{Q} \rangle$, where the brackets represent the ensemble average over the space $|\vec{Q}| < Q_0$, i.e. $\langle \vec{Q} \otimes \vec{Q} \rangle \equiv \int_{|\vec{Q}|<Q_0} \vec{Q} \otimes \vec{Q} f(\vec{x}, \vec{Q}, t) d\vec{Q}$, we get

$$\frac{\partial A}{\partial t} + (\vec{u} \cdot \nabla)A - (\nabla u)A - A(\nabla u)^T = -\frac{4}{\zeta} \int \frac{H\vec{Q} \otimes \vec{Q}}{1 - Q^2/Q_0^2} f(x, \vec{Q}, t) d\vec{Q} + \frac{4kT}{\zeta}.$$

If we select the conformation tensor as the only *state variable*, to close the equation, it is necessary to find an approximate relation to represent the ensemble average term on the right-hand side in terms of $A$. If done properly, this closure approach reduces the needs to resolve the probability space and thus results in dramatic savings computationally. Certainly, the challenge is to find a *good* closure approximation, which is often not an easy task, as a good closure model would be one that agrees quantitatively with the FENE model at least for certain regimes of the physical parameters. Among the well known ad-hoc closure models, the FENE-P model and FENE-L model have been used extensively in the literature (see for example [91] and references therein).

As opposed to the ad hoc closure, in [90], we adopted a more systematic strategy to derive closure approximations by making an ansatz on the class of functions that the PDF solution to the Fokker-Planck equation can take. In particular, our closure model is backed up by some underlying PDF and thus is kept from producing totally unrealistic predictions. Moreover, the resulting system still keeps certain energy law. On the other hand, the success of such methods clearly depends on how well the assumed class can capture the actual solution to the Fokker-Planck equation. It is shown in [90] that excellent agreement is obtained with the FENE model for shear flow with small shear rates, and for coupled simulation of two-dimensional driven-cavity flow. However, it is also observed that the approximation deteriorates as the distortion on the PDF by the flow becomes stronger. In order to improve the closure approximation at larger shear or extension rates, an improved closure is proposed in [32] that allows systematic inclusion of higher order moments. The

higher moment closure relation is derived based on a similar framework, except that the dependence of the ensemble average term needing closure on moments is linear, while the relation is nonlinear in all the previous closure models. It has been shown that, for moderate shear and extension rates, the closure based on higher order moments produces much better stress predictions than the existing closure models, and increasing the order of moments leads to better quantitative agreement with the FENE model.

Regarding the above systematic closure approximation approach, though effective to some extent, its success is heavily dependent on the prior knowledge and a good understanding of the models. As the ansatz is predefined, a rigid structure is mostly employed for the model reduction. In the realm of intelligent computing, we strive for more automated, adaptive and flexible approaches. One of our ongoing work in this direction is to couple a dynamically defined nonlinear Galerkin approximations to improve the static closure approximations. Progress along this line will be reported in the future.

4.2. **Manifold learning and clustering techniques.** Model reductions in connection with large scale computation may be performed both in an off-line mode or an on-line mode. While systematic moment closure scheme can often be viewed as reduction in the off-line mode, there are many applications where on-line or real time model reductions are needed, especially those being relevant to the concept of cognitive computing. Some examples arise in Molecular Dynamics (MD) and Monte Carlo (MC) simulations which are used to study problems of materials science, nanoscience, and biology [41]. MD solves the equation of motion and MC samples a statistical ensemble based on a microscopic description of the interaction between particles.

In order to simulate bigger systems for longer times, both the MD and MC techniques need to be improved to maximize their numerical efficiency. While the small time step in MD often limits the simulation time, it is not the only concern. The interactions between atoms, molecules, or other aggregates modelled by MD or MC simulations are consequences of a fascinating quantum mechanical description, via the Schrödinger equation, of the electrons and nuclei dominated by electrostatic repulsions between like particles and attractions between different ones, as in the first-principles or ab-initio methods. The complexity for evaluating the energy and the forces of one configuration grows super-linearly with the number of particles in most computational schemes which also leads to the size limitation.

A variety of methods have been created to deal with the limitation of ab-initio methods. In the battle for efficiency, a constant pattern emerges: efficiency is paid out of transferability and realism. The popular usage of pseudo-potential approximations is a typical example of model reduction, again, based on prior knowledge. In searching for more intelligent approaches for reducing the complexity, recently, with colleagues from Penn State, we have adopted techniques developed for data-mining to probe the potential energy surface (PES) via nonlinear manifold learning [84] and clustering analysis [26]. In this learning process, when the data interpolated PES enjoys good enough precision, it is to be used to replace the costly ab-initio evaluation. We coined our approach CAMLET which stands for a Combined Ab-initio Manifold LEarning Toolbox.

There are several critical issues in the development of the manifold learning toolbox, such as dimension estimation, global manifold alignment, energy surface interpolation, and clustering analysis in reduced dimensions. We leave systematic discussions to our future works, while only discuss some existing works on the clustering methods here. In the attempt to find important modes of a complex dynamic system, the so called Proper orthogonal decompositions (POD) have been used to systematically extract the most energetic modes. POD is closely related to the statistical method known as Karhunen-Loève analysis or the method of empirical orthogonal eigenfunctions, and it is intimately associated with the more well-known concept of singular value decomposition. There have been many studies devoted to the use of POD for obtaining low-dimensional dynamical system approximations; see, for example [3, 4, 9, 12, 42, 46, 47, 65, 78, 79, 80, 83]. Such popularity is mostly associated with its simplicity while it offers no assurance to success. In [20], we introduced a new clustering approach by combing the POD with the concept of centroidal Voronoi tessellations (CVTs) into a hybrid method for model reduction. The optimality of such an approach and various practical implementation strategies have also been discussed.

