NONLINEAR PROGRAMMING TECHNIQUES APPLIED TO STOCHASTIC PROGRAMS WITH RECOURSE

J.L. Nazareth
R.J-B. Wets

October, 1985
WP-85-062

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

ABSTRACT

Stochastic convex programs with recourse can equivalently be formulated as nonlinear convex programming problems. These possess some rather marked characteristics. Firstly, the proportion of linear to nonlinear variables is often large and leads to a natural partition of the constraints and objective. Secondly, the objective function corresponding to the nonlinear variables can vary over a wide range of possibilities; under appropriate assumptions about the underlying stochastic program it could be, for example, a smooth function, a separable polyhedral function or a nonsmooth function whose values and gradients are very expensive to compute. Thirdly, the problems are often large-scale and linearly constrained with special structure in the constraints.

This paper is a comprehensive study of solution methods for stochastic programs with recourse viewed from the above standpoint. We describe a number of promising algorithmic approaches that are derived from methods of nonlinear programming. The discussion is a fairly general one, but the solution of two classes of stochastic programs with recourse are of particular interest. The first corresponds to stochastic linear programs with simple recourse and stochastic right-hand-side elements with given discrete probability distribution. The second corresponds to stochastic linear programs with complete recourse and stochastic right-hand-side vectors defined by a limited number of scenarios, each with given probability. A repeated theme is the use of the MINOS code of Murtagh and Saunders as a basis for developing suitable implementations.

This paper is a draft for Chapter 4 of *Numerical Techniques for Stochastic Optimization Problems*, Y. Ermoliev and R.J.-B. Wets, eds., Springer-Verlag, to appear.

CONTENTS

NONLINEAR PROGRAMMING TECHNIQUES APPLIED TO STOCHASTIC PROGRAMS WITH RECOURSE

J.L. Nazareth and R.J-B. Wets

1. Introduction

We consider stochastic linear programs of the type

$$\text{Find } x \in R^{n_1} \text{ such that} \tag{1.1}$$

$$Ax = b, x \geq 0$$

$$\text{and } z = E_w[c(w)x + Q(x,w)] \text{ is minimized}$$

where Q is calculated by finding for given decision x and event w, an optimal recourse $y \in R^{n_2}$, viz.

$$Q(x,w) = \inf_{y \in C} [q(y,w) \mid Wy = h(w) - Tx] \quad . \tag{1.2}$$

Here $A(m_1 \times n_1)$, $T(m_2 \times n_1)$, $W(m_2 \times n_2)$ and $b(m_1)$ are given (fixed) matrices, $c(\cdot)(n_1)$ and $h(\cdot)(m_2)$ are random vectors, $y \to q(y,\cdot): R^{n_2} \to R$ is a random finite-valued convex function and C is a convex polyhedral subset of R^{n_2}, usually $C = R^{n_2}_+$. E denotes expectation.

With $c = E_w[c(w)]$, an equivalent form to (1.1) is

$$\text{minimize } cx + Q(x) \tag{1.3}$$

subject to

$$Ax = b$$

$$x \geq 0$$

where $Q(x) = E_w[Q(x,w)]$. Usually $q(y,w)$ will also be a linear nonstochastic

function qy. (For convenience, we shall, throughout this paper, write cx and qy instead of c^Tx and q^Ty.)

Two instances of the above problem are of particular interest:

(C1) Problems with simple recourse i.e. with $W = [I, -I]$, stochastic right-hand-side elements with given discrete probability distribution and penalty vectors q^+ and q^- associated with shortage and surplus in the recourse stage (1.2).

(C2) Problems with complete recourse and stochastic right-hand-side vectors defined by a limited number of scenarios, each with given probability.

Henceforth, for convenience ,we shall refer to these as C1 *and* C2 *problems respectively.* They can be regarded as a natural extension of linear and nonlinear programming models into the domain of stochastic programming. More general stochastic programs with recourse can sometimes be solved by an iterative procedure involving definition (for example, using approximation or sampling) of a sequence of C1 or C2 problems.

