

CeMEAI: The Brazilian Center and Its Mathematics Research for Industry

E. G. Birgin, J. A. Cuminato, J. M. Martínez, and T. Pereira

Abstract. We describe the Brazilian applied mathematics research center CeMEAI and some of the work being done there.

Introduction

CeMEAI, Center for Research in Mathematical Sciences Applied to Industry, hosted at the Institute of Mathematics and Computer Sciences of the University of São Paulo in São Carlos, is one of seventeen cross-disciplinary research, innovation, and dissemination centers funded by FAPESP, the São Paulo state research funding agency. CeMEAI started its activities in 2011 and has received substantial financial support from FAPESP since 2013. The present funding will last until June 2018, and it can be renewed for another two periods of three years each, pending the outcome of international assessment. CeMEAI involves five research institutions in São Paulo state with twenty-five principal investigators, around forty associated researchers, fifteen postdocs, twenty-five PhD students, and some master's students. In total more than one hundred researchers contribute directly to technological innovation projects in the main concentration areas of the center: applied optimization and operational research, computational fluid mechanics, risk analysis,

and computational intelligence and software engineering. The main purpose of CeMEAI is to provide a broad and well-resourced mechanism for connecting scientists, engineers, mathematicians, and computer experts in order to address scientific and technological challenges in a collaborative, engaging environment, developing transformative new mathematical techniques and exploring their applications. CeMEAI is also highly engaged in the training of the next generation of researchers, educators, and industry and government personnel. In 2013 CeMEAI started an MSc in industrial mathematics that has taken students from the banking and hospital industries for retraining in risk technology analysis and data systems.

Some research projects begin with industry visits: once industry's problems are identified, representatives are put in contact with researchers who may be able to help address them. Other strategies like an Oxford-style study group with industry and modeling camps have been organized annually since 2015.¹ Since the center started its activities in 2011, many industrial projects have been contracted: multiscale methods for reservoir simulation, broiler house environment control in fluid dynamics, mammal cancer treatment protocol, and a computer system for helping high school sports coaches rank their students in statistics. Our researchers have also worked on projects that have made an impact on a larger community, such as the systems Packmol, PUMA, and TANGO, described in more detail in what follows. Other researchers have worked on complex systems such as power grids, lasers, and biological oscillators.

Industrial Applications

Building Initial Configurations for Molecular Dynamics Simulations—Packmol

Molecular dynamics simulations consist of arrangements of molecules distributed in space in such a way as to approximately represent the system's overall structure. For the simulations not to be disrupted by large Van

E. G. Birgin is professor of mathematics at the University of São Paulo, Brazil. His e-mail address is egbirgin@ime.usp.br.

J. A. Cuminato is professor of mathematics at the University of São Paulo, Brazil, and coordinator of CeMEAI. His e-mail address is jacumina@icmc.usp.br.

J. M. Martínez is professor of mathematics at the State University of Campinas, Brazil, and deputy coordinator of CeMEAI. His e-mail address is martinez@ime.unicamp.br.

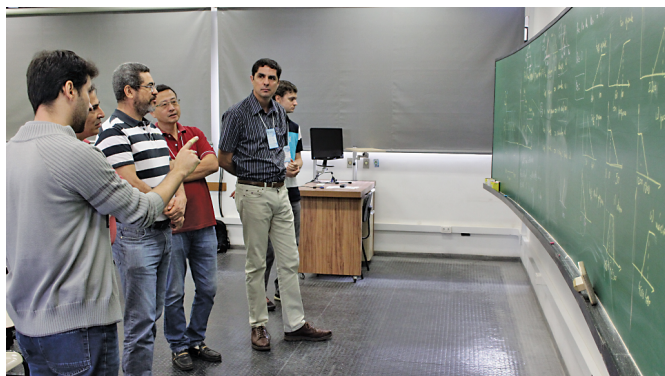
T. Pereira is associate professor of mathematics at the University of São Paulo, Brazil. His e-mail address is tiago@icmc.usp.br.

