Methods for Nonlinear Constraints in Optimization Calculations

by A.R. Conn[1], N. Gould[2] and Ph.L. Toint[3]

Report 96/6 June 25, 1996

[1] IBM T.J. Watson Research Center, P.O.Box 218, Yorktown Heights, NY 10598, USA
Email: arconn@watson.ibm.com

[2] Rutherford Appleton Laboratory,
Chilton, Oxfordshire, England
Email: N.Gould@letterbox.rl.ac.uk
Current reports available by anonymous ftp from the directory
"pub/reports" on joyous-gard.cc.rl.ac.uk (internet 130.246.9.91)

[3] Department of Mathematics, Facultés Universitaires ND de la Paix, 61, rue de Bruxelles, B-5000
Namur, Belgium, EU
Email: pht@math.fundp.ac.be
Current reports available by anonymous ftp from the directory
"pub/reports" on thales.math.fundp.ac.be (internet 138.48.20.102)

Invited presentation at the IMA conference on the
State of the Art in Numerical Analysis
(York, UK, March 1996).
To appear in "The State of the Art in Numerical Analysis",
A. Watson and I. Duff, eds., Oxford University Press, Oxford, 1996.

Methods for Nonlinear Constraints in Optimization Calculations

Andrew R. Conn Nicholas I. M. Gould Philippe L. Toint

June 25, 1996

Abstract

Ten years ago, the broad consensus among researchers in constrained optimization was that sequential quadratic programming (SQP) methods were the methods of choice. While, in the long term, this position may be justified, the past ten years have exposed a number of difficulties with the SQP approach. Moreover, alternative methods have shown themselves capable of solving large-scale problems. In this paper, we shall outline the defects with SQP methods, and discuss the alternatives. In particular, we shall indicate how our understanding of the subproblems which inevitably arise in constrained optimization calculations has improved. We shall also consider the impact of interior-point methods for inequality constrained problems, described elsewhere in this volume, and argue that these methods likely provide a more useful Newton model for such problems than do traditional SQP methods. Finally, we shall consider trust-region methods for constrained problems, and the impact of automatic differentiation on algorithm design.

1 Introduction

In the previous assessment of the state of the art of constrained optimization, Powell (1987) presented powerful evidence that the future lay with sequential quadratic programming (SQP) methods. Powell's article focused on methods for problems with equality constraints. Perhaps, and with hindsight, it is possible to foresee the difficulties which arise as soon as inequality constraints are admitted. While we may be optimistic that SQP methods will still be the future methods of choice, the past decade has been a slightly sobering experience for those researchers working in constrained optimization, particularly for those interested in implementing algorithms. The overwhelming research thrust in optimization circles over the past ten years has been on interior-point methods for linear and, more recently, nonlinear programs. These methods offer an exciting alternative to the active-set methods which preceded them, but more importantly allow us to examine SQP methods in a new light.

In this paper, we shall try to outline the main advances that have taken place over the past decade. Following a brief review, we shall start where Powell (1987) left off, with methods for equality constraints. We then embark on a description of non-interior methods for handling inequality constraints. We caution the reader that the distinction between interior and non-interior methods is somewhat hazy, and we will sometimes delve into interior territory. Interior point methods are described elsewhere in this volume. For simplicity, we shall deliberately consider equality and inequality constraints separately, but remark that algorithms for problems with a mixture of constraints are normally a hybrid of those for the separate problems.

2 SQP methods

A thorough treatment of the history, theory and practice of SQP methods is given by Boggs and Tolle (1995).

2.1 Methods for equality constraints

We are concerned with finding the smallest value of the function $f(\boldsymbol{x})$ of the n real variables $\boldsymbol{x}$ in the case where $\boldsymbol{x}$ is required to satisfy a set of m equality constraints $\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{0}$. The first-order optimality or, as they are often known, Karush-Kuhn-Tucker (KKT) conditions for this problem are that

$$\nabla_x \ell(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}, \quad \text{and} \quad \boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{0}, \tag{2.1}$$

where the Lagrangian function $\ell(\boldsymbol{x}, \boldsymbol{y}) \stackrel{\text{def}}{=} f(\boldsymbol{x}) - \boldsymbol{c}(\boldsymbol{x})^T \boldsymbol{y}$ and where the components of the m-vector $\boldsymbol{y}$ are Lagrange multipliers.

A *sequential*, or *recursive*, quadratic programming (SQP) method is a method which seeks to improve an estimate $(\boldsymbol{x}, \boldsymbol{y})$ of the solution to (2.1) by finding corrections $(\boldsymbol{\Delta x}, \boldsymbol{\Delta y})$ by solving one (or more) quadratic programming problems. The next estimate of the required solution will be

$$\begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{y}^+ \end{pmatrix} = \begin{pmatrix} \boldsymbol{x} + \alpha_x \boldsymbol{\Delta x} \\ \boldsymbol{y} + \alpha_y \boldsymbol{\Delta y} \end{pmatrix}, \tag{2.2}$$

where the nonnegative stepsizes α_x and α_x may or may not be equal. The prototypical SQP method (see Pschenichny, 1970) finds $\boldsymbol{\Delta x}$ as the solution of the quadratic program

$$\begin{array}{lll} \underset{\Delta x \in \Re^n}{\text{minimize}} & \frac{1}{2} \boldsymbol{\Delta x}^T \boldsymbol{H} \boldsymbol{\Delta x} + \boldsymbol{\Delta x}^T \nabla_x \ell(\boldsymbol{x}, \boldsymbol{y}) & (2.3a) \\ \text{subject to} & \boldsymbol{A}(\boldsymbol{x}) \boldsymbol{\Delta x} + \boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{0}, & (2.3b) \end{array}$$

where $\boldsymbol{A}(\boldsymbol{x})$ is the Jacobian $\nabla_x \boldsymbol{c}(\boldsymbol{x})$, and $\boldsymbol{H}$ is a symmetric *approximation* to the Hessian of the Lagrangian function; $\boldsymbol{\Delta y}$ are taken as the Lagrange multipliers for (2.3). The step size α_x is found by requiring that $\psi(\boldsymbol{x} + \alpha_x \boldsymbol{\Delta x})$ is sufficiently smaller than $\psi(\boldsymbol{x})$ for some suitable *merit* function, and this is achieved by performing a backtracking (Armijo) *linesearch* with a unit initial stepsize. Finally α_y is either set to one or to α_x. The missing ingredients here are the choices of $\boldsymbol{H}$ and ψ, and it is mostly in these that the many proposed methods differ.

2.1.1 Hessian approximations

The first-order optimality conditions for (2.3) are that

$$\begin{pmatrix} \boldsymbol{H} & \boldsymbol{A}(\boldsymbol{x})^T \\ \boldsymbol{A}(\boldsymbol{x}) & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Delta x} \\ -\boldsymbol{\Delta y} \end{pmatrix} = - \begin{pmatrix} \nabla_x \ell(\boldsymbol{x}, \boldsymbol{y}) \\ \boldsymbol{c}(\boldsymbol{x}) \end{pmatrix}. \tag{2.4}$$

Assuming that $\boldsymbol{A}(\boldsymbol{x})$ is of full rank, and letting $\boldsymbol{Y}(\boldsymbol{x})$ and $\boldsymbol{Z}(\boldsymbol{x})$ be matrices whose columns span, respectively, the range and null spaces of $\boldsymbol{A}(\boldsymbol{x})$, we may write $\boldsymbol{\Delta x} = \boldsymbol{Y}(\boldsymbol{x}) \boldsymbol{\Delta x}_y + \boldsymbol{Z}(\boldsymbol{x}) \boldsymbol{\Delta x}_z$. On substituting into (2.4) $\boldsymbol{\Delta x}_y$ is completely determined by the constraints, as the solution to the non-singular system

$$\boldsymbol{A}(\boldsymbol{x}) \boldsymbol{Y}(\boldsymbol{x}) \boldsymbol{\Delta x}_y = -\boldsymbol{c}(\boldsymbol{x}), \tag{2.5}$$

while $\boldsymbol{\Delta x}_z$ then satisfies

$$\boldsymbol{H}_{zz} \boldsymbol{\Delta x}_z = -\boldsymbol{Z}^T \nabla_x \ell(\boldsymbol{x}, \boldsymbol{y}) - \boldsymbol{H}_{zy} \boldsymbol{\Delta x}_y, \tag{2.6}$$

where $\boldsymbol{H}_{zz} = \boldsymbol{Z}(\boldsymbol{x})^T \boldsymbol{H} \boldsymbol{Z}(\boldsymbol{x})$ and $\boldsymbol{H}_{zy} = \boldsymbol{Z}(\boldsymbol{x})^T \boldsymbol{H} \boldsymbol{Y}(\boldsymbol{x})$. Returning to (2.3) and performing the same substitution for $\boldsymbol{\Delta x}$ also yields (2.6), but the further requirement in (2.3) that a minimizer be sought suggests that the reduced Hessian $\boldsymbol{H}_{zz}$ should be positive semi-definite; to make this solution unique, the requirement is normally strengthened to insist that the reduced Hessian be positive definite.

Early SQP methods assumed that $\boldsymbol{H}$ was itself positive definite. This is clearly stronger than requiring that $\boldsymbol{H}_{zz}$ be definite. Most importantly, second-order optimality conditions for the original problem suggest that the reduced Hessian of the Lagrangian should be at least positive semi-definite, but that there is no reason for the Hessian of the Lagrangian itself to be definite. Advocates of this assumption cite simplicity, and were clearly keen to define $\boldsymbol{H}$ via one of the positive definite secant updating formulae which had proven so successful in unconstrained optimization. However, in our opinion, the contortions that were necessary to bend the secant updates into a suitable form (see, for example, Powell (1978)) underline the difficulties with the approach. In mitigation, when inequality constraints are introduced, the dimension of $\boldsymbol{Z}(\boldsymbol{x})$ may change dramatically from one iteration to the next, and it is then certainly convenient that $\boldsymbol{H}$ is positive definite. Remarkably, the very first SQP method (Wilson, 1963) used the exact Hessian of the Lagrangian, but until recently very few authors considered this choice (see Boggs, Tolle and Kearsley, 1994 and also Bonnans and Launay (1995) who sometimes modify the exact Hessian).