Within each of several categories of nonlinear programming methods, we summarize briefly the main underlying approach for smooth problems, give where appropriate extensions to solve nonsmooth problems and then discuss how these lead to methods for solving stochastic programs with recourse. Thus, in each case, we begin with a rather broadly based statement of the solution strategy, and then narrow down the discussion to focus on methods and computational considerations for stochastic programs with recourse, where the special structure of the problem is now always in the background. (During the course of the discussion we occasionally consider other related formulations, in particular the model with probabilistic constraints. However it is our intention to concentrate upon the recourse model. (We do not discuss questions concerning approximation of distribution functions, except very briefly at one or two points in the text). This paper is *not* intended to provide a complete survey. Rather, our aim is to establish some framework of discussion within the theme set by the title of this paper and within it to concentrate on a number of promising lines of algorithmic development. We try to strike a balance between the specific (what is practicable using current techniques, in particular, for C1 and C2 problems) and the speculative (what should be possible by extending current techniques). *An important theme will be the use of MINOS* (the Mathematical Programming System of Murtagh and Saunders [49,50]) *as a basis for implementation.* Finally we seek to set the stage for the description of an optimization system based upon MINOS for solving C1 problems, see Nazareth [55].

We shall assume that the reader is acquainted with the main families of optimization methods, in particular,

(a) univariate minimization.

(b) Newton, quasi-Newton and Lagrangian methods for nonlinear minimization.

(c) subgradient (nonmonotonic) minimization of nonsmooth functions, possibly using space dilation (variable metric), and the main descent methods of nonsmooth minimization.

(d) stochastic quasi-gradient methods

(e) the simplex method of linear programming and its reduced-gradient extensions.

Good references for background material are Fletcher [20], Gill et al. [23], Bertsekas [4], Lemarechal [42], Shor [66], Ermoliev [16], Dantzig [11], Murtagh & Saunders [49].

We shall concentrate upon methods of nonlinear programming which seem to us to be of particular relevance to stochastic programming with recourse and discuss them under the following main headings:

1. Problem Redefinition
2. Linearization Methods
3. Variable Reduction (Partitioning) Methods
4. Lagrange Multiplier Methods

A nonlinear programming algorithm will often draw upon more than one of these groups and there is, in fact, significant overlap between them. However, for purposes of discussion, the above categorization is useful.

2. PROBLEM REDEFINITION

By problem redefinition we mean a restructuring of a nonlinear programming problem to obtain a new problem which is then addressed in place of the original one. This redefinition may be achieved by introducing new variables, exploiting separability, dualizing the original problem and so on. For example, consider the minimization of a polyhedral function given by

$$\min_{x \in R^n} \max_{j=1,\ldots,m} [(a^j)^T x + b^j] \tag{2.1a}$$

This can be accomplished by transforming the problem into a linear program

$$\text{minimize } v \quad \text{such that} \quad v \geq (a^j)^T x + b^j \quad , j = 1,\ldots,m \tag{2.1b}$$

which can then be solved by the simplex method.

Problem redefinition often precedes the application of other solution methods discussed in later sections of this paper.

2.1 Application to Recourse Problems

The following two transformations of recourse problems will prove useful:

(a) When the technology matrix is fixed, new variables χ, termed *tenders*, can be introduced into (1.3). This gives an equivalent form as follows:

$$\begin{array}{ll} \text{minimize} & cx + \Psi(\chi) \\ \text{subject to} & Ax = b \\ & Tx - \chi = 0 \\ & x \geq 0 \quad . \end{array} \tag{2.2}$$

(2.2) is useful because it is a nonlinear program in which the number of variables occurring nonlinearly is m_2 instead of n_1 and usually $m_2 << n_1$. For a more detailed discussion of the use of tenders in algorithms for solving stochastic linear programs with recourse, see Nazareth and Wets [56].

(b) Another useful transformation involves introducing second stage activities into the first stage. It is shown in Nazareth [51] that that an alternative form equivalent to (2.2) is

$$\begin{array}{ll} \text{minimize} & cx + qy + \Psi(\chi) \\ \text{subject to} & Ax = b \\ & Tx + Wy - \chi = 0 \\ & x \geq 0 \ , \ y \geq 0 \quad . \end{array} \tag{2.3}$$

This transformation also has significant advantages from a computational standpoint, as we shall see below. These stem, in part, from the fact that dual feasible variables, say (ρ, π) satisfy $W^{\mathrm{T}}\pi \leq q$.