This work has been partially supported by FAPESP (grants 2010/10133-0, 2013/03447-6, 2013/05475-7, 2013/07375-0, and 2014/18711-3) and CNPq (grants 309517/2014-1 and 303750/2014-6).

For permission to reprint this article, please contact: reprint-permission@ams.org.

DOI: <http://dx.doi.org/10.1090/noti1513>

¹For further information on the study group reports and activities please visit the site www.cemeai.icmc.usp.br/2WSMPI.



An engineer from industry and problem presenter (Y. C. S. Ribeiro from Tecumseh), three academic mathematicians (A. C. P. L. F. Carvalho from ICMC-USP, G. N. Silva from IBILCE-UNESP, and Z. Liang from FFCLRP-USP), a postdoctoral scholar (C. Affonso from ICMC-USP), and a PhD student (P. H. Pisani from ICMC-USP) participating in the activities of the Second Brazilian Study Group with Industry held at the Institute of Mathematics and Computer Sciences of the University of São Paulo July 11–15, 2016.

der Waals repulsive interactions, atoms from different molecules must keep safe pairwise distances. Obtaining such a molecular arrangement can be considered a packing problem: Each type of molecule must satisfy spatial constraints related to the geometry of the system, and the distance between atoms of different molecules must be greater than some specified tolerance. Packmol [3] is software able to pack millions of atoms, grouped in arbitrarily complex molecules, inside a variety of three-dimensional regions. The regions may be intersections of spheres, ellipses, cylinders, planes, or boxes. The user must provide only the structure and geometrical constraints for a molecule of each type. Building complex mixtures and interfaces and placing biomolecules in water or other solvents is straightforward. In addition, different atoms belonging to the same molecule may also be restricted to different spatial regions in such a way that more ordered molecular arrangements can be built, such as micelles or lipid double-layers. The packing time for state-of-the-art molecular dynamics systems varies from a few seconds to a few minutes on a personal computer. The input files are simple and currently compatible with PDB, Tinker, Molden, or Moldy coordinate files.

Let us call $nmol$ the total number of molecules that we want to place in a region of the three-dimensional space. For each $i = 1, \dots, nmol$, let $natom(i)$ be the number of atoms of the i th molecule. Each molecule is represented by the Cartesian coordinates of its atoms. The point whose coordinates are the arithmetic averages of the coordinates of the atoms is called the barycenter. To facilitate visualization, assume that the origin is the barycenter of all the molecules. For all $i = 1, \dots, nmol$, $j = 1, \dots, natom(i)$, let $a^{ij} = (a_1^{ij}, a_2^{ij}, a_3^{ij})$ be the coordinates of the j th atom of the i th molecule.



Students and mentors discussing alternatives for solving a problem (at the front table: Professors J. Ueyema and M. O. Santos and students A. C. V. Figur, E. D. Bernardes, and F. A. Aureliano; in the back: students H. F. Vieira and L. Y. M. Alves).

Suppose that the i th molecule is sequentially rotated around the axes x_1 , x_2 , and x_3 at angles $\theta^i = (\theta_1^i, \theta_2^i, \theta_3^i)$. Moreover, suppose that, after these rotations, the whole molecule is displaced so that its barycenter becomes $c^i = (c_1^i, c_2^i, c_3^i)$. These movements transform the coordinates a^{ij} to $p^{ij} = (p_1^{ij}, p_2^{ij}, p_3^{ij})$. Observe that p^{ij} is always a function of (c^i, θ^i) :

$$(1) \quad p^{ij} = c^i + R(\theta^i)a^{ij},$$

where $R(\theta^i)$ is the rotation matrix given by

$$(2) \quad R(\theta^i) = \begin{pmatrix} c_1^i c_2^i c_3^i - s_1^i s_3^i & s_1^i c_2^i c_3^i + c_1^i s_3^i & -s_2^i c_3^i \\ -c_1^i c_2^i s_3^i - s_1^i c_3^i & -s_1^i c_2^i s_3^i + c_1^i c_3^i & -s_2^i s_3^i \\ c_1^i s_2^i & s_1^i s_2^i & c_2^i \end{pmatrix},$$

in which $s_k^i \equiv \sin \theta_k^i$ and $c_k^i \equiv \cos \theta_k^i$, for $k = 1, 2, 3$.

