Mathematics of Information

Arieh Iserles, and Carola-Bibiane Schönlieb

Department for Applied Mathematics and Theoretical Physics,
University of Cambridge, UK

Wednesday 2$^{\text{nd}}$ October, 2013

1 The second industrial revolution

Our world has been shaped by the outcome of the scientific revolution and its offspring, the industrial revolution. The essence of this scientific revolution, shaped by its pioneers – Newton, Gallileo, Descartes, Kepler, Huygens, Boyle, Leibnitz – was in the rigorous understanding of the physical world, the laws underlying matter, energy and their interaction. The word "rigorous" is a hint to the fundamental role of mathematics in this endeavour but, one way or the other, the process has been driven by physics. Ultimately, physics led to applied physics (also known as engineering) and to the industrial revolution. Humanity has been changed forever.

We are living in interesting times, arguably at a juncture of an equal importance and impact to the original scientific revolution. The world around us is changing and it is clear that the future will be as unrecognisable to us as our world – the Internet, mobile telephony, medical imaging, satellite navigation, social networks, the entire panoply of personal computers, laptops, tablets, smart phones so ubiquitous in our daily life of work and leisure... – would have been incomprehensible to previous generations.

In as much as physics is fundamental to these developments, from quantum to solid state physics to nanotechnology, the common denominator underlying them is not physics but *information.* We are surrounded by technology that collects, transmits, manipulates and ultimately needs to understand reams of information, of an order of magnitude which is hard to comprehend: in 2012, in every single minute the Internet processed 2×10^8 emails, 2×10^6 GOOGLE search queries, 7×10^5 FACEBOOK postings and 48 hours of YOUTUBE uploads [18]. At the same minute CERN computers would have generated up to 600 gigabytes of data: ostensibly a theoretical physics centre, its main tool of trade is information. Even such emblematic 'physics machines' like cars are increasingly crammed with information systems, while avionics dominate the cost of modern jet fighters.

Mathematics has historically been the language of physical reality and its success there underpins all of modern science and technology [19]. Given the very nature of information, there is every reason to believe that mathematics is, to an even greater extent, the language of data. This has profound implications to the future of applied mathematics, as is the ever-present tension between *data* (values of qualitative or quantitative variables) and *information* (the content and meaning present in data).

Impact on mathematics Applied mathematics represents the broad interface between mathematical methodology and application areas. It is responsive, constantly changing to address challenges from applications, but also proactive, because mathematical tools ceaselessly reshape applications in science, technology and medicine. At a greater remove, pure mathematics is also shaped by challenges originating in application areas: thus, differential geometry had responded to general relativity, functional analysis to quantum mechanics and algebraic geometry to string theory. And the theory of differential equations has responded to endless problems occurring in a wide range of applications...

What are the new challenges, originating in the information revolution? These are early days and it is premature and foolish to seek a comprehensive list. Tentatively, the following themes have emerged in the last decades:

- ***Imaging.*** Modern medicine is increasingly based on the use of scanning devices. The challenge here is to produce a high quality description of patients and their ailments from data which is necessarily limited by the capability of scanners and the need to minimise exposure to harmful radiation. Imaging, however, ranges way beyond medicine, e.g. to reservoir modelling. Oil companies consider themselves lucky if the manage to recover 35% of reserves in an oil field: the key to better exploitation is improved imaging of oil reservoirs from seismic data.

- ***Data mining.*** The following situation is pervasive across wide fields of applications, from bio-informatics (in particular, genomics) to search engines, from forensics to marketing analytics: we are presented with huge quantities of data and need to 'mine' some deeper meaning. Most of the data is likely to be irrelevant or redundant in the underlying context (recall the tension between data and information!) and the little which is significant is invariably contaminated by noise, yet the imperative is to make sense out of it.

- ***Networks.*** Data flows across networks of an increasing complexity. It is switched and rerouted repeatedly (and automatically) to increase speed and reliability and in response to varying loads. The challenge is deceptively similar, from the flow of data on the Internet to mobile telephony to the electricity grid to the distribution of goods, but each medium has its own rules and characteristics.

- ***Signal processing.*** Most data is communicated as a long sequence of bits. How to do this efficiently and reliably? How to clean up noise without distorting meaning? How to represent data, indeed how to sample it so that it can be communicated efficiently? Inasmuch as this range of problems has been with us for almost a century, its very urgency is underlied by a step change in the amount of data that need be transmitted.