For C1 problems, $\Psi(\chi)$ in (2.2) is separable, i.e. $\Psi(\chi) = \sum_{i=1}^{m_2} \Psi_i(\chi_i)$. In such problems, each component of $h(\cdot)$ is assumed to be discretely distributed, say with $h_i(\cdot)$ given by levels $h_{i1},\ldots,h_{ik_i}$ and associated probabilities $p_{i1},\ldots,p_{ik_i}$; also $q(y)$ in (1.2) is two-piece linear and can be replaced by $q^+y^+ + q^-y^-$, $y^+ \geq 0$, $y^- \geq 0$ in (2.3). This implies that each $\Psi_i(\chi_i)$ is piecewise linear with slopes, say, s_{il}, $l = 0,\ldots,k_i$. By introducing new bounded variables z_{il}, $l = 0,\ldots,k_i$ we can reexpress χ_i as

$$\chi_i = h_{i0} + \sum_{l=0}^{k_i} z_{il}$$

where h_{i0} is the i-th component of h_0 the *base* tender. Then (2.2) takes the form:

$$\begin{aligned} \text{minimize} \quad & cx + \sum_{i=1}^{m_2} \sum_{l=0}^{k_i} s_{il} z_{il} \\ \text{subject to} \quad & Ax = b \\ & T^i x - \sum_{l=0}^{k_i} z_{il} = h_{i0} \ , \ i = 1,\ldots,m_2 \\ & x \geq 0 \ , \ 0 \leq z_{il} \leq d_{il} \ , \ l = 0,\ldots,k_i \text{ with } d_{il} = h_{i,l+1} - h_{il} \ . \end{aligned} \tag{2.4}$$

T^i denotes the i-th row of T. Optionally we can use the transformation (2.3) to introduce $W = [I, -I]$ into the first stage. Details of an algorithm based upon (2.4) can be found in Wets [72] and an alternative simpler version of this algorithm can be found in Nazareth & Wets [56]. The latter algorithm is implemented in the optimization system described [55], where further discussion and computational considerations may be found.

2.2 Extensions

The device of introducing new bounded variables, which was used to obtain (2.4), can be applied to a wider class of recourse problems. The assumptions of discrete distribution of $h(\cdot)$ and of two (or more) piece linearity of recourse objective are not central, although one must still retain the assumptions of simple recourse and separable recourse objective. Suppose, for example, the distribution function of $h_i(\cdot)$ which need not be continuous, is piecewise linear with knots $h_{i1},\ldots,h_{ik_i}$, and $q = (q^+, q^-)$. Then $\Psi_i(\chi_i)$ is piecewise quadratic. In general, if the distribution is defined in terms of splines of order s at knots $h_{i1},\ldots,h_{ik_i}$ and $q(y)$ is separable, say, $\sum_{i=1}^{n_2} q_i(y_i)$ with each $q_i(y_i)$ convex, then $\Psi_i(\chi_i)$ can be shown to be convex and piecewise smooth. Suppose it is given by pieces $\Psi_{il}(\chi_i)$ over intervals $(h_{il}, h_{i,l+1})$. Then, analogously to (2.4) we can transform the problem (2.2) into the structured and *smooth nonlinear* program

$$\begin{aligned} \text{minimize} \quad & cx + \sum_{i=1}^{m_2} \sum_{l=0}^{k_i} \Psi_{il}(h_{il} + z_{il}) - \Psi_{il}(h_{il}) \\ \text{subject to} \quad & Ax = b \\ & T^i x - \sum_{l=0}^{k_i} z_{il} = h_{i0} \ , \ i = 1,\ldots,m_2 \\ & x \geq 0 \ , \ 0 \leq z_{il} \leq d_{il} \ , \ l = 0,\ldots,k_i \text{ with } d_{il} = h_{i,l+1} - h_{il} \ . \end{aligned} \tag{2.5}$$

(Here again we could use the transformation (2.3) to introduce $W = [I, -I]$ into the first stage). Note that (2.4) is a special case of (2.5). The optimal solution of (2.5) has an important property which is easy to prove. This result makes the nonlinear program (2.5) very amenable to solution by MINOS-like techniques and it is given by the following proposition:

Proposition: *In the optimal solution of (2.5), say* (x^*, z_{il}^*), *if for some* t , $z_{it}^* < d_{it}^*$ *then* $z_{il}^* = d_{il}$ *for all* $l < t$.