Our objective is to find angles θ^i and displacements c^i , $i = 1, \dots, nmol$, in such a way that, for all $i = 1, \dots, nmol$, $j = 1, \dots, natom(i)$, the point whose coordinates are $(p_1^{ij}, p_2^{ij}, p_3^{ij})$ satisfies the constraints imposed on the atom j of the molecule i . In addition, we wish that for all $i \neq i'$, $j = 1, \dots, natom(i)$, $j' = 1, \dots, natom(i')$,

$$(3) \quad \|p^{ij} - p^{i'j'}\| \geq d_{tol},$$

where $d_{tol} > 0$ is a user-specified tolerance. The symbol $\|\cdot\|$ stands for the usual Euclidean distance. In other words, the rotated and displaced molecules must remain in the desired region, and the distance between any pair of atoms of different molecules must not be less than d_{tol} .

A large variety of positioning constraints may be imposed individually on the atoms. Let r^{ij} be the number of constraints that apply to the j th atom of the i th molecule. These constraints can be represented as

$$(4) \quad g_\ell^{ij}(p^{ij}) \leq 0, \quad \ell = 1, \dots, r^{ij}.$$

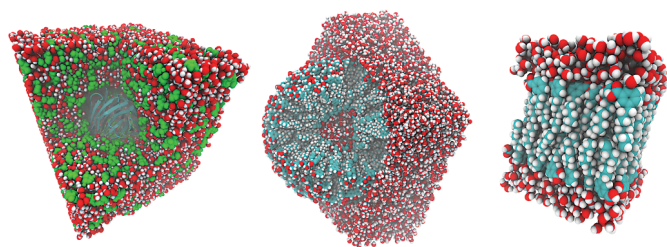


Figure 1. Molecular systems built with Packmol for molecular dynamics simulations: Left: a protein solvated with a mixture of water (red) and glucose (green); Center: a spherical lipid vesicle with water inside and outside; Right: a lipid bilayer with water above and below.

The objectives (3)–(4) lead us to define the following merit function f :

$$(5) \quad f(c, \theta) = \sum_{i=1}^{nmol} \sum_{j=1}^{natom(i)} \left(\sum_{i'=i+1}^{nmol} \sum_{j'=1}^{natom(i')} \max\{0, d_{tol}^2 - \|p^{ij} - p^{i'j'}\|^2\}^2 \right) + \sum_{i=1}^{nmol} \sum_{j=1}^{natom(i)} \left(\sum_{\ell=1}^{r^{ij}} \max\{0, g_{\ell}^{ij}(p^{ij})\}^2 \right),$$

where $c = (c^1, \dots, c^{nmol}) \in \mathbb{R}^{3 \times nmol}$ and $\theta = (\theta^1, \dots, \theta^{nmol}) \in \mathbb{R}^{3 \times nmol}$. (Remember the dependence of p^{ij} on the variables (c^i, θ^i) and the constants a^{ij} given by (1)–(2).) Note that $f(c, \theta)$ is nonnegative for all angles and displacements. Moreover, f vanishes if and only if the objectives (3)–(4) are fulfilled. This means that if we find displacements and angles where $f = 0$, the atoms of the resulting molecules fit the desired region and are sufficiently separated. This leads us to define the following unconstrained minimization problem:

$$(6) \quad \text{Minimize } f(c, \theta).$$

For solving (6) a combination of the local optimization solver Gencan [2] with heuristics devised to enhance the convergence to global (i.e. $f = 0$) minimizers is used.

Figure 1 presents examples of systems built with Packmol which illustrate some of the capabilities of the package. The corresponding input files can be obtained at the Packmol site or upon request to the authors.

Packmol is distributed as free software and can be downloaded from www.ime.unicamp.br/~martinez/packmol. Packmol has been downloaded more than 15,000 times by at least 9,000 different users. As a reflection of this, its main publication [3] has more than 800 citations registered in the Web of Science core collection (half of them in 2016), being one of the most cited articles written by Brazilian researchers only in all fields of research.

