How Mature is Nonlinear Optimization?

by N.I.M. Gould[1] and Ph. L. Toint[2]

Report 03/04 April 29, 2003

[1] Computational Science and Engineering Department
Rutherford Appleton Laboratory,
Chilton, Oxfordshire, England
Email: gould@rl.ac.uk

[2] Department of Mathematics,
University of Namur,
61, rue de Bruxelles, B-5000 Namur, Belgium,
Email: philippe.toint@fundp.ac.be

How Mature is Nonlinear Optimization?

Nick Gould and Philippe L. Toint

28 April 2003

Abstract

Numerical methods for solving nonlinear optimization problems have been developed for over 50 years. Has this field reached maturity? What are the current research frontiers and ongoing challenges? These are questions that this paper attempts to clarify, if not fully answer. The discussion does not explore the technical intricacies of nonlinear optimization techniques, but instead focusses on concepts and research practice. The field's vibrant nature is illustrated by a number of applications, such as adaptive lens design for spectacles, the piloting of injection of dangerous drugs to patients, the identification of parameters in biochemical models of neurons, food sterilization or animation techniques for video games.

1 Introduction

Numerical optimization, that is the corpus of methods and techniques for the solution of mathematically posed problems where one wishes to optimize one "objective" subject to a number of "constraints", has a rich history and continues to be an active research field. The purpose of the present paper is to consider an important subfield, nonlinear optimization, and to propose some thoughts about its level of maturity, both from the research and applications points of view.

Nonlinear optimization is concerned with the solution of continuous problems expressed in the form

$$\begin{array}{rl} \min\limits_{x} & f(x) \\ \text{subject to} & c_{\mathcal{E}}(x) = 0, \\ & c_{\mathcal{I}}(x) \geq 0, \end{array} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $c_{\mathcal{E}} : \mathbb{R}^n \to \mathbb{R}^e$ and $c_{\mathcal{I}} : \mathbb{R}^n \to \mathbb{R}^i$ are smooth. This description is somewhat simplified, as, for instance, the level of smoothness of the involved functions may vary or the constraints may involve convex sets... but it is adequate for the purpose of our discussion. The formulation (1.1) also hides a number of interesting special cases and distinctions, the most important being that between convex and nonconvex problems. We will discuss these issues in due course. At this point, we simply note that, except in a few isolated cases, all methods for solving (1.1) are iterative in nature, in the sense that they produces a potentially infinite sequence of iterates that (hopefully) converges to a desired solution.

2 A field no longer in infancy

That nonlinear optimization is no longer a new or young field of scientific activity barely needs discussion. Its rich history includes, for instance, the famous paper by Cauchy (1847), in the middle of the 19th century, but is also associated with other luminaries such as Euler, Gauss and Lagrange, for instance. It really became a field of its own immediately after World War II, along with the birth of the new field of "operations research". The development of the nonlinear least-squares fitting techniques by Levenberg (1944) may be considered as seminal from this point of view. The field had (and still has) special connections with linear algebra (the Cauchy paper being a good example, as are other important contributions such as the Conjugate-Gradient method by Hestenes and Stiefel, 1952 or the exact solution of the trust-region subproblem by Moré and Sorensen, 1983, and many others). Since our purpose is elsewhere, we will thus simply note that the field has a respectable history, a clear sign of being no longer in infancy.

At what we think is a more fundamental level, the methodological focus has also evolved from immediate (and vital) needs, such as effectively solving small problems, to more "long-term" questions. For instance, the early concern of designing methods that are asymptotically fast (that is converge quickly when started in a, possibly very small, neighbourhood of the solution) has progressively shifted to that of methods that are robustly *globally convergent*, in the sense that they are guaranteed to converge to a solution irrespective of the chosen starting point.

The field has become more conscious of itself, as new journals published a growing number of specialized contributions. Although some of the high quality publications of the early days such as *Mathematical Programming* continue to play an important role, it is telling that the long association of nonlinear optimization research with the *SIAM Journal on Numerical Analysis* has been mostly replaced by the highly successful *SIAM Journal on Optimization*. But, as also happens with teenagers, this growing self awareness went along with a clearer and more urgent realization of the dependence on the rest of the world. In particular, the strong and fruitful interaction of nonlinear optimization with a number of scientific domains where its techniques are applied has become even more crucial. The links between good numerical optimization methods and good software have also emerged as an important research topic. Other new journals, such as, for instance, *Optimization Methods and Software* and *Computational Optimization and Applications*, testify to this evolution, along with the continuing success of older sources like the *Journal of Optimization Theory and Applications* and the *Transactions of the ACM on Mathematical Software*.

Finally, the focus of the problems being solved has evolved from "toy problems", typically involving a very small number (typically less than 10) of variables and/or constraints, to larger and often more realistic instances, that currently feature possibly hundreds of thousands or even millions of variables and constraints. This is not to say that all small problems are uninteresting or easy, but the increasing size of the problems that can realistically be solved is, in our view, indicative of the field's evolution.

3 Some signs of maturity

We next review some elements that we believe testify of the maturity of nonlinear optimization.

3.1 An adequate theoretical understanding

We first look at the state of the theoretical understanding of the problems and numerical procedures to solve them. It is to us very noticeable that the role of theory itself has evolved to occupy a place which we believe is well balanced with practice. In what concerns the problems themselves, the gap between necessary and sufficient optimality conditions has now been shown to be tiny in general, and non-existent for problems such as quadratic programming. When applied to the theory of numerical algorithms, this balance manisfests itself in two complementary developments.

We first note that most of today's best practical algorithms are backed with a suitable convergence theory. This trend is not new, since it started with the convergence studies of variable-metric and quasi-Newton algorithms for unconstrained optimization in the 1970's (see, for instance Powell (1970, 1976) and with the analysis of penalty methods for constrained problems (see Fiacco and McCormick, 1968), soon followed by augmented Lagrangian (see Powell, 1969, Rockafellar, 1974 and Tapia, 1977) and sequential quadratic programming (SQP) (see Han, 1977, and Powell, 1978) methods. Irrespective of what these methods actually are, it is enough to say that they were (and for some, still are) at the leading edge of numerical nonlinear optimization at the time where they were studied. Thus researchers in the field have come to agree that providing convergence theory for successful algorithms is a very important part of making them even more robust and reliable. We continue today to hold the view that such a theory is a *necessary*[1]*, while by no means sufficient, condition* for a successful algorithm. Remarkably, today's best algorithms and packages (see next paragraph) are also supported by an adequate convergence theory. In a number of cases, this theory provides results on the crucial issue of global convergence to critical points, but also on the ultimate speed at which this convergence occurs. We also note that the two traditionally distinct (or, even, competing) algorithmic paradigms, known as linesearch and trust-region methods, may today be viewed in a unifying framework (see Section 10.3 of Conn, Gould and Toint, 2000*a*), which we also consider as a sign of maturity.

