

SIAG/OPT Views-and-News

A Forum for the SIAM Activity Group on Optimization

No. 2

Spring 1993

SOUL-SEARCHING IN THE MATHEMATICAL, COMPUTER AND DECISION SCIENCES

Computational mathematics is at least as old as the abacus, but it was only with the advent of the electronic computer that this previously stunted leg of the mathematical tripod was able to grow to full stature. The other two legs of the tripod are, of course, pure mathematics (the poetry of the subject) and applied mathematics (models and their analysis). Indeed, during the first part of the century, mathematics aspired to a precarious stability on just one leg, with G.H. Hardy atop, performing his graceful and acrobatic balancing act. Nowadays, however, the subject rests more secure with the help of the other two legs.

Computational mathematics is the study, in its broadest sense, of *algorithms* for solving problems of a *primarily mathematical nature*, for example, systems of equations or extrema of constrained functions, over finite and infinite-dimensional domains. Practitioners, most notably numerical analysts and optimization specialists (mathematical programmers), often had to seek refuge within newly-founded disciplines like computer science, operations research, mathematical sciences, decision science, management science, systems and control engineering, or industrial and operations engineering. In some cases, their mathematically and algorithmically-oriented work was able to coexist happily with the larger discipline. For example, business administration has never been distracted from its primary objective of *making money* by its more mathematically-oriented subdiscipline, management science. Similarly, engineering has always remained true to its primary objective of *building things*, and continues to extensively use and contribute to mathematics and algorithmics through systems theory, control theory, and other subdisciplines. In other cases, an excessive mathematical orientation has caused a distortion of the initial intent of a field, most notably, in the areas of economics and operations research. The latter, in particular, has experienced a crisis of identity, as evidenced by several recent soul-searching articles and a member survey in *Operations Research* and *OR/MS Today*.

Increasingly, mathematics is welcoming back her prodigal sons, and mathematics departments are beginning to view themselves as pure, applied *and* computational in nature, in a re-merging of the disciplines. Led by S. Smale

and his co-workers, a complexity theory for algorithms is being developed that is better suited to the real number field. And mathematicians are finally recognizing that computer scientists have provided them with an incomparable *laboratory* tool. Calculus courses aided by, for example, *Mathematica* or *Maple* (termed computer-assisted analysis or, less charitably, analysis without proofs) are now quite common. As a humorist puts it, soon algebra and geometry may enjoy popularizations as *calgebra* (aided by, for example, *Matlab*) and *calemetry* (aided by new scientific visualization tools).

A corresponding reappraisal is occurring in the sister disciplines. For example, the department of mathematical sciences, always a nebulous umbrella perhaps better suited to naming research institutes than university departments, has been recently reconstituted as a computational and applied mathematics department at Rice University. When the dust settles on the debate within the OR community, that subject may well become more oriented towards industrial and operations engineering (cf. U. Michigan). Numerical analysts, who have always maintained an uneasy truce with computer scientists, have also begun a re-examination of the undue influence of machines and finite-precision arithmetic on their subject. Indeed, in a recent very stimulating *SIAM News* article, Trefethen would even redefine numerical analysis from the popularly perceived 'study of rounding errors' to 'the study of algorithms for the problems of continuous mathematics'.

Computer science, on the other hand, is increasingly willing to be allied with engineering under the banner CS&E, and to explicitly recognize that its primary motivation is the *tool* itself, the 'care, feeding and procreation of computers' and attendant information processing needs of society at large. This has led to considerable debate

Contents:

Feature	1
Chairman's Column	2
Forum Essays	3
Bulletin Board	9

within the community centered on the recent *Computing the Future* report of the National Research Council. And, there are related rumours about the creation of computational science or scientific computing departments, although these may well be pipped at the post by a more embracing pure, applied and computational mathematics, or even a newly-minted *Algorithmic Science*. Time will tell!

Meanwhile, anyone interested in keeping track of these 'weighty' matters and their potential impact on optimization will find a short list of relevant references posted in the Bulletin Board Section.

CHAIRMAN'S COLUMN

by A.R. Conn

Dear Colleagues:

As I pointed out in the last issue, one of the advantages of being the chairman is that it gives me the opportunity to air some of my thoughts. Since I am also interested in encouraging feedback from members of this special interest group it does not hurt to be (at least a little) controversial.

Last month I was fortunate enough to attend the Conference on Large-Scale Optimization, at the University of Florida, organised by Bill Hager, Don Hearn and Panos Pardalos, (see elsewhere in this issue) and this led me to pondering about conferences and their role in our profession.

It seems to me that almost all the most satisfying conferences I have attended have been relatively specialised, have had no parallel sessions and have been worthwhile from a social as well as an intellectual point of view. For example, why do we ever have (or attend) large meetings with many parallel sessions in large, over-priced and relatively inhospitable hotels? I realise that it may well be selfish to insist upon small meetings, but at least if the number of attendees is large I still feel it is preferable for the number of talks to be small. One often hears the argument that unless one gives a talk how does one obtain funding to attend, but then that is exactly the point. Is it reasonable that one can only obtain funding for a meeting by providing a means to have that meeting be less productive?

It also seems to me that one of the groups most likely to benefit from a meeting are graduate students and junior researchers — exactly the groups least likely to have funding to attend. In fact, the Gainesville meeting did arrange special support for graduate students and their enthusiasm was in evidence. I think that there should always be support available for graduate students.

I also believe that invited speakers should be chosen primarily for their ability to deliver good presentations. In my opinion there is no excuse for poorly prepared transparencies and sloppily constructed talks. After all the speaker is imposing on a primarily captive audience. Moreover, the format that I think has been the most productive is to have a reasonable break after lunch and a later termination time (for example, as at Oberwolfach). If you want attendees to maintain an interest it seems obvious that talks from 8.30 am until 5.30 pm with only a one hour lunch break is not the correct formula. Speaking about formats, I have found that somehow poster sessions have never really worked for me. What do others think of this method of presentation?

On the same principle that says "however much I enjoy Bach, an ideal concert is not a program that consists only of Johann Sebastian", I want to broaden my horizons too, or at least have the opportunity to do so (even though sometimes this guarantees I will hear something that doesn't interest me at all). In the case of a meeting, is the best way to do this to have tutorial sessions (that is sessions meant to introduce a research area to a broader audience)?