Outline of Proof: Regard each $\Psi_{il}(\chi_i)$ as the limit of a piecewise linear function, and then appeal to the standard argument used in the piecewise-linear case.

The above proposition tells us that there are, at most, m_2 superbasic variables (see Section 4 for terminology) in the optimal solution of (2.5). This would be to the advantage of a routine like MINOS, which thrives on keeping the number of superbasics low. These remarks will become clearer after looking at Section 4. Note also that Wets [70] discusses a special case of (2.5) when $\Psi_{il}(\chi_i)$ are piecewise quadratic. A well-structured code for solving (2.4), which uses only the LP facilities of MINOS, would be capable of a natural extension to solve nonlinear problems of the form (2.5). MINOS was really designed to solve problem of this type.

The above approach remains limited in scope, because of the need to assume that recourse is simple and that the recourse objective is separable. Therefore we would not expect it to be useful for C2 problems.

The transformations given by (2.2) and (2.3) are very useful prior to the application of other techniques discussed in the following sections of this chapter. Let us consider some possibilities.

1. When T is nonstochastic, use of the transformation (2.2) in the methods described by Kall [30] or the L-shaped algorithm of Van Slyke and Wets [68] (see also Birge [6]) would lead to *fewer non-zero elements* in the representation of the associated large-scale linear programs.

2. When T is not a fixed matrix, typically only a few columns (activities), say $T_2(w)$, would be stochastic. Say these correspond to variables $\hat{x}$, with $x = (\tilde{x}, \hat{x})$. We could then introduce a redefinition of the problem in which a tender is associated with the non-stochastic columns, say T_1 of T; then the degree of nonlinearity of the equivalent deterministic nonlinear programming problem would be m_2 + dimension $(\hat{x})$ instead of n_1. For example, for simple recourse with $q = (q^+, q^-)$ we would have

$$\psi(\chi,\hat{x},w) = \min_{y^+,y^- \geq 0} [q^+y^+ + q^-y^- \mid y^+ - y^- = h(w) - \chi - T_2(w)\hat{x}]$$

$$\Psi(\chi,\hat{x}) = E_w[\psi(\chi,\hat{x},w)]$$

Note that $\Psi(\chi,\hat{x})$ continues to be separable in χ, i.e. $\Psi(\chi,\hat{x}) = \sum_{i=1}^{m_2} \Psi_i(\chi_i,\hat{x})$.

These observations and the further developments that they imply would be useful in a practical implementation.

3. Another interesting example of the use of the transformations involving tenders is given in Nazareth [52] where they are used in the solution of deterministic staircase-structured linear programs.

3. Linearization Methods

A prominent feature of methods in this group is that they solve sequences of linear programs. One can distinguish single-point and multi-point linearization. In both approaches convexity of functions is normally assumed.

3.1 Single-Point Linearization Methods

We discuss this case very briefly.

Consider the problem $\underset{x \in K}{\text{minimize}}\, f(x)$, where K is polyhedral and $f(x)$ is smooth. The approach consists of solving a sequence of problems of the form:

$$\underset{x \in \bar{K}}{\text{minimize}}\ \nabla f(x_k)^T(x - x_k) \tag{3.1}$$

where $\bar{K}$ is the original polyhedral set K, possibly augmented by some additional constraints. This leads to a variety of methods. When $K = \bar{K}$ we obtain the Frank–Wolfe [21] method, in which the solution, say x_k^*, defines a search direction $d_k = x_k^* - x_k$. The method has the virtue that the solution is found in one step if the original problem is linear. If K is augmented by the constraints $\|x - x_k\|_\infty \leq \delta$ for some small positive constant δ we obtain the Griffith & Stewart [26] method of approximate programming (MAP); for minimax applications see Madsen & Schjaer-Jacobsen [46] and for extensions to the domain of general nonsmooth optimization see the monograph of Demyanov & Vasiliev [12].

