Future Challenges for Variational Analysis

ANZMC2008

Christchurch Dec. 10, 2008

Nonlinear Optimization

ABSTRACT Modern non-smooth analysis is now roughly thirty-five years old. I shall briefly assess where the subject stands today both as theory and regarding applications.

I will also discuss open problems and current challenges for the subject.

Jonathan Borwein, FRSC

www.cs.dal.ca/~jborwein

Canada Research Chair

and

Laureate Professor

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Aust MS

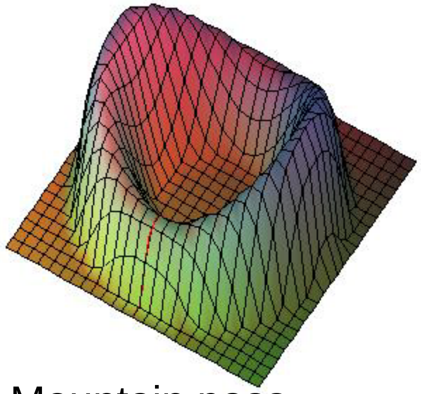
Australian Mathematical Society Inc.

Revised 05/12/2008

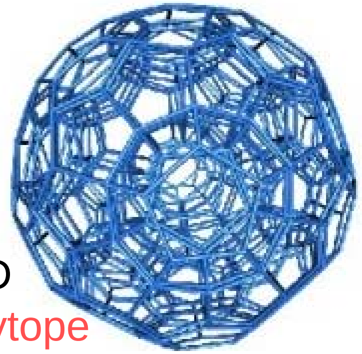
OUTLINE

- **First Order Theory**
- **High Order Theory**
  - second (and higher)
- **Applications**
  - inside Mathematics
  - outside Mathematics
- I'll mention **successes** and **failures**

- Each item has open questions even in the convex case (CA)
  - some technical and specialized
  - some broader and general
- To work fruitfully in VA it helps to understand CA and smooth analysis (SA)
  - they are the motivating foundations
  - and often provide the key technical tools



Mountain pass



A facet of Coxeter's favourite 4D convex polytope

LIPSCHITZ PRECURSORS

Pshenichnyi (1968)

- **descriptive**
- **Large class of good "**quasi-differentiable**" functions**

$$f'(x;h) := \limsup_{t \to 0^+} \frac{f(x+th) - f(x)}{t}$$

- required to be convex in h

Clarke (1972)

- **prescriptive**
- **New directional derivative**

$$f^\circ(x;h) := \limsup_{t \to 0^+, y \to x} \frac{f(y+th) - f(y)}{t}$$

- built to be convex in h

Both capture smooth and convex functions and are closed under + and $\vee$

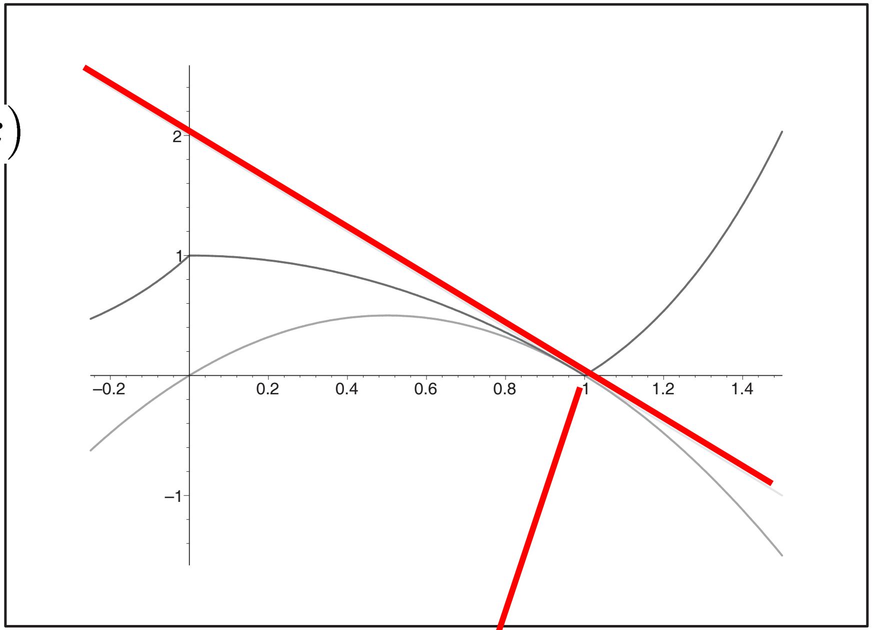
FIRST ORDER THEORY

- **Subgradient**
  - one-sided Fréchet $\partial_F f(x)$
    (Gâteaux or Hadamard)
- Viscosity subgradient $\partial_F^v f(x)$
  - derivative of smooth (local) minorants
  - in nice space $\partial_F f(x) = \partial_F^v f(x)$
- (Fuzzy) Sum Rule