In the proper orthogonal decompositions (POD) technique, dominant features from experimental or numerical data are extracted through a set of orthogonal functions which are related to the eigenfunctions of the correlation matrix of the data.

For $n$ snapshots $\widetilde{\mathbf{x}}_j \in \mathbb{R}^N$, $j = 1, \ldots, n$, let $\{\mathbf{x}_j = \widetilde{\mathbf{x}}_j - \widetilde{\boldsymbol{\mu}}\}_{j=1}^n$ be a set of modified snapshots where $\widetilde{\boldsymbol{\mu}} = (\sum_{j=1}^n \widetilde{\mathbf{x}}_j)/n$. Let $d \leq n$, the POD basis $\{\boldsymbol{\phi}_i\}_{i=1}^d$ of cardinality

$d$ is then found by successively solving, for $i = 1, \ldots, d$,

$$\lambda_i = \max_{\|\boldsymbol{\phi}_i\|=1} \frac{1}{n} \sum_{j=1}^{n} |\boldsymbol{\phi}_i^T \mathbf{x}_j|^2 \qquad \text{and} \qquad \boldsymbol{\phi}_i^T \boldsymbol{\phi}_\ell = 0 \quad \text{for } \ell \le i-1\,.$$

If $n \ge N$, this decomposition is known as the *direct* method; if $n < N$, then it is known as the *snapshot* method. For the latter case, $\boldsymbol{\phi}_i = \frac{1}{\sqrt{n\lambda_i}} A\boldsymbol{\chi}_i$, where $\boldsymbol{\chi}_i$ with $\|\boldsymbol{\chi}_i\| = 1$ denotes the eigenvector corresponding to the $i$-th largest eigenvalue $\lambda_i$ of the $n \times n$ correlation matrix $K = (K_{j\ell})$, where $K_{j\ell} = \mathbf{x}_j^T \mathbf{x}_\ell / n$. From now on, we will only consider the case $n < N$.

The POD basis is optimal in the following sense [46]. Let $\{\boldsymbol{\psi}_i\}_{i=1}^n$ denote an arbitrary orthonormal basis for the span of the modified snapshot set $\{\mathbf{x}_j\}_{j=1}^n$. Let $P_{\boldsymbol{\psi},d}\mathbf{x}_j$ be the projection of $\mathbf{x}_j$ in the subspace spanned by $\{\boldsymbol{\psi}_i\}_{i=1}^d$ and let the *error* be defined by $\mathcal{E} = \sum_{j=1}^n \|\mathbf{x}_j - P_{\boldsymbol{\psi},d}\mathbf{x}_j\|^2$. Then, the minimum error is obtained when $\boldsymbol{\psi}_i = \boldsymbol{\phi}_i$ for $i = 1, \ldots, d$, i.e., when the $\boldsymbol{\psi}_i$'s are the POD basis vectors.

A centroidal Voronoi tessellation (CVT) is a Voronoi tessellation of a given set such that the associated generating points are centroids, i.e., the centers of mass with respect to a given density function, of the corresponding Voronoi regions. Given the *discrete* set of modified snapshots $W = \{\mathbf{x}_j\}_{j=1}^n$ belonging to $\mathbb{R}^N$, a set $\{V_i\}_{i=1}^k$ is a *tessellation* of $W$ if $V_i \subset W$ for $i = 1, \ldots, k$, $V_i \cap V_j = \emptyset$ for $i \ne j$, and $\cup_{i=1}^k V_i = W$. Given a set of points $\{\mathbf{z}_i\}_{i=1}^k$ belonging to $\mathbb{R}^N$ (but not necessarily to $W$), the *Voronoi set* corresponding to the point $\mathbf{z}_i$ is defined by

$$\widehat{V}_i = \{\, \mathbf{x} \in W \quad | \quad \|\mathbf{x} - \mathbf{z}_i\| \le \|\mathbf{x} - \mathbf{z}_j\| \quad \text{for } j = 1, \ldots, k,\ j \ne i\}$$

where equality holds only for $i < j$. Other tie-breaking rules for points equidistant to two or more of the $\mathbf{z}_i$'s can also be used. The set $\{\widehat{V}_i\}_{i=1}^k$ is called a *Voronoi tessellation* or *Voronoi diagram* of $W$ [72].

Given a density function $\rho(\mathbf{y}) \ge 0$, defined for $\mathbf{y} \in W$, the mass centroid $\mathbf{z}^*$ of any subset $V \subset W$ is defined by

$$\sum_{\mathbf{y}\in V} \rho(\mathbf{y})|\mathbf{y} - \mathbf{z}^*|^2 = \inf_{\mathbf{z}\in V^*} \sum_{\mathbf{y}\in V} \rho(\mathbf{y})|\mathbf{y} - \mathbf{z}|^2\,,$$

where the sums extend over the points belonging to $V$ and the set $V^*$ can be taken to be $V$ or it can be an even larger set such as $\mathbb{R}^N$. In the latter case, $\mathbf{z}^*$ is the ordinary mean

$$\mathbf{z}^* = \{\sum_{\mathbf{y}\in V} \rho(\mathbf{y})\mathbf{y}\} / \{\sum_{\mathbf{y}\in V} \rho(\mathbf{y})\}\,,$$

in this case, $\mathbf{z}^* \notin W$ in general.