Why do we have published proceedings? They are consistently over-priced, under-refereed and rarely read. One often has the impression that the only people who have a personal copy are the attendees at the conference. If it is only for them, then maybe the rationale for their existence is correct. Also, even though it is not possible to arrange first rate displays at all meetings, I do find it useful to have extensive displays of journals, books and software.

Finally, in these times of tight budgets are conferences and travel a luxury that should be curtailed? Personally, I believe otherwise. By virtue of the nature of our work, our colleagues are spread over the globe. Even with the wonders of electronic mail and the rapid dissemination of latex files (these days one can hardly speak of fast distribution of ideas via journal papers) it is *essential* that we regularly have the opportunity to meet with our colleagues. I am frequently impressed by the level of conscientiousness of my fellow researchers in attending the sessions at meetings and making the most of the opportunity to consult and discuss issues of mutual interest.

I am interested in hearing your comments and advice for what a good conference should be. I should hasten to add that the Florida conference was certainly an example of the type of meeting that was very worthwhile to attend.

I hope that I can look forward to your active participation in meetings, the group and this newsletter. If this column doesn't excite any response (my email address is arconn@watson.ibm.com, our editor's is nazareth@amath.washington.edu) I will have to start talking about health care or gun control!

FORUM ESSAYS

TENSOR METHOD SOFTWARE PACKAGES FOR NONLINEAR EQUATIONS AND LEAST SQUARES, AND UNCONSTRAINED OPTIMIZATION

by Robert B. Schnabel¹

The purpose of this article is to mention the (free) availability of two new software packages, one for solving systems of nonlinear equations and nonlinear least squares problems, the second for solving unconstrained optimization problems. The packages implement the tensor methods that have been developed over the past decade by the author and collaborators Ali Bouaricha, Tatung Chow, Dan Feng, and Paul Frank. They also provide the option of using standard Newton or Gauss-Newton based methods. The software packages are intended for problems where the Jacobian or Hessian matrix is available either analytically or by finite differences, and the number of parameters is not too large, say less than 200. In tests so far, the tensor methods implemented in these packages appear to be significantly more robust and efficient than standard methods, especially on problems where the Jacobian or Hessian matrix at the solution has a small rank deficiency. The remainder of this article very briefly discusses the problems addressed by this software, the algorithmic approaches used, the features of the software packages, and how the packages may be obtained.

The first software package, TENSOLVE, is intended to find the simultaneous solution of n nonlinear equations in n unknowns,

$$\begin{aligned} &\text{given } F(x) : R^n \rightarrow R^n, \\ &\text{find } x_* \in R^n \text{ for which } F(x_*) = 0 \end{aligned}$$

or to solve the closely related nonlinear least squares problem,

$$\begin{aligned} &\text{given } F(x) : R^n \rightarrow R^m, \quad m > n, \\ &\text{find } x_* \in R^n \text{ that minimizes } \|F(x)\|_2. \end{aligned}$$

In both cases it is intended for problems where $F(x)$ is at least once continuously differentiable, and where the

Jacobian matrix $F'(x)$ is available analytically or by finite difference approximation. Standard methods for solving such problems base each iteration upon a linear Taylor series model of $F(x)$ around the current iterate x_c ,

$$M(x_c + \delta) = F(x_c) + F'(x_c)\delta,$$

resulting in the local method being Newton's method for the nonlinear equations problems, and the Gauss-Newton method for nonlinear least squares.

Tensor methods for nonlinear equations or nonlinear least squares instead base each iteration upon the quadratic model

$$M(x_c + \delta) = F(x_c) + F'(x_c)\delta + \frac{1}{2}T_c\delta\delta,$$

where T_c is a very low rank three dimensional tensor, most often rank one. No second order derivatives are calculated; rather, the tensor term T_c is formed so that the model interpolates $F(x)$ at the most recent previous iterate, and sometimes at additional previous iterates. This causes the model to have good second derivative information in the directions to one or more recent iterates, while requiring only a few vectors of additional storage, and allowing the model to be formed and solved using only a small multiple of mn operations more than the $O(mn^2)$ cost of forming and solving the standard model (here $m = n$ for nonlinear equations) [7,2]. That is, the additional costs per iteration of using the tensor method are insignificant.

The package TENSOLVE allows the user to choose a method based upon either the tensor model or the standard linear model, with either a line search or trust region global strategy in either case. In our tests on standard test problems, the tensor method using either global strategy solves significantly more problems than the standard method, and on problems solved by both methods, the tensor method is generally about 30-40% more efficient than the corresponding standard method on problems where $F(x_*)$ is nonsingular, and about 30-50% more efficient on problems where $F(x_*)$ has rank $n-1$ or $n-2$ [2,1]. In the case when $F(x_*)$ has rank $n-1$, the performance of the tensor method is explained in part by a (3-step) superlinear local convergence rate as opposed to the linear convergence rate (with constant 0.5) of the standard method [5], but empirically the advantage seems to come more generally from having a better model in the most recent step direction. In tests in [1], the efficiency of the tensor method for nonlinear least squares also compares favorably with the well-known package NL2SOL [4].

The second software package, TENMIN, is intended to solve the unconstrained optimization problem

$$\min_{x \in R^n} f(x) : R^n \rightarrow R.$$

¹Research supported by Air Force Office of Scientific Research grant AFOSR-90-0109, ARO grant DAAL 03-91-G-0151, NSF grant CCR-9101795. Author's address: Department of Computer Science, University of Colorado at Boulder, Boulder, Colorado 80309-0430.

In this case it is assumed that $f(x)$ is at least twice continuously differentiable, and that both the gradient vector $\nabla f(x)$ and the Hessian matrix $\nabla^2 f(x)$ are available analytically or by finite difference approximation. Whereas standard methods for solving these problems base each iteration upon a quadratic Taylor series model of $F(x)$ around the current iterate x_c ,

$$m(x_c + \delta) = f(x_c) + \nabla f(x_c)\delta + \frac{1}{2}\delta^T \nabla^2 f(x_c)\delta,$$

tensor methods augment this model by the third and fourth order terms

$$\frac{1}{6}T_c\delta\delta\delta + \frac{1}{24}V_c\delta\delta\delta\delta$$

where T_c and V_c again are very low rank tensors that are computed without calculating any higher order derivatives. In the software package TENMIN, T_c and V_c are calculated so that the tensor model interpolates the function and gradient value at the previous iterate. This results in T_c having rank two and V_c having rank one. It causes the tensor model to have good second derivative information in the direction to the most recent iterate, while again requiring only a few vectors of additional storage, and allowing the model to be formed and solved using only a small multiple of n^2 operations more than the $O(n^3)$ cost of forming and solving the standard model [6,3].