More recent methods have aimed at ensuring that $\boldsymbol{H}_{zz}$ is positive definite using positive-definite secant updates. However, this leads one to wonder how to handle the other matrices $\boldsymbol{H}_{zy}$ and $\boldsymbol{H}_{yy} \stackrel{\text{def}}{=} \boldsymbol{Y}(\boldsymbol{x})^T \boldsymbol{H} \boldsymbol{Y}(\boldsymbol{x})$, and it is here that most of the current proposals vary. Murray and Wright (1978) suggested that $\boldsymbol{H}_{zy}$ and $\boldsymbol{H}_{yy}$ should be set to zero. This gives what is known as a *reduced* Hessian method. With an appropriate secant update formula, such a scheme is two-step superlinearly convergent method so long as $\alpha_x = 1$ (Nocedal and Overton, 1985). A related reduced Hessian method, due to Coleman and Conn (1982*a*), replaces (2.5) by

$$\boldsymbol{A}(\boldsymbol{x}) \boldsymbol{Y}(\boldsymbol{x}) \boldsymbol{\Delta x}_y = -\boldsymbol{c}(\boldsymbol{x} + \boldsymbol{Z} \boldsymbol{\Delta x}_z), \tag{2.7}$$

in the vicinity of a stationary point or $\boldsymbol{\Delta x}_y = \boldsymbol{0}$ elsewhere. This method is also two-step superlinearly convergent. Perhaps more surprisingly, Byrd (1990) shows that the iterates $\boldsymbol{x} + \boldsymbol{Z} \boldsymbol{\Delta x}_z$ have in fact a one-step superlinear rate, and that such a rate is common for many SQP methods which involve the correction (2.7). Byrd and Nocedal (1991) show that these local results are not affected by global convergence concerns (see Section 2.1.2). Another possibility with the same theoretical convergence properties, proposed by Gilbert (1991), is to maintain a secant approximation to the inverse of $\boldsymbol{H}_{zz}$. A similar convergence rate is also achieved by methods which use Broyden-type secant methods to approximate the rectangular matrix $\boldsymbol{Z}(\boldsymbol{x})^T \boldsymbol{H} \left(\boldsymbol{Y}(\boldsymbol{x}) \quad \boldsymbol{Z}(\boldsymbol{x}) \right)$ (see, Nocedal and Overton, 1985, or Fontecilla, Steihaug and Tapia, 1987). Gurwitz (1994) prefers updates which treat the portions $\boldsymbol{H}_{zz}$ and $\boldsymbol{H}_{zy}$ separately while maintaining a positive definite approximation to the former. Coleman and Fenyes (1992) propose a similar method, and also a second method which additionally maintains an approximation to $\boldsymbol{H}_{yy}$. These methods appear to perform slightly better than those which merely maintain a nonzero $\boldsymbol{H}_{zz}$. Finally, an interesting new proposal by Biegler, Nocedal and Schmid (1995) notes that (2.6) does not actually require $\boldsymbol{H}_{zy}$ but rather the vector $\boldsymbol{H}_{zy} \boldsymbol{\Delta x}_y$. They thus propose to approximate this term directly by either finite differences or by a Broyden update.

And what of the Lagrange multiplier estimates? The values $\boldsymbol{\Delta y}$ from (2.4) satisfy

$$\boldsymbol{Y}(\boldsymbol{x})^T \boldsymbol{A}(\boldsymbol{x})^T \boldsymbol{\Delta y} = \boldsymbol{Y}^T \nabla_x \ell(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{H}_{zy}^T \boldsymbol{\Delta x}_z + \boldsymbol{H}_{yy} \boldsymbol{\Delta x}_y. \tag{2.8}$$

Clearly, setting $\boldsymbol{H}_{zy}$ and $\boldsymbol{H}_{yy}$ to zero imply that $\boldsymbol{y} + \boldsymbol{\Delta y}$ are least-squares multiplier estimates evaluated at $\boldsymbol{x}$. If included, these neglected terms would result in $\boldsymbol{y} + \boldsymbol{\Delta y}$ being approximations to least-squares multiplier estimates at $\boldsymbol{x} + \boldsymbol{\Delta x}$. Thus, rather than use these approximations, many authors prefer to use the current least-squares estimates

$$\boldsymbol{Y}(\boldsymbol{x}^+)^T \boldsymbol{A}(\boldsymbol{x}^+)^T \boldsymbol{y}^+ = \boldsymbol{Y}(\boldsymbol{x}^+)^T \nabla_x f(\boldsymbol{x}^+) \tag{2.9}$$

directly.

Finally, although we have argued that maintaining a positive definite approximation to the Hessian of the Lagrangian function is unnecessarily restrictive, it is more reasonable when approximating the Hessian of the augmented Lagrangian. Indeed, if ρ is a scalar, we can add the term $\rho||\boldsymbol{A}(x)\boldsymbol{\Delta x} + \boldsymbol{c}(\boldsymbol{x})||_2^2$ to the objective function of (2.3) without changing its solution. But, the Hessian of this modified problem is $\boldsymbol{H} + \rho\boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{A}(\boldsymbol{x})$ which can be expected to be positive definite for sufficiently large ρ. Such a method was first proposed by Tapia (1977) and suitable secant update formulae for $\boldsymbol{H} + \rho\boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{A}(\boldsymbol{x})$ are discussed by Byrd, Tapia and Zhang (1992).

2.1.2 Merit functions

The role of the merit function is to ensure convergence of the basic iteration (2.2) from arbitrary starting points. In unconstrained optimization, there is a natural merit function, the objective function. When there are constraints present, the conflicting goals of feasibility and optimality often preclude a natural choice, and most merit functions are attempts to balance these goals.

Early globally convergent SQP methods were based upon the l_1 exact penalty function

$$\psi_1(\boldsymbol{x}) = f(\boldsymbol{x}) + \rho||\boldsymbol{c}(\boldsymbol{x})||_1 \tag{2.10}$$

(see Pschenichny, 1970, Han, 1977, and Powell, 1978). So long as the penalty parameter ρ is sufficiently large, the iteration (2.2) converges globally with many of the Hessian approximations discussed in Section 2.1.1. However, despite its simplicity, the iteration has one serious drawback, namely that the merit function (2.10) (or indeed any function of the form $f(\boldsymbol{x}) + w(\boldsymbol{c}(\boldsymbol{x}))$ where $w \geq 0$ and $w(\boldsymbol{0}) = 0$) may prohibit the step $\alpha_x = 1$ arbitrarily close to a KKT point. Thus the promise of a fast asymptotic rate may be denied by the merit function. This defect was first observed by Maratos (1978). A number of remedies have been proposed, falling broadly into two camps: modify the search direction or change the merit function.

The Maratos "effect" arises when the curvature of the constraints is not adequately represented by the linearized model (2.3b). Recognizing this, Mayne and Polak (1982) propose adding a second-order correction to the standard SQP direction (2.4); Coleman and Conn (1982*a*) prefer to use the step (2.7) directly. Both techniques allow asymptotic unit steps and hence encourage superlinear convergence. A variation on this theme has also been suggested by Fukushima (1986).

A variety of alternatives to (2.10) have been considered. Perhaps the simplest suggestion is that by Chamberlain, Lemaréchal, Pedersen and Powell (1982) in which the requirement that $\psi_1(\boldsymbol{x}^+)$ be smaller than $\psi_1(\boldsymbol{x})$ every iteration is replaced by the requirement that this should happen at least once every $t > 1$ iterations. Remarkably, this is sufficient to ensure that, so long as a unit step is always attempted and provided the iterate is reset to the last "satisfactory" value if more than t iterations pass without a "satisfactory" reduction, a unit step will eventually be "satisfactory" at least every other iteration. A related recent proposal by Panier and Tits (1991) is to replace the linesearch requirement that $\psi_1(\boldsymbol{x}^+)$ be sufficiently smaller than $\psi_1(\boldsymbol{x})$ by the weaker requirement that the new value be smaller than $\max\{\psi_1(\boldsymbol{x}), \psi_1(\boldsymbol{x}^-), \psi_1(\boldsymbol{x}^=)\}$, where $\boldsymbol{x}^-$ and $\boldsymbol{x}^=$ are the previous two iterates. They show that such a strategy does not asymptotically prevent unit steps; the use of a non-monotonic linesearch is reminiscent of the overlooked procedure for unconstrained minimization by Grippo, Lampariello and Lucidi (1986).

More recently, Fletcher's (1970) differentiable exact penalty function

$$\psi_d(\boldsymbol{x}) = f(\boldsymbol{x}) - \boldsymbol{c}(\boldsymbol{x})^T\boldsymbol{y}(\boldsymbol{x}) + \rho||\boldsymbol{c}(\boldsymbol{x})||_2^2, \tag{2.11}$$

where

$$\boldsymbol{A}(\boldsymbol{x})\boldsymbol{A}(\boldsymbol{x})^T\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x})\nabla_x f(\boldsymbol{x}), \tag{2.12}$$

has been considered as a merit function by Powell and Yuan (1986). The main theoretical drawbacks with this function are the expense of computing its derivatives as well as the danger that the multiplier function $\boldsymbol{y}(\boldsymbol{x})$ is not uniquely defined whenever $\boldsymbol{A}(\boldsymbol{x})$ is less than full rank. To circumvent the former problem, Powell and Yuan (1986) show that it is possible to replace $\boldsymbol{y}(\boldsymbol{x}+\alpha_x\boldsymbol{\Delta x})$ in the linesearch by the interpolant $\boldsymbol{y}(\boldsymbol{x})+\alpha_x(\boldsymbol{y}(\boldsymbol{x}+\boldsymbol{\Delta x})-\boldsymbol{y}(\boldsymbol{x}))$ while ensuring global convergence at a superlinear rate. Fletcher (1973) prefers the variant

$$\psi_f(\boldsymbol{x}) = f(\boldsymbol{x}) - \boldsymbol{c}(\boldsymbol{x})^T\boldsymbol{y}(\boldsymbol{x}) + \rho\boldsymbol{c}(\boldsymbol{x})^T\left(\boldsymbol{A}(\boldsymbol{x})\boldsymbol{A}(\boldsymbol{x})^T\right)^{-1}\boldsymbol{c}(\boldsymbol{x}) \tag{2.13}$$

of (2.11), which he shows can be rewritten as

$$\psi_f(\boldsymbol{x}) = f(\boldsymbol{x}) - \boldsymbol{c}(\boldsymbol{x})^T\boldsymbol{\lambda}(\boldsymbol{x}) \tag{2.14}$$

where

$$\boldsymbol{\lambda}(\boldsymbol{x}) = \underset{\lambda}{\arg\min}\ \tfrac{1}{2}\|\boldsymbol{A}^T(\boldsymbol{x})\boldsymbol{\lambda} - \nabla_x f(\boldsymbol{x})\|_2^2 + \rho\boldsymbol{\lambda}^T\boldsymbol{c}(\boldsymbol{x}). \tag{2.15}$$

Because of the expense of inverting $\boldsymbol{A}(\boldsymbol{x}+\alpha_x\boldsymbol{\Delta x})\boldsymbol{A}(\boldsymbol{x}+\alpha_x\boldsymbol{\Delta x})^T$ at trial points, Boggs and Tolle (1989) propose using a variant of (2.13) in which this term is approximated in the linesearch by $\boldsymbol{A}(\boldsymbol{x})\boldsymbol{A}(\boldsymbol{x})^T$. Further generalizations of (2.11) are possible, and most are covered by the function

$$\psi_g(\boldsymbol{x}) = f(\boldsymbol{x}) - \boldsymbol{c}(\boldsymbol{x})^T\boldsymbol{u}(\boldsymbol{x}) + \rho\|\boldsymbol{c}(\boldsymbol{x})\|_2^2/a(\boldsymbol{x}), \tag{2.16}$$

where

$$\left(\boldsymbol{A}(\boldsymbol{x})\boldsymbol{A}(\boldsymbol{x})^T + \gamma\boldsymbol{I}\right)\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x})\nabla_x f(\boldsymbol{x}) \tag{2.17}$$

and $0 < a(\boldsymbol{x}) \leq \alpha$, for some $\alpha > 0$ and $\gamma \geq 0$. Facchinei and Lucidi (1994) show that this merit function does not impede the quadratic convergence of SQP methods for which the exact derivatives $\boldsymbol{H} = \nabla_{xx}\ell(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{x}))$ are used.