If $\mathbf{z}_i = \mathbf{z}_i^*$ for $i = 1, \ldots, k$, where $\{\mathbf{z}_i\}_{i=1}^k$ is the set of generating points for the Voronoi tessellation $\{\widehat{V}_i\}_{i=1}^k$ and $\{\mathbf{z}_i^*\}_{i=1}^k$ are the set of mass centroids of the Voronoi

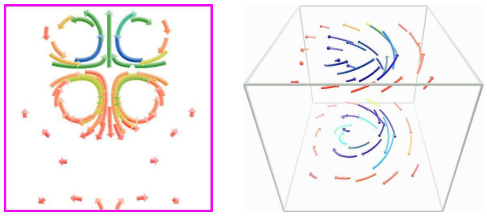

FIGURE 4. Simplified vector field representation based on the CVTs: 2d and 3d examples.

regions, we refer to the Voronoi tessellation as a *centroidal Voronoi tessellation.* The concept of CVT's can be extended to more general sets, including regions in Euclidean spaces, and more general metrics. They have a variety of applications including data compression, image compression, optimal allocations of resources, territorial behavior of animals, optimal sensor and actuator location, and numerical analysis including both grid-based and meshfree algorithms for interpolation, multi-dimensional integration, and partial differential equations; see [6, 8, 10, 11, 16, 17, 18, 19, 20, 21, 22, 24, 25, 34, 35, 36, 37, 38, 45, 55, 56, 67, 85, 87].

Given a discrete set of points $W = \{\mathbf{x}_j\}_{j=1}^n$ belonging to $\mathbb{R}^N$, we define the error with respect to a tessellation $\{V_i\}_{i=1}^k$ of $W$ and a set of points $\{\mathbf{z}_i\}_{i=1}^k$ belonging to $W$ or, more generally, belonging to $\mathbb{R}^N$ by

$$\mathcal{F}\big((\mathbf{z}_i, V_i), i = 1, \ldots, k\big) = \sum_{i=1}^{k} \sum_{\mathbf{y} \in V_i} \rho(\mathbf{y}) |\mathbf{y} - \mathbf{z}_i|^2 .$$

It can be shown that a necessary condition for the error $\mathcal{F}$ to be minimized is that the pair $\{\mathbf{z}_i, V_i\}_{i=1}^k$ form a CVT of $W$. We note that the above error is also often referred to as the *variance, cost, distortion error,* or *mean square error.* The CVTs of discrete sets are closely related to optimal *k-means clusters* and Voronoi regions and centroids are referred to as *clusters* and *cluster centers*, respectively. Clustering analysis provides a selection of a finite collection of templates that well represent, in some sense, a large collection of data as illustrated in [20]. It has been shown that, using the variance-based criteria to define optimality, the optimal clustering corresponds to a centroidal Voronoi tessellation. In figure 4, we provide some examples of using CVTs for vector field simplification and representation which again are useful in dealing with complex flow fields [37]. For algorithms to compute the centroidal Voronoi tessellations of a given set, we refer to [16, 17, 18, 55, 64, 66] and the references cited therein.

4.3. **CVOD.** To generalize the CVT concept so as to combine it with POD, there are essentially two ingredients that we need to redefine, namely, the concept of

distance (which appears in the definition of and thus serves to define Voronoi tessellations) and the concept of centroid.

First, the square of the distance from a one dimensional subspace spanned by a vector $\mathbf{x}$ to a $d$-dimensional subspace $\mathcal{Z}$ can be defined by

$$\delta^2(\mathbf{x}, \mathcal{Z}) = 1 - \frac{1}{\|\mathbf{x}\|^2} \sum_{i=1}^{d} |\mathbf{x}^T \boldsymbol{\theta}_i|^2 \,,$$

where $\{\boldsymbol{\theta}_i\}_{i=1}^d$ forms an orthonormal basis for $\mathcal{Z}$. Then, given a set of vectors (e.g., modified snapshots) $W = \{\mathbf{x}_j\}_{j=1}^n$ and a set of $d$-dimensional subspaces $\{\mathcal{Z}_i\}_{i=1}^k$ (which are called the generators), we define the generalized Voronoi tessellation of $W$ by

$$\mathcal{V}_i = \{\mathbf{x}_j \in W \ \mid \ \delta^2(\mathbf{x}_j, \mathcal{Z}_i) \le \delta^2(\mathbf{x}_j, \mathcal{Z}_\ell), \qquad \forall\, \ell \neq i \ \} \qquad \text{for } i = 1, \ldots, k \,.$$

A tie-breaking rule may be applied to insure that in the case equality holds in the above definition, each modified snapshot only belongs to one generalized Voronoi region.