3.1.1 **Applications to Recourse Problems**

For simple recourse when the equivalent (deterministic) nonlinear program is smooth, algorithms are given, for example, by Ziemba [77]. Kallberg and Ziemba [33] use the Frank–Wolfe method in a setting where only estimates of functions and gradients can be obtained. The approach has been widely studied within the context of the general expectation model, see Ermoliev [16] and models with probabilistic constraints, see Komaroni [37] and references cited there. In this latter context, however, one needs to rely on a variant of the standard Frank–Wolfe method to take into account nondifferentiability (infinite slope) of the objective at the boundary of the feasible region. Given a stochastic program with probabilistic constraints of the type

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & Ax = b \\ & \text{Prob}\,[w \mid Tx > h(w)] \geq \alpha \\ & x \geq 0 \end{array}$$

we see that it is equivalent to

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & Ax = b \\ & Tx - \chi \geq 0 \\ & g(\chi) \geq 0 \\ & x \geq 0 \\ \text{where} & g(\chi) = \ln\,(\text{Prob}\,[w \mid \chi > h(w)] - \alpha). \end{array}$$

Assuming that the probability measure is log-concave, it follows that g is concave and thus we are dealing with a convex optimization problem with one nonlinear constraint. Its dual is

$$\begin{array}{ll} \text{maximize} & ub + \rho(v) \\ \text{subject to} & uA + vT \leq c \\ & v \geq 0 \\ \text{where} & \rho(v) = \inf\,[v\chi \mid g(\chi) \geq 0] \quad . \end{array}$$

The function ρ is a sublinear (concave and positively homogeneous) finite-valued (only) on the positive orthant. If the probability measure is strictly log-concave, the function ρ is differentiable on the interior of the positive orthant and thus we could use the Frank–Wolfe procedure to solve this dual problem as long as the iterates (u^s, v^s) are such that $v^s \in \text{interior } R_+^{m_2}$; when v^s is on the boundary of $R_+^{m_2}$, the standard procedure must be modified to handle the 'infinite' slope case, see Komaroni [37].

3.2 Multi-Point Linearization Methods

Consider the problem

$$\text{minimize } f(x) \quad \text{where } g_i(x) \le 0\,,\ i=1,\ldots,m\,,\ x \in X \tag{3.2}$$

where all functions are convex, but not necessarily differentiable, and X is a compact set. We shall concentrate in this section on the generalized linear programming method (GLP) of Wolfe (see Dantzig [11], Shapiro [65]) which solves a sequence of problems obtained by *inner (or grid) linearization* of (3.2) over the convex hull of a set of points $x^1,\ldots,x^K$, to give the following master program:

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{K} \lambda_i f(x^i) \\ \text{subject to} \quad & u_i^{(K)}\colon \sum_{i=1}^{K} \lambda_i g_i(x^i) \le 0,\ i = 1,\ldots,m \\ & w^{(K)}\colon \sum_{i=1}^{K} \lambda_i = 1\,,\ \lambda_i \ge 0 \end{aligned} \tag{3.3a}$$

where $u^{(K)}$ and $w^{(K)}$ represent the dual variables associated with the optimal solution of (3.3a). The dual of (3.3a) is

$$\begin{aligned} \text{maximize} \quad & w \\ \text{subject to} \quad & w \le f(x^{(k)}) + ug(x^{(k)})\,,\ k = 1,\ldots,K \\ & u \ge 0\ . \end{aligned} \tag{3.3b}$$

The next grid point x^{K+1} is obtained by solving the Lagrangian subproblem

$$\underset{x \in X}{\text{minimize}}\ [f(x) + u^{(K)} g(x)] \tag{3.4}$$

where $u^{(K)}$ is also the optimal solution of (3.3b). Convergence is obtained when

$$f(x^{(K+1)}) + u^{(K)} g(x^{(K+1)}) \ge w^{(K)}\ .$$

Since the dual of (3.2) is

$$\underset{u \ge 0}{\text{maximize}}\ h(u)\,,\ \text{where } h(u) = \min_{x \in X} (f(x) + ug(x)) \tag{3.5}$$

and $h(u)$ is readily shown to be concave, an alternative viewpoint is to regard the GLP method as a *dual* cutting plane (or outer linearization) method on (3.5) yielding (3.3b); new grid points obtained from (3.4) yield a supporting hyperplane to $h(u)$ at $u^{(K)}$.

