from *SIAM News*, Volume 33, Number 10

IPAM Gets in Gear with a Year of Genes and Geometry

By Barry A. Cipra

Mathematicians jonesin' for interdisciplinary research have a new place to go for it: the Institute for Pure and Applied Mathematics, located on the campus of the University of California at Los Angeles. IPAM has opened with a fall program on functional genomics, to be followed next spring with a program on geometrically based motions. IPAM will also offer short programs on financial mathematics in January, and on oscillatory integrals and dispersive equations in March.

Director Tony Chan and co-directors Mark Green and Eitan Tadmor hope the format of the new institute will expand the role of mathematics in other disciplines. "One of IPAM's goals is to break down barriers between mathematicians and other scientists," Chan says. "We like to think IPAM is the right idea at the right time."

One of the emphases of the fall program is on mathematical and statistical aspects of a new biotechnology known as gene expression microarrays. The spring program will also include some high-biotech, with a workshop on medical imaging.

Computational Genomegnosis

Don't it make my brown eyes blue?
—Crystal Gale

Functional genomics is an especially timely subject, says IPAM co-director Mark Green. With the recent joint announcement by Celera Genomics and the government-sponsored Human Genome Project of the "complete" sequencing of the human genome, researchers' attention has increasingly shifted to the even more demanding task of sifting the gigabytes of data, trying to figure out what the tens of thousands of genes do and how they work together—the "functional" part of genomics.

"Science now has a huge body of genetic information, and researchers need methods—algorithms to search the data, clustering methods and computer models, among others—to interpret it," Green says. "The convergence of mathematics and the life sciences, which was not foreseen a generation ago, is a tremendous opportunity."

Roughly speaking, the human condition is defined by upwards of a hundred thousand genes, scattered across 23 chromosomes. (The exact number of genes is yet to be determined. There is, in fact, a friendly "gene pool" going, with scientists wagering on the exact number. Current estimates range from forty to more than a hundred thousand.)

Each gene contains a "blueprint" for a specific protein and a "program" for controlling its production. The gene itself is a string of DNA, and the protein it produces is a string of amino acids. When a gene is expressed, it is transcribed into a complementary molecule of messenger RNA (mRNA), which is then used to build the protein. Every cell contains the entire genome, but different cells express different genes at different times. Just how all that expression is orchestrated is one of the great challenges for modern science.

One of the first jobs is to determine which genes are being expressed, and to what extent. That's where gene expression microarrays come in.

A microarray consists of tens of thousands of tiny dots of DNA on a glass slide. Each dot represents a particular gene. The slide is washed with a target cell's mRNA, to which a fluorescent dye has been attached. To be more precise, the microarray is washed with equal concentrations of mRNA from two different cells, one labelled with red dye and the other with green. The ratio of the intensities of the two colors at each DNA dot is a measure of the comparative level of expression of the corresponding gene. The color-coded results can highlight, for example, the differences between a cancerous and a noncancerous liver cell, or the effects of a cancer-fighting drug. Similarly, a sequence of microarray comparisons can demonstrate the timing of gene expression.

In a project called NCI60, The National Cancer Institute has been compiling microarray (and other) data on 60 lines of cancer cells. Researchers have also found evidence that some of the variability in outcome—why one patient responds to chemotherapy and another doesn't—is correlated with differences in gene expression. In effect, two people with the "same" cancer may actually have different diseases. If such findings hold up, microarray testing could become a powerful diagnostic tool.

Microarrays "have a lot of potential to give tons and tons of information," says Kathleen Kerr, a biostatistician at Jackson Laboratories in Bar Harbor, Maine, who is spending the fall at IPAM. But the data are extremely noisy, she adds. "That's where statistics comes in."

The mathematical and statistical problems associated with microarrays are formidable. Image analysis of the fluorescent signals is one: Those tens of thousands of dots have to be identified. Indeed, deciding what DNA to put at which dot is a problem in itself, with overtones of combinatorics and optimization. And once obtained, the data from a microarray become, in effect, a single point in a very high dimensional vector space, with all the associated—and largely unsolved—problems of multivariate analysis in very high dimensions.