Second, given a set of vectors $\mathcal{V} = \{\mathbf{x}_j\}$ that span an $m$-dimensional subspace of $\mathbb{R}^N$ (e.g., again, for us these are a subset of cardinality $m$ of the modified snapshots), the concept of $d$-generalized centroid ($d \le m$) of $\mathcal{V}$ may be defined by an orthonormal basis $\{\boldsymbol{\phi}_i\}_{i=1}^d$ which minimizes

$$\mathcal{D} = \sum_{\mathbf{x}_j \in \mathcal{V}} \|\mathbf{x}_j - P\mathbf{x}_j\|^2 \,,$$

where $P$ denotes the projection operator into the $d$-dimensional subspace spanned by $\{\boldsymbol{\phi}_i\}_{i=1}^d$. For simplicity, we call such a centroid or basis the *$d$-$g$ centroid of $\mathcal{V}$*. Note that, the optimal basis $\{\boldsymbol{\phi}_i\}_{i=1}^d$ is the $d$-dimensional POD basis for the set $\mathcal{V}$.

Note that the generators $\{\mathcal{Z}_j\}$ in general may not be required to have the same dimension. Thus, if $k$ denotes the number of generators, we may use a multi-index $\mathbf{d} = \{d_i\}_{i=1}^k$ to replace the scalar index $d$. Then, the CVT based POD can be properly defined as follows [20]: a set of finite subspaces $\{\mathcal{Z}_j\}_{i=1}^k$ with dimensions $\mathbf{d} = \{d_i\}_{i=1}^k$, respectively, along with the corresponding generalized Voronoi tessellation $\{\mathcal{V}_j\}_{i=1}^k$ is called a *$d$-$g$ CVT* if and only if the $\mathcal{Z}_i$'s are themselves the $d$-$g$ centroids of the $\mathcal{V}_i$'s. The union of basis vectors corresponding to a $d$-$g$ CVT is called a *CVT based POD* or a *centroidal Voronoi orthogonal decomposition (CVOD).*

To recapitulate, CVOD can be viewed as a generalization of CVT for which the set $W$ of modified snapshots is divided into $k$ clusters or generalized Voronoi regions $\{\mathcal{V}_i\}_{i=1}^k$ and for which the generators are $d_i$-dimensional spaces each of which is spanned by the $d_i$-dimensional POD basis for the cluster. CVOD can also

be viewed as a generalization of POD for which the set of modified snapshots is divided into $k$ clusters and then a POD basis is separately determined for each cluster. In fact, if $d_i = 1$ for $i = 1, \ldots, k$, then CVT based POD reduces to the standard CVT. On the other hand, if $k = 1$, then CVT based POD reduces to the standard POD.

Algebraically, one may also interpret CVOD as follows. First, the original correlation matrix for the whole set of snapshots $W$ is replaced by a block diagonal matrix with diagonal blocks being the correlation matrices for the snapshots in individual Voronoi sets $\{V_i\}$; then, the POD analysis is separately performed on each of the blocks. These Voronoi sets form a generalized centroidal Voronoi tessellation of $W$. Thus, the role of CVT within CVOD may be viewed as providing, in some sense, an optimal clustering of the modified snapshots; the role of POD is then to provide an optimal reduced basis for each cluster.

There are cases where certain snapshots need to be weighted more heavily; thus, weighted POD's have been defined [9]. In light of the fact that a nonuniform density function can be used in the standard CVT construction, we may also define the weighted CVOD with a prescribed *discrete density* or a set of weights, i.e., we may minimize

$$\sum_{\mathbf{x}_j \in \mathcal{V}} \rho(\mathbf{x}_j) \delta^2(\mathbf{x}_j, \mathcal{Z}_i)$$

over a $d_i$-dimensional subspace of $\mathcal{V}$ for a given density function $\rho$.

Similar to the original CVT, we can easily prove[20]:

**Theorem 4.1.** *For a density function with values $\{\rho_j\}_{j=1}^n$, the d-g CVT minimizes the functional*

$$\mathcal{G}\Big((\mathcal{Z}_i, \mathcal{V}_i),\, i = 1 \ldots, k\Big) = \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in \mathcal{V}_i} \rho_j \delta^2(\mathbf{x}_j, \mathcal{Z}_i)\,.$$

The functional $\mathcal{G}$ also provides a natural error tolerance measure in the sense that

$$\mathcal{G}\Big((\mathcal{Z}_i, \mathcal{V}_i),\, i = 1 \ldots, k\Big) = \sum_{i=1}^{k} |\mathcal{V}_i| \sum_{j=d_i+1}^{|\mathcal{V}_i|} \lambda_{i_j}\,,$$

where $|\mathcal{V}_i|$ denotes the cardinality of the Voronoi set or cluster $\mathcal{V}_i$ and the $\lambda_{i_j}$'s are the eigenvalues (in decreasing order) of the (weighted) local correlation matrix of the snapshots in the cluster.

For $k$ large, it has been conjectured [18] that CVT's enjoys the equi-partition of error property; it is natural to extend such a conjecture to CVT based POD. Such

an error equi-partition property leads naturally to adaptive strategies to refine the CVOD analysis.

A reduced basis, being it POD or CVT or CVOD, can be used to define a low-order model for a complex system in a more automated fashion, in comparing with the closure approach discussed earlier. For instance, let $F(t, X, u(X, t)) = 0$ be a system of partial differential equations with suitable boundary and/or initial conditions for the unknown function $u$. Here, $t$ could be the time variable or some system parameter. Then, the CVOD based model reduction is performed as follows.