[1] Honesty forces us to acknowledge a few remarkable exceptions to this rule, like the BFGS variable-metric algorithm for nonconvex unconstrained minimization (Broyden, 1970, Fletcher, 1970, Goldfarb, 1970 and Shanno, 1970) or the MINOS algorithm (Murtagh and Saunders, 1978).

The second important such sign, as far as theory is concerned, is the development of an improved theory for the simpler but very important subclass of convex problems. For a long time, and although the best available algorithms were often more efficient for such problems, their supporting theory was typically unable to provide stronger or finer results for the convex case. The theory of self-scaling functions pioneered by Nesterov and Nemirovsky (1993) has changed this state of affairs considerably for the better. It indeed gives a much better insight on the *global speed of convergence*, that is even from the early iterations, when the iterates may still be far away from the (in that case, unique) solution. This has allowed the development of very efficient methods that are specific to convex problems. Again we see as a sign of maturity that the arguedly most important distinction between nonlinear problems (conxevity vs. non-convexity) is now reflected in our theoretical understanding. Equally important is that theoretical and practical improvements for the convex case are distilling into the non-convex world, a prime example being the global adoption of primal-dual rather than primal models in interior point algorithms for constrained optimization.

3.2 Improved software testing

A second important element in our analysis is the clear improvement in the quality of software testing, itself resulting in better software reliability. At first sight, software testing and comparision may seem a rather mundane and unchallenging part of the algorithmic development process, but fortunately this view has now been widely replaced with the realization of its crucial nature.

Testing nonlinear optimization software rests on two important and complementary topics: test problems and comparison methododogy. Both of these have matured considerably over the past ten years.

When nonlinear optimization was young, and most problems treated were small-scale "toys", exchanging the formulation of test cases was easy, as one could write their analytic description "on the back of an envelope", or publish them in a paper (see Hock and Schittkowski, 1981 and Moré, Garbow and Hillstrom, 1981 for influential publications of that kind). When larger problems became the norm, the likelihood of introducing coding errors or slight variations in test problems grew and made software comparison very awkward. An electronically transferable format for test problems was therefore desirable. The first such widely used format originated as a by-product of the development of the nonlinear programming package LANCELOT (see Conn, Gould and Toint, 1992) in the early 1990s. This format, ambitiously (and, with hindsight, perhaps rather arogantly) called the Standard Input Format, or SIF, was designed as a direct extension to nonlinear problems of the highly successful MPS format for linear programs. As such, it missed several features of more advanced modelling languages (such as sets), but had and continues to have the advantages of merely existing and of coming with free decoding programs. A complete testing environment, the Constrained and Unconstrained Testing Environment, or CUTE, was made available (without cost) to the research community by Bongartz, Conn, Gould and Toint (1995), with a large collection of test problems already coded in SIF, interfacing tools between this format and a number of existing packages and an extensive set of tools to facilitate testing of codes still at the development stage. These combined adavantages have probably contributed to outweight SIF's limitations and the use of the CUTE test problems and environment quickly became ubiquitous. For having talked with package developers, we believe that CUTE has increased the level of testing of software packages significantly, helping to track down coding bugs and providing a better assessment of code reliability. It is important to note that the CUTE test problem collection has continued to grow to include other sets (see Moré, 1989, Averick and Moré, 1992, Bondarenko, Bortz and Moré, 1999, Maros and Meszaros, 1999) and a number of problems arising directly from applications. It currently contains over a thousand problems of varying size, structure and difficulty. The CUTE environment has recently been superseded by a substantially improved avatar, named CUTEr (see Gould, Orban and Toint, 2003*b*).

Another positive development along this line is the growing success of the more complete modelling languages AMPL (see Fourer, Gay and Kernighan, 2003) and GAMS (see Brooke, Kendrick and Meeraus, 1988). There is no doubt that their modelling power considerably exceed that of SIF, but their generalization remains, in our view, somewhat hampered by their non-trivial cost.

The second pilar of nonlinear optimization software testing is the methodology used for comparing algorithms. This has long been a matter of debate, as providing combined measures of both reliability (the capacity of a package to effectively solve a problem) and efficiency (its speed in obtaining the solution) has always been difficult. The initial attempts by the Mathematical Programming Committe on Algorithms (COAL) did not result in any consensus in the community, and reporting of numerical experience with new algorithms has been *adhoc* for a long time. It is only recently that Dolan and Moré (2001) have proposed the concept of a *performance profile*, which seems to have gained increasing acceptance as a suitable way to compare reliability and efficiency of different algorithms. Suppose that a given algorithm i from a set $\mathcal{A}$ reports a statistic $s_{ij} \geq 0$ when run on example j from a problem test set $\mathcal{T}$, and that the smaller this statistic the better the variant is considered. Let

$$k(s, s^*, \sigma) = \begin{cases} 1 & \text{if } s \leq \sigma s^* \\ 0 & \text{otherwise.} \end{cases}$$

Then, the performance profile of algorithm i is the function

$$p_i(\sigma) = \frac{\sum_{j \in \mathcal{T}} k(s_{i,j}, s_j^*, \sigma)}{|\mathcal{T}|} \qquad (\sigma \geq 1),$$

where $s_j^* = \min_{i \in \mathcal{A}} s_{ij}$. Thus $p_i(1)$ gives the fraction of the number of examples for which algorithm i was the most effective (according to statistics s_{ij}), $p_i(2)$ gives the fraction of the number for which algorithm i is within a factor of 2 of the best, and $\lim_{\sigma \longrightarrow \infty} p_i(\sigma)$ gives the fraction of the examples for which the algorithm succeeded. Thus the performance profile gives comparative information on both efficiency and reliability. We believe that such profiles provide a very effective means of comparing the relative merits of different algorithms. This is important when designing new algorithms or improved variants, and clearly helps in establishing a more balanced (shall we say mature?) relative appraisal of today's nonlinear optimization packages, like KNITRO (Byrd, Hribar and Nocedal, 2000*b*), LOQO (Vanderbei and Shanno, 1999), SNOPT (Gill, Murray and Saunders, 2002), IPOPT (Wächter, 2002), filterSQP (Fletcher and Leyffer, 1998) or the GALAHAD library (Gould, Orban and Toint, 2003*c*). This appraisal is further clarified by independent comparative algorithms benchmarking, as H. Mittelman's initiative (see `http://plato.asu.edu/bench.html`).