One issue we have not considered so far is that the linearized constraints (2.3b) may be inconsistent. Fletcher (1981, Section 14.4) provides a useful alternative in which the model problem (2.3) is replaced by the problem of minimizing

$$m_1(\boldsymbol{\Delta x}) \stackrel{\text{def}}{=} \tfrac{1}{2}\boldsymbol{\Delta x}^T\boldsymbol{H}\boldsymbol{\Delta x} + \boldsymbol{\Delta x}^T\nabla_x f(\boldsymbol{x}) + \rho\|\boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Delta x}\|_1 \tag{2.18}$$

of the merit function (2.10). The problem of minimizing (2.18) may be reformulated as a quadratic programming problem, and has the desirable property that the subproblem is always consistent (this is also implicit in the algorithm of Coleman and Conn, 1982*b*). Nonetheless, Fletcher (1982) and Yuan (1985*a*) observe that the Maratos "effect" may still occur if the search direction is computed by minimizing (2.18) but can be prevented if a second-order correction of the form (2.7) is made. Fast local convergence properties of such a method are examined by Womersley (1985) and Yuan (1985*b*), while Wright (1987, 1989*b*) shows that it is not necessary to minimize (2.18) to full accuracy to achieve fast convergence.

2.2 Trust region methods

Linesearch methods aim *a posteriori* to control a bad choice of step $\boldsymbol{\Delta x}$ by comparing $\psi(\boldsymbol{x}+\alpha\boldsymbol{\Delta x})$ with $\psi(\boldsymbol{x})$. Trust-region methods, on the other hand, aim *a priori* to ensure that the step is adequate by imposing extra restrictions on the model from which the step is derived. The simplest example would be to consider the model problem (2.3) but to require additionally that

$$\|\boldsymbol{\Delta x}\| \leq \Delta \tag{2.19}$$

for some scalar $\Delta > 0$. The extra constraint (2.19) is known as the *Trust-region* constraint and the scalar Δ is the trust-region *radius*. The size of the radius is controlled by comparing the actual reduction in the merit function when the step is taken with the value predicted by a model of this function for which $\boldsymbol{\Delta x}$ is a good step. Normally Δ will be increased if there is good agreement and the trust-region constraint is active, and decreased when the agreement is poor. The introduction of a trust region allows considerable extra freedom when specifying $\boldsymbol{H}$ as (2.19) stops inappropriate choices of $\boldsymbol{H}$ leading to unbounded steps.

Fletcher (1982) includes a trust-region constraint when minimizing the model function (2.18) of (2.10). It is particularly convenient in this case to choose the infinity norm for (2.19) as the resulting model problem may then still be posed as a quadratic program. There are a number of problems however if we try to impose a trust region on (2.3).

Firstly, the linear constraints (2.3b) and the trust-region (2.19) may have no common feasible point. A number of attempts have been made to overcome this defect. Vardi (1985) and Byrd, Schnabel and Schultz (1987) suggests replacing (2.3b) by constraints of the form

$$\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Delta x} + \theta \boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{0}, \tag{2.20}$$

where $\theta \in (0, 1]$ is chosen so that the new constraints and the trust region have a common feasible point. (A similar device was proposed by Powell, 1978, to handle inconsistent constraints in the basic subproblem (2.3).) Another possibility is to replace (2.3b) by

$$\|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Delta x} + \boldsymbol{c}(\boldsymbol{x})\| \leq \theta, \tag{2.21}$$

where θ is chosen so that the intersection of (2.19) and (2.21) has a solution. (Once again, a similar device was proposed by Burke and Han (1989) to handle inconsistent constraints in the basic subproblem (2.3).) Celis, Dennis and Tapia (1985) choose

$$\min_{\|d\| \leq \Delta} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{d} + \boldsymbol{c}(\boldsymbol{x})\| \leq \theta \leq \|\boldsymbol{c}(\boldsymbol{x})\|, \tag{2.22}$$

while Powell and Yuan (1990) prefer

$$\min_{\|d\| \leq \beta_1 \Delta} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{d} + \boldsymbol{c}(\boldsymbol{x})\| \leq \theta \leq \min_{\|d\| \leq \beta_2 \Delta} \|\boldsymbol{A}(\boldsymbol{x})\boldsymbol{d} + \boldsymbol{c}(\boldsymbol{x})\|, \tag{2.23}$$

where $0 < \beta_2 \leq \beta_1 < 1$. Both sets of authors suggest particular choices of θ which satisfy their restrictions when the two-norm is used. It is important to note that it often suffices to obtain an approximate solution to the given model problem, and a general theory which covers this possibility is given by Dennis, El-Alem and Maciel (1992). Burke (1992) studies methods of this sort in a very general setting, and allows for the possibility that the original problem may be infeasible by showing convergence to a "nearest" infeasible KKT point.

Secondly it is not clear what model of the merit function should be used. The most common approach is to model the merit function by taking first or second-order approximations of constituent terms. Both El-Alem (1995), for the subproblem based on (2.20), and El-Alem (1991) and Powell and Yuan (1990), for that based on (2.21), use the merit function (2.11) and model this by

$$\begin{aligned} \psi_d(\boldsymbol{x}) + \boldsymbol{\Delta x}^T \nabla_x \ell(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x})) - (\boldsymbol{y}(\boldsymbol{x} + \boldsymbol{\Delta x}) - \boldsymbol{y}(\boldsymbol{x}))^T \left(\boldsymbol{c}(\boldsymbol{x}) + \tfrac{1}{2}\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Delta x}\right) \\ + \tfrac{1}{2}\boldsymbol{\Delta x}^T \boldsymbol{H} \boldsymbol{\Delta x}_n + \rho \left(\|\boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Delta x}\|_2^2 - \|\boldsymbol{c}(\boldsymbol{x})\|_2^2\right), \end{aligned} \tag{2.24}$$

where $\boldsymbol{\Delta x}_n$ is the orthogonal projection of $\boldsymbol{\Delta x}$ into the null-space of $\boldsymbol{A}(\boldsymbol{x})$. The authors provide schemes for automatically adjusting the penalty parameter ρ and establish global and locally superlinear convergence.

Thirdly, even when the constraints are consistent, the resulting subproblem may not be easy to solve. When the infinity norm is used, the resulting subproblem is inevitably a quadratic program. When the two-norm is used, Yuan (1990) gives an algorithm for solving the subproblem involving (2.19) and (2.21), that is of minimizing a quadratic function in a region defined by the intersection of two balls. Zhang (1992) simplifies this scheme in the case where $\boldsymbol{H}_{zz}$ is positive semi-definite, while Heinkenschloss (1994) does the same in the general case. Finally, recent work by Moré (1993) and Stern and Wolkowicz (1995) has generalized this to the case where (2.19) is replaced by the condition $\Delta_1 \leq \boldsymbol{\Delta x}^T \boldsymbol{C} \boldsymbol{\Delta x} \leq \Delta_2$ and where $\boldsymbol{C}$ may be indefinite.

When a reduced Hessian method is used, the obvious trust-region generalization is to choose the step $\boldsymbol{\Delta x}_z$ by solving a subproblem of the form

$$\underset{\|\Delta x_z\| \leq \Delta}{\text{minimize}} \quad \tfrac{1}{2} \boldsymbol{\Delta x}_z^T \boldsymbol{H}_{zz} \boldsymbol{\Delta x}_z + \boldsymbol{\Delta x}_z^T \boldsymbol{Z}^T \nabla_x \ell(\boldsymbol{x}, \boldsymbol{y}). \tag{2.25}$$

Such a scheme is proposed by Zhang and Zhu (1990), while methods which allow approximate solutions of (2.25) are considered by Zhang, Zhu and Fan (1993).

2.3 Methods for inequality constraints

Nonlinear programming problems rarely exclusively involve equality constraints, but typically involve a mixture of equations and inequalities. We thus turn to the inequality constrained problem. The Karush-Kuhn-Tucker (first-order optimality) conditions for the inequality problem

$$\underset{x \in \Re^n}{\text{minimize}} \quad f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{c}(\boldsymbol{x}) \geq \boldsymbol{0} \tag{2.26}$$

are that

$$\nabla_x \ell(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{0}, \;\; \boldsymbol{c}(\boldsymbol{x}) \geq \boldsymbol{0}, \;\; \boldsymbol{\lambda} \geq \boldsymbol{0}, \quad \text{and} \quad \boldsymbol{c}(\boldsymbol{x})^T \boldsymbol{\lambda} = 0, \tag{2.27}$$

where the Lagrangian function $\ell(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \boldsymbol{c}(\boldsymbol{x})^T \boldsymbol{\lambda}$.

At the time of the last conference, algorithms for (2.26) were primarily of the active-set variety. An *active set* method is a method which aims to solve (2.26) by predicting which of the inequalities will be active (ie, which of the $c_i(\boldsymbol{x}) = 0$) and which are inactive (ie, $c_i(\boldsymbol{x}) > 0$) at the solution. Once these sets are known, the problem can be solved as if it involves only equality constraints, namely those deemed to be active at the solution. The main justification, therefore for much of the work described in Section 2.1 is as a tool for analyzing active-set methods.

The principal differences between active set methods is in the way that the active set is assigned. In inequality-(constrained) quadratic programming (IQP) methods, no *a priori* choice of the active set is made when choosing the correction $\boldsymbol{\Delta x}$, rather $\boldsymbol{\Delta x}$ is obtained by solving the quadratic programming problem of minimizing (2.3a) subject to a linear approximation $\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Delta x} + \boldsymbol{c}(\boldsymbol{x}) \geq \boldsymbol{0}$ of all of the constraints. The active set for this problem is taken as a prediction of that for (2.26). Robinson (1974) provides theoretical justification for such an approach. In equality-(constrained) quadratic programming (EQP) methods, the active set is assigned prior to the selection of $\boldsymbol{\Delta x}$ (primarily on the basis of inequalities which are close to being active and whose Lagrange multiplier estimates are positive) and $\boldsymbol{\Delta x}$ is found directly by solving (2.3a) subject to the linear equality constraints $\boldsymbol{A}(\boldsymbol{x})_{\mathcal{A}} \boldsymbol{\Delta x} + \boldsymbol{c}_{\mathcal{A}}(\boldsymbol{x}) = \boldsymbol{0}$, where the subscript $\mathcal{A}$ denotes those constraints which are considered to be active. Which strategy is preferable is a matter for some debate (see, for example, Murray and Wright, 1982).