- ***Machine learning.*** Since the dawn of computing (or even earlier, if you include fiction) humankind has been concerned with 'smart machines': artificial intelligence, robots, automata…'Smart' is all about endowing information with meaning in an organised manner. Machines need to *acquire* information – thus, not just 'read' an image as an aggregate of pixels but 'understand' it as a collection of moving objects – and *comprehend* it: understand the unfolding scene and decide how to react to it subject to predetermined goals.

What does it mean for mathematics? Here our guesses are, if at all, even more tentative. Clearly, entire new mathematical subjects are bound to evolve in response to the challenge of information and it is unwise to speculate too much. By this stage all we can do is sketch crude outlines, based on the snapshot of our current understanding.

The challenges posed by information have several broad mathematical organising principles:

1. Both the process of measuring information and the imperfection of our mathematical models mean that data is invariably contaminated by noise and uncertainty. Thus, *statistics* is an essential tool.

2. A major feature of information is that there is plenty of it! Thus, its understanding calls for large scale *computation*: numerical linear algebra, computational differential equations and, in particular, large-scale optimization. Much of this computation is new in kind, shifting the classical centre of gravity of numerical analysis.

3. Signals need be represented and approximated in the correct mathematical framework. In practical terms, this means *harmonic analysis* and *approximation theory.*

4. A major challenge is to recover a *sparse structure* hidden in data: to describe 'big data' using a small number of variables – ideally, reduced to a small number of observations. This calls for techniques from *combinatorics, functional analysis* and, again, *computation.*

5. To recover and understand spatial structure, an image hiding in a long sequence of bits, we need geometric insight, hence *geometric* and *topological analysis.*

6. Mathematical analysis of data rapidly leads to problems defined in a very large number of dimensions. This is a challenge for a number of reasons, not least the well-known *curse of dimensionality* [3] in computation. Yet,

it is also an opportunity, because techniques of *functional analysis, measure theory* and *convexity theory* can sometimes turn this curse into a blessing of dimensionality.

7. A typical problem arising in the analysis of information admits many solutions: it is an *inverse problem.* It has no 'right' or 'wrong' solution in a standard scientific sense. For example: we may have an image except that small bit is missing and there is an infinite number of ways to 'inpaint' it. The human eye (to be more exact, the human mind) is quite good at this sort of task, the challenge here is to 'educate' a computer to do this just as well – or even better.

The impact of the information revolution on applied mathematics is bound to be profound, much more than just a change in the sort of problems attracting the attention of academic and industrial mathematicians of the next generation. The entire 'graph of connectivity' of mathematics, a graph that has been shaped since the dawn of the 'physics revolution' of Galileo and Newton, is likely to change. A significant proportion of information-heavy problems arises in areas which are new to the experience of mathematicians, e.g. medicine, biology or social sciences, as well as in the less classical areas of engineering: information engineering, software engineering and electronics. We are, indeed, living in interesting times.

To flesh out the dry concepts and convey some of the excitement of mathematics of information, we briefly review two lively current themes of research: image processing and sparse recovery.

2 Image Processing

Digital images are one of the main sources of information. As in our other examples before – the plain vastness of images and videos that exist in our digital system nowadays makes their unaided processing and interpretation by humans impossible. Automatic storage management, processing and analysis algorithms are needed to be able to retrieve only the essence of what the visual world has up its sleeve. Moreover, certain acquisition devices – such as magnetic resonance tomography or remote sensing of the atmosphere – do not immediately provide us with the kind of information relevant to our needs. Mathematical inversion algorithms are needed to extract this information from the physical and statistical laws that relate the measurements with the image.