3.3 A world of applications

While giving all the above considerations their proper place in the argument, the most obvious sign of maturity of nonlinear optimization remains the vast range of its applications to various branches of scientific research. Reviewing them, even briefly, is totally impossible here. A limited list of references to applications (of trust-region methods only) is available in Section 1.3 of Conn et al. (2000*a*). It is enough to mention here these applications cover fields as diverse as applied mathematics, physics, chemistry, biology, geology, engineering, computer science, medicine, economics, finance, sociology, transportation, ... and the enumeration is far from being exhaustive.

In what follows, we briefly outline five applications that we find interesting. We do not expect the reader to follow every detail of these problems (as we do not supply it), but their description or mathematical formulation is intended to illustrate the diversity of applications being considered, as well as the level of complexity that can be tackled with today's techniques. The interested reader is also invited to consult Averick and Moré (1992), Bondarenko et al. (1999) or R. Vanderbei's facinating Web site `http://www.princeton.edu/~rvdb`.

3.3.1 Progressive adaptive lens design

Our first application is the use of nonlinear optimization for the design of "progressive adaptive lenses" (PAL). In its simplest form, the PAL problem is to design the surface of a lens whose optical power must be smooth and is specified in different parts of the lens (low for far vision in the middle and high for near vision in the bottom part, see Figure 3.1), while at the same time minimizing astigmatim. Different formulations of the problem are possible (constrained or unconstrained), but they are all strongly nonlinear and nonconvex. Indeed, if the equation of the lens surface is given as the smooth function $z(x,y)$, then the optical power at (x,y) is given by

$$p(x,y) = \frac{N(x,y)^3}{2}\left[\left(1+\left[\frac{\partial z}{\partial x}(x,y)\right]^2\right)\frac{\partial^2 z}{\partial y^2}(x,y) + \left(1+\left[\frac{\partial z}{\partial y}(x,y)\right]^2\right)\frac{\partial^2 z}{\partial x^2}(x,y)\right.$$
$$\left. -2\frac{\partial z}{\partial x}(x,y)\frac{\partial z}{\partial y}(x,y)\frac{\partial^2 z}{\partial x \partial y}(x,y)\right],$$

where $N(x,y)$ is the z component of the vector normal to the surface, that is

$$N(x,y) = \frac{1}{\sqrt{1 + \left[\frac{\partial z}{\partial x}(x,y)\right]^2 + \left[\frac{\partial z}{\partial y}(x,y)\right]^2}}.$$

The surface astigmatism at (x,y) is then given by

$$a(x,y) = -2\sqrt{p(x,y) - N(x,y)^4\left(\frac{\partial z}{\partial x}(x,y)\frac{\partial z}{\partial y}(x,y) - \left[\frac{\partial^2 z}{\partial x \partial y}(x,y)\right]^2\right)},$$

which is even more nonlinear than the optical power.

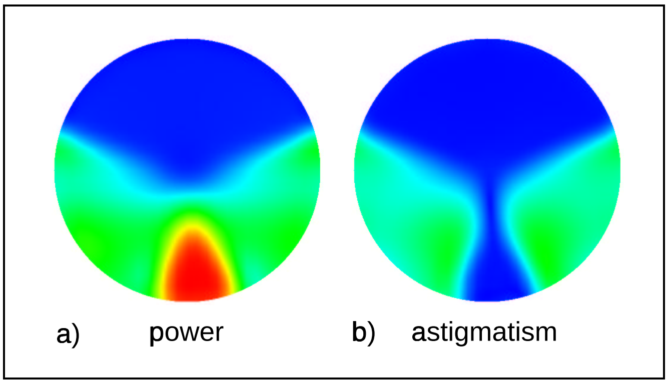

Figure 3.1: Optical power and astigmatism in a typical PAL design, with a smooth transition from low values in blue to high values in red (source: Loos et al., 1997)

3.3.2 Controlled drug injection

Discretized optimal control problem also constitute a growing source of applications for nonlinear optimization. Problems that involve constraints on the state variables (as opposed to constraints on the control variables only) are of special interest.

The controlled drug injection problem, whose full description can be found in of Maurer and Wiegand (1992), is a control problem based on the kinetic model of Aarons and Rowland for drug displacement, which simulates the interaction of the two drugs (warfarin and phenylnutazone) in a patient bloodstream. The state variable are the concentrations of unbound warfarin and phenylbutazone. The problem is to control the rate of injection of the pain-killing phenylbutazone so that both drugs reach a specified steady-state in minimum time and the concentration of warfarin does not rise above a given toxicity level. This last constraint therefore applies to the state variables of the problem, making the use of nonlinear programming techniques attractive. The differential equation describing the evolution of the drug concentrations in the bloodstream is discretized using a simple trapezoidal rule. The intrinsic nonlinearities of the model are non-convex.

3.3.3 Food sterilization

Another interesting discretized control problem is that of piloting the process of food sterilization in industrial autoclaves, as described in Kleis and Sachs (2000), where a full discussion of the problem and its solution can be found.. The idea is that the food to be sterilized is placed in closed autoclaves (see Figure 3.2) where it is heated (typically by hot water or steam).

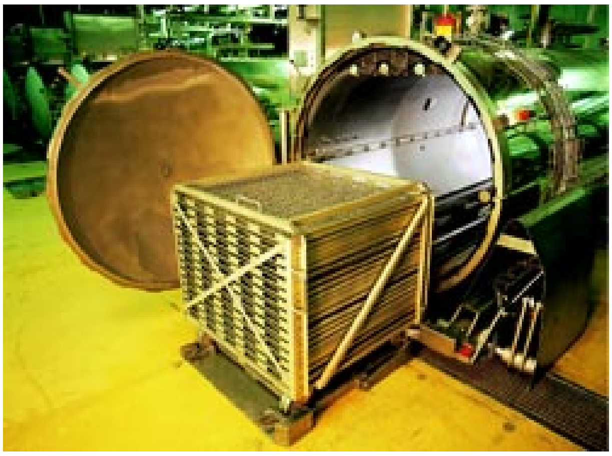

Figure 3.2: An autoclave for food sterilization

The question is then to optimize this heating in order to minimize the loss of vitamins but subject to the constraint that a certain fraction of undesired micro-organisms are killed and that every part of the food must reach a minimum temperature and not exceed a maximal one. The destruction of micro-organims and other nutrients of interest is described by

$$\frac{\partial C}{\partial t}(x,t) = -K[\theta(x,t)]C(x,t),$$

where $C(x,t)$ is the concentration of living micro-organisms or nutrients and $\theta(x,t)$ is the absolute temperature, at point x and time t. We also have that the function K depends on the temperature via the Arrhenius equation, that is

$$K[\theta] = K_1 e^{-K_2\left(\frac{1}{\theta}-\frac{1}{\theta_r}\right)},$$

where K_1, K_2 and θ_r are suitable constants. The evolution of temperature in the food container within the autoclave is described by a nonlinear heat equation of the form

$$\rho c(\theta)\frac{\partial \theta}{\partial t} = \nabla \cdot [k(\theta)\nabla\theta],$$

with suitable boundary conditions. Due to symmetry of the autoclaves, this 3D-problem can be reduced to 2D. The heat equation is discretized using finite elements for the spatial variables and the backward Euler method for time. This problem is also mentioned in Sachs (2003), where the reader will find an interesting discussion of PDE constrained optimization. There is an increasing awareness in the PDE community of the power of optimization, and an ongoing project to foster further links in this direction (see `http://plato.asu.edu/pdecon.html`).