The package TENMIN also allows the user to choose a method based upon either the tensor model or the standard quadratic model, using a line search global strategy in either case. In our tests on standard test problems, the tensor method solves considerably more problems than the standard method. On problems solved by both methods, on the average the tensor method is about 20-30% more efficient than the corresponding standard method on problems where $\nabla^2(x)$ is nonsingular at the solution, and about 30-40% more efficient on problems where $\nabla^2(x)$ has rank $n - 1$ or $n - 2$ at the solution [3].

It should be noted that since tensor methods use the Jacobian or Hessian matrix, they may be less efficient than quasi-Newton methods for nonlinear equations or unconstrained optimization (e.g. Broyden's method or the BFGS method) in cases where this matrix must be calculated by finite differences and function evaluation is expensive. However in practice it appears that for many (perhaps most) problems, the Jacobian or Hessian is available analytically, often at not much more cost than the function or gradient. For such problems, tensor methods should be of interest. For nonlinear least squares problems, almost all methods compute or approximate the Jacobian at each iteration (to assess whether one is close to a solution), so tensor methods appear to clearly be of interest.

Thus, we have created the TENSOLVE and TENMIN software packages because the tensor methods may offer gains in robustness and efficiency over existing methods on a reasonable selection of problems. We have included the option of using standard methods in both packages so that the user may, with one package, compare and choose between the two approaches for any particular class of problems. The style and options of both software packages is similar to the UNCMIN unconstrained minimization package [8]. Both provide the user with the choice of either a short calling sequence where the user supplies only the function, number of parameters, and a starting point, or a longer calling sequence where the user can override any of the default parameter values. Using the longer calling sequence the user can choose whether to use the standard or tensor method, the line search or trust region strategy in TENSOLVE, and whether to supply analytic derivatives or calculate derivatives by finite differences. With the longer calling sequence the user can also supply stopping tolerances, scaling information, and control the output from the package. Both packages are coded in Fortran 77. The TENSOLVE package consists of approximately 8800 lines of code (25% comments) of which 3500 are the UNCMIN software package that occasionally is called to solve a small subproblem in the tensor method. The TENMIN package consists of approximately 2800 lines of code of which about 35% are comments.

To obtain either of these packages, please contact the author, preferably by email at bobby@cs.colorado.edu, or Betty Eskow (a research staff member at U. Colorado) at eskow@cs.colorado.edu. There is no charge, although the packages can not be used in a commercial product without prior consent of the developers. We do ask that users report any interesting experiences with these packages to the developers, such as interesting comparisons between tensor methods and either the standard methods in these packages or other software.

It should finally be mentioned that in [1], tensor methods were extended to solve large, sparse systems of nonlinear equations and nonlinear least squares problems in an efficient manner. Software that implements these methods and standard methods, using efficient sparse linear algebra techniques, has been developed and tested. This software is not yet available for distribution but is likely to be within about a year.

References

- [1] A. Bouaricha, "Solving large sparse systems of nonlinear equations and nonlinear least squares problems using tensor methods on sequential and parallel computers", Ph.D. dissertation, Department of Computer Science, University of Colorado, 1992.

- [2] A. Bouaricha and R. B. Schnabel, "A software package for tensor methods for nonlinear equations and nonlinear least squares", in preparation.
- [3] T. Chow, E. Eskow, and R. B. Schnabel, "A software package for unconstrained optimization using tensor methods", Technical Report CU-CS-491-90, Department of Computer Science, University of Colorado, Dec. 1990, to appear in *ACM Trans. Math. Soft.*
- [4] J. E. Dennis, D. M. Gay, and R. E. Welsch, "NL2SOL - An adaptive nonlinear least squares algorithm", *ACM Trans. Math. Soft.* 7, 1981, pp. 369-383.
- [5] D. Feng, P. D. Frank, and R. B. Schnabel, "Local convergence analysis of tensor methods for nonlinear equations", Technical Report CU-CS-591-92, Department of Computer Science, University of Colorado, Apr. 1992, to appear in *Mathematical Programming*, 1993.
- [6] R. B. Schnabel and T. Chow, "Tensor methods for unconstrained optimization using second derivatives", *SIAM J. Opt.* 1, 1991, pp. 293-315.
- [7] R. B. Schnabel and P. D. Frank, "Tensor methods for nonlinear equations", *SIAM J. Numer. Anal.* 21, 1984, pp. 815-843.
- [8] R. B. Schnabel, J. E. Koontz, and B. E. Weiss, "A modular system of algorithms for unconstrained minimization", *ACM Trans. Math. Soft.* 11, 1985, pp. 419-440.

GLOBAL CONFUSION

by Jorge Nocedal¹

One of the most interesting results one can prove about an optimization algorithm is that it is globally convergent. By this I mean that, for any starting point, the sequence of iterates (or a subsequence of iterates) converges to a solution. A good example is the following well-known result for the steepest descent method for the minimization of a smooth function f . It states that, if implemented with an appropriate line search, and if f is bounded below, then $\lim_{k \rightarrow \infty} \nabla f(x_k) = 0$. Thus the steepest descent method is globally convergent to stationary points of f . This is a useful global convergence result since it applies to a wide class of functions occurring in practice; its only weakness is that it does not guarantee that the iterates approach minimizers, but only stationary points - which for the purpose of this result are the "solution

points". Other very interesting convergence results have been established, for example, for trust region methods for unconstrained minimization.

Since global convergence is such an important property, it has also been considered in many studies of algorithms for nonlinearly *constrained* optimization. However, a reader who consulted the literature to find out which algorithms are robust and likely to converge from remote starting points would be confused by many articles. This is because various "global convergence results" in the literature are not at all global; in fact some of these results could not even be characterized as local. An example given later on will clearly illustrate this.