Even denoised data can be misleading, Kerr points out. For example, the ratio of the color intensities at a given dot doesn't automatically translate into a ratio of gene expression. Researchers have found that two runs of the same experiment, with the colors reversed (e.g., the red dye attached to the cancerous cells in the first run, and to the noncancerous cells in the second, and vice versa for the green dye), often give different results. “There’s no clear reason it should happen, but we keep seeing it,” Kerr says.

Mark Ettinger, a mathematician at Los Alamos National Laboratory who is also at IPAM for the fall program, says that data from microarrays can be viewed, “to the zero-th order approximation,” as generated by a large dynamical system. (Ettinger is typical among mathematicians for the way he came to mathematical biology. After completing a PhD dissertation on combinatorial games, he “switched fields completely,” first to quantum computing and more recently to biological problems. He credits colleagues at Los Alamos with getting him interested in complexity issues of computational biology.)

Suppose, Ettinger suggests, you have microarray data on, say, 5000 genes for 50 time points during cell division. If the data were perfect—if you could tell exactly which genes are turned on at each point in time, so that the whole system could be viewed as a 50×5000 matrix of 0’s and 1’s—then you could hope to see how the expression of one gene, or cluster of genes, promotes or inhibits the expression of other genes. In the process, you would have mapped out a flow chart of gene regulation. But “there’s a certain amount of crudeness to the dynamical systems point of view,” he says.

Kerr concurs. “Some people have way too much faith in what a computer program can do for them,” she says. Microarray data can help form hypotheses, but they’re not enough to identify genetic pathways unambiguously. “It’s not because the software isn’t good enough,” Kerr says. “It’s just a fact of science and of life.”

Progress in genomics, then, will continue to depend on cleverly designed lab experiments with a healthy ratio of signal to noise. But with the help of mathematical and statistical analyses, good data will give researchers a shot at untangling the myriad mysteries of functional genomics. And it may not take too long. The field “is moving incredibly quickly,” Kerr says. “It’s incredibly exciting.”

Curves Ahead

There’s something in the way she moves
—James Taylor

The spring 2001 program on geometrically based motions will also address some pressing problems in modern technology, in particular materials science, geophysics, and medical imaging. “Motion driven by the geometry of interfaces is ubiquitous in many areas of science, from growing crystals for manufacturing semiconductors to tracking tumors in biomedical images,” Tadmor explains.

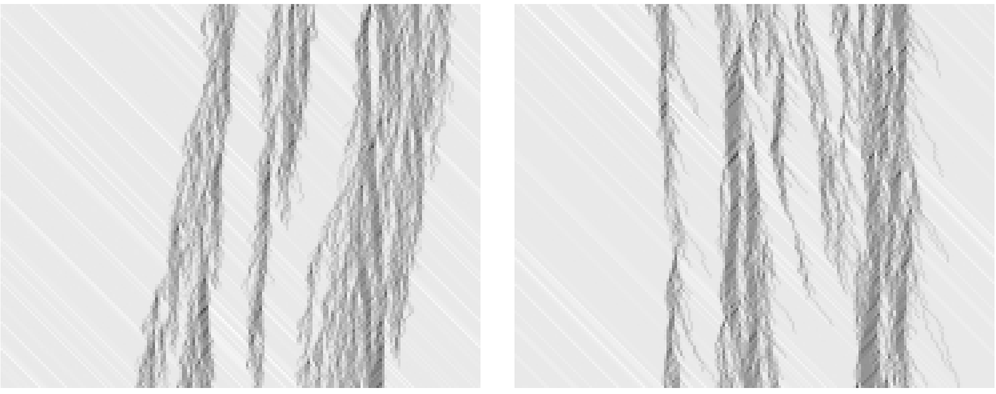
Level sets and viscosity solutions for PDEs are among the mathematical methods that have emerged over the last decade as powerful tools for analyzing geometric objects with mobile boundaries. Stochastic PDEs and threshold dynamics are also being used to study what goes on at moving interfaces. The IPAM program will include workshops on material interfaces (April 9–13), geometrically based high-frequency waves (April 18–20), moving interfaces and threshold dynamics (May 9–11), and image processing in medicine and the neurosciences (May 21–25). There was also a preliminary workshop in September, sponsored by the Office of Naval Research, on wavelet- and PDE-based imaging techniques.