3.3.4 Biological parameters estimation

We next consider a biological parameter identification problem discussed in Toint and Willms (2003). The problem is to identify parameters in a model of the voltage across a neuron membrane in the presence of a single passive current and a single voltage-activated current with Hodgkin-Huxley channel gating (see Figure 3.3). That is, the activation of p independent gates and total inactivation divided into n_h groups of partial activations with identical steady-state characteristics but different kinetic properties to give multi-exponential decay characteristics.

The ODEs for the voltage $v(t)$, the activation $m(t)$ and the partial inactivations $h_i(t)$ are

$$C\frac{dv}{dt}(t) = -g_a m(t)^p h(t)(v(t) - E_a) - g_\ell(v(t) - E_\ell) + I(t),$$
$$\frac{dm}{dt} = \alpha_m[v(t)](1 - m(t)) - \beta_m[v(t)]m(t),$$
$$\frac{dh_i}{dt} = \alpha_{h_i}[v(t)](1 - m(t)) - \beta_{h_i}[v(t))]m(t), \quad (i = 1, \dots, n_h),$$

where C is the membrane capacitance, g_a is the (time independent) active conductance, g_ℓ is the (time independent) passive conductance, E_a is the (time independent) active current reversal potential, E_ℓ is the (time independent) passive current reversal potential, $I(t)$ is the injected current, and where the total inactivation $h(t)$ is the sum of the different partial inactivations

$$h(t) = \sum_{i=1}^{n_h} f_i h_i(t) + f_{n_h+1}$$

for all t, and where the inactivation fractions f_i satisfy

$$0 \le f_i \le 1 \quad (i = 1, \dots, n_h) \text{ and } \sum_{i=1}^{n_h+1} f_i = 1.$$

The functions $\alpha_*(v)$ and $\beta_*(v)$ are Boltzmann functions of the form

$$\alpha_*[v] = \frac{1}{\tau_{\alpha,*}(1 - e^{(v-u_{\alpha,*})/\sigma_{\alpha,*}})}$$

and

$$\beta_*[v] = \frac{1}{\tau_{\beta,*}(1 - e^{(v-u_{\beta,*})/\sigma_{\beta,*}})}$$

with $*$ being m or h_i $(i = 1, \dots, n_h)$. Additionally, the parameters of the Boltzmann functions have to satisfy, for $i = 1, \dots, n_h$,

$$\tau_{\alpha,h_i} = \epsilon_i \tau_{\alpha,h}, \quad u_{\alpha,h_i} = u_{\alpha,h}, \quad \sigma_{\alpha,h_i} = \sigma_{\alpha,h}$$

$$\tau_{\beta,h_i} = \epsilon_i \tau_{\beta,h}, \quad u_{\beta,h_i} = u_{\beta,h}, \quad \sigma_{\beta,h_i} = \sigma_{\beta,h}$$

where the scaling factors ϵ_i are constrained by

$$1 = \epsilon_1 < \epsilon_2 < \ldots < \epsilon_{n_h}.$$

The ODE's are discretized using a 5 steps Backward Differentiation Formula with constant time stepping. The objective function is to minimize the least-squares distance between the voltages satisfying those equations and observed voltage values for a number of experiments (or sweeps). The experimental data is for a potassium A current in a pyloric dilator cell of the stomatogastric ganglion of the Pacific spiny lobster (see Figure 3.4). As can be seen from the equations, the problem is non-convex.

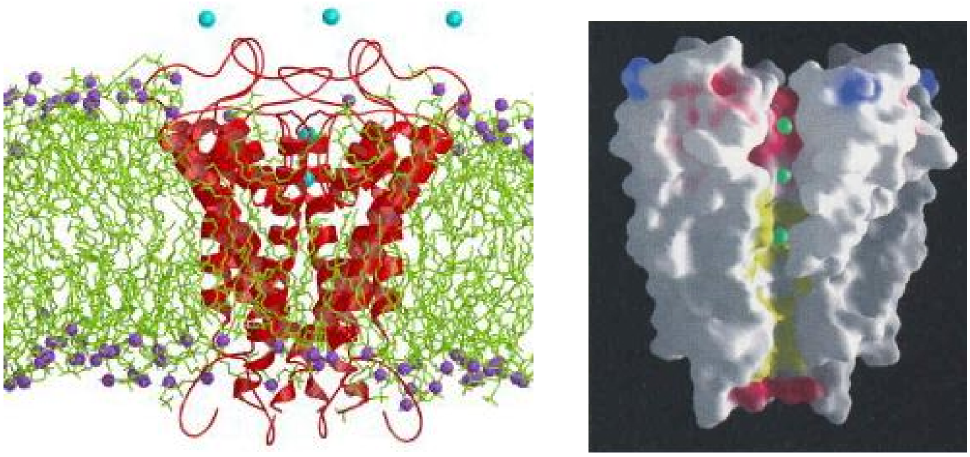

Figure 3.3: The ribbon structure of the K^+ channel molecule showing its insertion the membrane (the blue ions on top are at the exterior of the cell) and a solid rendering of this molecule (source: right picture from Sansom, 2001, left picture from Doyle et al., 1998)

In its current formulation, the problem uses fours experimental sweeps and involves around 16,000 variables and about the same number of constraints, only one of which is linear.