The misuse of the term "global convergence" in nonlinearly programming is common, and has spread to the engineering literature. Readers should be warned that many "global" convergence results do not imply that the algorithm will converge from any starting point, or from a large set of starting points, but only establish certain properties about cluster points (which may or may not exist), or establish point of attraction results. For general nonlinearly constrained problems it is very difficult, if not impossible, to establish convergence from remote starting points. This has led authors to study several other types of results, which are often called "global". Unfortunately, many papers do not clearly spell out what the convergence results actually are, and it is not easy to find this out.

Let me be more concrete. Suppose that we had shown that, if the iterates generated by an algorithm remain in a compact set, there is a subsequence that converges to a solution. Is this a good result, a weak result, a global result or what? Well it depends: it could be a very interesting result or it could be quite useless. I would not call it a global convergence result until I had seen the conditions under which it had been established. Many merit functions used in practice can guide the iterates away from solution points, and make the sequence of iterates diverge. So the basic assumption may not hold. Would the inclusion of simple bounds on all the variables remove any doubts about the usefulness of the result? Not necessarily. Consider the following example due to Richard Byrd (private communication).

Example: Suppose that f is a nonlinear scalar function of several variables, and that we wish to find a stationary point of f , i.e. a point x_* such that $\nabla f(x_*) = 0$. This problem can be attacked, for example, by applying Newton's method with a trust region to the system of equations $\nabla f(x) = 0$.

We will, however, use a different (and most inappropriate) method, and will show that it is "globally convergent". We will employ the steepest descent method, with an exact line search, to minimize f . The reader will surely agree that this is a very inadequate method for the

¹Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208, USA.

problem of finding stationary points of f . If the only stationary point is a saddle point, then even if we start the method very close to the solution, the iterates are likely to move away from it, and in fact diverge to infinity.

Nevertheless we prove our “global convergence result” by assuming that the iterates stay in a compact set, and that f is smooth and bounded in this set. Then by the result stated in the first paragraph, for any starting point, we have that $\{\nabla f(x_k)\} \rightarrow 0$. Thus we have proved that the method, which we know must fail on many problems, is globally convergent!

The trouble is that assuming the boundedness of the iterates is not reasonable. Therefore to try to rescue our result we now force the iterates to stay in a compact set by imposing bounds on the variables. The compactness assumption is now valid, but the analysis of the method in this constrained setting cannot possibly give us a good result since we know that the method is still terribly wrong.

If instead of obscuring matters by assuming that the iterates stay in a compact set we had assumed that the method is to be applied only to functions f that are bounded below, the discussion would have been clear from the outset.

□

I hope that these remarks draw some attention to what I believe is an undesirable situation. If convergence studies are to be intelligible to more than a handful of people, they should state or summarize in simple terms the convergence results proved and the assumptions made. They should also discuss the validity of the assumptions, if they are not obvious. And perhaps these remarks will convince some authors not to describe certain convergence results as “global”, and find a more suitable name for them.

FROM DIKIN TO KARMARKAR VIA SCALINGS, MODELS AND TRUST REGIONS

by Larry Nazareth¹

The search direction used in Dikin’s interior-point algorithm [3] for linear programming (rediscovered during the past decade by several researchers and strengthened through extensive convergence analysis and computational investigation on practical problems) can be derived in three different, but equivalent ways, involving respectively:

- a rescaling of variables

¹Department of Pure and Applied Mathematics, Washington State University, Pullman, WA 99164-3113.

- a quadratic regularizing model
- an ellipsoidal trust region.

Each approach can be generalized using conic functions and collinear transformations, which were originally introduced into nonlinear minimization by Davidon [2]. In each case, one obtains an augmentation of Dikin’s direction by a ‘centering’ direction. This is the distinguishing characteristic of all Karmarkar-type interior-point algorithms.

In this mathematical essay, we outline these three approaches and their extensions with the objective of stimulating interest in this point of view (particularly the trust region approach) and perhaps motivating further research along these lines. Further detail and references to related work are given in a short bibliography, which is necessarily incomplete in the interests of brevity.

1. RESCALINGS

Consider the linear program

$$(\mathcal{LP}) : \text{minimize } c^T x \text{ s.t. } Ax = b, x \geq 0,$$

where $x \in R^n$, A is an $m \times n$ matrix and the other quantities are of matching dimensions. Let $x^k > 0$ be a feasible interior point and let $D \equiv \text{diag}[x_1^k, \dots, x_n^k]$.

Rescaling of variables is the most direct way to derive the search direction used in Dikin’s algorithm, namely, define $w = D^{-1}x$ and form a projected steepest descent direction at the transformed iterate $e = (1, 1, \dots, 1)^T$. In the original variables, it is the direction $d_p = -DPDc \equiv -c_p$, where $P = I - DA^T(AD^2A^T)^{-1}AD$. Note that the boundary hyperplanes are invariant, i.e., $u_j^T x = 0$, where u_j denotes the j ’th unit vector, becomes $u_j^T w = 0$ under the diagonal transformation. Also, the term ‘linear scaling’ is more accurate than ‘affine scaling’ (see (3)) in this derivation.

Collinear transformations provide a natural extension, because they are the most general transformations that map straight lines to straight lines, affine subspaces (hyperplanes) to affine subspaces and convex sets (that do not intersect the horizon) to convex sets. Consider therefore the following *collinear scaling* and its inverse:

$$w = \frac{D^{-1}x}{1 + e^T D^{-1}x}, \quad x = \frac{Dw}{1 - e^T w}. \quad (1)$$

Observe that $x \geq 0$ if and only if $w \geq 0$ and $\gamma \equiv (1 - e^T w) \geq 0$. Since D is diagonal, the boundary hyperplanes are again invariant. (\mathcal{LP}) transforms to the following fractional program:

$$\text{minimize } \frac{c^T Dw}{\gamma}$$

$$\text{s.t. } [AD| -b] \begin{bmatrix} w \\ \gamma \end{bmatrix} = 0, \quad [e^T | 1] \begin{bmatrix} w \\ \gamma \end{bmatrix} = 1,$$

$$w \geq 0, \gamma \geq 0.$$

In the transformed space, the search direction may be defined by the projected steepest descent direction $-\bar{P}[Dc/\gamma, -c^T Dw/\gamma^2]^T$, evaluated at the point $[\frac{1}{n+1}, \dots, \frac{1}{n+1}]^T$, where \bar{P} is the projection matrix $\bar{P} = I - \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{A}$ and $\bar{A} = \begin{bmatrix} AD & -b \\ e^T & 1 \end{bmatrix}$. It can then be expressed in the original variables using the collinear transformation.