Before we go any further we first need to understand what a digital image really is. Roughly speaking it is obtained from an analogue image (representing the continuous world) by sampling and quantization. Basically this means that the digital camera superimposes a regular grid on an analogue image and assigns a value, e.g., the mean brightness in this field, to each grid element. In the terminology of digital images these grid elements are called pixels. The image content is then described by grey values or colour values in each pixel. The grey values are scalar values ranging between 0 (black) and 255 (white). The colour values are vector values, e.g., (r, g, b), where each channel r, g and b represents the red, green, and blue component of the colour and ranges, as the grey values, from 0 to 255. The mathematical representation of a digital image is a so-called image function u defined (for now) on a two dimensional (in general rectangular) image domain, the grid. Indeed, in some applications, images are three dimensional (e.g. videos, 3D medical imaging) or even four dimensional (involving three spatial dimensions and time) objects, but for simplicity we focus on the two dimensional case for the following conceptual presentation. This function is either scalar valued in the case of a grey value image, or vector valued in the case of a colour image. Here the function value $u(x, y)$ denotes the grey value, i.e., colour value, of the image in the pixel (x, y) of the image domain. Figure 1 visualizes the connection between the digital image and its image function for a grey value image.

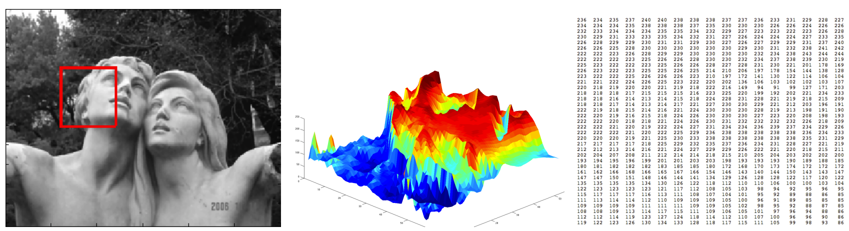


Figure 1: Digital image versus image function: On the very left a grey value photograph; in the middle the image function within a small selection of the digital photograph is shown where the grey value $u(x, y)$ is plotted as the height over the (x, y) - plane; on the very right the grey values for a small detail of the digital photograph are displayed in matrix form.

Typical sizes of digital images range from 2000×2000 pixels in images taken with a simple digital camera, to 10000×10000 pixels in images taken with high-resolution cameras used by professional photographers. The size of images in medical imaging applications depends on the task at hand. Positron emission tomography (PET) for example produces three-dimensional image data, where a full-length body scan has a typical size of $175 \times 175 \times 500$ pixels.

Image de-noising In most acquisition processes for digital images wrong information is added to an image. Even modern cameras which are able to acquire high-resolution images produce noisy outputs, cf. Figure 2. In fact, the appearance of noise is an intrinsic problem in image processing. When presented with a noisy image or data the task is to identify and remove the noise while preserving the most important information and structures. While for the human eye, noise is an easy problem to cope with – indeed if the noise is not too strong we are still able to analyse an image for its contents – this is not the case for the computer. This is an important insight when aiming for an automated analysis of an image.

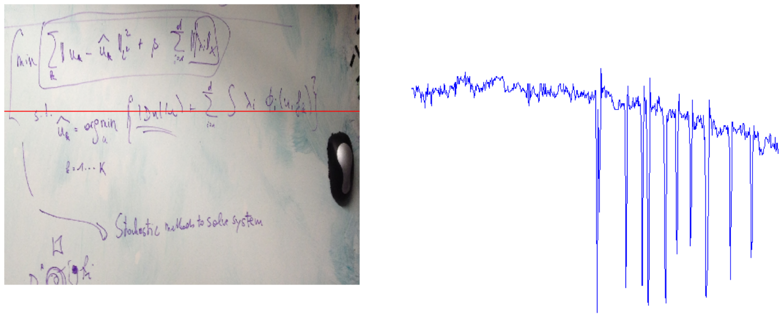


Figure 2: Bad lighting conditions may result into noisy images. Left: A digital photo which has been acquired under poor lighting. Right: Plot of the grey values of the red channel along the straight line in the photograph.

In medical, seismic, and biological imaging or for certain visualisation tasks in chemistry, physical imaging tools are employed to visualise the inside of the body, the earth, a cell or chemical reactions. In such applications, it is usually not the image that is measured but noisy samples of its Fourier or Radon transform for instance. In these cases, we want to reconstruct an approximation of the original image density from (usually under-) sampled and noisy transform data by 'smoothly' inverting the transformation, cf. Figure 3.