$$\partial_F(f+g)(x) \subseteq \partial_F f(x_1) + \partial_F g(x_2) + \varepsilon B_{X^*}$$

- Normal cones $N_{\mathrm{epi}\,f} := \partial \iota_{\mathrm{epi}\,f}$
- Variational principles (VP)

A lsc function f,
a viscosity subgradient in red,
and a smooth minorant g



$$f(y) - g(y) \geq f(x) - g(x) \text{ for } y \text{ near } x$$

STATE OF THE THEORY

ACHIEVEMENTS

- Limiting subgradients

$$\partial f(x) := \lim_{y \to_f x} \partial_F f(x)$$

- Coderivatives of multis

$$D^*\Omega(x,y)(y^*) = \{x^*: (x^*, -y^*) \in N_{\mathrm{gph}(\Omega)}(x,y)\}$$

- VP+Viscosity+Sum Rule
  - yield **fine** 1st order theory for real-valued functions
  - esp. in Lipschitz case or in finite dimensions
  - seq normal compactness needed more generally; also for
- Metric regularity: locally

$$Kd(\Omega(x), y) \geq d(x, \Omega^{-1}(y))$$

  - and its extensions
  - implicit functions (Dontchev-Rock)
  - alternating projections (Bauschke-Combettes)

I now ignore epsilons and the like

LIMITATIONS

- Inapplicable outside of Asplund space (reflexive, …)
  - $\partial_H f$ extensions fiddly, limited
- Theoretically very beautiful
  - hard to compute even for 'nice' functions

Generically non-expansive functions have $\partial f(x) \equiv B_{X^*}$

  - **SNC** restriction is fundamental not technical

- better results rely on restricting classes of functions (and spaces)
  - E.g., prox-normal, lower C_2,
  - essentially smooth (B-Moors)

B-Sciffer (08) separable explicit construction for $\partial f(x) \equiv B_{X^*}$

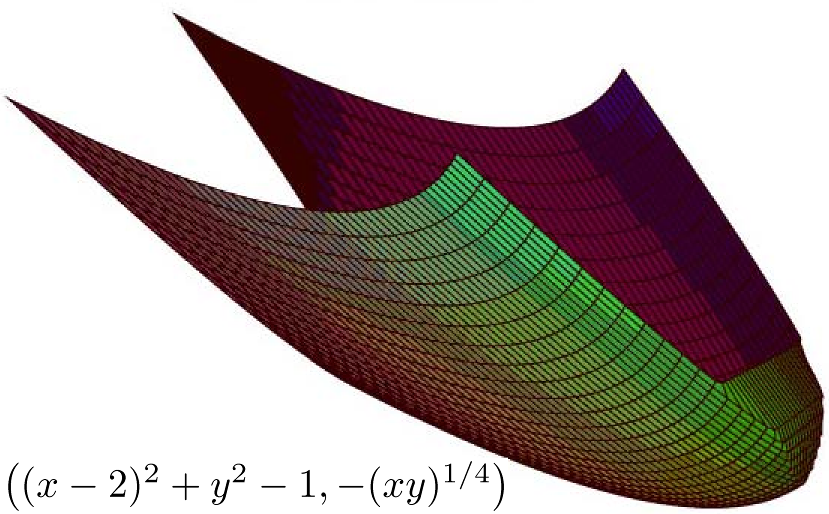
OPEN QUESTIONS

CONVEX

- Find a reflexive Legendre function generalization for convex functions without points of continuity such (neg) Shannon entropy