The merit function must also account for inequality constraints. Normally, this is merely a matter of replacing the term(s) which handle equality constraints with a similar term for the inequalities. For instance, the analog of (2.10) for the problem (2.26) is the function

$$f(\boldsymbol{x}) + \rho \|\boldsymbol{c}(\boldsymbol{x})_-\|_1, \tag{2.28}$$

where $\boldsymbol{c}_-$ gives, componentwise, the smaller of c_i and zero. Pantoja and Mayne (1991) and Heinz and Spellucci (1994) prefer the infinity to the one-norm in (2.28), and propose a line-search method based on the model problem

$$\underset{\Delta x \in \Re^n}{\text{minimize}} \ \tfrac{1}{2}\boldsymbol{\Delta x}^T \boldsymbol{H} \boldsymbol{\Delta x} + \boldsymbol{\Delta x}^T \nabla_x f(\boldsymbol{x}) + \rho \|(\boldsymbol{c}(\boldsymbol{x}) + \boldsymbol{A}(\boldsymbol{x})\boldsymbol{\Delta x})_-\|_\infty. \tag{2.29}$$

Yuan (1995) analyses a similar method in which a trust region (2.19) is imposed on the model (2.29).

Generalizations of the functions (2.11) or (2.13) are more interesting. Fletcher (1973) gives a simple generalization of the function (2.13) in the inequality case by replacing (2.15) with

$$\boldsymbol{\lambda}(\boldsymbol{x}) = \underset{\lambda \geq 0}{\arg\min} \ \tfrac{1}{2}\|\boldsymbol{A}^T(\boldsymbol{x})\boldsymbol{\lambda} - \nabla_x f(\boldsymbol{x})\|_2^2 + \rho \boldsymbol{\lambda}^T \boldsymbol{c}(\boldsymbol{x}). \tag{2.30}$$

Boggs, Tolle and Kearsley (1991) prefer to introduce extra slack variables $\boldsymbol{s}$ in order to replace the inequality constraints with the equations $c_i(\boldsymbol{x}) - \frac{1}{4}s_i^2 = 0$. Introducing these variables into (2.13) and setting $z_i = \frac{1}{4}s_i^2 \geq 0$, they propose the merit function

$$f(\boldsymbol{x}) - (\boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{z})^T \boldsymbol{\pi}(\boldsymbol{x}) + \rho(\boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{z})^T \left(\boldsymbol{A}(\boldsymbol{x})\boldsymbol{A}(\boldsymbol{x})^T + \boldsymbol{Z}\right)^{-1} (\boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{z}), \tag{2.31}$$

where $\boldsymbol{Z}$ is the diagonal matrix with entries z_i and

$$(\boldsymbol{A}(\boldsymbol{x})\boldsymbol{A}(\boldsymbol{x})^T + \boldsymbol{Z})\boldsymbol{\pi}(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x})\nabla_x f(\boldsymbol{x}). \tag{2.32}$$

It is straightforward to design updates for $\boldsymbol{z}$ which ensure that $\boldsymbol{z} \geq 0$. A version of (2.16) appropriate for inequality constraints is given by Di Pillo, Facchinei and Grippo (1992).

2.4 Difficulties

Having surveyed the main developments in SQP methods, we now consider the difficulties with the approach. Many of the difficulties are directly attributable to a lack of coherence between the step calculation and the merit function. In unconstrained optimization, there is a direct relationship between the merit function — in this case, invariably the objective function — and the calculation of the step. The prototypical method, Newton's method, may be viewed both as a method which attempts to satisfy the first-order optimality conditions *and* as a method which aims to reduce the merit function through a Taylor's series approximation. In the constrained case, while the SQP direction (2.3) may be viewed as an attempt to satisfy the KKT conditions, it is not directly related to any of the merit functions that have been proposed, although it does often provide a descent direction for them. Perhaps the only satisfactory methods from the point of consistency are those which directly attempt to link the merit function and the step. Fletcher's (1982) and Coleman and Conn's (1982*b*) methods based on (2.10) and (2.18) and their generalizations are in this class, but, as we have noted, even these have disadvantages.

A second drawback is that so few of the suggestions we have considered are appropriate if the number of variables is large. In particular, unless function values are expensive, the dominant cost of the methods tends to be in solving linear systems; for inequality problems this may be particularly acute as each subproblem may require the solution of a sequence of such systems. If n is large, there is little hope unless either the required systems are small or sparse. There are two important cases where this is so. Firstly, if the number of equality or active constraints is close to n, reduced Hessian methods, such as those proposed by Coleman and Conn (1982*a*), Gilbert (1991), and Biegler et al. (1995), which maintain the matrix $\boldsymbol{H}_{zz}$ but ignore $\boldsymbol{H}_{zy}$ and $\boldsymbol{H}_{yy}$, may be successful. The only systems which need to be solved involve $\boldsymbol{H}_{zz}$ (small) and $\boldsymbol{A}(\boldsymbol{x})\boldsymbol{Y}(\boldsymbol{x})$ and its transpose (sparse, we hope). Secondly, if the matrix $\boldsymbol{H}$ is sparse, sparse methods for linear systems (when an EQP method is used) or quadratic programming (for IQP methods) may be employed. This will often be the case if $\boldsymbol{H}$ is chosen as the Hessian of the Lagrangian function, or from a structured or sparse secant updating formula (see, for example, Toint, 1977, Conn, Gould and Toint, 1990, and Fletcher, 1995).

Finally, one of the main advances in methods for the unconstrained minimization of large problems was the recognition that a Newton-like direction need not computed very accurately when far from a stationary point (see Dembo, Eisenstat and Steihaug, 1982). Clearly, when equality constraints are present a similar result would be valuable, but there has been remarkably little work on this topic. For equality constrained problems involving relatively few constraints, Fontecilla (1990) notes that (2.5) and (2.9) are small systems, and the only large system is (2.6). He thus proposes a method which initially solves (2.6) to low accuracy. Alas, for fast asymptotic convergence (2.6) must eventually be solved to high accuracy which limits the effectiveness of this proposal. When the constraints are inequalities and an active-set IQP method is used, Murray and Prieto (1995) show that it is possible to stop the solution of the QP subproblem at the first stationary point encountered rather than solving the problem to completion.

3 Optimality-condition based methods

Since the mid 1980s there has been a revolution in the way in which the optimality conditions (2.27) have been viewed. In essence, active set methods ultimately aim to satisfy the dual feasibility requirement $\nabla_x \ell(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{0}$, while ensuring that the remaining feasibility requirements, $\boldsymbol{c}(\boldsymbol{x}) \geq \boldsymbol{0}$ and $\boldsymbol{\lambda} \geq \boldsymbol{0}$, and complementarity condition, $\boldsymbol{c}(\boldsymbol{x})^T \boldsymbol{\lambda} = 0$, are always (effectively) satisfied. This is achieved by the simple combinatorial expedient of ensuring that, for each constraint, either $c_i(\boldsymbol{x}) = 0$ or $\lambda_i = 0$, but suggests that in the worst case, all combinations may be examined. By contrast, the newer interior-point methods try for optimality by ensuring that the feasibility requirements are always (effectively) satisfied while aiming ultimately to satisfy the complementarity condition. If we consider, for a moment, the case where (2.26) is a linear or quadratic program, the feasibility requirements are linear while the complementarity condition is nonlinear (quadratic). Thus the active set methods may be viewed as trying to hide this nonlinearity within a combinatorial problem, while the interior-point methods confront the nonlinearity directly. Significantly, the interior-point approach has been shown to have strong complexity advantages on many classes of convex problems (see Nesterov and Nemirovskii, 1994), and there is some evidence that these advantages transfer to improved practical performance (see, for instance, Jarre and Saunders, 1995). Gill, Murray, Saunders, Tomlin and Wright (1986) were quick to make the connection between the new-looking interior-point methods and the long-discarded barrier-function methods, and thus lead the community to re-examine barrier methods in this new light. Clearly, IQP based methods are already able to take advantage of interior-point technology as the quadratic subproblems are often convex. It remains to be seen whether the same is true of EQP based methods.

While barrier function and other interior-point methods are discussed elsewhere in this volume, we briefly mention a couple of other possibilities. Friedlander, Martínez and Santos (1994) propose a method for solving linearly-constrained problems by minimizing

$$\|\nabla_x \ell(\boldsymbol{x}, \boldsymbol{\lambda})\|_2^2 + (\boldsymbol{c}(\boldsymbol{x})^T \boldsymbol{\lambda})^2$$

over the feasible set $\boldsymbol{c}(\boldsymbol{x}) \geq \boldsymbol{0}$. They show that this formulation does not introduce unnecessary local minimizers when f is convex.

Kanzow and Kleinmichel (1995) observe that the optimality conditions (2.27) may be replaced by the nonlinear system

$$\nabla_x \ell(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{0}, \quad \text{and} \quad \phi(c_i(\boldsymbol{x}), \lambda_i) = 0, \tag{3.1}$$

where ϕ is any function for which

$$\phi(u, v) = 0 \quad \text{if and only if} \quad u \geq 0, \;\; v \geq 0 \quad \text{and} \quad uv = 0. \tag{3.2}$$

A simple example is the function $\phi(u, v) = \sqrt{u^2 + v^2} - u - v$. Their proposal is now to apply Newton's method to (3.1), and they provide a local analysis of such a method. However, the Jacobian of such a system will be singular whenever the solution is degenerate. Pang (1991, 1994) prefers the choice $\phi(u, v) = \min\{u, v\}$, and shows that this choice may be made the basis of a globally convergent method using the merit function

$$f(\boldsymbol{x}) + \mu \left(\|\nabla_x \ell(\boldsymbol{x}, \boldsymbol{\lambda})\|_2^2 + \|\phi(c_i(\boldsymbol{x}), \lambda_i)\|_2^2 \right).$$

4 Other methods for nonlinear constraints

Now that barrier functions are back in fashion, it is worth evaluating the status of other methods which SQP algorithms were supposed to have succeeded. One of the earliest methods for solving equality constrained problems was to minimize the quadratic penalty function

$$\psi_2(\boldsymbol{x}) = f(\boldsymbol{x}) + \rho \|\boldsymbol{c}(\boldsymbol{x})\|_2^2, \tag{4.1}$$

for a sequence of scalars ρ approaching infinity. Although this method was dismissed in the 1970s, it has been seen in a more favourable light since then. Firstly, perceived difficulties with ill-conditioning were shown to be benign provided sufficient care is taken (Broyden and Attia, 1984, Gould, 1986, Coleman and Hempel, 1990). Secondly, the requirement that (4.1) be minimized is easily relaxed. Moreover, Gould (1989) shows that asymptotically at most two Newton-like steps are required for each value of ρ and that this results in a globally and (two-step) superlinearly convergent method. This result is generalized by Dussault (1995) to problems involving both equations and inequalities.

These methods and the succeeding augmented Lagrangian methods have three significant advantages over most SQP methods. Firstly, the choice of the correction $\boldsymbol{\Delta x}$ is normally intimately connected to the merit function — it is usually obtained by minimizing a second-order model of the function. Secondly, the second derivative matrices of these functions can normally be expected to be positive (semi-)definite in some neighbourhood of the solution, and thus it is reasonable to approximate these derivatives via positive-definite secant formulae. Thirdly, the Newton-like systems which arise are usually easier to handle when n is large and the problem sparse than for SQP methods — this is a consequence of the matrices being (relatively) sparse and definite. It is this third point which may explain why many of the currently most successful codes for large-scale nonlinear programming are based on these methods (see Section 7). However, these methods do have some drawbacks, namely that a sequence of problems has to be solved, the iterates will often initially move away from the solution, and, significantly, no advantage is taken of linear or other simple constraints.