The approach is closely related to the original projective approach of Karmarkar [7]. Connections between Karmarkar's projective transformations and collinear scalings have been noticed early on, see, for example, Lagarias [8]. Also related are the projective transformations of Freund [4] derived from convex analysis. It would be interesting to work out the precise connections between these different approaches.

Collinear transformations are taken up again later (see (4)).

2. MODELS

A second derivation of Dikin's algorithm has antecedents in the proximal-point method of Rockafellar [14] and the quadratic regularization of Pschenichny and Danilin [13]. The bound constraints $x \geq 0$ of (\mathcal{LP}) are replaced by a *quadratic regularization of the objective function in the metric defined by D^{-2}* . The resulting *quadratic approximating (Dikin) model* is

$$\text{minimize } c^T d + \frac{1}{2} d^T D^{-2} d \text{ s.t. } Ad = 0, \quad (2)$$

where $d \equiv x - x^k$. It has only *equality* constraints and can thus be solved explicitly. The vector from x^k to its minimizing point is along the Dikin direction. This approach is discussed in more detail in Nazareth [11] where it is also shown that *primal-dual affine scaling* and the relation defining the *central path* can be derived from a simultaneous consideration of quadratic models for (\mathcal{LP}) and its dual.

A very natural generalization of the foregoing approach is to substitute a conic approximating model that matches the *quadratic approximating (Dikin) model upto second order terms*. Thus one defines a model that locally approximates (\mathcal{LP}) at x^k as follows:

$$\min \frac{c^T d}{1 + h^T D^{-1} d} + \frac{1}{2} \frac{d^T (D^{-2} + hc^T + ch^T) d}{(1 + h^T D^{-1} d)^2} \text{ s.t. } Ad = 0.$$

where $d \equiv x - x^k$, $h \equiv \mu e$ and $\mu \geq 0$ scales the gauge vector of the conic model and provides the parameterization for the model.

Make the transformation $x = x^k + Zw$, where Z is a matrix of full rank that spans the null space of A , thus obtaining an unconstrained problem. By setting

its gradient to zero, solving and reverting to the original variables, it is immediately evident that the search direction is a linear combination of Dikin's direction $d_p = -Z(Z^T D^{-2} Z)^{-1} Z^T c = -DPDc = -c_p$ and the centering direction $e_p = Z(Z^T D^{-2} Z)^{-1} Z^T D^{-1} e = DPe$, where P is defined in Section 1. The gauge parameter μ , in turn, determines the relative weight given to each one. See Nazareth [11] for more detail.

The model-based approach is taken up again below (see 5)).

3. TRUST REGIONS

Quadratic-based trust regions have enjoyed substantial use in nonlinear minimization (Moré [9]) and more recently in linear programming (Ye [17], Gonzaga [6]). They are especially appealing from a geometric point of view.

In Dikin's algorithm, the bound constraints $x \geq 0$ of (\mathcal{LP}) are replaced by a quadratic (ellipsoidal) trust region $(x - x^k)^T D^{-2} (x - x^k) \leq 1$ and a step taken to or in the direction of the minimizing point of the resulting local approximating problem. Make the *affine scaling* $w = D^{-1}(x - x^k)$, i.e., a shift plus linear scaling $x = x^k + Dw$ to obtain

$$\text{minimize } (Dc)^T w \text{ s.t. } ADw = 0, w^T w \leq 1, \quad (3)$$

whose solution is obviously the projected steepest descent direction $-PDc$ with P defined in Section 1. In the original variables, this corresponds to the usual Dikin direction $c_p = -DPDc$.

It is a reasonable conjecture, premised on Davidon's [2] techniques, that something useful might be gained from substituting a conic trust region for the quadratic trust regions currently in use. Thus let us locally approximate (\mathcal{LP}) by the following problem:

$$(CTR\mathcal{P}) : \text{minimize } c^T d \text{ s.t. } Ad = 0, \frac{d^T D^{-2} d}{(1 + e^T D^{-1} d)^2} \leq \rho^2,$$

where $d \equiv x - x^k$, $D^{-1}e$ represents a particular choice of gauge vector and ρ is a scalar parameter. (For necessary background on conic functions, consult Davidon [2]. For more recent developments, see, for example, Ariyawansa and Lau [1] and references cited therein.)

The solution of $(CTR\mathcal{P})$ is discussed, in detail, in Nazareth [12]. In particular, the conic trust region is ellipsoidal whenever $0 \leq \rho < 1/\|e\|_2$. At the upper bound, the trust region becomes a paraboloid. (For example, when $n = 2$ and $x^k = e$, the Dikin trust region is a circle centered on e and the conic trust region is an ellipse with principal axis along the vector e and, in the limiting case, a parabola with focus at the point e .) $(CTR\mathcal{P})$ is a convex programming problem and the search vector obtained from it is always a direction of descent. The

latter has interesting properties. When $\rho \rightarrow 0$, this direction approaches Dikin's direction and when $\rho \rightarrow 1/\|e\|_2$, it approaches the centering direction, $e_p \equiv DPe$ whenever $c^T e_p \leq 0$, and the descent direction parallel to $-c_p + \left(\frac{c^T c_p}{2c^T e_p}\right) e_p$ otherwise.

When x^k has some very small components, the Dikin ellipsoidal trust region hugs the boundary along the corresponding axes and in consequence, the next iterate can also be very close to a boundary. The conic trust region, in effect, allows the Dikin ellipsoid to be reoriented along a preferred direction and reshaped. Optimizing over this new ellipsoid can give greater improvement as measured by objective function value and/or 'centeredness' of the next iterate.

The defining relation of the *spherical cone* or *conical* trust region formulated by Todd [16] (see page 3 of his article), when squared and inverted, closely resembles our conic trust region. There are, however, also significant differences as can be seen by looking at the limiting cases and also noting that the direction in [16] is not always a direction of descent. Todd [16] mentions that his approach is strongly motivated by the conical projections of Gonzaga [5], which in turn has antecedents and makes reference to Davidon [2]. Our approach is much more directly based on Davidon's techniques and it would be very interesting to explore the precise connections between the different approaches.