Estimation of the Thickness and the Optical Parameters of Superimposed Thin Films

The software PUMA [1] addresses the reverse engineering problem of estimating the thicknesses and the optical constants of a single thin film or superimposed thin

films deposited on a transparent substrate using only transmittance data through the whole stack. Estimations may also be done if the available data correspond to normal reflectance.

The problem of estimating the thickness and the optical constants of thin films using only transmittance (or reflectance) data is very challenging from the mathematical point of view and has a technological and an economic importance. It always represents a very ill-conditioned inverse problem with many local-nonglobal solutions. The ill condition of this reverse engineering process stems from the fact that the available transmittance data for retrieving the structure is incomplete and frequently noisy. So, as in highly underdetermined problems, extremely unstable or physically meaningless solutions are expected.

Several variants of the presented problem have been solved using mathematical optimization models and techniques. The inverse problem was solved for thin and for very thin films using spectral transmittance data only. An extension of the solution of the inverse problem using only reflectance data was also reported. The problem of retrieving the properties of an optical structure that includes more than one dielectric film, deposited either on one side or onto both sides of a transparent substrate of known optical properties, was addressed as well.

PUMA is freely available at www.ime.usp.br/~egbirgin/puma. The PUMA Project webpages have been visited more than 20,000 times by 8,000 different users. As a whole, papers related to PUMA have almost 300 citations registered in the Web of Science core collection. In response to users' requests, PUMA has evolved from a software that applies to the recovery of optical constants from transmittance data of systems with a single film to a widely applicable software able to tackle multilayer systems and to recover optical constants from reflectance data as well as from transmittance data.

Synchronization of Complex Systems

In many real-world complex systems, the ability to synchronize is a key property. Without a central controller or external driver, the system organizes itself in a collective state. Spontaneous synchronization is explored in our technologies. In power grids, power stations must keep proper synchronization to avoid blackouts. Wireless networks rely on synchronization among sensors to transmit information. Our everyday life is no different. Pacemaker cells in our hearts adjust their rhythm and behave in unison to deliver the strong electrical pulses that make our hearts beat.

Synchronization is also a double-edged sword. In the brain, epileptic seizures and Parkinson's disease are a strong manifestation of synchronization. Many engineering problems have their root cause in synchronization. A famous example is the London Millennium Bridge, when synchronization between pedestrians and the bridge's lateral motion led to strong bridge oscillations. Only in recent years have we been able to probe the delicate linking structure of real-world networks and its impact on network performance.

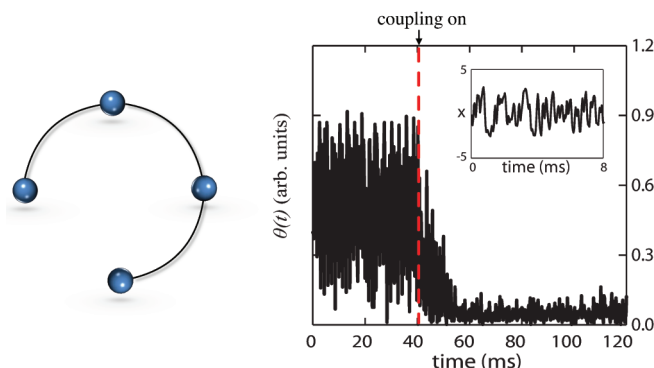


Figure 2. Dynamics of diffusively coupled lasers. In the beginning, the lasers are uncorrelated and voltage fluctuates wildly. When we turn on the coupling, we observe a rapid decay towards synchronization. In the inset, we show the voltage of the laser after they synchronize. The time delay induces a chaotic motion in the synchronized dynamics.