**Algorithm 4.1.** *CVOD based model reduction procedure*

**1:** *Construct a set of modified snapshots* $\{u_j\}_1^n$ *by solving the system of differential equations for different values of* $t$.

**2:** *Calculate the CVOD for the set* $\{u_j\}_1^n$ *for some integer* $k$ *and multi-index* $\{d_j\}_{j=1}^k$ *to obtain a set of basis vectors* $\{\phi_m\}_{m=1}^{|\mathbf{d}|}$.

**3:** *With a suitable inner product, solve the reduced system:*

$$\left\langle \phi_m, F(t, X, \sum_{l=1}^{|\mathbf{d}|} \beta_l \phi_l) \right\rangle = 0 \qquad \text{for } m = 1, 2, \ldots, |\mathbf{d}| \ .$$

The advantage of CVOD over POD is clear, POD mostly is a linear decomposition technique, while CVOD is nonlinear in nature and it introduces the concept of clustering into the decomposition. Imagine situations where intermittency is important or the dynamics are described by somewhat less related modes. By imposing a clustering, each sub-CVOD basis for a specific cluster can be used to capture the dynamics of that cluster. CVOD also reduces the amount of work relative to the full POD analysis. POD involves the solution of an $n \times n$ eigenproblem, where $n$ is the number of snapshots; CVOD instead requires the solution of several smaller eigenproblems.

Another interesting feature of CVOD which has been observed in other contexts, e.g., image processing [18], is that it avoids the over-crowding of the reduced basis into a few dominant modes which is viewed as one of the drawbacks of POD in some practical simulations. There have been applications of the CVOD concept to fluid control problems [6] and problems in homogenization. The application of CVOD in MD and ab-initio simulations is also currently under investigation.

Yet, perhaps the important feature of CVOD is its simplicity, it can be developed as a on-line toolbox that can constantly probe the computation results and seek for clustering patterns and identifying dominant modes. This serves the goal of manifold learning, model reduction and knowledge discovery. For a complex evolving system, how to efficiently compute the CVOD dynamically through some

adaptive and hierarchical processes becomes important. In this regard, there has been some recent works on multilevel methods for computing the CVTs [16] that may be further extended in the future.

## 5. Conclusion

In the last couple of decades, we have witnessed the exhilarating development in scientific computation which is widely viewed as the third approach in scientific research, along with theory and experiments. The recent emphasis on virtual science or cyber-science research symbolizes the central role now played by scientific computation. It is expected that with the increasing demand of making scientific computing as a virtual laboratory to replace physical experiments in many key areas of sciences, the validation and verification ($V\&V$) process in scientific computation is going to be more and more important. This calls more intelligent and informative (or cognitive) scientific computing to provide not only robustness but also confidence, assurance and intelligence on the computational findings. A closer link between computational science and cognition science becomes a necessity. The relevant modelling and algorithmic issues provide challenges to mathematicians and computational scientists and they lay the foundation for technological breakthroughs in this direction.

In this paper, we discussed several examples taken from our recent research projects. While their background are different, they all share features that are consistent to our objective: making the scientific computation process more intelligent and more informative (thus leading to the cognitive scientific computation) by retrieving useful statistics and recognizing and extracting important modes. The discussion is certainly rudimentary, yet, it is our intension to take this opportunity to discuss some of the issues involved. We conclude with the observation that it is absolutely critical that the research on mathematical analysis, algorithm innovation, software design and practical applications are to be integrated together in order to make scientific computation fulfill its great promise.

## Acknowledgment

The research is supported in part by NSF DMS 0409297, NSF CCF 0430349 and NSF ITR 0205232. The works discussed in the manuscript were carried out in collaboration with Drs. Longqing Chen, Max Gunzburger, Lili Ju, Tianjiang Li, Chun Liu, Rolf Ryham, Xiaoqiang Wang, Peng Yu, and Hongyuan Zha.

## References


[1] H. Abelson, M. Eisenberg, M. Halfant, J. Katzenelson, E. Sacks, G. Sussman, J. Wisdom and K. Yip, Intelligence in scientific computing, *Communications of the ACM*, 32, pp.546–562, 1989.

[2] I. Babuska and T. Oden, Verification and validation in comptuational engineering and science. Part I: Basic concepts. *Comp. Meth. Appl. Mech. Engrg.*, 193, pp.4057–4066, 2004.

[3] G. Berkooz, P. Holmes and J. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, *Ann. Rev. Fluid. Mech.*, 25, pp.539–575, 1993.

[4] G. Berkooz and E. Titi, Galerkin projections and the proper orthogonal decomposition for equivariant equations, *Physics Letters A*, 174, pp.94–102, 1993.

[5] R. Bird, O. Hassager, R. C. Armstrong and C. F. Curtiss, *Dynamics of Polymeric Fluids, Vol. 2, Kinetic Theory*. John Wiley & Sons, New York, 1977.

[6] J. Burkardt, M Gunzburger and H.-C. Lee, Centroidal Voronoi Tessellation-Based Reduced-Order Modeling of Complex Systems, to *SIAM J. Sci Comp*, 28, pp.459-484, 2006.

[7] G. Caginalp and X. F. Chen, Phase field equations in the singular limit of sharp interface problems, in *On the evolution of phase boundaries* (Minneapolis, MN, 1990–91), Springer, New York, 1992, pp. 1–27.