The resulting conic-based trust region LP algorithm also resembles that of Todd [15] and using proof techniques developed there and a suitable strategy for choosing ρ at each iteration, it is likely that polynomial complexity can be established. This is currently being studied.

Conic trust regions can be formulated in *dual and primal-dual* settings in an analogous way by making an appropriate choice of diagonal scaling matrix.

The use of *gauge vectors* other than $D^{-1}e$ is an interesting topic for further study (see also [11]).

Conic trust regions can also be used for nonlinear minimization leading to conjugate-gradient-like algorithms, as discussed, in detail, in Nazareth [12].

4. TRANSFORMATIONS AND MODELS MOTIVATED BY TRUST REGIONS

A variant on solving (CTRP) is to make a collinear transformation (shift plus collinear scaling) of the conic trust region based on the following mapping and its inverse (cf. (1)):

$$w = \frac{D^{-1}(x - x^k)}{1 + e^T D^{-1}(x - x^k)}, \quad x = x^k + \frac{Dw}{1 - e^T w}, \quad (4)$$

where the Jacobian of the second transformation is given by $J = (1/\gamma)(D + Dwe^T/\gamma)$, $\gamma = 1 - e^T w$. Thus (CTRP)

becomes

$$\text{minimize } \frac{c^T Dw}{1 - e^T w} \quad \text{s.t. } ADw = 0, \quad w^T w \leq \rho^2,$$

and ρ can be restricted to ensure that $1 - e^T w \geq 0$. If the objective function of this transformed problem is linearized then the solution is the projected steepest descent direction $-PJ^T c$, where P is defined earlier. This is a reasonable *approximation* to the solution of (CTRP) and it is easily seen to be a linear combination of the Dikin and centering directions. The approach is again closely related to the original approach of Karmarkar (see earlier rescaling discussion).

Another variant, motivated by the quadratic regularization (2) and the Lagrangian equations that can be associated with (CTRP), is to replace the bound constraints of (LP) by a *conic regularization* of its objective function, i.e., define a model that locally approximates (LP) at x^k as follows:

$$\text{minimize } c^T d + \frac{1}{2} \frac{d^T D^{-2} d}{(1 + \mu e^T D^{-1} d)^2} \quad \text{s.t. } Ad = 0, \quad (5)$$

where again $d \equiv x - x^k$ and $\mu \geq 0$ is a parameter. As before, make the transformation $d = Zw$. Then the foregoing model can be reexpressed as:

$$\text{minimize } \bar{c}^T w + \frac{1}{2} \frac{w^T \bar{D} w}{(1 + \bar{e}^T w)^2},$$

where $\bar{c} = Z^T c$, $\bar{D} = Z^T D^{-2} Z$ and $\bar{e} = \mu Z^T D^{-1} e$. The gradient of this function is $\frac{1}{\gamma^2} (I + \bar{e} w^T)^{-1} \bar{D} w + \bar{c}$ with $\gamma_* = (1 - \bar{e}^T w)$. By setting this gradient to zero, solving and reverting to the original variables, it is immediately evident that the resulting search direction is also a linear combination of Dikin's direction d_p and the centering direction e_p . This direction and the resulting algorithm have interesting properties that we are currently studying.

Finally, we may note that *the use of conic regularization within proximal-point algorithms* (Rockafellar [14]) offers an intriguing potential application to convex programming.

References

- [1] Ariyawansa, K.A. and Lau, D.T.M. (1992), "Local and q-superlinear convergence of a class of collinear scaling algorithms that extends quasi-Newton methods with Broyden's bounded- ϕ class of updates", *Optimization*, 23, 323-339.
- [2] Davidon, W.C. (1980), "Conic approximations and collinear scalings for optimizers", *SIAM J. Numerical Analysis*, 17, 268-281.

BULLETIN BOARD

- [3] Dikin, I.I. (1967), "Iterative solution of problems of linear and quadratic programming", *Soviet Math. Doklady*, 8, 674-675.
- [4] Freund, R. (1988), "An analog of Karmarkar's algorithm for inequality constrained linear programs with a 'new' class of projective transformations for centering a polytope", *Op. Res. Letters*, 7, 9-14.
- [5] Gonzaga, C.C. (1989), "Conical projection methods for linear programming", *Math. Prog.*, 43, 151-173.
- [6] Gonzaga, C.C. (1991), "An interior trust region method for linearly constrained optimization", *MPS-COAL Newsletter*, 19, 55-65.
- [7] Karmarkar, N. (1984), "A new polynomial-time algorithm for linear programming", *Combinatorica*, 4, 373-395.
- [8] Lagarias, J.C. (1991), "A collinear scaling interpretation of Karmarkar's linear programming algorithm", *SIAM J. Optimization*, to appear.
- [9] Moré, J.J. (1983), "Recent developments in algorithms and software for trust region methods", in A. Bachem, M. Grottschel and B. Korte (Eds.), *Mathematical Programming: The State of the Art, Bonn, 1982*, Springer-Verlag, 258-287.
- [10] Nazareth, J.L. and Ariyawansa, K.A. (1989), "On accelerating Newton's method based on a conic model", *Information Processing Letters*, 30, 277-281.
- [11] Nazareth, J.L. (1992), "Quadratic and conic approximating models in linear programming", *MPS-COAL Bulletin*, to appear.
- [12] Nazareth, J.L. (1992), "Trust regions based on conic functions in linear and nonlinear programming", *J. Num. Lin. Alg. Applics.*, to appear.
- [13] Pschenichny, B.N. and Danilin, Y.M. (1978), *Numerical Methods in Extremum Problems*, M.I.R. Publishers, Moscow, 1975 (English translation, 1978).
- [14] Rockafellar, R.T. (1976), "Augmented lagrangians and applications of the proximal point algorithm in convex programming", *Math. Oper. Res.*, 1, 97-116.
- [15] Todd, M.J. (1990), "A low complexity interior-point algorithm for linear programming", Technical Report 903, School of OR&IE, Cornell University, Ithaca, NY.
- [16] Todd, M.J. (1992), "Another derivation of the Karmarkar direction for linear programming", Technical Report 995, School of OR&IE, Cornell University, Ithaca, NY.
- [17] Ye, Y. (1989), "An extension of Karmarkar's algorithm and the trust region method for quadratic programming", in N. Megiddo (Ed.), *Progress in Mathematical Programming*, Springer-Verlag, 49-64.