Synchronization of Lasers. When laser power is increased, the laser stability decreases. To create a high-power laser system, one technology is combining many low-power lasers into a single device. The question is how to make the lasers spontaneously behave in unison. We will illustrate the main characteristics of the phenomenon in an experiment described in [4]. The experiment consists of four coupled time-delayed optoelectronic oscillators (lasers) interacting according to a given network motif, as shown in Figure 2. We measure the voltage $x_i(t)$ of each laser and calculate the voltage fluctuation

$$\theta(t) = \sum_{i,j} |x_i(t) - x_j(t)|$$

as a measure of synchronization. When the lasers are uncoupled they behave asynchronously. However, when the coupling is increased above a critical threshold the lasers synchronize, as observed in Figure 2.

Modelling and Results. We now discuss a model that has illustrated how the linking structure affects synchronization. We consider an undirected network of identical elements with diffusive interaction

$$(7) \quad \dot{\mathbf{x}}_i = \mathbf{f}(\mathbf{x}_i) + \alpha \sum_{j=1}^n A_{ij} \mathbf{H}(\mathbf{x}_j - \mathbf{x}_i),$$

where $i = 1, 2, \dots, n$, $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a smooth vector field, and $\alpha \geq 0$ is the overall coupling strength representing an energy cost per link. The matrix \mathbf{A} describes the network structure; i.e. $A_{ij} = 1$ if i and j are connected and 0 otherwise. The local coupling function \mathbf{H} is smooth satisfying $\mathbf{H}(\mathbf{0}) = \mathbf{0}$. This last condition guarantees that the synchronization subspace $\mathbf{x}_1 = \mathbf{x}_2 = \dots = \mathbf{x}_n$ is invariant for all values of α . The key question is whether the subspace is attractive. That is, given that the initial difference between states is small, we wish to know whether solutions of the network synchronize:

$$\lim_{t \rightarrow \infty} \|\mathbf{x}_i(t) - \mathbf{x}_j(t)\| = 0 \quad \text{for any } i \text{ and } j.$$

The model in (7) captures many aspects of the applications we have discussed so far. For instance, in neuron networks \mathbf{f} represents the isolated neuron dynamics and

\mathbf{H} the electric synaptic coupling. The model may also represent networks of biological oscillators, networks of electric circuits with resistive coupling, or networks of lasers. For lasers we should include a time-delay coupling in the function \mathbf{H} . This is no restriction, and the present analysis cascades to the time-delay under minor modifications.

An undirected network has a symmetric adjacency matrix. The *degree* k_i of the i th node is the number of connections it receives, $k_i = \sum_j A_{ij}$. The Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D} = \text{diag}(k_1, \dots, k_n)$, plays a major role in synchronization analysis. \mathbf{L} is positive semidefinite; its eigenvalues, enumerated in increasing order and repeated according to their multiplicity, $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$, dictate the onset of synchronization. The second eigenvalue, λ_2 , is bounded away from zero whenever the network is connected. The smallest nonzero eigenvalue is known as the spectral gap. To avoid technical discussions, we won't state a theorem on synchronization, but rather we will discuss the key ideas. Under certain mild assumptions there are $\beta_1 = \beta_1(\mathbf{f}, \mathbf{H})$ and $\beta_2 = \beta_2(\mathbf{f}, \mathbf{H})$ such that if

$$(8) \quad \beta_1 < \alpha \lambda_2 \leq \alpha \lambda_n < \beta_2$$

the network synchronizes. Hence, the condition in (8) provides a sufficient condition for synchronization, as well as a splitting of the problem in terms of the dynamical quantities (\mathbf{f}, \mathbf{H}) and network structure by means of the eigenvalues of the Laplacian. Obtaining β_1 and β_2 is delicate, but it can be done in an algorithmic way.

This field has been very prolific in recent years, and we have come to understand many relations between the network structure and the stability of synchronization. Some of the main things we have learned are:

- *Typically, random networks synchronize better than regular networks.* We compare networks with the mean same degree m and same size n . Let's consider a $2K$ regular network; that is, each node has two $2K$ connections to its nearest neighbors. If $K \ll n$, then $\lambda_2 \rightarrow 0$, whereas in the random network $\lambda_2 \approx m$. Moreover, adding a small fraction of links at random favors synchronization, as it increases λ_2 significantly while keeping λ_n basically unchanged.