[8] M. Cappellari and Y. Copin, Adaptive spatial binning of integral-field spectroscopic data using Voronoi tessellations, *Monthly Notices Royal Astronomical Soc.*, 342, pp. 345-354, 2003.

[9] E. Christensen, M. Brons and J. Sorensen, Evaluation of proper orthogonal decomposition based decomposition techniques applied to parameter dependent nonturbulent flows, *SIAM J. Sci. Comp.*, 21, pp.1419–1434, 2000.

[10] D. Cohen-Steiner, P. Alliez and M. Desbrun, Variational shape approximation, *ACM Transactions on Graphics*, 23, pp.905-914, 2004.

[11] J. Cortes, S. Martinez, T. Karatas and F. Bullo, Coverage control for mobile sensing networks *IEEE Tran. Robotics and Automation*, 20, pp.243-255, 2004.

[12] A. Deane, I. Kevrekidis, G. Karniadakis and S. Orszag, Low-dimensional models for complex geometry flows: applications to grooved channels and circular cylinders, *Phys. Fluids A*, 3, pp.2337–2354, 1991.

[13] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*. Clarendon Press, Oxford, 1986.

[14] Q. Du, Optimization based nonoverlapping domain decomposition algorithms and their convergence, *SIAM J. Numer. Analysis*, 39, pp.1056-1077, 2001

[15] Q. Du, Some perspective on the large scale scientific computation research, *Science in China A*, 47 (2004), Supp., pp.1-4.

[16] Q. Du and M. Emelianenko, Acceleration schemes for computing centroidal Voronoi tessellations, *Numer Linear Algebra and Appl*, 13, pp.173 - 192, 2006

[17] Q. Du, M. Emelianenko and L. Ju, Convergence properties of the Lloyd algorithm for computing the centroidal Voronoi tessellations, *SIAM J. Num. Anal.*, 44, pp.102-119, 2006

[18] Q. Du, V. Faber and M. Gunzburger, Centroidal Voronoi tessellations: applications and algorithms, *SIAM Review*, 41, pp. 637–676, 1999.

[19] Q. Du and M. Gunzburger, Grid Generation and Optimization Based on Centroidal Voronoi Tessellations, *Appl. Math. and Comput.*, 133, pp.591–607, 2002.

[20] Q. Du and M. Gunzburger, Centroidal Voronoi tessellation based proper orthogonal decomposition analysis, *International series of Numerical Mathematics*, Birkhauser, 143, pp.137-150, 2002.

[21] Q. Du, M. Gunzburger and L. Ju, Constrained centroidal Voronoi tessellations on general surfaces, *SIAM J. Sci. Comput.*, 24, pp. 1488–1506, 2003.

[22] Q. Du, M. Gunzburger, L. Ju and X. Wang, Centroidal Voronoi tessellation algorithms for image processing, *J. of Mathematical Imaging and Vision*, 24, pp.177-194, 2006

[23] Q. Du, Z. Huang and D. Wang, Mesh and solver coadaptation in finite element methods for anisotropic problems, *Numerical Methods for Differential Equations*, 21, pp.859-874, 2005.

[24] Q. Du and L. Ju, Finite volume methods and spherical centroidal Voronoi tessellations, *SIAM J. Numer. Anal.*, **43**, pp.1673-1692, 2005.

[25] Q. Du and L. Ju, Approximations of a Ginzburg-Landau model for superconducting hollow spheres based on spherical centroidal Voronoi tessellations, *Mathematics of Computation*, 74, pp.1257-1280, 2005.

[26] Q. Du, T. Li and H. Zha, Eckart conditions, SVD and quaternions, Grassmann and Clifford algebras: concepts and methods for data preprocessing in manifold learning, preprint, 2007

[27] Q. Du, C. Liu, R. Ryham and X. Wang, A phase field formulation of the Willmore problem, *Nonlinearity*, 18, pp.1249-1267, 2005.

[28] Q. Du, C. Liu, R. Ryham and X. Wang, Energetic variational approaches to modeling vesicle and fluid interactions, preprint, 2006

[29] Q. Du, C. Liu and X. Wang, A phase field approach in the numerical study of the elastic bending energy for vesicle membranes, *Journal of Computational Physics*, 198, pp.450–468, 2004.

[30] Q. Du, C. Liu and X. Wang, Retrieving topological information for phase field models, *SIAM Journal on Applied Mathematics*, 65, pp.1913-1932, 2005.

[31] Q. Du, C. Liu and X. Wang, Simulating the deformation of vesicle membranes under elastic bending energy in three dimensions, *J. Computational Phys.*, 212, pp.757-777, 2006.

[32] Q. Du, C. Liu and P. Yu, FENE dumbbell models and its several linear and nonlinear closure approximations, *SIAM Multiscale Modelling and Simulations*, 4, pp.709-731, 2005.

[33] Q. Du, C. Liu, R. Ryham and X. Wang, Phase field modeling of the spontaneous curvature effect in cell membranes, *Comm Pure Appl Anal*, 4, pp.537-548, 2005.

[34] Q. Du and D. Wang, Tetrahedral mesh generation and optimization based on centroidal Voronoi tessellations, *Int. J. Numer. Meth. Eng.*, 56, pp.1355–1373, 2002.

