[25] R. Johnson and T. Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 315–323. 2013.

[26] Y. Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.

[27] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[28] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Neurocomputing: Foundations of research. chapter Learning Representations by Back-Propagating Errors, pages 696–699. MIT Press, Cambridge, MA, 1988.

[29] B. A. Pearlmutter. Fast exact multiplication by the Hessian. *Neural Computation*, 6(1):147–160, 1994.

[30] W. Zhou, J. D. Griffin, and I. G. Akrotirianakis. A globally convergent modified conjugate-gradient line-search algorithm with inertia controlling. Technical Report 2009-01, 2009.

[31] J. Martens. Deep learning via Hessian-free optimization. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 735–742, 2010.

[32] Y. N. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, NIPS'14, pages 2933–2941, Cambridge, MA, 2014. MIT Press.

[33] R. Pascanu, Y. N. Dauphin, S. Ganguli, and Y. Bengio. On the saddle point problem for non-convex optimization. *CoRR*, abs/1405.4604, 2014.

[34] S. Ghadimi and G. Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

[35] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust Region Methods*. MPS/SIAM Series on Optimization. SIAM, Philadelphia, 2000.

[36] C. Cartis, N. I. M. Gould, and P. L. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Math. Program.*, 127(2):245–295, 2011.

[37] C. Cartis, N. I. M. Gould, and P. L. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity. *Math. Program.*, 130(2):295–319, 2011.

[38] C. Cartis, N. I. M. Gould, and P. L. Toint. How much patience do you have? A worst-case perspective on smooth nonconvex optimization. *Optima*, 88, 2012.

[39] Y. Nesterov. *Introductory Lectures on Convex Optimization*. Kluwer Academic Publishers, Boston, MA, 2004.

[40] S. J. Reddi, A. Hefny, S. Sra, B. Poczos, and A. Smola. Stochastic variance reduction for nonconvex optimization. Technical Report arXiv:1603.06160, 2016.

[41] C. Cartis and K. Scheinberg. Global convergence rate analysis of unconstrained optimization methods based on probabilistic models. Technical Report, ISE, Lehigh, 2015.

[42] C. Chen, M. Menickelly, and K. Scheinberg. Stochastic optimization using a trust-region method and random models. Technical Report arXiv:1504.04231, 2015.

[43] J. Larson and S.C. Billups. Stochastic derivative-free optimization using a trust region framework. *Computational Optimization and Applications*, 64(3):619–645, 2016.

[44] S. Shashaani, F. S. Hashemi, and R. Pasupathy. ASTRO-DF: a class of adaptive sampling trust-region algorithms for derivative-free simulation optimization. Technical report, Purdue University, 2015.

[45] J. Blanchet, C. Cartis, M. Menickelly, and K. Scheinberg. Convergence rate analysis of a stochastic trust region method for nonconvex optimization. Technical Report arXiv:1609.07428, 2016.

In Memoriam

Jon Borwein: a personal reflection

Borwein at a 2006 summer school in Paseky.

Jon Borwein died on August 2, 2016. His untimely passing has deprived us all of a singular and brilliant mind and an inspirational intellectual leader, and I have lost a close personal friend. Rather than a formal memorial, my words are a personal reflection on my coauthor (of fifteen papers and a book [46]), a mentor to whom I owe my career.

Jon's mathematical breadth and energy make a fascinating but bewildering picture, extending far beyond traditional optimization, and challenging to sketch. He delighted in collaboration, and many of us knew first-hand his research style: whirling, exuberant, defamiliarizing, endlessly curious, elegant, scholarly, generous, and honest. He made time for everyone, no matter their rank or eccentricity. Shortly after I met Jon, at the height of the public prominence of his work around pi with his brother Peter, highlighted in their book [47] and a Scientific American article [48], he remarked to me how he kept in mind the eminent English mathematician G.H. Hardy, the sole reader of Ramanujan's first terse but prophetic notes.