One of the goals in medical imaging is to automate processes traditionally done by humans, such as identifying the boundary of a tumor or specific regions in a brain scan, says Guillermo Sapiro of the University of Minnesota, one of the organizers for the late-May workshop. Part of the motivation is economic: “The more automatically you do things, the cheaper they become,” Sapiro points out. But automation also opens up new possibilities. For example, “we can do statistics on thousands of brains,” he says. “Sometimes just by computer mathematical analysis, we end up discovering things that we couldn’t anticipate before.”

Geometrically based motion is a means to an end in image processing, Sapiro says. Finding the extent of a tumor, for example, can be accomplished by “growing” a curve that starts out well within the affected region, using algorithms that respond to gradients within the image. Image registration—comparing, say, two brain scans—can be accomplished by a mathematical deformation that carries one image smoothly to the other.

Even denoising an image can be conceived of as a geometric motion. “Imagine that you have a very noisy image as the initial condition of a differential equation,” Sapiro explains. “We could design the differential equation in such a form that the steady state [solution] is a clean version of the image.” The differential equation prescribes a geometric deformation based on curvature motion. “The deformation is what gets us from the ‘bad’ original image or original curve to the good one.”

That’s easier said than done, of course. Despite the high-tech veneer of image processing, much work in the field is still done on an ad hoc basis. “In a certain sense, we are trying to put in a lot of mathematics, but there is no universal thing that we can say this is what we need to optimize



Clear phase separation is seen in traffic jams, both backward-moving (left) and forward-moving (right). Jams can occur even when there’s plenty of room on the road. Courtesy of David Griffeath.

IPAM Basics

With a five-year, $12.5 million grant from the National Science Foundation, IPAM joins the NSF-funded Institute for Mathematics and Its Applications at the University of Minnesota and the Berkeley-based Mathematical Sciences Research Institute, which have been operating since 1982. The awards to the three centers were announced last year by NSF's Division of Mathematical Sciences, which is currently soliciting proposals for as many as four more mathematical research institutes.

DMS director Philippe Tondeur, Rolf Jeltsch, president of the European Mathematics Society, and Felix Browder, president of the American Mathematical Society, were on hand in August for the new center's inauguration. SIAM president Gilbert Strang was unable to attend, but sent a cordial letter in which he called the new institute "a great event." There's an increasing need for teamwork in mathematics, he wrote, and "out of that need comes the creation of an Institute like this." The opening ceremony was followed by a reception at IPAM's new digs, in a building near the mathematics department designed by the renowned architect Frank Gehry. (Compared with Gehry's recent designs, such as the Experience Music Project in Seattle or the Guggenheim Museum in Bilbao, Spain, not to mention the new academic center planned for MIT, the IPAM building, built in 1973 as UCLA's Career and Placement Center, is a relatively conventional structure.)

The IPAM format, Chan explains, is to maintain a mix of mathematicians and other scientists. Each program will have a core group of four senior researchers and eight postdoctoral visitors, with approximately half from each side of the interdisciplinary chasm. In addition, IPAM will support an average of 20 other visitors at any given time, for periods ranging from days to weeks. Chan estimates that each program will engage about 200 participants altogether.

The postdoctoral portion of the program is broadly construed to include recently tenured faculty. A key provision of the institute is that IPAM will not offer multi- or even full-year postdoc positions. Instead, visitors are expected to have home institutions. The center's main goal, Chan stresses, is to establish research contacts that will continue beyond people's stay at IPAM.

Toward that goal, IPAM plans to hold "reunion" conferences for each major program a year or so after its conclusion. The reunions will be held at UCLA's conference center at Lake Arrowhead, a lakeside resort in the San Bernardino mountains. "We think it's a good experiment to try," Chan says. "We hope it'll work."

As they get the current year's programs up and running, Chan and co-directors Green and Tadmor are busy organizing the schedule for 2001–02. The major programs on tap are conformal field theory for the fall and communication networks for the spring. As an institute oriented toward interdisciplinary research, IPAM's job is not just to find areas that are mathematically interesting. "We also try to look at what is driving science," Chan says. "We want to be at the forefront of that."

for," Sapiro says. "Many times it happens that one algorithm is the best for one group of applications, and another algorithm is the best for another group of applications."