[35] Q. Du and D. Wang, Anisotropic centroidal Voronoi tessellations and their applications, *SIAM J. Sci. Comput.*, 26, pp.737-761, 2005.

[36] Q. Du and D. Wang, New progress in robust and quality Delaunay mesh generation, *J. Computational Applied Mathematics*, 195, pp.8-23, 2006.

[37] Q. Du and X. Wang, Centroidal Voronoi tessellation based algorithms for vector fields visualization and segmentation, *IEEE Proceedings of Visualization 2004 (VIS2004)*, IEEE, pp43-50, Austin, TX, 2004.

[38] Q. Du and X. Wang, Tessellation and clustering by mixture models and their parallel implementations, *Proceeding of the fourth SIAM international conference on Data Mining*, Lake Buena Vista, FL, 2004, SIAM, pp257-268.

[39] W. E and B. Engquist, Heterogeneous multiscale methods, *Comm. Math. Sci*, 1, pp.87-132, 2003.

[40] W. E, T. Li and P. Zhang, Well-posedness for the dumbbell model of polymeric liquids. *Comp. Math. Phys.*, 248, pp. 409-427, 2004.

[41] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, Academic Press, San Diego, 2002.

[42] M. Graham and I. Kevrekidis, Pattern analysis and model reduction: some alternative approaches to the Karhunen-Lóeve decomposition, *Comp. Chem. Engng.*, 20, pp.495–506, 1996.

[43] R. Gray and D. Neuhoff, Quantization, *IEEE Trans. Inform. Theory*, 44, pp.2325–2383, 1998.

[44] W. Helfrich, Elastic properties of lipid bilayers: theory and possible experiments, *Z. Naturforsch. C*, 28, pp.693–703, 1973.

[45] S. Hiller, H. Hellwig and O. Deussen, Beyond stippling - Methods for distributing objects on the plane, *Computer Graphics Forum*, 22, pp. 515-522, 2003.

[46] P. Holmes, J. Lumley and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, 1996.

[47] P. Holmes, J. Lumley, G. Berkooz, J. Mattingly and R. Wittenberg, Low-dimensional models of coherent structures in turbulence, *Phys. Rep.*, 287, pp.337–384, 1997.

[48] T. Hou and X. Wu, A multiscale finite element method for elliptic problems in composite materials and porous media, *J. Comput. Phys.*, 134, pp.169–189, 1997.

[49] R. Huber, C. Kulikowski, J. Krivine and J. David, *Artificial Intelligence in Scientific Computation: Towards Second Generation Systems*, 1989, JC Baltzer AG, Scientific Publishing Co, Basel.

[50] T. Hughes, G. Feijoo, J. Mazzei and J. Quincy, Variational multiscale method - a paradigm for computational mechanics, *Computer Methods in Applied Mechanics and Engineering*, 166, pp.3-24, 1998.

[51] M. Hulsen, A. van Heel and B. van den Brule, Simulation of viscoelastic flows using Brownian configuration fields. *J of Non-Newtonian Fluid Mechanics*, 70, pp.79–101, 1997.

[52] Y. Jie, Q. Liu, J. Liu and Z. Ou-Yang, Numerical observation of nonaxisymmetric vesicles in fluid membranes, *Phys. Rev. E*, 58, pp.4730–4736, 1998.

[53] B. Jourdain, C. Le Bris and T. Lilievre, On a variance reduction technique for the micro-macro simulations of polymeric fluids, *Journal of Non-Newtonian Fluid Mechanics*, 122, pp.91-106, 2004.

[54] B. Jourdain, T. Lilievre and C. Le Bris, Existence of solution for a micro-macro model of polymeric fluid: the FENE model, *Journal of Functional Analysis*, 209, pp.162-193, 2004.

[55] L. Ju, Q. Du and M. Gunzburger, Probablistic methods for centroidal Voronoi tessellations and their parallel implementations, *Parallel Computing*, 28, pp. 1477–1500, 2002.

[56] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman and A. Wu, An efficient k-means clustering algorithm: Analysis and implementation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24, pp.881-892, 2002.

[57] M. Laso and H. C. Ottinger Calculation of viscoelastic flow using molecular models: The connffessit approach. *J of Non-Newtonian Fluid Mechanics*, 47, pp.1–20, 1993.

[58] T. Li, H. Zhang and P. Zhang, Local existence for the dumbbell model of polymeric fluids. *Communications in Partial Differential Equations*, 29, pp.903 - 923, 2004

[59] G. Lielens, P. Halin, I. Jaumain, R. Keunings and V. Legat, New closure approximations for the kinetic theory of finitely extensible dumbbells. *J of Non-Newtonian Fluid Mechanics*, 76, pp.249–279, 1998.

[60] G. Lielens, R. Keunings and V. Legat, The FENE-L and FENE-LS closure approximations to the kinetic theory of finitely extensible dumbbells. *J of Non-Newtonian Fluid Mechanics*, 87, pp.179–196, 1999.

[61] G. Lim, M. Wortis and R. Mukhopadhyay, Stomatocyte-discocyte-echinocyte sequence of the human red blood cell: Evidence for the bilayer-couple hypothesis from membrane mechanics, *PNAS*, 99, pp. 16766-16769, 2002.

[62] F. Lin, C. Liu and P. Zhang, On a micro-macro model for polymeric fluids near equilibrium, *Comm Pure Appl Math*, 60, pp.838-866, 2006.