Augmented Lagrangian methods solve the equality constrained problem by minimizing a sequence of problems of the form

$$\psi_a(\boldsymbol{x}) = f(\boldsymbol{x}) - \boldsymbol{c}(\boldsymbol{x})^T \boldsymbol{y} + \rho \|\boldsymbol{c}(\boldsymbol{x})\|_2^2, \tag{4.2}$$

where $\boldsymbol{y}$ are estimates of the Lagrange multipliers and ρ a positive penalty parameter (see, Hestenes, 1969, and Powell, 1970). Convergence is assured by adjusting $\boldsymbol{y}$ and ρ, and it is not necessary for ρ to approach infinity. When first-order multiplier updates are used, the minimizers of (4.2) converge linearly, the rate being proportional to $1/\rho$. Faster rates are possible using higher-order multiplier updates, but first-order updates are convenient and effective for large-scale problems. Bartholomew-Biggs (1987) notes that the Newton equations for (4.2) may be reformulated as a quadratic programming problem, and this problem is a perturbation of (2.3). The method is easily generalized for inequality constraints (see, Rockafellar, 1974).

Realistic nonlinear programming problems often involve a mixture of linear and nonlinear constraints. Conn, Gould and Toint (1991) consider the case where there are simple bounds $\boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}$ on the variables in addition to equality constraints. Their algorithm finds a sequence of minimizers of (4.2) where the simple bounds are explicitly enforced and only approximate minimizations are performed. The convergence results obtained match that for the case without simple bounds. Inequality constraints are handled by introducing slack variables, but Conn, Gould and Toint (1994) show that these slack variables need not affect the linear algebra costs. More significantly, Conn, Gould and Toint (1992*b*) show that a single Newton-like step eventually suffices in the approximate minimization, and thus the iterates are globally convergent at a (fast) linear rate. This strategy is generalized by Conn, Gould, Sartenaer and Toint (1995) for the case where the constraint set is a mixture of equality and linear inequality constraints; a sequence of approximate minimizers of (4.2) subject to the linear constraints are sought. This theory also allows for independent penalty parameters for each of the penalized constraints.

Another class of important methods, the direct ancestors of SQP methods, are the sequential *linear* programming (SLP) methods (see Griffith and Stewart, 1961) in which linear approximations are taken of both objective and constraint functions. The resulting linear program may then be solved using either simplex or interior point methods. Modern versions are based on linear approximations of the merit function (2.10) (see, for example, Zhang, Kim and Lasdon, 1985), while Fletcher and Sainz de la Maza (1989) propose a hybrid method which tries a Newton-like (Coleman and Conn, 1982*a*) step for (2.10) but falls back on the SLP correction whenever the Newton step is unsuccessful. Clearly, unless some form of second-order acceleration is used the convergence of these methods will typically be rather slow, but the mature state of linear programming algorithms means that the subproblems can be solved efficiently.

At the other extreme, Maany (1987) proposes a method in which quadratic approximations are taken of both objective and constraints. Such an approach is appropriate for highly-curved constraints, but has the disadvantage that the subproblems are hard to solve.

Finally, an interesting class of feasible-point SQP methods have been developed by Panier and Tits (1987, 1993) and Bonnans, Panier, Tits and Zhou (1992) for inequality constrained problems. As the iterates are feasible, the objective function may be used as a merit function. The methods require the solution of two linear or quadratic programming problems at each iteration to generate corrections; a backtracking linesearch is performed along a quadratic arc defined by these directions, and the first step which sufficiently reduces f and satisfies the constraints is accepted. The methods are shown to be globally and two-step superlinearly convergent.

5 Linear and convex constraints

We have seen that most methods for nonlinear constraints solve a sequence of subproblems involving simpler constraints. For instance, SLP and SQP methods solve linear and quadratic programs, while other methods differentiate between "difficult" constraints (normally the nonlinear ones) and "easy" ones (normally the linear or convex ones), treating the easy constraints directly in the subproblem. In this section, we consider methods specifically designed for problems involving linear or convex constraints; we exclude a discussion of interior-point methods as they are covered elsewhere in this volume.

5.1 Simple bounds, gradient projection and projected gradients

Two classes of problems have attracted attention here, those for which the objective function is a quadratic and those with a general objective. Standard active-set methods for linearly constrained optimization problems typically refine the active set slowly, perhaps one constraint leaving or entering the active set at each iteration. When the constraint set only involves simple bounds, however, it is far easier to add or delete many constraints at each iteration, and the best mechanism for achieving this is the gradient projection algorithm.

The *gradient projection* algorithm (Levitin and Polyak, 1966) simply chooses iterates according to

$$\boldsymbol{x}^{+} = P_{\Omega}[\boldsymbol{x} - \alpha_x \nabla_x f(\boldsymbol{x})], \tag{5.1}$$

where Ω is the set of feasible points, $P_{\Omega}[\boldsymbol{v}]$ is the projection of $\boldsymbol{v}$ into Ω and α_x is a suitable stepsize (see, for instance, Bertsekas, 1976, or Dunn, 1981). When the constraints are simple bounds, $\Omega = \{\boldsymbol{x} : \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}\}$, and the projection is easily computed as $P_{\Omega}[\boldsymbol{v}] = \text{mid}(\boldsymbol{l}, \boldsymbol{v}, \boldsymbol{u})$, where mid denotes the vector whose components are the medians of l_i, v_i and u_i. However, as the gradient projection algorithm is just a constrained variant of the method of steepest descent, it is clear that some form of acceleration is needed if the method is to be practical.

Moré and Toraldo (1991) propose that the active sets for consecutive iterates be compared, and if the sets are identical, or if little progress is being made on the "face" defined by the current set, the current face should be explored using a higher-order method. To this end, the authors propose that the conjugate-gradient method should be used to find an approximate stationary point $\boldsymbol{x}^c$ on the face, and then a linesearch performed on the piecewise-linear arc $P_{\Omega}[\boldsymbol{x} + \alpha_x(\boldsymbol{x}^c - \boldsymbol{x})]$.

The justification for such a scheme is that the gradient projection algorithm is guaranteed to determine the optimal active set for nondegenerate problems in a finite number of iterations (see Bertsekas, 1976). Calamai and Moré (1987) improve this considerably by noticing that the result is true for *any* algorithm for which the projected gradient converges to zero; the *projected gradient* is

$$\nabla_{\Omega} f(\boldsymbol{x}) = P_{T(x)}[-\nabla_x f(\boldsymbol{x})], \tag{5.2}$$

where the tangent cone $T(\boldsymbol{x})$ is the closure of the cone of all feasible directions at $\boldsymbol{x}$. While the projected gradient may be difficult to calculate in general, for simple bounds it is given (componentwise) by

$$\nabla_{\Omega} f(\boldsymbol{x})_i = \begin{cases} \min\{\nabla_x f(\boldsymbol{x})_i, 0\} & \text{if} \quad x_i = l_i \\ \nabla_x f(\boldsymbol{x})_i & \text{if} \quad l_i < x_i < u_i \\ \max\{\nabla_x f(\boldsymbol{x})_i, 0\} & \text{if} \quad x_i = u_i \end{cases} \tag{5.3}$$

This result has subsequently been generalized. Dunn (1987) and Burke and Moré (1988) relax the non-degeneracy assumption to one of requiring that there is a set of strictly complementary Lagrange multipliers while Burke and Moré (1994) obtain similar results in a more general geometric setting. De Angelis and Toraldo (1993) show that the simplified gradient projection scheme of Dem'yanov and Rubinov (1970), in which (5.1) is replaced by an iteration of the form $\boldsymbol{x}^{+} = \boldsymbol{x} + \alpha_x(P_{\Omega}[\boldsymbol{x} - \eta \nabla_x f(\boldsymbol{x})] - \boldsymbol{x})$ for some η, inherits the active-constraint identification property of its predecessor. Burke (1990) generalizes the previous analysis for nonconvex problems, and shows that projected gradient of a suitable linearization of the problem correctly identifies the optimal active set. This result, and a similar analysis by Wright (1989*a*), suggests why EQP strategies for SQP and linearizations of (2.28) are successful.

Moré and Toraldo (1991) show that their algorithm for quadratic objectives is necessarily finite under a nondegeneracy assumption. Friedlander and Martínez (1994) provide a mechanism for leaving an unpromising face and ensuring that their algorithm will not return to the face unless a substantial improvement in the value of the objective is possible.

An interesting observation drives the method of Coleman and Hulbert (1993). They note that the optimality conditions may be expressed as

$$\nabla_x f(\boldsymbol{x}) \cdot (\boldsymbol{x} - \tfrac{1}{2}(\boldsymbol{l} + \boldsymbol{u}) + \tfrac{1}{2}(\boldsymbol{u} - \boldsymbol{l}) \cdot \text{ sign } (\nabla_x f(\boldsymbol{x}))) = \boldsymbol{0}, \tag{5.4}$$

where $\text{sign}(\boldsymbol{v})_i$ is -1 if $v_i \leq 0$ and 1 otherwise, and $\boldsymbol{v} \cdot \boldsymbol{w}$ is the vector whose components are $v_i w_i$. They then derive a Newton-like correction for the nonlinear system (5.4), while recognizing that the function is non-differentiable whenever a component of $\nabla_x f(\boldsymbol{x}))$ is zero. A related merit function is provided, and they are able to show global and superlinear convergence under a suitable nondegeneracy assumption.

For general objective functions, Conn, Gould and Toint (1988) provide a class of trust-region methods in which a quadratic model is minimized within a region defined by the intersection of the simple bounds and a trust region. An approximate solution of the model problem is found, of which only a local minimizer of the model along the arc

$$\boldsymbol{x}(\alpha) = P_{\Omega \cap \{y : \|y - x\| \leq \Delta\}}[\boldsymbol{x} - \alpha \nabla_x f(\boldsymbol{x})] \tag{5.5}$$

is required for global convergence. Convergence is accelerated by continuing the model minimization using conjugate gradients in the face determined by the solution to (5.5). As before, so long as the problem is non-degenerate, the active set at the solution is identified by that of (5.5) after a finite number of iterations, and thus the speed of convergence is determined by the accuracy required in the conjugate-gradient step. Lescrenier (1991) shows that the non-degeneracy assumption is unnecessary. Byrd, Lu, Nocedal and Zhu (1995) give a line-search variant of Conn et al.'s (1988) algorithm, using an efficient limited-memory Hessian approximation. Finally, Toint (1988) shows that the frameworks of Calamai and Moré (1987) and Conn et al. (1988) may be extended to cover infinite-dimensional problems in a general Hilbert-space setting.

5.2 General linear and convex constraints

One of the earliest active set methods for linearly constrained minimization was Rosen's (1960) gradient projection algorithm (not to be confused with the method (5.1) of the same name). At each stage of this prototypical feasible-point active-set method, a step is taken in the direction of the gradient projected orthogonally into the null-space of those constraints which are currently considered active. When this projected gradient is small compared to the largest of the least-squares multiplier estimate (2.9), the step may leave the current face. Although the method was succeeded long ago by SQP and other methods which incorporate curvature, remarkably the method has only recently been shown to be globally convergent (Du and Zhang, 1986, 1989).