3.3.5 Mechanics and video games

Finally, we would like to mention here an application in a fairly different area: that of video animation and video-games. In an intereting paper, Anitescu and Potra (1996) have formulated the problem of representing the motion of multiple rigid objects in space, including their interaction (friction) when they hit each other. The formulation used is that of a time-dependent linear complementarity problem. While this problem is at the boundary of linear and nonlinear problems (it is solved by a variant of Lemke's algorithm), it is nevertheless of interest to us because it can be seen as the problem of finding a feasible solution, at each time t, of the nonlinear set of inequalities

$$\nabla_q \Phi[q(t)]v(t) \geq 0, \quad \Phi(q(t)) \geq 0$$

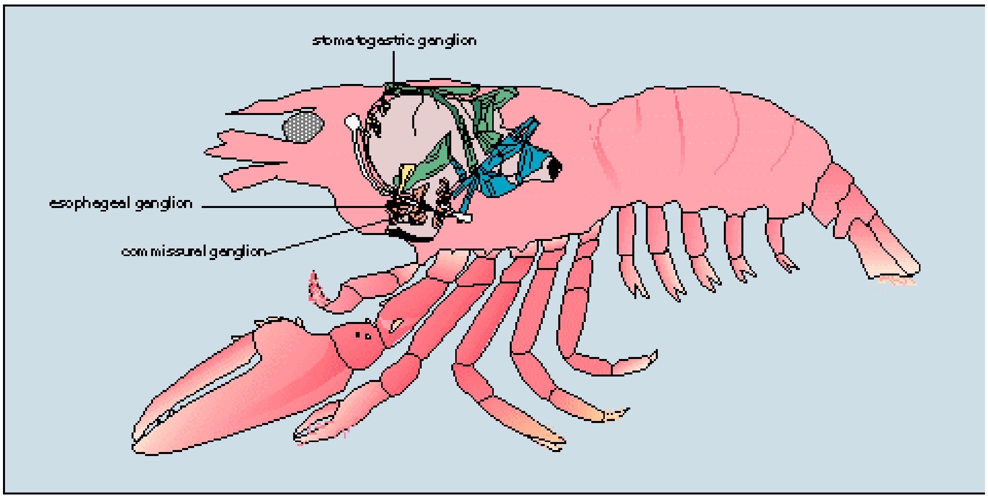

Figure 3.4: The position of the stomatogastric ganglion within the Pacific spiny lobster (source: Simmers et al., 1995)

where $q(t)$ is vector of states (positions) of the multi-body system at time t, $v(t) = \frac{dq}{dt}(t)$ is the vector of velocities, and the second inequality expresses the contact constraints (the fact that the problem bodies do not interpenetrate) for some smooth function Φ. This formulation is not only elegant, but is also amenable to practical implementation. It is in fact, in an implementation by MathEngine, at the heart of video-games such as the Vivid Image Actor, and provides a very realistic real-time simulation of shocks between rigid objects. The hidden presence of nonlinear problems in environments as ubiquitous as video-games also testify of its interest and reinforce our argument.

4 Is senility lurking?

Cynical observers may thus accept the maturity of nonlinear optimization as a discipline. They might also wonder if it already shows dangerous signs of aging and osbsolescence...these signs typically include a a more self-centered discourse or the repetition of older ideas instead of the creation of new ones. Although we acknowledge that self-centered contributions do exist(2), we hope that the variety of applications we have exposed in the previous section is convincing enough to dismiss the case of a narrower interaction with the world at large. We therefore focus, in what follows, on indicating that new directions and ideas continue to sustain the field's creativity.

4.1 The continuing impact of interior point methods

The first active current of research was initiated by the revival of interior point methods in linear and semi-definite programming. This generated a number of new contribu-

(2)There are, in our view, too many papers presenting convergence proofs for algorithms that have never been and will probably never be properly implemented, or even tried on simple examples...

tions that attempted to adapt these ideas initially to nonlinear convex problems, and subsequently to nonconvex ones. The main difficulty in adapting to the latter is that the first-order optimality conditions, for minimization, which are necessary and sufficient for linear and convex problems, are insufficient for nonconvex ones. Indeed, they can be satisfied at saddle points or even at maximizers.

We believe it is fair to say that the numerous contributions[(3)] on this topic are far from having exhausted the question or solved all practical problems. Oustanding issues include the efficient handling of nonlinear equality constraints, the effect of constraint scaling, suitable preconditioning techniques and extrapolation along the (possibly bizarre) central path for nonconvex problems. Moreover, the relative merits of interior point methods compared to more traditional SQP approaches are still a matter of lively research and debate (for a recent non-technical discussion of this topic, see Gould, 2003).

4.2 The revival of derivative free optimization

Algorithms for nonlinear programming that do not make use of derivative information have also come back in the foreground of research, after a long eclipse. Very popular in the infancy of the field, with classics like the simplex method of Nelder and Mead (1965), interest in these methods has been revived by significant recent progress in two different directions: interpolations methods and pattern search methods.

The first class of methods attempts to build a (typically quadratic) model of the function to be minimized, using multivariate interpolation techniques. The resulting algorithms (see Powell, 1994, 2000, 2002, or Conn, Scheinberg and Toint, 1997, 1998) are typically very efficient, and exploitation of problem structure is currently being successfully experimented (Colson and Toint, 2001, 2002, 2003).

The second class of derivative free methods use a prespecified or adaptive "pattern" to sample the variable space and compute minimizers. These methods are also the subject of much ongoing research (see Dennis and Torczon, 1991, Torczon, 1997, Coope and Price, 2000 and 2001, or Audet and Dennis, 2003). Extension of these techniques to large-scale problems is also being investigated (see Price and Toint, 2003).

Much remains to be done in this challenging sector, including better algorithms to handle larger problems with constraints.

4.3 Filter methods

We could not conclude this section of the new exciting ideas in nonlinear programming without briefly covering the filter methodology introduced by Fletcher and Leyffer

[(3)]See, for instance, Bonnans and Bouhtou (1995), Lasdon, Plummer and Yu (1995), Coleman and Li (1996*a*, 1996*b*), Bonnans and Pola (1997), Simantiraki and Shanno (1997), Forsgren and Gill (1998), Dennis, Heinkenschloss and Vicente (1998), Gay, Overton and Wright (1998), Vanderbei and Shanno (1999), Byrd et al. (2000*b*), Conn, Gould, Orban and Toint (2000*b*), Byrd, Gilbert and Nocedal (2000*a*), Chapter 13 of Conn et al. (2000*a*), or Gould, Orban, Sartenaer and Toint (2001), amongst many others.

(2002). This technique aims at promoting global convergence to minimizers of constrained problems without the need for a penalty function. Instead, the new concept of a "filter" is introduced which allows a step to be accepted if it reduces either the objective function or the constraint violation function. This simple yet powerful idea may be, in our view, the most significant progress in the past five years, and has already generated, in a very short time, a flurry of related research, both on algorithmic aspects (Ulbrich, Ulbrich and Vicente, 2000, Chin and Fletcher, 2001, Fletcher and Leyffer, 2003, Gonzaga, Karas and Vanti, 2002, Gould and Toint, 2002, Gould, Leyffer and Toint, 2003*a*) and on its theoretical underpinnings (Wächter and Biegler, 2001, Fletcher, Leyffer and Toint, 2002*b*, Fletcher, Gould, Leyffer, Toint and Wächter, 2002*a*) and inspired the organization of conferences and workshops devoted to this topic.