[63] Z. Liu, L. Chen, P. Raghavan, Q. Du, J. Sofo, S. Langer and C. Wolverton, An integrated Framework for multi-scale materials simulation and design, *J. Computer Aided Material Design*, 11, pp.183-199, 2004

[64] S. Lloyd; Least squares quantization in PCM, *IEEE Trans. Infor. Theory*, 28, pp.129–137, 1982.

[65] J. Lumley, *Stochastic Tools in Turbulence*, Academic, New York, (1971).

[66] J. MacQueen; Some methods for classification and analysis of multivariate observations, in *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, I,* Ed. by L. Le Cam and J. Neyman, 1967, pp. 281–297.

[67] A. Mendes and I. Themido, Multi-outlet retail site location assessment. *Inter. Trans. in Operational Research*, 11, pp.1-18, 2004.

[68] U. Moller, M. Galicki, E. Baresova and H. Witte, An efficient vector quantizer providing globally optimal-solutions, *IEEE Trans. Signal Proc.*, 46, pp.2515–2529, 1998.

[69] A. Newell, Fairy tales, *AI Magazine*, 13(4), 46–48, 1992,

[70] W. Oberkampf, Bibliography for Verification and Validation in Computational Simulation, Sandia Report SAND98-2041, 1998,

[71] W. Oberkampf, T. Trucano and C. Hirsch, Verification, validation and predictive capabilities in computational engineering and physics, *Applied Mechanics Reviews*, 57, pp.345-384, 2004.

[72] A. Okabe, B. Boots, K. Sugihara and S. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd edition, Wiley, Chichester, 2000.

[73] N. Oreskes, K. Shrader-Frechette and K. Belitz, Verification, validation, and confirmation of numerical models in the earth sciences, *Science*, 263, 641-646, 1994.

[74] S. Osher and R. Fedkiw, *The Level Set Method and Dynamic Implicit Surfaces*, Springer-Verlag, 2002.

[75] H. Ottinger, *Stochastic Processes in Polymeric Fluids, Tools and Examples for Developing Simulation Algorithms*. Springer-Verlag, Berlin, 1996.

[76] Z. Ou-Yang, Anchor ring-vesicle membranes, *Physical Review A*, 41, pp.4517-4520, 1990,

[77] Z. Ou-Yang, J. Liu and Y. Xie, *Geometric Methods in the Elastic Theory of Membranes in Liquid Crystal Phases*, World Scientific, Singapore, 1999.

[78] M. Rathinam and L. Petzold, A new look at proper orthogonal decomposition, *SIAM J. Numer. Anal.*, 41, pp.1893-1925, 2003

[79] S. Ravindran, Proper orthogonal decomposition in optimal control of fluids, *Int. J. Numer. Meth. Fluids*, 34, pp.425–448, 2000.

[80] J. Rodríguez and L. Sirovich, Low-dimensional dynamics for the complex Ginzburg-Landau equations, *Physica D,* 43, pp.77–86, 1990.

[81] U. Seifert, K. Berndl and R. Lipowsky, Configurations of fluid membranes and Vesicles, *Physical Rev A*, 44, pp.1182-1202, 1991.

[82] J. Sethian, *Level Set Methods and Fast Marching Methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. Cambridge University Press, New York, 2nd edition, 1999.

[83] N. Smaoui and D. Armbruster, Symmetry and the Karhunen-Loève analysis, *SIAM J. Sci. Comput.*, 18, pp. 1526–1532, 1997.

[84] J. Tenenbaum, V. De Silva and J. Langford, A global geometric framework for nonlinear dimension reduction, *Science*, 290, pp.2319–2323, 2000.

[85] S. Valette and J. Chassery, Approximated centroidal Voronoi Diagrams for uniform polygonal mesh coarsening, *Computer Graphics Forum*, 23, pp.381–390, 2004.

[86] S. Volkwein, Optimal control of a phase field model using the proper orthogonal decomposition, *ZAMM*, 81, pp.83–97, 2001.

[87] C. Wager, B. Coulland and N. Lange, Modelling spatial intensity for replicated inhomogeneous point patterns in brain imaging, *J. Royal Stat. Soc. B*, 66, pp.429–446, 2004.

[88] X. Wang and Q. Du, Modelling and simulations of multi-component lipid membranes and open membranes via diffuse interface approaches, *J. Mathematical Biology*, 56, 2007.

[89] S. Weerawarana, E. Houstis, J. Rice, A. Joshi and C. Houstis, PYTHIA: a knowledge based system for intelligent scientific computing, *ACM Transactions on Mathematical Software*, 22, pp447–468, 1996.

[90] P. Yu, Q. Du and C. Liu, From micro to macro dynamics via a new moment closure approximation to the FENE model of polymeric fluids. *SIAM Journal of Multiscale Modeling and Simulation*, 3, pp.895-917, 2005

[91] Q. Zhou and R. Akhavan, A comparison of FENE and FENE-P dumbbell and chain models in turbulent flow, *J. Non-Newtonian Fluid Mech.*, 109, pp.115-155, 2003.

[92] J. Zhu, T. Wang, A. Ardell, S. Zhou, Z. Liu and L. Chen, Three-dimensional phase-field simulations of coarsening kinetics of $\gamma'$ particles in binary Ni-Al alloys, *Acta Mater.*, 52, pp.2837-2845, 2004.