$$\mathcal{B}(x) := \int_0^1 x(t)\ \log x(t)\,\mathrm{d}t$$

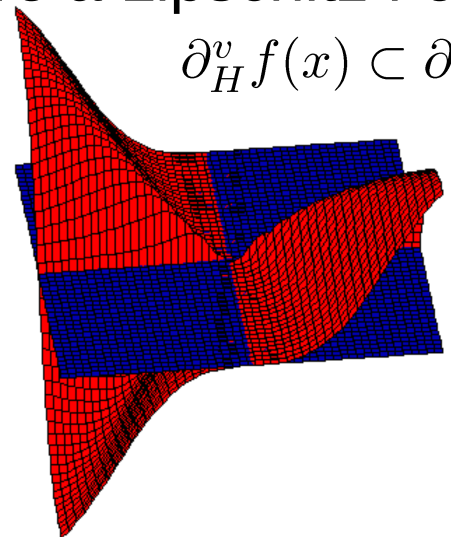
AN ESSENTIALLY STRICTLY CONVEX FUNCTION WITH NONCONVEX SUBGRADIENT DOMAIN AND WHICH IS NOT STRICTLY CONVEX



$\max\left((x-2)^2+y^2-1, -(xy)^{1/4}\right)$

NONCONVEX

Is there a Lipschitz f on ℓ^2 with $\partial_H^v f(x) \subset \partial_H f(x)$?



$\frac{xy^3}{x^2+y^4}$ has $0 \in \partial_H f(0)$ but $0 \notin \partial_H^v f(0)$

- Is the product of a separable and a weak Asplund space still weak Asplund?
  - true for GDS, but WASP is the largest class on which refined NA calculus could be expected to work

SECOND ORDER THEORY

Recall for **convex functions** the *difference quotient* of f by

$$\Delta_t f(x) : h \mapsto \frac{f(x+th) - f(x)}{t};$$

and the *second-order difference quotient* of f by

$$\Delta_t^2 f(x) : h \mapsto \frac{f(x+th) - f(x) - t\langle \nabla f(x), h\rangle}{\frac{1}{2}t^2}.$$

Analogously let

$$\Delta_t[\partial f](x) : h \mapsto \frac{\partial f(x+th) - \nabla f(x)}{t}.$$

For any $t > 0$,

$$\Delta_t f(x)$$

is closed, proper, convex and nonnegative.

Quite beautifully

$$\partial \left[\frac{1}{2}\Delta_t^2 f(x)\right] = \Delta_t[\partial f](x).$$

Theorem Alexandrov (1939) *In Euclidean space a real-valued continuous convex function admits a second order Taylor expansion at almost all points (w.r.t. Lebesgue measure).*

My favourite proof is a specialization of Mignot's 1976 extension to monotone operators [RW98, BV09]

The result relies on many happy coincidences in Euclidean space

The convex case is quite subtle and so the paucity of definitive non-convex results is no surprise

STATE OF THE THEORY

ACHIEVEMENTS

- Some lovely patterns and fine theorems in Euclidean space
  - but no definitive corpus of results
  - nor even canonical definitions outside of the convex case
  - interesting work by Jeyakumar 2006, Duta and others.
  - Many have noted

$$\partial^2 f(x) := \partial \nabla_G f(x)$$

is a fine object when function is Lipschitz smooth in separable Banach space (Rademacher's Thm. applies)

- Fundamental results by Ioffe-Penot (TAMS 1997) on limiting 2-subjets (and coderivatives)
- Fine refined calculus of 'efficient subset' of sub-hessians (Eberhard-Wenczel, SVA 2007)

LIMITATIONS

- Little 'deep' work in infinite dim.
  - i.e., if obvious extensions fail
  - even in Hilbert space
- Outside separable Hilbert space general results are not to be expected [BV09]
  - research should focus on structured classes: integral functionals (Moussaoui-Seeger TAMS 1999)
  - composite convex functions

2-subjet: all order two expansions of C^2 minorants agreeing at x. Now take limits

$$\overline{\partial}^2 f(x) := \limsup_{y \to_f x} \underline{\partial}^2 f(y)$$

yields sum rule in para-convex case,etc.