Noise in an image usually constitutes a highly oscillatory (high frequency) component of the acquired image data. One way to think about de-noising an image is to smooth (or regularise) that image, aiming to 'smooth' away the noise. The simplest and best investigated method for regularising images is to apply a linear filter. One example of such a filter is Gaussian regularisation. However, linear filters such as the Gaussian are not recommended when aiming for noise reduction and structure preservation at the same time. While the Gaussian filter removes the noise it also blurs the intrinsic image structures, see Figure 4b. One of the most successful image denoising approaches counteracting the introduction of blur is nonlinear PDEs and non-smooth variational models [2]. Here, the denoised image is modelled as a function in the continuum and computed as a solution of a differential equation or a minimiser of a convex (or even non-convex) functional. Among the pioneers of these approaches are Rudin, Osher and Fatemi who introduced in 1992 the total variation for image regularisation [17]. In particular, for a noisy image g defined on a rectangular image domain $\Omega \subset \mathbb{R}^2$ the standard total variation denoising approach computes

[2] Data provided by the MPI for biophysical chemistry Göttingen

[3] Photo courtesy of Kostas Papafitsoros

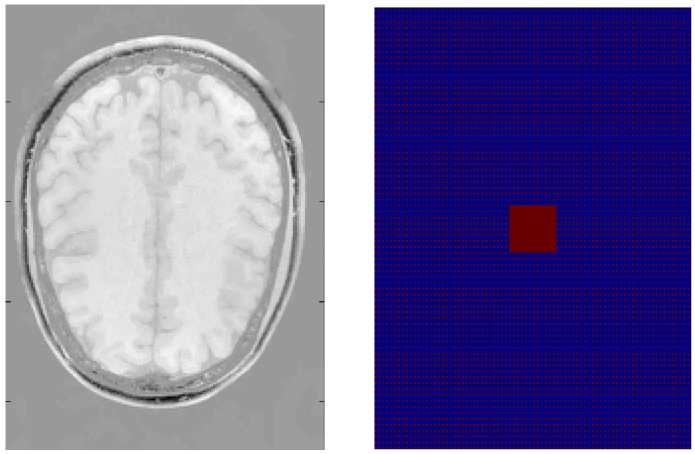
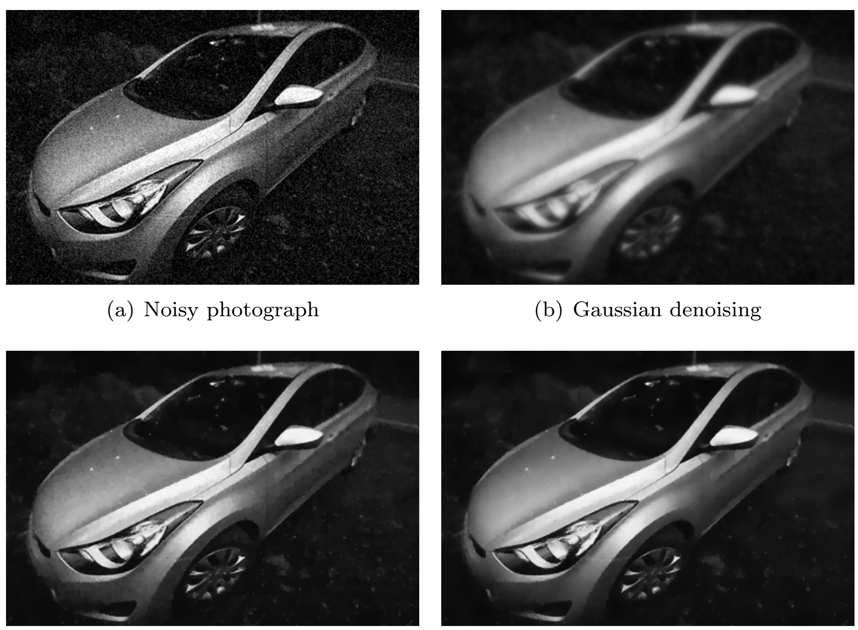


Figure 3: Reconstruction of a slice of a brain scan (left image) acquired with a magnetic resonance tomograph[2] from just 12% noisy samples of the Fourier transform data (right image).