THE NEXT INTERNATIONAL SYMPOSIUM ON MATHEMATICAL PROGRAMMING

The 15th International Symposium on Mathematical Programming sponsored by the Mathematical Programming Society will be held in Ann Arbor, Michigan, USA on the campus of the University of Michigan during 15-19 August 1994. Interested persons who have not received an announcement can request one by sending an e-mail message to xvismp@um.cc.umich.edu, or by writing to "15th International Symposium on Mathematical Programming, Conferences and Seminars, 541 Thompson St., Room 112, University of Michigan, Ann Arbor, MI 48109-1360, USA, FAX: 313-764-2990." All interested persons are requested to fill in the preregistration form on this announcement, or a Xerox copy of it, and send it to this address soon. In September 1993, a detailed 2nd announcement will be sent to all those who return the preregistration form.

POSTSCRIPT: CONFERENCE ON LARGE-SCALE OPTIMIZATION

On February 15-17, 1993, a conference on Large Scale Optimization, hosted by the Center for Applied Optimization (organized by W. Hager, D. Hearn and P. Pardalos), was held at the University of Florida. The conference was supported by the National Science Foundation and the U. S. Army Research Office and was endorsed from SIAM, MPS, ORSA and IMACS. Forty one invited speakers from all over the world presented papers on topics on mathematical programming and optimal control with an emphasis on algorithms, software development, and numerical experimentations. A wide range of practical applications such as airline crew scheduling, network problems, protein folding, location problems, multitarget tracking, and database design problems were discussed.

Attendees also included representatives from IBM, American Airlines, US Air, UPS, AT & T Bell Labs, Thinking Machines, and Argonne Labs. A unique feature of the meeting was the NSF sponsored attendance of thirteen graduate students from universities in the United States.

A conference publication on large-scale optimization will be published by Kluwer Academic Press.

Panos M. Pardalos, University of Florida, Gainesville.

LINEAR AND NONLINEAR CG-RELATED METHODS

Researchers in the area of CG-related methods for solving large-scale linear systems are often surprisingly isolated from researchers in the area of CG-related methods for large-scale nonlinear optimization. With this in mind, Loyce Adams (UW), Randy LeVeque (UW), Larry Nazareth (WSU) and Dave Watkins (WSU) are planning to organize a (small-scale) workshop at the University of Washington, Seattle, which is intended to bring together researchers from these two communities for an exchange of ideas. Summer, '95 seems a good time for optimizers, since it is between the Mathematical Programming Symposium (Ann Arbor) in '94 and the next SIAM Optimization Conference in '96. This also seems suitable from the point of view of the computational linear algebra schedule of events. We'd welcome hearing from people who are interested in participating and/or have preconditioning opinions that they might care to share during the planning stage as we put together a proposal.

SUPPLEMENTARY REFERENCES FOR FEATURE ARTICLE

- [1] Lloyd N. Trefethen, The Definition of Numerical Analysis, *SIAM News*, 25, p. 6, November, 1992.
- [2] Computing the Future: A Broader Agenda for Computer Science and Engineering, Report of the National Research Council, National Academy Press, 1992.
- [3] John A. White, An Existence Theorem for OR/MS, *Operations Research*, 39, p. 183-193, 1991.
- [4] John D.C. Little, Operations Research in Industry: New Opportunities in a Changing World, *Operations Research*, 39, p. 531-542, 1991.
- [5] Richard Teach, State of the Profession, *OR/MS Today*, p. 26-30, October, 1991.

SELECTED UPCOMING ARTICLES FOR SIAM J. OPTIMIZATION

A Globally and Superlinearly Convergent Algorithm for Convex Quadratic Programs with Simple Bounds *Thomas F. Coleman and Laurie A. Hulbert*
A New Method for Optimal Truss Topology Design *Aharon Ben-Tal and Martin P. Bendsøe*
Quantitative Stability of Variational Systems II. A Framework for Nonlinear Conditioning *H. Attouch and R. Wets*
Partial-Update Newton Methods for Unary, Factorable, and Partially Separable Optimization *Donald Goldfarb*

and *Siyun Wang*

Large-Step Interior Point Algorithms for Linear Complementarity Problems *M. Kojima, Y. Kurita, and S. Mizuno*

A Finite Smoothing Algorithm for Linear L_1 Estimation *H. B. Nielsen and K. Madsen*

On the Superlinear Convergence of Interior-Point Algorithms for a General Class of Problems *Yin Zhang, Richard Tapia, and Florian Potra*

Pointwise Broyden Methods *C. T. Kelley and E. W. Sachs*

Partially Finite Programming in L_1 and the Existence of Maximum Entropy Estimates *J. M. Borwein and A. S. Lewis*

Reverse Auction and the Solution of Inequality Constrained Assignment Problems *D. P. Bertsekas, D. A. Castañon, and H. Tsaknakis*

Uniformly Extremal Solutions in Sobolev Function Spaces for the Quadratic Case: Characterization and Applications *L. S. Thakur*

Second-Order Multiplier Update Calculations for Optimal Control Problems and Related Large-Scale Nonlinear Programs *J. C. Dunn*

A Noninterior Continuation Method for Quadratic and Linear Programming *Bintong Chen and Patrick T. Harker*

A Globally Convergent Method for L_p Problems *Yuying Li*

Convergence Analysis of a Proximal-Like Minimization Algorithm Using Bregman Functions *Gong Chen and Marc Teboulle*

Nonsmooth Equations: Motivation and Algorithms *Jong-Shi Pang and Liqun Qi*

A Lagrangian Relaxation Algorithm for Multidimensional Assignment Problems Arising from Multitarget Tracking *Aubrey B. Poore and Nenad Rujavec*

An Implementation of the Dual Affine Scaling Algorithm for Minimum Cost Flow on Bipartite Uncapacitated Networks *Mauricio G. C. Resende and Geraldo Veiga*

Multi-Objective Control-Structure Optimization Via Homotopy Methods *Joanna Rakowska, Raphael T. Haftka, and Layne T. Watson*

A Newton Method for Convex Regression, Data Smoothing, and Quadratic Programming with Bounded Constraints *Wu Li and John Swetits*

Manifold Structure of the Karush-Kuhn-Tucker Stationary Solution Set with Two Parametrics *Ryuichi Hirabayashi, Masayuki Shida, and Susumu Shindoh*

Numerical Experience with Limited-Memory Quasi-Newton and Truncated Newton Methods *X. Zou, I. M. Navon, M. Berger, K. H. Phua, T. Schlick and F. X. LeDimet*

A Collinear Scaling Interpretation of Karmarkar's Linear Programming Algorithm *J. C. Lagarias*

CONTRIBUTIONS TO THE V&N

The essay section of the next issue (Fall, '93) will focus on two related and computationally extremely challenging research areas at the frontier: *optimization under uncertainty* and *nonsmooth optimization*. It will contain articles by George Dantzig/Gerd Infanger, Roger Wets and Claude Lemarechal.