- extensions by Eberhard-Ralph, others

OPEN QUESTIONS

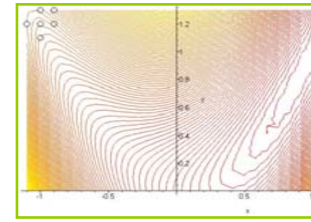
CONVEX

- Does every continuous convex function on separable Hilbert space admit a second order Gateaux expansion at at least one point (or on a dense set of points)?
  - fails in non-separable Hilbert space or in $\ell_p(\mathbf{N}), p \neq 2$
  - fails in the Fréchet sense even in $\ell_2(\mathbf{N})$

NONCONVEX

- Are there sizeable classes of functions for which subjets or other useful second order expansions can be built in separable Hilbert space?
  - I have no precise idea what "useful" means
  - even in convex case this is a tough request; then use
  - Lasry-Lions regularization (Penot, Eberhard, …)

APPLICATIONS



Pattern search

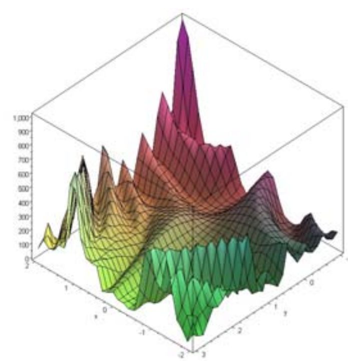
SUCCESSES

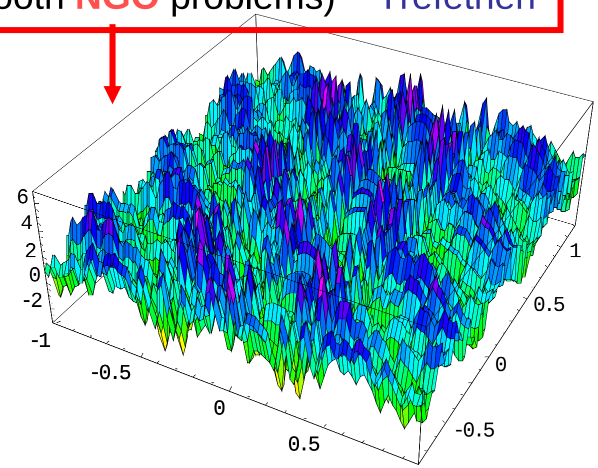
- Tools now part of pure non-linear and functional analysis
- Convergence theory for “pattern search” derivative-free optimization algorithms [Ma08]
- Eigenvalue and singular value optimization theory [BL05]
  - 2nd order (Lewis and Sendov)
- Differential Inclusions and Optimal Control
  - approx Maximum Principle
  - Hamilton-Jacobi equations
- Non-convex mathematical economics and MPECS
- Exact penalty and universal barrier methods [BV09]
  - oodles more counting convex analysis [BV04]

FAILURES

- Limited numeric successes
  - even in convex case excluding spectral & SDP code (somewhat)
  - bundle methods (Lemaréchal et al)
  - Vanderbei’s LOCO package
  - composite convex & smoothing
  - need for “structured nonsmooth” optimization (a la Boyd)

Pinter (Smooth **NGO** problems) Trefethen





$$\min \exp(\sin(50x)) + \sin(60e^y) + \sin(70\sin x) + \sin(\sin(80y)) - \sin(10(x+y)) + (x^2+y^2)/4?$$

$$(x^*, y^*) \approx (-0.024627, 0.211789)$$
with minimum ≈ -3.30687

THREE MORE OPEN QUESTIONS

1. The Chebyshev problem (Klee 1961)

   If every point in H has a unique nearest point in C is C convex? (Yes: Motzkin-Bunt in Euclidean space [BL05])

2. Existence of nearest points (proximal boundary?)

   Do some (many) points in H have a nearest point in C in every renorm of H ? (Yes: if the set is bounded or the norm is Kadec-Klee [BZ05])

3. Universal barrier functions in infinite dimensions

   Is there an analogue for H of the universal barrier function so important in Euclidean space? (I doubt it [BV09])

Convex Functions

The Chebyshev problem (Klee 1961) A set is Chebyshev if every point in H has a unique nearest point in C

Theorem If C is weakly closed and Chebyshev then C is convex. So in Euclidean space Chebyshev iff convex.

Four Euclidean variational proofs ([BL05], Opt Letters 07, [BV09])

1. **Brouwer's theorem** (Cheb. implies **sun** implies convex)
2. **Ekeland's theorem** (Cheb. implies **approx. convex** implies convex)
3. **Fenchel duality** (Cheb. iff d_C^2 is Frechet) use f^* smooth implies f convex for

$$\left(\frac{\iota_C + \|\cdot\|^2}{2}\right)^* = \frac{\|\cdot\|^2 + d_C^2}{2}$$

4. Inverse geometry also shows if there is a counter-example it can be a **Klee cavern** (Asplund) the closure of the complement of a convex body. **WEIRD**