(a) Noisy photograph

(b) Gaussian denoising

(c) Total variation denoising (1)

(d) Total generalised variation denoising

Figure 4: Different methods for image denoising.[3]

the denoised image u as a solution of

$$\min_u \left\{ \underbrace{\alpha|Du|(\Omega)}_{\substack{\text{Regularising term}\\\text{modelled by a multiple}\\\text{of the total variation}}} + \underbrace{\frac{1}{2}\|u-g\|_2^2}_{\substack{\text{Data fidelity term}\\\text{modelled by the squared}\\\text{distance between } u \text{ and } g}} \right\}. \tag{1}$$

The total variation $|Du|(\Omega)$ of the image function u on Ω is really the total variation (in the measure-theoretic sense) of the Radon measure given by the distributional derivative Du [1]. In Figure 4c an example for total variation denoising is shown. In comparison with the Gaussian filtered image in Figure 4b image structures, such as edges are much better preserved. Yet, total variation denoising is far from being the 'perfect' denoising method: it introduced the so-called staircasing effect in the parts of the image which undergo a linear change of grey values (such as on the bonnet of the car). In the last couple of years alternatives and extensions of total variation denoising have been proposed, which aim to improve upon the staircasing artefacts by introducing higher-order derivatives into the denoising model, compare e.g. [6, 16] and references therein. A very successful approach along these lines is total generalised variation denoising proposed in [4], compare Figure 4d. Other denoising approaches involve representing the image within a multi scale basis or a frame such as wavelets, shearlets and alike [9, 11].

Segmentation Going beyond image enhancement towards image analysis, a common question is what are the main objects in an image (or indeed video, which corresponds to a sequence of images) that encode the essential information given. In brain imaging for instance, a good indicator for certain deseases is the ratio of white to grey matter in the brain, which makes it necessary to separate these parts in the brain from the rest. Researchers who are interested in monitoring mitosis of cells in a sequence of microscopy frames aim for tracking separate cells and their division.

What is common to all these problems is the goal to segment an image into its different objects. The simplest situation is binary, a segmentation into object and background. Image segmentation aims to segment one or more objects of interest in an image, also under the presence of noise and blur. A large community of researchers including mathematicians, engineers and computer scientists have investigated image segmentation, proposing and analysing models for this task. The methodologies used have a wide range, from machine learning to geometric measure theory. As for image denoising variational segmentation models form a significant part of research in this area. Mumford and Shah [14] introduced in 1989 a segmentation model that is based on the idea of decomposing an image into piecewise smooth parts that are separated by an edge set Γ – the boundary of the object to be segmented. As before, let $\Omega \subset \mathbb{R}^2$ be a rectangular domain and g a given (possibly noisy) image. Further, define an edge set Γ to be a relatively closed subset of Ω, with finite one-dimensional Hausdorff measure.

We search for a pair (u, Γ) minimising

$$\frac{\alpha}{2} \int_{\Omega \setminus \Gamma} |\nabla u|^2 \, dx + \beta \mathcal{H}^1(\Gamma) + \frac{1}{2} \int_{\Omega} (u - f)^2 \, dx. \tag{2}$$

Here α and β are nonnegative constants and $\mathcal{H}^1(\Gamma)$ is the one-dimensional Hausdorff measure of Γ (which is the length of Γ if Γ is regular).

Since this pioneering proposal of David Mumford and Jayant Shah, many more variational models for image segmentation have been proposed, all based on the detection of boundaries between different objects as either locations of prominent edges in the image (edge-based segmentation like (2)) or borders of regions with a specific colour value (region-based segmentation like in [7]). In Figure 5 image segmentation is done by employing the total variation as an indicator for this detection. In general, analytic and numerical treatment of these segmentation approaches is difficult due to the measure-theoretic character of solutions as well as the usual non-convexity of the functional $\mathcal{J}$. This gives analysts and optimisers many beautiful mathematical problems to work on whose solution will have a huge impact in a wide range of image segmentation applications.

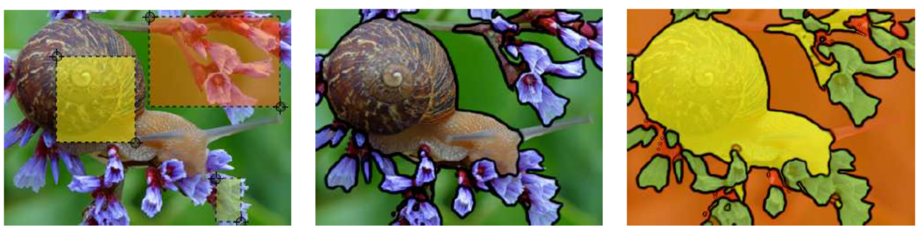


Figure 5: Image segmentation with total variation labelling [12]

Recapping the general message of mathematical image processing we see that a key ingredient in processing and analysing images is to reduce their information to just few essential features. In total variation denoising these features are the image edges, in Mumford–Shah segmentation these are the edges together with the smooth image parts in between. This demonstrates an important principle true across data analysis, namely the reduction of data size of acquired measurements.