Early in 1987 Jon had welcomed me to the delightful city of Halifax, Nova Scotia—then his home. During a two-year postdoctoral fellowship, I shadowed him closely on his travels. Astonishingly, among his many projects then was a Dictionary of Mathematics [49], and indeed I felt a kind of prosaic Boswell to his dizzying Samuel Johnson. In the decade that followed, we made our independent ways across Canada, through the University of Waterloo to Simon Fraser University. There, Jon founded the Center for Experimental and Computational Mathematics, a pioneering base for his international pre-eminence in experimental mathematics.

Jon laid down many roots. Wikipedia describes him as a "Scottish mathematician," born in St Andrews in 1951.

With his encyclopedic erudition, Jon probably knew Johnson's poke that "Much may be made of a Scotchman, if he be caught young"; if he did know, he took it to heart, receiving his doctorate as a Rhodes Scholar at Oxford. He went on to spend the core of his career in Canada, where he served as President of the Canadian Mathematical Society and was elected a Fellow of the Royal Society of Canada. Along the way, he was elected a Foreign Member of the Bulgarian Academy of Science, and Fellows of both the American Association for the Advancement of Science and the American Mathematical Society. He made his final home in 2009 at the University of Newcastle in Australia as Laureate Professor and was elected a Fellow of the Australian Academy of Science.

Jon's diverse honors make his generous and articulate collaborative style all the more striking. He worked and collaborated intensely but did not suffer fools gladly. I imagine he sympathized with another of Johnson's quips: "Sir, I have found you an argument; but I am not obliged to find you an understanding." He was nonetheless a painstaking and articulate stylist and in 1993 won (with his brother Peter and his long-time collaborator David Bailey) the mathematical world's highest honor for exposition, the Chauvenet Prize. (Sixty years earlier, the winner was G.H. Hardy.)

Jon and his family—his wife, Judi, and two young daughters, Rachel and Naomi (their sister Tova being yet to arrive)—more or less adopted me as a family member when I arrived in Canada. I essentially lived with them during month-long visits to Limoges, France (where Jon later received an honorary doctorate), to the Technion in Israel, and to Canberra and Newcastle, Australia. The sheer fun of that last visit probably inspired the Borweins' later choice of adopted home.

Life at the Borweins' home was an inspiring and exhausting blur. A typical evening involved prodigious and virtuoso culinary feats from Judi, feisty debates from Rachel and Naomi, and multiple simultaneous media playing at full volume. At a minimum, these included political news (Jon was intensely active, politically, serving for a while as treasurer of the Nova Scotia New Democratic Party), major league baseball (another domain of erudition), and music. All gradually dissolved into large glasses of Scotch (Jon's Scotchness, like Healey Willan's, was mostly "by absorbtion"), and then a call to arms from Jon to prove some reluctant theorem. The exuberant and dizzying environment mirrored Jon's mathematics, a style so appealing it quickly sealed my own career choice as a mathematician.

Jon left us too soon. I seek some small solace in that during any of his years, Jon's ideas were at least twice as good, twice as fast, twice as many, and twice as well shared as, say, mine. But for his beloved and devoted family, his death has been simply shocking and untimely.

Optimization theory was just one of Jon's arenas; but as the one I know best, I would like to pick out a few personal favorites, most from that same era. To Jon's extraordinary academic family of collaborators, many of whom I race by unmentioned, my apologies.

A theme running through much of Jon's work was his emphatic belief in optimization and analysis as a single discipline, often unified through the language of set-valued mappings. He recognized early, for example, the importance of characterizing "metric regularity" for constraint systems [50]–now commonly known as "error bounds," stably bounding the distance to the feasible region by a multiple of the constraint error. Such bounds are of widespread interest, in particular, in the convergence analysis of first-order methods. Jon and his student Heinz Bauschke used similar ideas in a deep and long-running study of von Neumann's alternating projection algorithm and its relatives [51]. Another theme underlying much of Jon's work on the interplay of analysis and optimization was his extensive use both of proximal analysis (a technique growing out of viscosity solutions of PDEs and optimal control) [52] and of generalized derivatives in surprising contexts, especially Banach space geometry [53].