Many active-set methods have been proposed for convex quadratic programming, most of which differ in their linear algebra requirements rather than in the iterates they generate (Best, 1984). There have been relatively fewer methods for non-convex problems. Fletcher (1987) assessed the state of the art ten years ago. The only real advances since then, aside from those with interior-point methods, are for large-scale problems. Gill, Murray, Saunders and Wright (1990, 1991) and Gould (1991) consider versions of Fletcher's (1971) method for non-convex problems which are able to exploit sparsity. Boggs, Domich and Rogers (1995) suggest investigating a low-dimensional subspace of "interesting" directions at each iteration. This reduction in dimension means that the subproblems are small and can thus be tackled with any of the good algorithms for small problems.

Concerns over the past decade for general linearly constrained problems have focused on how to cope with problems which involve considerably more inequality constraints than variables. Powell (1989), wishing to avoid the increase in dimensionality which would occur if slack variables were added to the inequalities, prefers to treat the inequalities directly. He proposes that constraints which are "close" to active are not allowed to approach the constraint boundaries until absolutely necessary, and this allows larger steps to be taken in early iterations than would be possible in a conventional active set method. The state of the art for large-scale linearly constrained minimization has changed very little, the method proposed by Murtagh and Saunders (1978) still being pre-eminent.

One of the main difficulties which arises in active-set methods for linearly constrained optimization may occur when the constraint gradients in the active set at $\boldsymbol{x}$ are linearly dependent. The correction $\boldsymbol{\Delta x}$ will normally be chosen orthogonal to a linearly independent subset of the the active constraints, but in the degenerate case a nonzero step along $\boldsymbol{\Delta x}$ may not be possible because of the remaining active constraints. In the worst case this can cause an algorithm to cycle infinitely through subsets of the active constraints. Thus finding a $\boldsymbol{\Delta x}$ which is capable of moving away from $\boldsymbol{x}$ is crucial, especially as degenerate active sets are extremely common in practice. Of late, attention has focused on methods which are capable of dealing with this situation when floating-point computations are performed, and the methods of Fletcher (1988), Ryan and Osborne (1988), Dax (1989) and Gill, Murray, Saunders and Wright (1989) have all proved effective in practice.

Burke, Moré and Toraldo (1990) and Conn, Gould, Sartenaer and Toint (1993) give a general theory of trust-region methods for problems involving convex constraints. As solving the trust-region problem may now be an expensive calculation, Conn et al. (1993) show that an approximate solution suffices, and provide an algorithm which delivers such an approximation. Sartenaer (1995) indicates that this approach is effective in the case of network constraints. Martinez and Santos (1995) extend this class of methods to handle general, nonconvex domains.

6 Other topics

In this short survey, we naturally have to be selective on what to include and what to leave out. One trend that has been noticeable over the past ten years has been the increasing cross-fertilization between nonlinear optimization and other branches of numerical analysis and applied mathematics. In this section, we briefly mention a few topics we feel deserve more attention from the numerical analysis community.

The promise of automatic differentiation, that is the automatic accumulation of derivatives directly from codes which provide function values (see Griewank, 1989), has been a long time coming. Dixon (1991) argues that automatic differentiation will revitalize second derivative methods, as there is then little reason to rely on secant approximations. This has profound implications as, for example, many SQP methods depend upon properties which do not hold for exact derivatives (see Section 2.1.1). The main drawback in the past was the lack of efficient software for automatic differentiation, but that has dramatically changed over the past five years (see the WWW page http://www.mcs.anl.gov/Projects/autodiff /AD_Tools for details of currently available packages). A number of optimization packages now make direct use of automatic derivatives.

A fundamental issue when using optimization algorithms is how important it is that function and derivative information is accurate. Kupferschmid and Ecker (1987) compare the Ellipsoid and SQP algorithms when function values and gradients are inaccurate, and observe that the former is much less susceptible to inaccuracies. Toint (1988) and Carter (1991) examine the convergence of trust-region methods when the gradient is inaccurate, while Conn et al. (1993) do the same for inaccurate function and gradient values.

An important class of problems not covered above are those for which derivatives may not exist but for which it is possible to calculate subgradients. These non-differentiable problems are frequently solved by methods which build local piecewise affine models of the functions concerned. For instance, each convex function $f(\boldsymbol{x})$ may be modelled as $\max f(\boldsymbol{x}_k) + \boldsymbol{s}_k^T(\boldsymbol{x} - \boldsymbol{x}_k)$, where the $\boldsymbol{x}_i$ are (not necessarily distinct) previous iterates and $\boldsymbol{s}_k$ is a subgradient evaluated at $\boldsymbol{x}_k$. Modelling each function in this way, we may then obtain a new trial iterate by minimizing the modelled objective subject to the modelled constraints and a stabilizing trust region (or sometimes a penalty on the objective). The set of previous values $(\boldsymbol{x}_k, \boldsymbol{s}_k)$ which are included is known as a *bundle* and the determination of consecutive bundles is fundamental to the success of the method. An introduction to such methods is provided by Kiwiel (1989), Hiriart-Urruty and Lemaréchal (1993) and Lemaréchal and Zowe (1994), while numerical evidence that they are effective is given by Schramm and Zowe (1992).

Finally, we have said very little about infinite-dimensional problems. We particularly regret having no space to mention the significant advances in algorithms for network optimization, optimal control, and parallel optimization, but merely point to Bertsekas (1991), Hager (1990) and Schnabel (1994) as "tasters" for progress in these fields.

7 Software

We should not forget that the main reason for designing and analyzing algorithms is to enable others to solve "real" optimization problems, and one of the best ways of doing this is for researchers to provide quality software which implements their ideas. Frankly, we were surprised when researching this paper quite how few papers contained numerical results which justified their author's optimistic analytic assessments, or indeed any numerical results at all! Fortunately, there is a fair amount of good software available particularly for small problems. Major sources are the Harwell, Hatfield, NAG, and Visual Numerics (formerly IMSL) subroutine libraries. The book by Moré and Wright (1993) provides a thorough assessment of the state and scope of optimization software.

Possibly the biggest change over the past ten years has been the size of nonlinear problems that can be, and are now being, solved. It was rare to find results for problems involving, say, more than 50 unknowns at the time of the last conference, but now highly nonlinear problems involving, say, 20,000 unknowns and similar numbers of constraints can be solved in reasonable times on current desktop computers. Of current codes capable of handling such problems CONOPT (Drud, 1985) and LSGRG2 (Smith and Lasdon, 1992) are generalized reduced gradient methods, MINOS (Murtagh and Saunders, 1982) and LANCELOT (Conn, Gould and Toint, 1992*a*) are based on augmented Lagrangian functions, ETR (Lalee, Nocedal and Plantega, 1993) is an SQP method for equality constraints, while that by Boggs et al. (1994) is a general SQP method.

8 Conclusion

The past ten years in nonlinear optimization have been a time of consolidation rather than inspiration. The energy that has been devoted to interior-point methods — particularly for linear programs — have left the community slightly exhausted when the challenges of nonlinearity arise. However, we certainly have a better understanding of when and why the methods we considered ten years ago work. Furthermore, we are now capable of solving far larger problems than before, primarily because of our better exploitation of problem structure.

It is not difficult to see how the field will develop in the short term. Interior-point methods will be extended to handle nonlinear and nonconvex problems, and many of the subproblems currently solved using active set methods will be tackled with interior methods. Moreover, the wider availability of second (and higher order) derivatives must result in a reappraisal of our current "favourite" approaches.

Some areas remain vastly understudied. The global effect of strong nonlinearity on algorithms has not been considered in any depth, most algorithms retreating to tiny steps and/or variants of (constrained) steepest descent under these circumstances. Little is really known about how modern algorithms compare, especially on large or highly nonlinear problems. We have many preconceptions but, as the resurrection of barrier methods shows, folklore should not necessarily be trusted. Another area which deserves more attention is the effects of noise on minimization algorithms, particularly as so many industrial problems involve noisy functions. Yet further topics which have only recently received attention are nonlinear mixed integer and global optimization problems. And finally, we tend to rely heavily on matrix factorization as a tool, but there are many classes of large problems for which this impossible. We must concern ourselves more in the future on methods for which approximate solutions to model problems are sought.

In our experience a considerable number of users want to solve large problems, while perfectly adequate methods are now available for small problems so long as derivatives are available. So we end with a plea to the optimization research community: if the "new" method you are considering is not applicable to large problems, consider seriously whether it really is worth investigating. Perhaps, in this way, in ten years time, we shall be able to report a narrowing of the gap between the needs of the user community and the provisions of researchers.

References

Bartholomew-Biggs, M. C. 1987 Recursive quadratic-programming methods based on the augmented Lagrangian. *Mathematical Programming Studies* **31**, 21–41.

Bertsekas, D. P. 1976 On the Goldstein-Levitin-Polyak gradient projection method. *IEEE Transactions on Automatic Control* **21**, 174–184.

Bertsekas, D. P. 1991 *Linear Network Optimization: Algorithms and Codes*. Cambridge, Massachusetts, USA: MIT Press.

Best, M. J. 1984 Equivalence of some quadratic-programming algorithms. *Mathematical Programming* **30**, 71–87.

Biegler, L., Nocedal, J., & Schmid, C. 1995 A reduced Hessian method for large-scale constrained optimization. *SIAM Journal on Optimization* **5**, 314–347.

Boggs, P. T., & Tolle, J. W. 1989 A strategy for global convergence in a sequential quadratic programming algorithm. *SIAM Journal on Numerical Analysis* **26**, 600–623.

Boggs, P. T., & Tolle, J. W. 1995 Sequential quadratic programming. *Acta Numerica* **4**, 1–51.

Boggs, P. T., Domich, P. D., & Rogers, J. E. 1995 An interior point method for general large-scale quadratic programming problems. Internal report NISTIR 5406 Applied and Computational Mathematics Division, National Institute of Standards and Technology, Gaithersburg, Maryland, USA.

Boggs, P. T., Tolle, J. W., & Kearsley, A. J. 1991 A merit function for inequality constrained nonlinear programming problems. Internal report NISTIR 4702 Applied and Computational Mathematics Division, National Institute of Standards and Technology, Gaithersburg, Maryland, USA.

Boggs, P. T., Tolle, J. W., & Kearsley, A. J. 1994 A practical algorithm for general large scale nonlinear optimization problems. Internal report NISTIR 5407 Applied and Computational Mathematics Division, National Institute of Standards and Technology, Gaithersburg, Maryland, USA.

Bonnans, J. F., & Launay, G. 1995 Sequential quadratic programming with penalization of the displacement. *SIAM Journal on Optimization* **5**, 792–812.

Bonnans, J. F., Panier, E. R., Tits, A. L., & Zhou, J. L. 1992 Avoiding the Maratos effect by means of a nonmonotone line search II. inequality constrained problems – feasible iterates. *SIAM Journal on Numerical Analysis* **29**, 1187–1208.