Issues of the Views-and-News will appear each Spring and Fall. Articles contributed by SIAG/OPT members are always welcome and can take one of two forms:

a) *Views*: short, scholarly, N^3 (Not Necessarily Noncontroversial) essay-type articles, say 2 to 4 pages long, on any topic in optimization and its interfaces with the sciences, engineering and education.

b) *News*: brief items for the Bulletin Board Section.

Author/developer previews of definitive optimization research monographs and software libraries, which are in the works or have just appeared, are also welcome for the essay section (space permitting). However, book reviews will not be published in order to avoid unnecessary overlap with the Mathematical Programming Society newsletter *Optima* nor short technical notes of the sort sought by the recently reorganized *MPS-COAL Bulletin*.

Our first preference is that a contribution take the form of a LaTeX file sent by email to the editor at the address given below. (If possible try it out in two-column format.) However, other forms of input are also acceptable.

The deadline for the Fall issue is September 1, 1993.

Larry Nazareth, Editor
Department of Pure and Applied Mathematics
Washington State University
Pullman, WA 99164-3113
email: nazareth@wsumath.bitnet
or nazareth@amath.washington.edu



PHILADELPHI

SIAM ANNUAL MEETING

JULY 12 - 16, 1993 WYNDHAM FRANKLIN PLAZA HOTEL

SELECTED THEMES

- Computational Fluid Dynamics
- Materials Science
- Biotechnology
- Industrial Problems
- Optimization
- Large Scale Computing
- Inverse Problems
- Nondestructive Testing
- Dynamical Systems
- Discrete Mathematics
- Finance

and in conjunction with the annual meeting

SYMPOSIUM July 8-10, 1993
Inverse Problems and Optimal Design in Industry

TUTORIALS July 11, 1993
Wavelets and Applications * Continuous Time
Finance * Structured Population Dynamics

WORKSHOP July 10, 1993
Making Mathematics Count:
A Workshop for High School Mathematics Teachers

For a copy of the preliminary program, please contact:

siam.
Conference Coordinator
Society for Industrial and Applied Mathematics
3600 University City Science Center
Philadelphia, PA 19104-2688
Telephone: 215-382-9800
Fax: 215-386-7999
E-mail: meetings@siam.org

Having trouble with your latest math paper?
Giving a presentation that you just can't pull together?
Struggling with your thesis or trying to get your first article published in a technical journal?
Handbook of Writing for the Mathematical Sciences is the book for you!



Handbook
of
Writing
for the
Mathematical
Sciences



NICHOLAS J. HIGHAM

This handy volume provides information on virtually every issue you will face when writing a technical paper or talk, from choosing the right journal to handling your references. You'll also get an overview of the entire publication process—invaluable for anyone hoping to publish in a technical journal.

To write a truly impressive paper, you'll need to understand the anatomy of a research paper and the steps involved in revising a draft. This book offers discussions of these fundamental topics, along with illustrative and provocative examples. Also included are chapters on standard English usage, using computers for writing and research, and writing technical material when English is a foreign language.

This handbook provides much-needed advice on handling the basic ingredients of a research paper, like definitions, theorems, examples, and equations. In addition, appendices provide essential reference material, including summaries of LaTeX symbols and Emacs commands, addresses of mathematical societies, and a list of papers that have won expository writing prizes.

This book is ideal for graduate students and teachers. Among its special features:

- Detailed and extensive use of examples make it an excellent teaching tool.
- Discusses TeX and other software tools for preparing publications.
- Thorough treatment, with examples, of how to write slides (transparencies) for a mathematical talk.
- Comprehensive index and extensive bibliography make it an excellent reference.
- Readable for everyone in the mathematical sciences—from undergraduates to seasoned professionals.

About the Author

Nicholas J. Higham is a Reader in Mathematics at the University of Manchester, England. He is the author of more than 40 publications and is a member of the editorial board of *SIAM Journal on Matrix Analysis and Applications*.

June 1993 / Approx. 224 pp. / Soft / 0-89871-314-5
List Price \$21.50 / SIAM Member Price \$17.20
SIAM Student Member Price \$12.00 / Order Code OT39

Contents

Preface; Chapter 1: *General Principles*; Chapter 2: *Writer's Tools and Recommended Reading*; Chapter 3: *Mathematical Writing*; Chapter 4: *English Usage*; Chapter 5: *When English is a Foreign Language*; Chapter 6: *Writing a Paper*; Chapter 7: *Revising a Draft*; Chapter 8: *Publishing a Paper*; Chapter 9: *Writing a Talk*; Chapter 10: *Computer Aids for Writing and Research*; Appendix A: *The Greek Alphabet*; Appendix B: *Summary of TeX and LaTeX Symbols*; Appendix C: *GNU Emacs—The Sixty+ Most Useful Commands*; Appendix D: *Mathematical Organizations*; Appendix E: *Winners of Prizes for Expository Writing*; Glossary; Bibliography; Index.

siam.

TO ORDER

Use your credit card (AMEX, MC, and VISA):
Call toll-free in USA: 800-447-SIAM
Outside USA call: 215-382-9800
Fax: 215-386-7999 / E-mail: service@siam.org

Or send check or money order to:
SIAM, Dept. BKAG93
P.O. Box 7260, Philadelphia, PA 19101-7260.

Shipping and Handling

USA: Add \$2.75 for first book and \$.50 for each additional book.
Canada: Add \$4.50 for first book and \$1.50 for each additional.
Outside USA/Canada: Add \$4.50 per book.