To illustrate its power, and at the same time that of the performance profiles of Dolan and Moré, in Figure 4.5 we present a CPU time comparison of a classical trust-region method and FILTRANE, a multidimensional filter method (Gould and Toint, 2003), on a large set of nonlinear feasibilty problems from the CUTEr collection.

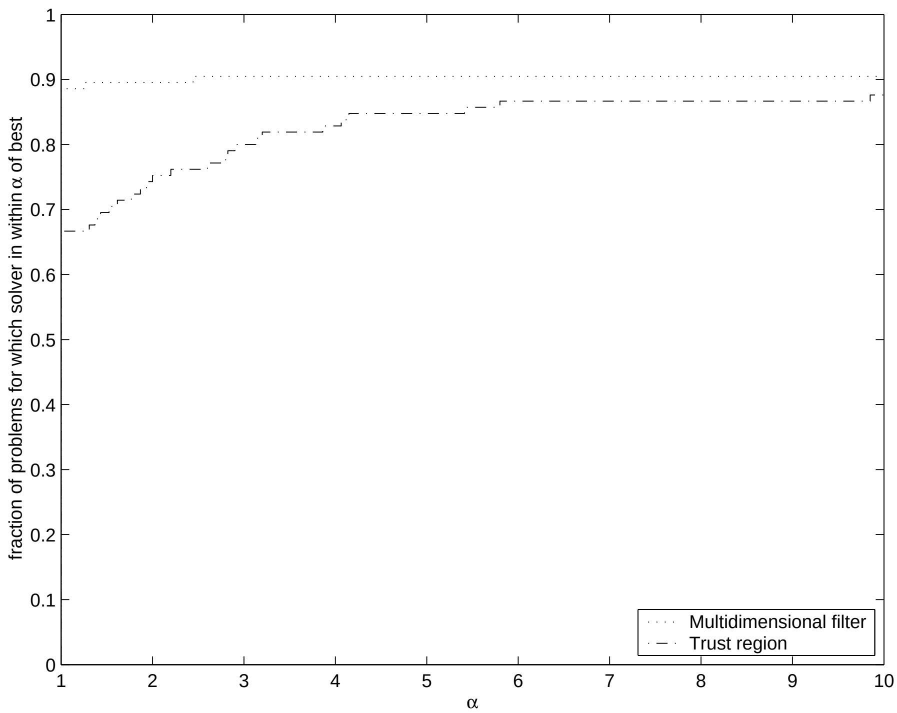

Figure 4.5: CPU time performance profile for multidimensional filter algorithm vs. classical trust-region algorithm on a set of 106 nonlinear feasibility problems

We see in this figure that the classical pure trust-region algorithm (one of the very best options before the filter idea) is slightly less reliabile than FILTRANE, and that the latter code is best (or tied best) on around 88% of the problems, a very significant advantage when compared to approximately 66% of the problems where the while the trust-region method is best. Furthermore, FILTRANE is within a factor 2 of the best on approximately 89% and within a factor 5 for approximately 91% of the problems, again an excellent performance. This kind of numerical results is really encouraging and stimulating, and one may therefore expect even more research activity in the domain of the filter methods. If it were only for that, it would already be enough to indicate the continuing vitality of nonlinear optimization...

5 Conclusion: the future's challenges

We have presented some arguments to vindicate our view that nonlinear optimization is a mature but not yet senile domain of research. Of course, these arguments are biased by our own experience and work, but we believe they are shared by a number of actors in the field. The last issue of the SIAG/OPT Views-and-News(4) provides additional elements that concur with ours, and also points to other domains where nonlinear optimization is strongly developing, like problems with equilibrium constraints, DAE-constrained problems, or, even more challengingly, nonlinear optimization with discrete variables.

What are the future's challenges? Besides the continuing improvement of methods and software, we feel that the successful specialization of nonlinear optimization to problem subclasses (like discretized optimal control problem or DAE constrained identification problems) constitutes a fruitful evolution and will in due course become important. The quest for methods that can solve problems that are intractable today, because of their size, nonlinearity or because they involve to many discrete variables) is not either anywhere near its end, a very invigourating perpective.

Acknowledgements

The second author is indebted to a number of colleagues who have helped supplying some of the material in this talk. In particular, thanks are due to Katia Demaseure, Mevlut Dinc at Vivid Image, Helmut Maurer, Ian Rycroft and Dilip Sequeira at MathEngine, Ekkehard Sachs, Andy Wathen and Allan Willms.

References

M. Anitescu and F. Potra. Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems. Reports on Computational Mathematics 93, Department of Mathematics, University of Iowa, Iowa City, USA, 1996.

C. Audet and J. E. Dennis. Analysis of generalized pattern searches. *SIAM Journal on Optimization*, **13**(3), 889–903, 2003.

(4) "Large-Scale Nonconvex Optimization", volume 14(1), April 2003, guest editors: S. Leyffer and J. Nocedal.

B. M. Averick and J. J. Moré. The Minpack-2 test problem collection. Technical Report ANL/MCS-P153-0694, Mathematics and Computer Science, Argonne National Laboratory, Argonne, Illinois, USA, 1992.

A. S. Bondarenko, D. M. Bortz, and J. J. Moré. COPS: Large-scale nonlinearly constrained optimization problems. Technical Report ANL/MCS-TM-237, Mathematics and Computer Science, Argonne National Laboratory, Argonne, Illinois, USA, 1999.

I. Bongartz, A. R. Conn, N. I. M. Gould, and Ph. L. Toint. CUTE: Constrained and Unconstrained Testing Environment. *Transactions of the ACM on Mathematical Software*, **21**(1), 123–160, 1995.

J. F. Bonnans and M. Bouhtou. The trust region affine interior-point algorithm for convex and nonconvex quadratic-programming. *RAIRO-Recherche Opérationnelle—Operations Research*, **29**(2), 195–217, 1995.

J. F. Bonnans and C. Pola. A trust region interior point algorithm for linearly constrained optimization. *SIAM Journal on Optimization*, **7**(3), 717–731, 1997.

A. Brooke, D. Kendrick, and A. Meeraus. *GAMS: a User's Guide*. The Scientific Press, Redwood City, USA., 1988.

C. G. Broyden. The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and its Applications*, **6**, 76–90, 1970.

R. H. Byrd, J. Ch. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming, Series A*, **89**(1), 149–186, 2000*a*.

R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large scale nonlinear programming. *SIAM Journal on Optimization*, **9**(4), 877–900, 2000*b*.

