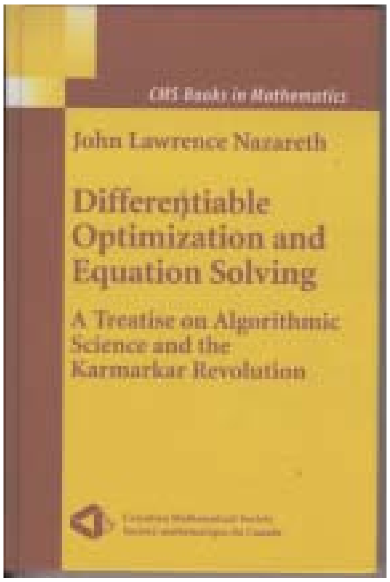
A DELIGHTFUL TREATISE FOR BOTH NOVICE AND EXPERT

Book review by Henry Wolkowicz, University of Waterloo

Differential Optimization and Equation Solving
by John Lawrence Nazareth
CMS Books in Mathematics 13
Springer 2003 xv1 + 256 pages



The new book by John Nazareth is a delightful novel treatise in Optimization and Equation Solving. In particular, this new book deals with the interior-point revolution that has changed the way optimizers look at optimization problems. (See e.g. [11] and the details below.) Nazareth has written an excellent book that includes both introductory and advanced topics. It provides a description of many of the techniques in this area. In addition, the book is sprinkled with beautiful analogies and insights. These insights make this book an interesting read and a learning experience for both the novice and the expert. (I will outline some of these enjoyable/novel insights below.) The book is based on Nazareth's extensive experience and publications in optimization and equation solving.

Consider the (differentiable) nonlinear programming model, CCP;

	minimize	$f(x)$	
	subject to	$H(x)$ # 0	(0 $\mathbb{R}^m$)
(NLP)		$G(x)$ 0	(0 $\mathbb{R}^p$)
		x 0 $\mathbb{R}^n$,	

i.e. the minimization of the function f subject to m equality constraints $h_i = 0$, $i = 1,\ldots,m$ and and p inequality constraints g_j # 0, $j = 1,\ldots,p$. This general model includes: if there are no constraints, then we get the unconstrained minimization model, UM; if there are no inequality constraints, we get the nonlinear equality-constrained programming model, NECP; if the equality constraints are linear and the inequality constraints convex, we get the convex-constrained programming model, CCP; etc... Nazareth includes: nonlinear least-squares, NLSQ; nonlinear equations, NEQ; linear convex-constrained, LCCP; and the classical linear programming problem, LP. In particular, the case NEQ is treated as a special case of NLP with an unspecified objective function, rather than an unconstrained minimization problem, e.g. sum of squares.

The early years following the introduction of the Simplex Method for LP in 1948 (e.g. [3]), i.e. the Dantzig modeling-and-algorithmic revolution, was characterized by a watershed between LP and NLP. The enormous prominence of LP was due in great part to the success of the simplex method. Whereas, the world being nonlinear, NLP provides better models in general, [4]. However, the introduction of primal-dual interior-point methods, for both LP and NLP, following the Karmarkar revolution, has shown that "The great watershed in optimization isn't between linearity and nonlinearity but convexity and nonconvexity", [9]. In addition, this revolution has brought to light the importance and centrality of Newton's method.

Nazareth concentrates on UM, NEQ, one-dimensional problems and LP and the algorithms used to solve them. Several themes are followed throughout the book. Comparisons are made between the model approach with Newton's method and the variable metric approach in the spirit of Cauchy. Arguments are presented to illustrate the flaws in a least squares approach in comparison to the homotopy approach followed in the modern interior-point methods. In addition, comparisons are made to illustrate the difference between algorithmic versus implementable methods.

The popular geometric view for UM is that of a marble on a mountain rolling downhill to some minimizing point. However, as Nazareth points out: "this ignores a central tenet of algorithmic optimization, namely, that the acquisition of information at any point x incurs a significant, nonzero cost. … Thus, a much better metaphor, …, is that of a small boat floating on an opaque lake that entirely covers the landscape." Thus experiments (costly) have to be made to estimate the depth/slope/curvature, whereas the marble samples these continuously at no expense.

Algorithms for UM are based on either the Newton (model based) or Cauchy (metric based) complementary perspectives. At the current estimate of the minimum, x_k, a direction finding problem, DfP, is solved to find a new improved point x_{k+1}. For Newton's method, one solves the so-called trust region subproblem, TRS, or quadratic model

$$\text{(TRS)} \quad \begin{array}{ll} \text{minimize} & g_k^T (x - x_k) + \tfrac{1}{2} (x - x_k)^T H_k (x - x_k) \\ \text{subject to} & \| (x - x_k) \|_{D^+} \leq \delta_k, \end{array}$$

where g_k is the gradient, H_k is an approximation of the Hessian (both at x_k), and D^+ scales the norm, i.e. the objective function is replaced by a (local) quadratic approximation and we restrict to the region where we trust the model. The optimal solution (approximated) is usually used as the new point x_{k+1}, or a line search is done in the direction $x_{k+1} - x_k$. These methods have proven to be robust and efficient and they can solve large scale problems, e.g. [8,2].

A simple algorithm for UM is: Cauchy's steepest descent method which uses the negative gradient as a search direction to find a new point x_{k+1}. Variable metric methods change the geometry by changing the metric/norm under consideration using information based on curvature considerations. This leads to the classical quasi-Newton methods, e.g. BFGS and DFP methods or updates. In these methods, first order (gradient) information is used to build up second order curvature information.