3 Sparsity

Compressed sensing, i.e., the nonadaptive compressed acquisition of data, comprises techniques which allow one to sense/acquire only the essential features of signals and to recover them by efficient algorithms. The two main practical approaches to sparse recovery are ℓ_1-minimization and greedy methods. In the last five to ten years there has been a significant progress in this field and its impact has grown on daily basis through many new applications. Given a discrete signal $s \in \mathbb{R}^N$, one searches for a so-called *sparse* vector representation of this signal with respect to a prescribed basis (or union of bases), in the sense that it should only contain as few as possible nonzero entries, i.e., the number of nonzero entries in this sparse representation should be $K \ll N$. In the ideal case, these nonzero entries perfectly constitute the features of this signal. In mathematical terms, given a prescribed basis or dictionary (i.e., a union of multiple bases) $\Phi \in \mathbb{R}^N \times \mathbb{R}^M$ with $M \geq N$ for the signal s, one seeks the sparsest coefficients $\alpha \in \mathbb{R}^M$ such that $s = \Phi\alpha$. Mathematically, the sparsest α is given by the solution of the following optimization problem:

$$\min_{\alpha} \|\alpha\|_0 , \qquad \text{s.t. } s = \Phi\alpha, \tag{3}$$

where $\|\alpha\|_0$ denotes the so-called "ℓ_0 norm" which simply counts the nonzero entries in α. Although, due to its combinatorial nature, in general this recovery is NP-hard (i.e., it is not expected to have polynomial complexity) [13], work in compressed sensing has shown that for certain measurement matrices Φ exact recovery is possible in polynomial time if the signal s is sparse enough. Random partial Fourier matrices are an example of such an admissible measurement matrix [5].

On the one hand, greedy solvers have been designed for its solution. These methods find the support of the signal s iteratively. *Orthogonal matching pursuit* (OMP) is such an algorithm, first proposed by Mallat and Zhang [13]. OMP was followed by a series of improved algorithms, such as *Regularized OMP* (ROMP), and *Compressive Sampling MP* (CoSaMP) providing stronger guarantees of recovery and better error estimates.

Facing the same problem of sparse recovery Chen, Donoho, and Saunders [8] proposed a convexified form of the minimization problem (3), called *basis pursuit*:

$$\min_{\alpha} \|\alpha\|_1 , \qquad \text{s.t. } s = \Phi\alpha, \tag{4}$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm, which, for a discrete coefficient $\alpha \in \mathbb{R}^M$, reads $\|\alpha\|_1 = \sum_{k=0}^{M-1} |\alpha_k|$.

By replacing the ℓ_0 norm by the ℓ_1 norm, the quest for the sparsest solution now amounts to solving a convex optimization problem, which is numerically affordable with polynomial complexity. In subsequent work Donoho [10] could show that for certain random measurements, basis pursuit (4) actually provides a unique solution, which equals the solution of (3). Candès, Romberg, and Tao also proposed ℓ_1 minimization for sparse signal recovery specifically in the case of random partial Fourier matrices [5].

Greedy algorithms on the one hand and ℓ_1-minimization on the other hand show certain advantages and challenges with respect to each other. First of all, the ℓ_1 minimization method provides more uniform guarantees for sparse recovery than OMP. In particular, having a suitable measurement matrix given, the former guarantees recovery for all sparse signals whereas OMP can only guarantee this for a fixed signal s. This problem with OMP was alleviated by its successors ROMP and CoSaMP. Still, in the presence of noise in the signal, ℓ_1 minimization performs more robustly than OMP and related methods with respect to the expected approximation error. On the other hand OMP is in general computationally faster (for which there are mathematical proofs and also clear empirical evidence). In fact, a very active area of research is the construction of accelerated algorithms for ℓ_1 minimization, which already promise to be competitive with OMP, see [15].