- **Counterexamples** exist in incomplete inner product spaces. #2 seems most likely to work in Hilbert space.
- **Euclidean case** is due to Motzkin-Bunt

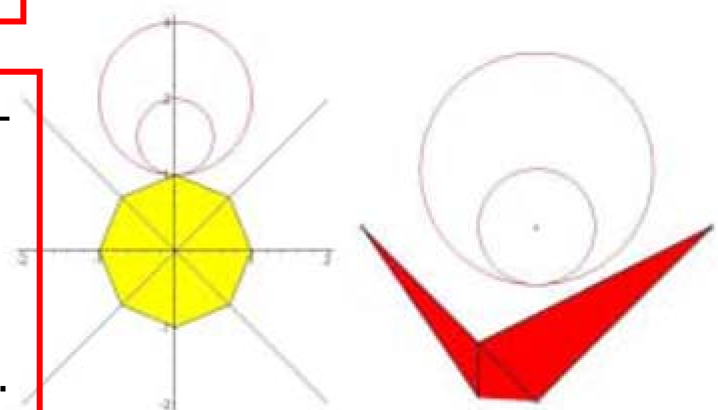


FIGURE 1. Suns and approximate convexity.

Existence of nearest points

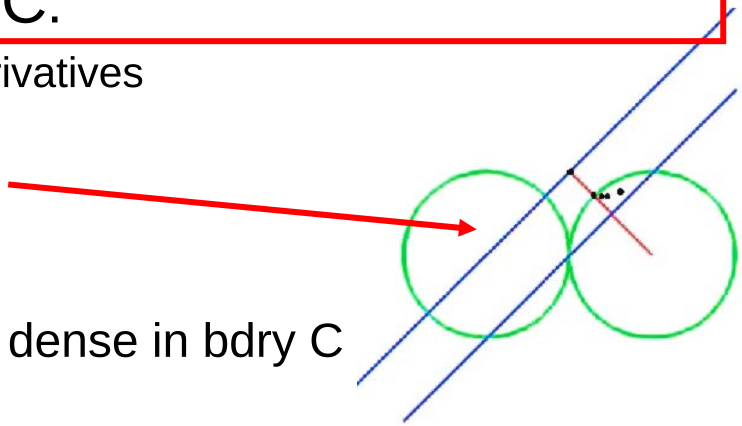
Do some (many) points in H have a nearest point in C in every renorm of H ?

Theorem (Lau-Konjagin, 76-86) A norm on a reflexive space is Kadec-Klee iff for every norm-closed C in X best approximations exist generically (densely) in X\ C.

Nicest proof is via dense existence of Frechet subderivatives

$$\phi \in \partial_F d_C(x)$$

The KK property forces approximate minimizers to line up.



- There are non KK norms with proximal points dense in bdry C
- If C is closed and bounded then there are some points with nearest points (RNP)
- So a counterexample has to be a weird unbounded set in a rotten renorm (BFitzpatrick 89, [BZ09])

A norm is **Kadec-Klee** norm if weak and norm topologies agree on the unit sphere.

Hence all LUR norms are Kadec-Klee.

Universal barrier functions in infinite dimensions

– Is there an analogue for H of the universal barrier function that is so important in Euclidean space?

Theorem (Nesterov-Nemirovskii) For any open convex set A in n-space, the function

$$F(x) := \lambda_N \left((A - x)^o \right)$$

is an essentially smooth, log-convex barrier function for A.

– This relies heavily on the existence of **Haar** measure (Lebesgue).
– Amazingly for A the semidefinite matrix cone we recover – log det, etc

In Hilbert space the only really nice examples I know are similar to:

$$\phi(T) := \mathrm{trace}(T) - \log\det(I + T)$$

is a strictly convex Frechet differentiable barrier function for the Hilbert-Schmidt operators with I+T > 0.

We are able to build barriers in great generality but not "universally" [BV09]

REFERENCES

[BL05] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization. Theory and Examples*, CMS-Springer Books, Second edition, 2005.

[BZ05] J.M. Borwein and Qiji Zhu, Techniques of Variational Analysis, CMS-Springer Books, 2005.

[BV09] J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples,* Cambridge University Press, 2009.

[BV04] S. Boyd. and L. Vanden-berghe, *Convex Optimization*, Cambridge University Press, 2004.

[Ma08] Mason, Macklem, *Parallel Continuous Optimization and Distributed Collaboration,* PhD Thesis, Dalhousie, 2008.