There is an ongoing debate on whether these methods are still needed following the introduction of automatic differentiation, see e.g. [1] and ADIFOR with URL: *www.cs.rice.edu/~ adifor/*.

Nazareth includes details on which choices of trust regions and Newton-Cauchy methods to chose in different settings.

The interior-point revolution has emphasized the importance of using Newton's method and solving a system of nonlinear equations based on the optimality conditions of an optimization problem. Nazareth presents two opposing views for solving NEQ, e.g. $H(x) = 0$, $H : \mathbb{R}^n \to \mathbb{R}^n$. Applying Newton's method directly is equivalent to applying the Gauss-Newton method, i.e. minimize the sum of squares $\min \| H(x) \|^2$ using a truncated quadratic model. However, this can lead to local minima which are not roots of $H(x)$. Nazareth calls this approach inherently flawed. Another approach uses homotopy or path-following to solve a parameterized problem that converges to a root as the parameter is varied. (This approach is the basis behind the successful modern primal-dual interior-point methods.)

Solving the one-dimensional root problem $h(x) = 0$ can be transformed using a potential function, i.e. we can integrate and find a function whose minimum coincides with $h(x) = 0$. However, this is not true for higher dimensions, since the Jacobian of h will not be symmetric. Thus, Nazareth makes the case that one-dimensional root finding is not the correct paradigm to lead to higher dimensional root finding. Rather a nonlinear least squares approach should be used. This leads to conjugate gradient methods for minimization. Included are discussions on the simplex and Nelder-Mead methods for nondifferentiable minimization.

As it was for the first revolution (led by Dantzig), the recent interior-point revolution (started by Karmarkar) originally focused on LP. Karmarkar's basic idea was to start at a central interior point of the feasible set and construct an ellipse around it within the feasible set. Optimizing the linear function over this ellipse is easy, thus yielding an improved point. Repeating this process can result in getting stuck near the boundary, as the new ellipse will have to be small. Therefore, the problem is rescaled so that the point is central again before constructing the ellipse. A potential function is used to ensure polynomial time convergence.

However, there have not been any practical numerical implementations of Karmarkar's original approach. (Connections between Karmarkar's approach and an implementable version called the affine scaling method have been made, see [5,6]). A breakthrough came when an equivalence was made with the classical log-barrier interior-point methods, [7] for a special choice of barrier parameters. This led to the introduction of the elegant primal-dual interior-point methods. These methods can be derived using the primal or dual log-barrier problem. They consist in applying Newton's method to the optimality conditions consisting of: (i) dual feasibility; (ii) primal feasibility; (iii) complementary slackness. Nazareth's preference for these methods is to focus on potential reduction and affine scaling. He includes a careful description of the path-following approach with the Mehrotra predictor-corrector modification. This builds on his previous work in the book on path-following. He also includes a chapter introducing the connection of log-barrier methods.

There are many excellent papers and books written describing the current interior-point revolution. Three recent books are [10,12,13].

The area of Optimization has reached a certain maturity. Problems of complexity/size undreamed of fifteen years ago are now solved as a matter of course. As stated by many numerical analysts: "I would rather be using today's theory and yesterday's computer than the reverse".

Nazareth has written a book that is both readable and covers many of the important new developments in Optimization.

References

[1] M. Berz, C. Bischof, and G. Corliss, editors. Computational differentiation, Philadelphia, PA, 1996. Society for Industrial and Applied Mathematics (SIAM). Techniques, applications, and tools.

[2] A.R. Conn, N.I.M. Gould, and P.L. Toint. Trust-Region Methods. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[3] G. Dantzig. Linear Programming and Extensions. Princeton University Press, Princeton, New Jersey, 1963.

[4] G. Dantzig. Linear programming. In History of Mathematical Programming: A Collection of Personal Reminiscences. CWI North-Holland, Amsterdam, 1991.

[5] I. I. Dikin. Iterative solution of problems of linear and quadratic programming. Doklady Akademii Nauk SSSR, 174:747-748, 1967. Translated in : Soviet Mathematics Doklady, 8:674-675, 1967.

[6] B.A. Freedman, M.S. Meketon, and R.J. Vanderbei. A modification of Karmarkar's linear programming algorithm. Technical report, AT&T Bell Laboratories, Holmdel, NJ, 1985.

[7] P.E. Gill, G.H. Golub, W. Murray, and M.A. Saunders. On projected newton barrier methods for linear programming and an equivalence to karmarkar's projective method. Math. Programming, 36(2):183-209, 1986.

[8] J. Nocedal and S.J. Wright. *Numerical optimization*. Springer-Verlag, New York, 1999.

[9] R.T. Rockafellar. Lagrange multipliers and optimality. SIAM Review, 35:183-238, 1993.

[10] R.J. Vanderbei. Linear Programming: Foundations and Extensions. Kluwer Acad. Publ., Dordrecht, 1998.

[11] M. Wright. The interior-point revolution in constrained optimization. In High performance algorithms and software in nonlinear optimization (Ischia, 1997), volume 24 of Appl. Optim., pages 359-381. Kluwer Acad. Publ., Dordrecht, 1998.

[12] S. Wright. Primal-Dual Interior-Point Methods. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, 1996.

[13] Y. Ye. Interior point algorithms. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1997. Theory and analysis, A Wiley-Interscience Publication.