A. Cauchy. Méthode générale pour la résolution des systèmes d'équations simultannées. *Comptes Rendus de l'Académie des Sciences*, pp. 536–538, 1847.

C. M. Chin and R. Fletcher. Convergence properties of SLP-filter algorithms that takes eqp steps. Numerical Analysis Report NA/199, Department of Mathematics, University of Dundee, Dundee, Scotland, 2001.

T. F. Coleman and Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, **6**(2), 418–445, 1996*a*.

T. F. Coleman and Y. Li. A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization*, **6**(4), 1040–1058, 1996*b*.

B. Colson and Ph. L. Toint. Exploiting band structure in unconstrained optimization without derivatives. *Optimization and Engineering*, **2**, 349–412, 2001.

B. Colson and Ph. L. Toint. A derivative-free algorithm for sparse unconstrained optimization problems. *in* A. H. Siddiqi and M. Kočvara, eds, 'Trends in Industrial and Applied Mathematics', pp. 131–149, Dordrecht, The Netherlands, 2002. Kluwer Academic Publishers.

B. Colson and Ph. L. Toint. Exploiting problem structure in derivative-free algorithms for unconstrained optimization. Technical Report (in preparation), Department of Mathematics, University of Namur, Namur, Belgium, 2003.

A. R. Conn, N. I. M. Gould, and Ph. L. Toint. LANCELOT*: a Fortran package for large-scale nonlinear optimization (Release A).* Number 17 *in* 'Springer Series in Computational Mathematics'. Springer Verlag, Heidelberg, Berlin, New York, 1992.

A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. Number 01 *in* 'MPS-SIAM Series on Optimization'. SIAM, Philadelphia, USA, 2000*a*.

A. R. Conn, N. I. M. Gould, D. Orban, and Ph. L. Toint. A primal-dual trust-region algorithm for minimizing a non-convex function subject to bound and linear equality constraints. *Mathematical Programming*, **87**(2), 215–249, 2000*b*.

A. R. Conn, K. Scheinberg, and Ph. L. Toint. Recent progress in unconstrained nonlinear optimization without derivatives. *Mathematical Programming, Series B*, **79**(3), 397–414, 1997.

A. R. Conn, K. Scheinberg, and Ph. L. Toint. A derivative free optimization algorithm in practice. Technical Report TR98/11, Department of Mathematics, University of Namur, Namur, Belgium, 1998.

I. D. Coope and C. J. Price. Frame-based methods for unconstrained optimization. *Journal of Optimization Theory and Applications*, **107**, 261–274, 2000.

I. D. Coope and C. J. Price. On the convergence of grid-based methods for unconstrained optimization. *SIAM Journal on Optimization*, **11**, 859–869, 2001.

J. E. Dennis and V. Torczon. Direct search methods on parallel machines. *SIAM Journal on Optimization*, **1**(4), 448–474, 1991.

J. E. Dennis, M. Heinkenschloss, and L. N. Vicente. Trust-region interior-point SQP algorithms for a class of nonlinear programming problems. *SIAM Journal on Control and Optimization*, **36**(5), 1750–1794, 1998.

E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. Technical Report ANL/MCS-P861-1200, Mathematics and Computer Science, Argonne National Laboratory, Argonne, Illinois, USA, 2001.

D. A. Doyle, J. Morais Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon. Molecular basis of K^+ conduction and selectivity. *Science*, **280**(5360), 69–77, 1998.

A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* J. Wiley and Sons, Chichester, England, 1968. Reprinted as *Classics in Applied Mathematics 4*, SIAM, 1990.

R. Fletcher. A new approach to variable metric algorithms. *Computer Journal*, **13**, 317–322, 1970.

R. Fletcher and S. Leyffer. User manual for filterSQP. Numerical Analysis Report NA/181, Department of Mathematics, University of Dundee, Dundee, Scotland, 1998.

R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, **91**(2), 239–269, 2002.

R. Fletcher and S. Leyffer. Filter-type algorithms for solving systems of algebraic equations and inequalities. *in* G. Di Pillo and A. Murli, eds, 'High Performance Algorithms and Software in Nonlinear Optimization', pp. 259–278, Dordrecht, The Netherlands, 2003. Kluwer Academic Publishers.

R. Fletcher, N. I. M. Gould, S. Leyffer, Ph. L. Toint, and A. Wächter. Global convergence of trust-region SQP-filter algorithms for nonlinear programming. *SIAM Journal on Optimization*, **13**(3), 635–659, 2002*a*.

R. Fletcher, S. Leyffer, and Ph. L. Toint. On the global convergence of a filter-SQP algorithm. *SIAM Journal on Optimization*, **13**(1), 44–59, 2002*b*.

A. Forsgren and P. E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM Journal on Optimization*, **8**(4), 1132–1152, 1998.

R. Fourer, D. M. Gay, and B. W. Kernighan. *AMPL: A modeling language for mathematical programming.* Brooks/Cole–Thompson Learning, Pacific Grove, California, USA, second edn, 2003.

D. M. Gay, M. L. Overton, and M. H. Wright. A primal-dual interior method for nonconvex nonlinear programming. *in* Y. Yuan, ed., 'Advances in Nonlinear Programming', pp. 31–56, Dordrecht, The Netherlands, 1998. Kluwer Academic Publishers.

P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Journal on Optimization*, **12**(4), 979–1006, 2002.

D. Goldfarb. A family of variable metric methods derived by variational means. *Mathematics of Computation*, **24**, 23–26, 1970.

C. C. Gonzaga, E. Karas, and M. Vanti. A globally convergent filter method for nonlinear programming. Technical report, Department of Mathematics, Federal University of Santa Catarina, Florianopolis, Brasil, 2002.

N. I. M. Gould. Some reflections on the current state of active-set and interior point methods for constrained optimization. *SIAG/OPT Views-and-News*, **14**(1), 2–7, 2003.

N. I. M. Gould and Ph. L. Toint. Global convergence of a non-monotone trust-region filter algorithm for nonlinear programming. Technical Report (in preparation), Department of Mathematics, University of Namur, Namur, Belgium, 2002.

N. I. M. Gould and Ph. L. Toint. FILTRANE, a Fortran 95 filter-trust-region package for solving systems of nonlinear equalities, nonlinear inequalities and nonlinear least-squares problems. Technical Report (in preparation), Rutherford Appleton Laboratory, Chilton, Oxfordshire, England, 2003.

N. I. M. Gould, S. Leyffer, and Ph. L. Toint. A multidimensional filter algorithm for nonlinear equations and nonlinear least-squares. Technical Report TR-2003-004, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England, 2003*a*.