Finally, note the connection to image processing algorithms such as total variation denoising (1) discussed earlier. In fact, for a discrete $M \times N$ image u, the total variation $|Du|(\Omega) = \|\nabla u\|_1$ is the ℓ_1 norm of the gradient of the image u. This means that computing the denoised image as the solution of (1) is equivalent to seeking an image u whose gradient is sparse, that is the image consists only of a small number of edges with constant colour values in between. This closes the circle to the image processing methods discussed before.

4 Conclusion

In modern, computerised and technology driven world an efficient acquisition, processing and interpretation of data is a central challenge of our times. In particular, the large-scale of acquired data (resulting from the improvement of hardware as well as in the increase in spatial density of acquisition devices – think of all the CCTV cameras watching over us!) is problematic for classical processing and analysis methodologies. While the first three hundred years of modern science engaged mainly with the natural world, current challenges are mostly concerned with the fruits of our technology.

This challenge has been addressed in the last few decades by mathematicians, computer scientists and information engineers, although it is fair to observe that it is present across all of scholarship, from natural sciences and engineering to medicine, social sciences and humanities. All generate data, all strive to read information hidden within the data.

Recent breakthroughs in mathematical methodology, associated with the work of Emmanuel Candès, Ingrid Daubechies, David Donoho, Stanley Osher and Terence Tao, are but a first step on a long journey. By this stage it is premature not just to sketch the roadmap for this journey but even to choose the vehicles. Looking at the history of mathematics – and, indeed, of scholarship – it is clear that the future belongs to theories and methodologies which are yet undiscovered. Yet, we need to start somewhere and visualise (ever so tentatively) the first few steps along this journey:

- Detecting the correct framework for 'sparsity' in different applications. What are the correct features (in terms of harmonic analysis, what is the correct basis) in which the solution is sparse?

- Understanding sparsity in terms of mathematical analysis and statistics;

- Advancing processing and analysis methodologies that are tuned to the application at hand – in other words, combining mathematical insight with knowledge gained in application areas;

- Developing efficient computational algorithms that can address the very large and demanding numerical problems emerging from data research;
- Addressing the need for real-time computation, of crucial importance e.g. in security and in medical diagnostics;
- Making data analysis more 'machine intelligent', combining artificial intelligence, machine learning, robotics and the understanding of data;
- Understanding inverse problems from both mathematical and phenomenological point of view. What does it really *mean* to 'solve' an inverse problem and how to go about it.

The world is changing and we, as mathematicians, must change with it – not just because this is important, not just because this is exciting and replete with wonderful intellectual challenges but also because, once we claim that mathematics is the language of science and technology, we must be ready to deploy it to sing new tunes.

References

[1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems.* Oxford University Press, Oxford, 2000.

[2] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147. Springer Science + Business Media, 2006.

[3] R.E. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, NJ, 1957.

[4] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM J. Imaging Sc.*, 3:492–526, 2011.

[5] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Information Theory*, 52(2):489–509, 2006.

[6] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.

[7] T.F. Chan and L.A. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2):266–277, 2001.

[8] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sc. Comp.*, 20(1):33–61, 1998.

[9] I. Daubechies. *Ten Lectures on Wavelets*, volume 61. SIAM, 1992.

[10] D.L. Donoho. Compressed sensing. *IEEE Trans. Information Theory*, 52(4):1289–1306, 2006.

[11] G. Kutyniok and D. Labate. *Shearlets: Multiscale Analysis for Multivariate Data*. Springer, 2012.

[12] J. Lellmann, D. Breitenreicher, and C. Schnörr. Fast and exact primal-dual iterations for variational problems in computer vision. In *Computer Vision–ECCV 2010*, pages 494–505. Springer, 2010.

[13] S.G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Processing*, 41(12):3397–3415, 1993.

[14] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[15] Y. Nesterov and A. Nemirovski. On first-order algorithms for ℓ_1/nuclear norm minimization. *Acta Numerica*, 22:509–575, 2013.

[16] K. Papafitsoros and C.-B. Schönlieb. A combined first and second order variational approach for image reconstruction. *J. Mathematical Imaging & Vision*, 2013. Published online.

[17] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[18] N. Spencer. How much data is created every minute? 2012. http://www.visualnews.com/2012/06/19/how-much-data-created-every-minute/f.

[19] E. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Comm. Pure & Applied Maths*, 13:1–14, 1960.