The effort to convert ad hoc approaches into methods based on fundamental principles makes image processing an attractive area for mathematicians. "Whatever your mathematical background, I think that you have something to contribute," Sapiro says. "We have been very successful for many areas of application, but we still have a long way to go."

The same is true in threshold dynamics, says David Griffeath of the University of Wisconsin, an organizer of the IPAM workshop planned for early May. Threshold-driven interfaces are crucial in a range of physical phenomena, from crystal growth to traffic jams. The key idea is an all-or-nothing nonlinearity: Some interactions occur only when local conditions cross a certain threshold. Once that happens, the interactions can propagate, often in unexpected ways.

"There's a very broad range of dynamics that have these threshold ingredients in them," Griffeath says. The IPAM workshop "is going to bring together people from different areas of mathematically related fields to share technologies for making progress on how these things behave. They're notoriously difficult systems to study."

Jam Session

Hit the road, Jack
—Ray Charles

Griffeath is especially keen these days on traffic jams. "This is my favorite thing in about 10 years," he says. Inspired by research done at Los Alamos National Laboratory by Kai Nagel (who is now at ETH Zurich), Griffeath and Larry Gray of the University of Minnesota have discovered an extremely simple abstract model that exhibits many of the essential features of traffic jam formation. In particular, the model indicates that jams are inevitable once traffic reaches a critical density—even when there's plenty of space for everyone to spread out and drive at the speed limit.

Griffeath and Gray's model posits identical cars on a one-lane road. For extra simplicity, the road is a discrete set of points, each of which is either empty or occupied by a car. (In theory there are infinitely many road points, but in computer simulations the road is turned into a big circle—think racetrack or belt-way—typically with ten to fifty thousand points.)

Time is also discrete. At each time step, each car either moves forward one unit or stays where it is, according to a simple set of rules. To begin with, if the point immediately in front of a car is occupied, the car stays put. Otherwise, it checks whether the next point (two units ahead) is occupied, and likewise the point immediately behind it—i.e., is there open road ahead, and am I being tailgated?

There are four possible scenarios, and a car moves one unit forward with probability α, β, γ, and δ, depending on which scenario it sees. In what Griffeath likes to call "a brilliant stroke of nomenclature," the parameters stand for acceleration, braking, congestion, and driving. You accelerate when there's clear road ahead and someone's on your tail; you brake in the opposite case; congestion is when you're crowded on both sides (but can still inch forward); and driving is what you do when there's open road ahead and behind. Typically, δ is set to 1, making one unit per time step the ideal speed at which traffic can flow. "We've studied how an initially random distribution of cars progresses down the road for different values of α, β, γ, and δ," Griffeath explains. Further simplifications turn the parameter space into an easy-to-visualize two-dimensional system, "which we have analyzed in gory detail in the last year and a half."

For starters, Griffeath and Gray have found that they can reproduce the qualitative features of Nagel's more complicated models. In particular, a key factor in traffic jams is seen to be what Nagel calls "slow to start" dynamics: Congestion tends to persist in large part because cars at the head of a jam are slow to accelerate. In their model, this corresponds to $\alpha < \delta$. "It basically means that if you've been in a jam and now see a bunch of wide-open space in front of you, it actually takes a little bit of time to ramp up in speed," Griffeath says. So next time you've been stuck in traffic, floor it as soon as you get the chance.

Their model also suggests a proper, albeit abstract, mathematical definition of a jam-prone system, Griffeath points out, because when congestion occurs, it often occurs on all scales. For many typical parameter values and traffic densities, the infinite system will, over time, evolve into arbitrarily large stretches of congested traffic separated by arbitrarily large stretches of free flow. In other words, their traffic model is a conservative system (there are no on or off ramps, so the number of cars is constant) that exhibits phase separation.

"That's a kind of behavior that people in physics are very interested in," Griffeath says: Although their model is much too simple to use in practical traffic management, Griffeath is "pretty convinced that this behavior that we're describing is in some sense fundamental to a lot of traffic patterns and is very robust." Workshop participants who come by car via the LA freeways may see just how robust it is.

Barry A. Cipra is a mathematician and writer based in Northfield, Minnesota.