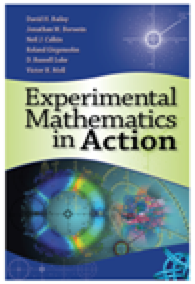
[M06a] B.S. Mordukhovich, Variational Analysis & Generalized Differentiation I. Basic Theory, Springer-Verlag Grundlehren, **330**, 2006.

[M06b] B.S. Mordukhovich, Variational Analysis & Generalized Differentiation II. Applications, Springer-Verlag Grundlehren, **331**, 2006.

[RW98] R.T. Rockafellar and RJ-B. Wets, Variational Analysis, Springer-Verlag Grundlehren, **317**, 2006.

Experiencing Experimental Mathematics

Experimental Mathematics in Action
David H. Bailey, Jonathan M. Borwein, Neil J. Calkin, Roland Girgensohn, D. Russell Luke, Victor H. Moll



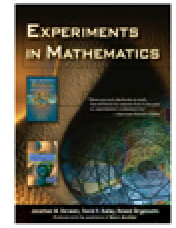
"David H. Bailey et al. have done a fantastic job to provide very comprehensive and fruitful examples and demonstrations on how experimental mathematics acts in a very broad area of both pure and applied mathematical research, in both academic and industry. Anyone who is interested in experimental mathematics should, without any doubt, read this book!"

—*Gazette* of the Australian Mathematical Society

978-1-56881-271-7; Hardcover; $49.00

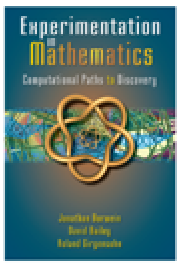
Experiments in Mathematics (CD)
Jonathan M. Borwein, David H. Bailey, Roland Girgensohn

In the short time since the first edition of *Mathematics by Experiment: Plausible Reasoning in the 21st Century* and *Experimentation in Mathematics: Computational Paths to Discovery*, there has been a noticeable upsurge in interest in using computers to do real mathematics. The authors have updated and enhanced the book files and are now making them available in PDF format on a CD-ROM. This CD provides several "smart" features, including hyperlinks for all numbered equations, all Internet URLs, bibliographic references, and an augmented search facility assists one with locating a particular mathematical formula or expression.



978-1-56881-283-0; CD; $49.00

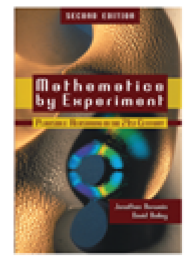
Experimentation in Mathematics
Computational Paths to Discovery
Jonathan M. Borwein, David H. Bailey, Roland Girgensohn



Mathematics by Experiment — Second Edition
Plausible Reasoning in the 21st Century
Jonathan M. Borwein, David H. Bailey

"These are such fun books to read! Actually, calling them books does not do them justice. They have the liveliness and feel of great Web sites, with their bite-size fascinating factoids and their many human- and math-interest stories and other gems. But do not be fooled by the lighthearted, immensely entertaining style. You are going to learn more math (experimental or otherwise) than you ever did from any two single volumes. Not only that, you will learn by osmosis how to become an experimental mathematician."

—*American Scientist Online*



978-1-56881-136-9; Hardcover; $59.00

978-1-56881-442-1; Hardcover; $69.00

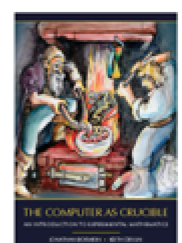
Communicating Mathematics in the Digital Era
Edited by J. M. Borwein, E. M. Rocha, J. F. Rodrigues

Digital technology has dramatically changed the ways in which scientific work is published, disseminated, archived, and accessed. This book is a collection of thought-provoking essays and reports on a number of projects discussing the paradigms and offering mechanisms for producing, searching, and exploiting scientific and technical scholarship in mathematics in the digital era.

978-1-56881-410-0; Hardcover; $49.00

The Computer as Crucible
An Introduction to Experimental Mathematics
Jonathan Borwein, Keith Devlin

Keith Devlin and Jonathan Borwein cover a variety of topics and examples to give the reader a good sense of the current state of play in the rapidly growing new field of experimental mathematics. The writing is clear and the explanations are enhanced by relevant historical facts and stories of mathematicians and their encounters with the field over time.



978-1-56881-343-1; Paperback; $29.95