Broyden, C. G., & Attia, N. F. 1984 A smooth sequential penalty function method for nonlinear programming. In: *11th IFIP Conference on System Modelling and Optimization* (A. V. Balarkishnan and M. Thomas, eds). Lecture Notes in Control and Information Sciences 59. Heidelberg, Berlin, New York: Springer Verlag 237–245.

Burke, J. V. 1990 On the identification of active constraints II: The nonconvex case. *SIAM Journal on Numerical Analysis* **27**, 1081–1102.

Burke, J. V. 1992 A robust trust region method for constrained nonlinear programming problems. *SIAM Journal on Optimization* **2**, 324–347.

Burke, J. V., & Han, S. P. 1989 A robust sequential quadratic-programming method. *Mathematical Programming* **43**, 277–303.

Burke, J. V., & Moré, J. J. 1988 On the identification of active constraints. *SIAM Journal on Numerical Analysis* **25**, 1197–1211.

Burke, J. V., & Moré, J. J. 1994 Exposing constraints. *SIAM Journal on Optimization* **4**, 573–595.

Burke, J. V., Moré, J. J., & Toraldo, G. 1990 Convergence properties of trust region methods for linear and convex constraints. *Mathematical Programming* **47**, 305–336.

Byrd, R. H. 1990 On the convergence of constrained optimization methods with accurate Hessian information on a subspace. *SIAM Journal on Numerical Analysis* **27**, 141–153.

Byrd, R. H., & Nocedal, J. 1991 An analysis of reduced Hessian methods for constrained optimization. *Mathematical Programming* **49**, 285–323.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. 1995 A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* **16**, 1190–1208.

Byrd, R. H., Schnabel, R. B., & Schultz, G. A. 1987 A trust region algorithm for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis* **24**, 1152–1170.

Byrd, R. H., Tapia, R. A., & Zhang, Y. 1992 An SQP augmented Lagrangian BFGS algorithm for constrained optimization. *SIAM Journal on Optimization* **2**, 210–241.

Calamai, P. H., & Moré, J. J. 1987 Projected gradient methods for linearly constrained problems. *Mathematical Programming* **39**, 93–116.

Carter, R. G. 1991 On the global convergence of trust region methods using inexact gradient information. *SIAM Journal on Numerical Analysis* **28**, 251–265.

Celis, M. R., Dennis, J. E., & Tapia, R. A. 1985 A trust region strategy for nonlinear equality constrained optimization. In: *Numerical Optimization 1984* (P. T. Boggs, R. H. Byrd and R. B. Schnabel, eds). Philadelphia: SIAM 71–82.

Chamberlain, R. M., Powell, M. J. D., Lemaréchal, C., & Pedersen, H. C., 1982 The watchdog technique for forcing convergence in algorithms for constrained optimization. *Mathematical Programming Studies* **16**, 1–17.

Coleman, T. F., & Conn, A. R. 1982*a* Nonlinear programming via an exact penalty function method: Asymptotic analysis. *Mathematical Programming* **24**, 123–136.

Coleman, T. F., & Conn, A. R. 1982*b* Non-linear programming via an exact penalty-function: Global analysis. *Mathematical Programming* **24**, 137–161.

Coleman, T. F., & Fenyes, P. A. 1992 Partitioned quasi-Newton methods for nonlinear equality constrained optimization. *Mathematical Programming* **53**, 17–44.

Coleman, T. F., & Hempel, C. 1990 Computing a trust region step for a penalty function. *SIAM Journal on Scientific and Statistical Computing* **11**, 180–201.

Coleman, T. F., & Hulbert, L. A. 1993 A globally and superlinearly convergent algorithm for convex quadratic programs with simple bounds. *SIAM Journal on Optimization* **3**, 298–321.

Conn, A. R., Gould, N. I. M., & Toint, Ph. L. 1988 Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM Journal on Numerical Analysis* **25**, 433–460. See also same journal **26**, 764-767, 1989.

Conn, A. R., Gould, N. I. M., & Toint, Ph. L. 1990 An introduction to the structure of large scale nonlinear optimization problems and the LANCELOT project. In: *Computing Methods in Applied Sciences and Engineering* (R. Glowinski and A. Lichnewsky, eds). Philadelphia: SIAM 42–54.

Conn, A. R., Gould, N. I. M., & Toint, Ph. L. 1991 A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis* **28**, 545–572.

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

Conn, A. R., Gould, N. I. M., & Toint, Ph. L. 1992*b* On the number of inner iterations per outer iteration of a globally convergent algorithm for optimization with general nonlinear equality constraints and simple bounds. In: *Proceedings of the 14th Biennial Numerical Analysis Conference Dundee 1991* (D. F. Griffiths and G. A. Watson, eds). Longmans 49–68.

Conn, A. R., Gould, N. I. M., & Toint, Ph. L. 1994 A note on exploiting structure when using slack variables. *Mathematical Programming* **67**, 89–97.

Conn, A. R., Gould, N. I. M., Sartenaer, A., & Toint, Ph. L. 1993 Global convergence of a class of trust region algorithms for optimization using inexact projections of convex constraints. *SIAM Journal on Optimization* **3**, 164–221.

Conn, A. R., Gould, N. I. M., Sartenaer, A., & Toint, Ph. L. 1995 Convergence properties of an augmented Lagrangian algorithms for optimization with a combination of general equality and linear constraints. Technical Report RAL-95-009 Rutherford Appleton Laboratory, Chilton, England. To appear in *SIAM Journal on Optimization*.

Dax, A. 1989 The minimax solution of linear equations subject to linear constraints. *IMA Journal of Numerical Analysis* **9**, 95–109.

De Angelis, P. L., & Toraldo, G. 1993 On the identification property of a projected gradient method. *SIAM Journal on Numerical Analysis* **30**, 1483–1497.

Dembo, R. S., Eisenstat, S. C., & Steihaug, T. 1982 Inexact-Newton methods. *SIAM Journal on Numerical Analysis* **19**, 400–408.

Dem'yanov, V. F., & Rubinov, A. M. 1970 *Approximation Methods in Optimization Problems*. Amsterdam: Elsevier.

Dennis, J. E, El-Alem, M., & Maciel, M. C. 1992 A global convergence theory for general trust-region based algorithms for equality constrained optimization. Technical Report CRPC-TR92247 Center for Research on Parallel Computers, Rice University, Houston, Texas, USA. To appear *SIAM Journal on Optimization*.

Di Pillo, G., Facchinei, F., & Grippo, L. 1992 An RQP algorithm using a differentiable exact penalty function for inequality constrained problems. *Mathematical Programming* **55**, 49–68.

Dixon, L. C. W. 1991 On the impact of automatic differentiation on the relative performance of parallel truncated Newton and variable metric algorithms. *SIAM Journal on Optimization* **1**, 475–486.

Drud, A. 1985 CONOPT - a GRG code for large sparse dynamic nonlinear optimization problems. *Mathematical Programming* **31**, 153–191.

Du, D. Z., & Zhang, X. S. 1986 A convergence theorem of Rosen's gradient projection method. *Mathematical Programming* **36**, 135–144.

Du, D. Z., & Zhang, X. S. 1989 Global convergence of Rosen's gradient projection method. *Mathematical Programming* **44**, 357–366.

Dunn, J. C. 1981 Global and asymptotic convergence rate estimates for a class of projected gradient processes. *SIAM Journal on Control and Optimization* **19**, 368–400.

Dunn, J. C. 1987 On the convergence of projected gradient processes to singular critical points. *Journal of Optimization Theory and Applications* **55**, 203–216.

Dussault, J.-P. 1995 Numerical stability and efficiency of penalty algorithms. *SIAM Journal on Numerical Analysis* **32**, 296–317.

El-Alem, M. 1991 A global convergence theory for the Celis-Dennis-Tapia trust region algorithm for constrained optimization. *SIAM Journal on Numerical Analysis* **28**, 266–290.

El-Alem, M. 1995 A robust trust-region algorithm with a nonmonotonic penalty parameter scheme for constrained optimization. *SIAM Journal on Optimization* **5**, 348–378.

Facchinei, F., & Lucidi, S. 1994 Local properties of a Newton-like direction for equality constrained minimization problems. *Optimization Methods and Software* **3**, 13–26.

Fletcher, R. 1970 A class of methods for nonlinear programming with termination and convergence properties. In: *Integer and nonlinear programming* (J. Abadie, ed). Amsterdam: North Holland.

Fletcher, R. 1971 A general quadratic programming algorithm. *Journal of the Institute of Mathematics and its Applications* **7**, 76–91.

Fletcher, R. 1973 An exact penalty function for nonlinear programming with inequalities. *Mathematical Programming* **5**, 129–150.

Fletcher, R. 1981 *Practical Methods of Optimization: Constrained Optimization.* New-York: J. Wiley and Sons.

Fletcher, R. 1982 Second order corrections for non-differentiable optimization. In: *Numerical Analysis, Proceedings Dundee 1981* (D. F. Griffiths, ed). Lecture Notes in Mathematics 912. Heidelberg, Berlin, New York: Springer Verlag 85–114.

Fletcher, R. 1987 Recent developments in linear and quadratic programming. *in* Iserles and Powell (1987) 213-243.

Fletcher, R. 1988 Degeneracy in the presence of roundoff. *Linear Algebra and its Applications* **106**, 149–183.

Fletcher, R. 1995 An optimal positive definite update for sparse Hessian matrices. *SIAM Journal on Optimization* **5**, 192–217.

Fletcher, R., & Sainz de la Maza, E. 1989 Nonlinear programming and nonsmooth optimization by successive linear programming. *Mathematical Programming* **43**, 235–256.

Fontecilla, R. 1990 Inexact secant methods for nonlinear constrained optimization. *SIAM Journal on Numerical Analysis* **27**, 154–165.

Fontecilla, R., Steihaug, T., & Tapia, R. A. 1987 A convergence theory for a class of quasi-Newton methods for constrained optimization. *SIAM Journal on Numerical Analysis* **24**, 1133–1151.

Friedlander, A., & Martínez, J. M. 1994 On the maximization of a concave quadratic function with box constraints. *SIAM Journal on Optimization* **4**, 177–192.

Friedlander, A., Martínez, J. M., & Santos, S. A. 1994 On the resolution of linearly constrained convex minimization problems. *SIAM Journal on Optimization* **4**, 331–339.

Fukushima, M. 1986 A successive quadratic-programming algorithm with global and superlinear convergence properties. *Mathematical Programming* **35**, 253–264.

Gilbert, J. C. 1991 Maintaining the positive definiteness of the matrices in reduced secant methods for equality constrained optimization. *Mathematical Programming* **50**, 1–28.

Gill, P. E., Murray, W., Saunders, M. A., & Wright, M. H. 1989 A practical anticycling procedure for linearly constrained optimization. *Mathematical Programming, Series B* **45**, 437–474.

Gill, P. E., Murray, W., Saunders, M. A., & Wright, M. H. 1990 A Schur-complement method for sparse quadratic programming. In: *Reliable Scientific Computation* (M. G. Cox and S. J. Hammarling, eds). Oxford: Oxford University Press 113–138.

Gill, P. E., Murray, W., Saunders, M. A., & Wright, M. H. 1991 Inertia-controlling methods for general quadratic programming. *SIAM Review* **33**, 1–36.