Perhaps Jon's most celebrated result in nonsmooth analysis and optimization is the Borwein-Preiss variational principle [54]. A ubiquitous technique throughout variational mathematics appeals to the existence of a minimizer of a function. Without some compactness, the argument breaks, but a famous result of Ekeland rescues it through a small perturbation to the original function. Ekeland's perturbation is, unfortunately, nonsmooth; but using a deep and surprising argument, Borwein and Preiss showed that a smooth perturbation will in fact suffice.

Much of Jon's broader mathematical opus is intertwined with computation, and he believed fundamentally in the computer as a tool for mathematical discovery. Many of his contributions to optimization were, by contrast, conceptual rather than computational. An interesting exception is the Barzilai-Borwein method [55], an odd and ingenious nonmonotone gradient-based minimization algorithm that has attracted growing attention during the big-data-driven resurgence of first-order methods.

I cannot resist a nod at my own postdoctoral work with Jon, much of which grew out of the maximum entropy methodology for estimating an unknown probability density from some of its moments. In one of my favorite results from that period, we showed that a sequence of densities, the kth of which agreeing with the unknown density up to the first k moments, need not converge weakly in the space L1, but nonetheless must do so if each has the maximum possible Shannon entropy [56, 57].

Jon somehow maintained a fresh, youthful intellectual style until the end. Sitting on my desktop, dated two weeks before he died, is his last paper [58], a lovely essay on the craft of mathematical research. He writes: "I can no longer resist making some observations... to a young mathematician... but we are nearly all young longer." Fortunately, he shared his advice in the nick of time. His final instruction is "Above all, be honest."

The book [46] that Jon and I published together in 2000 has found some popularity, even though we had intended the material just to be a quick introduction, Chapter 1 of a serious book. It exists only because I momentarily caught up: for a brief and happy time, I could scribe slightly faster than Jon's genius could create. The subsequent chapters will not be done before we meet again.

Adrian Lewis
School of Operations Research and Information Engineering, Cornell University, USA, adrian.lewis@cornell.edu

REFERENCES

[46] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer, New York (2000).

[47] J.M. Borwein and P.B. Borwein, *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*, Wiley, New York (1987).

[48] J.M. Borwein and P.B. Borwein, "Ramanujan and Pi," *Scientific American* (February 1988), 112–117.

[49] E.J. Borowski and J.M. Borwein, *Dictionary of Mathematics*, Collins, Glasgow (1989).

[50] J.M. Borwein, "Stability and regular points of inequality systems," *J. Optimization Theory and Applications* 48 (1986), 9–52.

[51] H.H. Bauschke and J.M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Review* 38 (1996), 367–426.

[52] J.M. Borwein and A. Ioffe, "Proximal analysis in smooth spaces," *Set-Valued Analysis* 4 (1996), 1–24.

[53] J.M. Borwein, S.P. Fitzpatrick, and J.R. Giles, "The differentiability of real functions on normed linear space using generalized gradients," *J. Optimization Theory and Applications* 128 (1987), 512–534.

[54] J.M. Borwein and D. Preiss, "A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions," *AMS Transactions* 303 (1987), 517–527.

[55] J. Barzilai and J.M. Borwein, "Two point step-size methods," *IMA J. Numerical Analysis* 8 (1988), 141–148.

[56] J.M. Borwein and A.S. Lewis, "Convergence of best entropy estimates," *SIAM J. Optimization* 1 (1991), 191–205.

[57] J.M. Borwein and A.S. Lewis, "On the convergence of moment problems," *AMS Transactions* 325 (1991), 249–271.

[58] J.M. Borwein, "Generalisations, examples and counter-examples in analysis and optimisation," *Set-Valued and Variational Analysis* (2016), in press. DOI:10.1007/s11228-016-0379-2