N. I. M. Gould, D. Orban, and Ph. L. Toint. CUTEr, a contrained and unconstrained testing environment, revisited. *Transactions of the ACM on Mathematical Software*, (to appear), 2003*b*.

N. I. M. Gould, D. Orban, and Ph. L. Toint. GALAHAD—a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization. *Transactions of the ACM on Mathematical Software*, (to appear), 2003*c*.

N. I. M. Gould, D. Orban, A. Sartenaer, and Ph. L. Toint. On the local convergence of a primal-dual trust-region interior-point algorithm for constrained nonlinear programming. *SIAM Journal on Optimization*, **11**(4), 974–1002, 2001.

S. P. Han. A globally convergent method for nonlinear programming. *Journal of Optimization Theory and Applications*, **15**, 319–342, 1977.

M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of the National Bureau of Standards*, **49**, 409–436, 1952.

W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes*. Springer Verlag, Heidelberg, Berlin, New York, 1981. Lectures Notes in Economics and Mathematical Systems 187.

D. Kleis and E. W. Sachs. Optimal control of the sterilization of prepackaged food. *SIAM Journal on Optimization*, **10**, 1180–1195, 2000.

L. S. Lasdon, J. Plummer, and G. Yu. Primal-dual and primal interior point algorithms for general nonlinear programs. *ORSA Journal on Computing*, **7**(3), 321–332, 1995.

K. Levenberg. A method for the solution of certain problems in least squares. *Quarterly Journal on Applied Mathematics*, **2**, 164–168, 1944.

J. Loos, G. Greiner, and H.-P. Seidel. Computer aided spectacle lens design. Technical Report 5, Department of Computer Science, University of Erlangen, Erlangen, Germany, 1997.

I. Maros and C. Meszaros. A repository of convex quadratic programming problems. *Optimization Methods and Software*, **11-12**, 671–681, 1999.

H. Maurer and M. Wiegand. Numerical solution of a drug displacement problem with bounded state variables. *Optimal Control Applications and Methods*, **13**, 43–55, 1992.

J. J. Moré. A collection of nonlinear model problems. Technical Report ANL/MCS-P60-0289, Mathematics and Computer Science, Argonne National Laboratory, Argonne, Illinois, USA, 1989.

J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, **4**(3), 553–572, 1983.

J. J. Moré, B. S. Garbow, and K. E. Hillstrom. Testing unconstrained optimization software. *Transactions of the ACM on Mathematical Software*, **7**(1), 17–41, 1981.

B. A. Murtagh and M. A. Saunders. Large-scale linearly constrained optimization. *Mathematical Programming*, **14**, 41–72, 1978.

J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, **7**, 308–313, 1965.

Y. Nesterov and A. Nemirovsky. *Self-concordant functions and polynomial-time methods in convex programming*. SIAM, Philadelphia, USA, 1993.

M. J. D. Powell. A method for nonlinear constraints in minimization problems. *in* R. Fletcher, ed., 'Optimization', pp. 283–298, London, 1969. Academic Press.

M. J. D. Powell. A new algorithm for unconstrained optimization. *in* J. B. Rosen, O. L. Mangasarian and K. Ritter, eds, 'Nonlinear Programming', pp. 31–65, London, 1970. Academic Press.

M. J. D. Powell. Some global convergence properties of a variable metric algorithm for minimization without exact line searches. *in* 'SIAM-AMS Proceedings 9', pp. 53–72, Philadelphia, USA, 1976. SIAM.

M. J. D. Powell. A fast algorithm for nonlinearly constrained optimization calculations. *in* G. A. Watson, ed., 'Numerical Analysis, Dundee 1977', number 630 *in* 'Lecture Notes in Mathematics', pp. 144–157, Heidelberg, Berlin, New York, 1978. Springer Verlag.

M. J. D. Powell. A direct search optimization method that models the objective by quadratic interpolation. Presentation at the 5th Stockholm Optimization Days, Stockholm, 1994.

M. J. D. Powell. UOBYQA: unconstrained optimization by quadratic interpolation. Technical Report NA14, Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, England, 2000.

M. J. D. Powell. Least Frobenius norm updating of quadratic models that satisfy interpolation conditions. Technical Report NA02, Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, England, 2002.

C. J. Price and Ph. L. Toint. Exploiting problem structure in pattern search methods for unconstrained optimization. Technical Report (in preparation), Department of Mathematics and Styatistics, University of Canterbury, Christchurch, New Zealand, 2003.

R. T. Rockafellar. Augmented Lagrangian multiplier functions and duality in nonconvex programming. *SIAM Journal on Control and Optimization*, **12**(2), 268–285, 1974.

E. W. Sachs. PDE constrained optimization. *SIAG/OPT News-and Views*, **14**(1), 7–10, 2003.

M. S. P. Sansom. Laboratory journal 2001 web page. `http://biop.ox.ac.uk/www/lj2001/sansom/sansom.html`, Laboratory of Molecular Biophysics, University of Oxford, Oxford, UK, 2001.

D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, **24**, 647–657, 1970.

E. M. Simantiraki and D. F. Shanno. An infeasible-interior-point method for linear complementarity problems. *in* I. Duff and A. Watson, eds, 'The State of the Art in Numerical Analysis', pp. 339–362, Oxford, England, 1997. Oxford University Press.

J. Simmers, P. Meyrand, and M. Moulins. Dynamic networks of neurons. *American Scientist*, **83**, 262–268, 1995.

R. A. Tapia. Diagonalized multiplier methods and quasi-Newton methods for constrained optimization. *Journal of Optimization Theory and Applications*, **22**, 135–194, 1977.

Ph. L. Toint and A. Willms. Numerical estimation of the parameters in a model of the voltage across a neuron membrane in the presence of a single passive current and a single voltage-activated current with Hodgkin-Huxley channel gating. Technical Report (in preparation), Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand, 2003.

V. Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, **7**(1), 1–25, 1997.

M. Ulbrich, S. Ulbrich, and L. Vicente. A globally convergent primal-dual interior point filter method for nonconvex nonlinear programming. Technical Report TR00-11, Department of Mathematics, University of Coimbra, Coimbra, Portugal, 2000.

R. J. Vanderbei and D. F. Shanno. An interior point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, **13**, 231–252, 1999.

A. Wächter. *An Interior Point Algorithm for Large-Scale Nonlinear Optimization with Applications in Process Engineering.* PhD thesis, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, USA, 2002.

A. Wächter and L. T. Biegler. Global and local convergence of line search filter methods for nonlinear programming. Technical Report CAPD B-01-09, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, USA, 2001.