Gill, P. E., Murray, W., Saunders, M. A., Tomlin, J. A., & Wright, M. H. 1986 On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method. *Mathematical Programming* **36**, 183–209.

Gould, N. I. M. 1986 On the accurate determination of search directions for simple differentiable penalty functions. *IMA Journal of Numerical Analysis* **6**, 357–372.

Gould, N. I. M. 1989 On the convergence of a sequential penalty function method for constrained minimization. *SIAM Journal on Numerical Analysis* **26**, 107–128.

Gould, N. I. M. 1991 An algorithm for large-scale quadratic programming. *IMA Journal of Numerical Analysis* **11**, 299–324.

Griewank, A. 1989 On automatic differentiation. *in* Iri and Tanabe (1989) 83–108.

Griffith, R. E., & Stewart, R. A. 1961 A nonlinear programming technique for the optimization of continuous processing systems. *Management Science* **7**, 379–392.

Grippo, L., Lampariello, F., & Lucidi, S. 1986 A nonmonotone line search technique for Newton's method. *SIAM Journal on Numerical Analysis* **23**, 707–716.

Gurwitz, C. 1994 Local convergence of a two-piece update of a projected Hessian matrix. *SIAM Journal on Optimization* **4**, 461–485.

Hager, W. W. 1990 Multiplier methods for nonlinear optimal control. *SIAM Journal on Numerical Analysis* **27**, 1061–1080.

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

Heinkenschloss, M. 1994 On the solution of a two ball trust region subproblem. *Mathematical Programming* **64**, 249–276.

Heinz, J., & Spellucci, P. 1994 A successful implementation of the Pantoja-Mayne SQP method. *Optimization Methods and Software* **4**, 1–28.

Hestenes, M. R. 1969 Multiplier and gradient methods. *Journal of Optimization Theory and Applications* **4**, 303–320.

Hiriart-Urruty, J.-B., & Lemaréchal, C. 1993 *Convex Analysis and Minimization Algorithms. Part 2: Advanced Theory and Bundle Methods.* Heidelberg, Berlin, New York: Springer Verlag.

Iri, M., & Tanabe, K., eds 1989 *Mathematical Programming: Recent Developments and Applications.* Dordrecht: Kluwer Academic Publishers.

Iserles, A., & Powell, M. J. D., eds 1987 *The State of the Art in Numerical Analysis.* Oxford: Oxford University Press.

Jarre, F., & Saunders, M. A. 1995 Practical interior-point method for convex programming. *SIAM Journal on Optimization* **5**, 149–171.

Kanzow, C., & Kleinmichel, H. 1995 A class of Newton-type methods for equality and inequality constrained optimization. *Optimization Methods and Software* **5**, 173–198.

Kiwiel, K. C. 1989 A survey of bundle methods for non-differentiable optimization. *in* Iri and Tanabe (1989) 263–282.

Kupferschmid, M., & Ecker, J. G. 1987 A note on solution of nonlinear-programming problems with imprecise function and gradient values. *Mathematical Programming Studies* **31**, 129–138.

Lalee, M., Nocedal, J., & Plantega, T. 1993 On the implementation of an algorithm for large-scale equality optimization. Technical Report NAM 09-93 Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois.

Lemaréchal, C., & Zowe, J. 1994 A condensed introduction to bundle methods in nonsmooth optimizations. *in* Spedicato (1994) 357–382.

Lescrenier, M. 1991 Convergence of trust region algorithms for optimization with bounds when strict complementarity does not hold. *SIAM Journal on Numerical Analysis* **28**, 476–495.

Levitin, E. S., & Polyak, B. T. 1966 Constrained minimization problems. *USSR Computational Mathematics and Mathematical Physics* **6**, 1–50.

Maany, Z. A. 1987 A new algorithm for highly curved constrained optimization. *Mathematical Programming Studies* **31**, 139–154.

Maratos, N. 1978 Exact penalty function algorithms for finite-dimensional and control optimization problems. PhD thesis University of London England.

Martinez, J. M., & Santos, S. A. 1995 A trust-region strategy for minimization on arbitrary domains. *Mathematical Programming* **68**, 267–301.

Mayne, D. Q., & Polak, E. 1982 A superlinearly convergent algorithm for constrained optimization problems. *Mathematical Programming Studies* **16**, 45–61.

Moré, J. J. 1993 Generalizations of the trust region problem. *Optimization Methods and Software* **2**, 189–209.

Moré, J. J., & Toraldo, G. 1991 On the solution of large quadratic programming problems with bound constraints. *SIAM Journal on Optimization* **1**, 93–113.

Moré, J. J., & Wright, S. J. 1993 *Optimization Software Guide.* Frontiers in Applied Mathematics 14. Philadelphia: SIAM.

Murray, W., & Prieto, F. J. 1995 A sequential quadratic programming algorithm using an incomplete solution of the subproblem. *SIAM Journal on Optimization* **5**, 590–640.

Murray, W., & Wright, M. H. 1978 Projected Lagrangian methods based on the trajectories of penalty and barrier functions. Technical Report SOL78-23 Department of Operations Research, Stanford Universit, Stanford, California 94305, USA.

Murray, W., & Wright, M. H. 1982 Computation of a search direction in constrained optimization algorithms. *Mathematical Programming Studies* **16**, 62–83.

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

Murtagh, B. A., & Saunders, M. A. 1982 A projected Lagrangian algorithm and its implementation for sparse non-linear constraints. *Mathematical Programming Studies* **16**, 84–117.

Nesterov, Y., & Nemirovskii, A. 1994 *Interior-point polynomial algorithms in convex programming.* Philadelphia: SIAM.

Nocedal, J., & Overton, M. L. 1985 Projected Hessian updating algorithms for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis* **22**, 821–850.

Pang, J. S. 1991 A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. *Mathematical Programming* **51**, 101–131.

Pang, J.-S. 1994 Serial and parallel computation of Karush-Kuhn-Tucker points via nonsmooth equations. *SIAM Journal on Optimization* **4**, 872–893.

Panier, E. R., & Tits, A. L. 1987 A superlinearly convergence feasible method for the solution of inequality constrained optimization problems. *SIAM Journal on Control and Optimization* **25**, 934–950.

Panier, E. R., & Tits, A. L. 1991 Avoiding the Maratos effect by means of a nonmontone line search I. general constrained problems. *SIAM Journal on Numerical Analysis* **28**, 1183–1195.

Panier, E. R., & Tits, A. L. 1993 On combining feasibility, descent and superlinear convergence in inequality constrained optimization. *Mathematical Programming* **59**, 261–276.

Pantoja, J. F. A., & Mayne, D. Q. 1991 Exact penalty functions with simple updating of the penalty parameter. *Journal of Optimization Theory and Applications* **69**, 441–467.

Powell, M. J. D. 1970 A new algorithm for unconstrained optimization. In: *Nonlinear Programming* (J. B. Rose, O. L. Mangasarian and K. Ritter, eds). London and New York: Academic Press.

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

Powell, M. J. D. 1987 Methods for nonlinear constraints in optimization calculations. *in* Iserles and Powell (1987) 325–358.

Powell, M. J. D. 1989 A tolerant algorithm for linearly constrained optimization calculations. *Mathematical Programming, Series B* **45**, 547–566.

Powell, M. J. D., & Yuan, Y. 1986 A recursive quadratic-programming algorithm that uses differentiable exact penalty-functions. *Mathematical Programming* **35**, 265–278.

Powell, M. J. D., & Yuan, Y. 1990 A trust region algorithm for equality constrained optimization. *Mathematical Programming* **49**, 189–211.

Pschenichny, B. N. 1970 Algorithms for general problems of mathematical programming. *Kibernetica* **6**, 120–125.

Robinson, S. M. 1974 Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms. *Mathematical Programming* **7**, 1–16.

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

Rosen, J. B. 1960 The gradient projection method for nonlinear programming, Part I - linear constraints. *SIAM Journal on Applied Mathematics* **8**, 181–217.

Ryan, D. M., & Osborne, M. R. 1988 On the solution of highly degenerate linear programs. *Mathematical Programming* **41**, 385–392.

Sartenaer, A. 1995 A class of trust region methods for nonlinear network optimization problems. *SIAM Journal on Optimization* **5**, 379–407.

Schnabel, R. B. 1994 Parallel nonlinear optimization: limitations, challenges and opportunities. *in* Spedicato (1994) 531–559.

Schramm, H., & Zowe, J. 1992 A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. *SIAM Journal on Optimization* **2**, 121–152.

Smith, S., & Lasdon, L. 1992 Solving large-sparse nonlinear programs using GRG. *ORSA Journal of Computing* **4**, 1–15.

Spedicato, E., ed 1994 *Algorithms for continuous optimization. The state of the art.* Dordrecht: Kluwer Academic Publishers.

Stern, R. J., & Wolkowicz, H. 1995 Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. *SIAM Journal on Optimization* **5**, 286–313.

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

Toint, Ph. L. 1977 On sparse and symmetric matrix updating subject to a linear equation. *Mathematics of Computation* **31**, 954–961.

Toint, Ph. L. 1988 Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space. *IMA Journal of Numerical Analysis* **8**, 231–252.

Vardi, A. 1985 A trust region algorithm for equality constrained minimization: convergence properties and implementation. *SIAM Journal on Numerical Analysis* **22**, 575–591.

Wilson, R. B. 1963 A simplicial algorithm for concave programming. PhD thesis Harvard University Massachusetts, USA.

Womersley, R. S. 1985 Local properties of algorithms for solving nonsmooth composite functions. *Mathematical Programming* **32**, 69–89.

Wright, S. J. 1987 Local properties of inexact methods for minimizing nonsmooth composite functions. *Mathematical Programming* **37**, 232–252.

Wright, S. J 1989*a* Convergence of SQP-like methods for constrained optimization. *SIAM Journal on Control and Optimization* **27**, 13–26.

Wright, S. J. 1989*b* An inexact algorithm for composite nondifferentiable optimization. *Mathematical Programming* **44**, 221–234.

Yuan, Y. 1985*a* Conditions for convergence of trust region algorithms for nonsmooth optimization. *Mathematical Programming* **31**, 220–228.

Yuan, Y. 1985*b* On the superlinear convergence of a trust region algorithm for nonsmooth optimization. *Mathematical Programming* **31**, 269–285.

Yuan, Y. 1990 On a subproblem of trust region algorithms for constrained optimization. *Mathematical Programming* **47**, 53–63.

Yuan, Y. 1995 On the convergence of a new trust region algorithm. *Numerische Mathematik* **70**, 515–539.

Zhang, J., Kim, N.-H., & Lasdon, L. 1985 An improved successive quadratic programming algorithm. *Management Science* **31**, 1312–1331.

Zhang, J., & Zhu, D. 1990 Projected quasi-Newton algorithm with trust-region for constrained optimization. *Journal of Optimization Theory and Applications* **67**, 369–393.

Zhang, J., Zhu, D., & Fan, Y. 1993 A practical trust region method for equality constrained optimization problems. *Optimization Methods and Software* **2**, 45–68.

Zhang, Y. 1992 Computing a Celis-Dennis-Tapia trust-region step for equality constrained optimization. *Mathematical Programming* **55**, 109–124.