- *Networks with heterogeneous degree synchronize poorly.* For example, if the network is random, then roughly $\lambda_n \approx d_{\max}$ and $\lambda_2 \approx d_{\min}$ so the condition reduces to $d_{\max}/d_{\min} < \beta_1/\beta_2$. If the network degrees are very disparate it won't be possible to achieve synchronization. Typically, the maximum degree is a function of the network size $d_{\max} = d_{\max}(n)$. So, this class of system will present a maximum system size for which synchronization is attainable.

These results allow identifying the dynamical importance of a link and thereby have a major impact on the design and control of directed networks. They also shed light on how to plan and design network modifications without destroying the network performance, for instance in power grids.

Generalizations. Although we started with identical nodes, this is no severe restriction, as synchronization is persistent under perturbations of the vector field [5]. An important generalization is to consider directed and weighted networks. Recent work has provided sufficient conditions to guarantee the stability of synchronization in directed networks and weighted networks [5]. In this case, small perturbations in the structure can cause major changes in the dynamics. Understanding the impact of structural modifications, such as changing weights and adding or deleting links, on synchronization remains an open problem.

Concluding Remarks

In this note we described a few applications which, having matured over considerable time, have shown their innovation potential. Currently, researchers of the center are working on other applications in medicine, engineering, and sports. We hope for improved processes in these fields and useful feedback for the development of new mathematics.

References

- [1] E. G. BIRGIN, I. CHAMBOULEYRON, and J. M. MARTÍNEZ, Estimation of the optical constants and the thickness of thin films using unconstrained optimization, *Journal of Computational Physics* **151** (1999), 862–880.
- [2] E. G. BIRGIN and J. M. MARTÍNEZ, *Practical Augmented Lagrangian Methods for Constrained Optimization*, *Fundamentals of Algorithms* **10** (2014), SIAM, Philadelphia, PA. MR3186234
- [3] L. MARTÍNEZ, R. ANDRADE, E. G. BIRGIN, and J. M. MARTÍNEZ, PACKMOL: A package for building initial configurations for molecular dynamics simulations, *Journal of Computational Chemistry* **30** (2009), 2157–2164.
- [4] J. D. HART, J. P. PADE, T. PEREIRA, T. E. MURPHY, and R. ROY, Adding connections can hinder network synchronization of time-delayed oscillators, *Physical Review E* **92** (2015), 022804.
- [5] T. PEREIRA, J. ELDERING, M. RASMUSSEN, and A. VENEZIANI, Towards a theory for diffusive coupling functions allowing persistent synchronization, *Nonlinearity* **27** (2014), 501–525. MR3168262

Credits

Photo of Second Brazilian Study Group is courtesy of B. D. Grachet.

Photo of problem-solving group is courtesy of B. D. Grachet and L. Sarno.

Figure 1 is courtesy of L. Martínez.

ABOUT THE AUTHORS

Ernesto G. Birgin is a member of the editorial boards of *Journal of Global Optimization*, *Computational and Applied Mathematics*, *Bulletin of Computational Applied Mathematics*, and *Pesquisa Operacional*. He has published over seventy papers on computational optimization and applications.



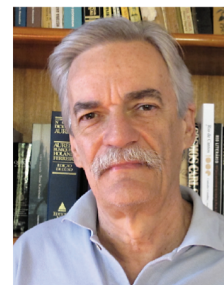
Ernesto G. Birgin

Jose Alberto Cuminato, coordinator of CeMEAI, is a member of the Brazilian Academy of Science and corresponding fellow of the Royal Society of Edinburgh (FRSE), member of the editorial board of *Computational and Applied Mathematics*, and author of over 50 papers on CFD, numerical mathematics, and applications.



Jose Alberto Cuminato

José Mario Martínez, deputy coordinator of CeMEAI, is a member of the Brazilian Academy of Science, former editor in chief of *Computational and Applied Mathematics*, member of the editorial board of *Numerical Algorithms*, and the author of over 180 papers on numerical mathematics, optimization, and applications.



José Mario Martínez

Tiago Pereira is a member of SIAM and author of over thirty papers on dynamical systems and complex networks.



Tiago Pereira