It is worth emphasizing again that an important advantage of the inner-linearization approach is that it can be directly applied to the solution of non-smooth convex problems without extensions.

Outer linearization could be applied directly to the functions in (3.2) to give a *primal* cutting plane method which also solves sequences of linear programs. For details, see Kelley [34], Zangwill [76] and Eaves & Zangwill [13].

3.2.1 Applications to Recourse Problems

For recourse problems, particularly with the form (2.2) using tenders, the GLP approach looks very promising.

Using GLP to solve *simple recourse* problems has an early history. It was first suggested by Williams [74], in the context of computation of error bounds and also used at an early date by Beale [2]. Parikh [57] describes many algorithmic details. The method has also been implemented for specialized applications (e.g. see Ziemba [78], for an application to portfolio selection). However, as a general computational technique in particular, for *non-simple recourse* it has apparently not been studied until recently, see Nazareth and Wets [56] and Nazareth [51].

The GLP method applied to (2.2) yields the following master program:

$$\begin{aligned} \text{minimize} \quad & cx + \sum_{i=1}^{K} \lambda_k \Psi(\chi^{(k)}) \\ \text{subject to} \quad & \rho^{(K)}: \; Ax = b \\ & \pi^{(K)}: \; Tx - \sum_{k=1}^{K} \lambda_k \chi^k = 0 \\ & \vartheta^{(K)}: \; \sum_{k=1}^{K} \lambda_k = 1 \\ & x \geq 0 \,, \; \lambda_k \geq 0 \quad . \end{aligned} \tag{3.6a}$$

The associated subproblem is

$$\underset{\chi \in X}{\text{minimize}} \; \Psi(\chi) + \pi^{(K)}\chi \tag{3.6b}$$

In order to complete the description of the algorithm it is necessary to specify X, and if this is not a compact set, to extend the master program (3.6a) by introducing directions of recession (whose associated variables do not appear in the convexity row). In addition, a suitable set of starting tenders which span R^{m_2} should be specified. As discussed in Nazareth [51] these considerations can be largely circumvented by using the equivalent form (2.3) and solving master

programs of the form:

$$\begin{aligned} \text{minimize} \quad & cx + qy + \sum_{k=1}^{K} \lambda_k \Psi(\chi^{(k)}) \\ \text{subject to} \quad & Ax = b \\ & Tx + Wy - \sum_{k=1}^{K} \lambda_k \chi^{(k)} = 0 \\ & \sum_{k=1}^{K} \lambda_k = 1 \\ & x \geq 0 \,, \; y \geq 0 \,, \; \lambda_k \geq 0 \quad . \end{aligned} \tag{3.7}$$

As discussed in more detail in Nazareth & Wets [56], we expect the above algorithm to perform well because normally only a few tenders will have non-zero coefficients in the optimal solution and because one can expect to obtain a good set of starting tenders from the underlying recourse program.

Still at issue is how readily one can compute $\Psi(\chi^{(k)})$ and its subgradients at a given point χ^k. This in turn determines the ease with which one can solve the subproblem (3.6b) and obtain coefficients in the objective row of the master.

For C1 problems $\Psi(\chi)$ is separable and easy to specify explicitly (see Wets [72]). Algorithms have been given by Parikh [57] and Nazareth [51]. A practical implementation is given in Nazareth [55] where further details may be found.

For C2 problems (i.e. with complete recourse and a relatively small set of scenarios say, $h^l, l = 1,\ldots,L$ with known probabilities $f_l, l = 1,\ldots,L$) one can solve the subproblem (3.6b) and compute $\Psi(\chi^{(k)})$ in one of two ways, as discussed in Nazareth [51]:

(i) Formulate (3.6b) as a linear program which can be efficiently solved by Schur-Complement techniques, see Bisschop & Meeraus [9], and Gill et al. [24]. The values $\Psi(\chi^{(k)})$ are a part of the solution of this linear program.

(ii) Use unconstrained non-smooth optimization techniques, see Lemarechal [40,41], Kiwiel [35] and Shor [66]. Information needed by such methods is $\Psi(\chi^{(k)})$ and its subgradient $g(\chi^{(k)})$ and this can be computed by solving a set of linear programs of the form:

$$\psi(\chi^{(k)},h^l) = \min_{y \geq 0} [qy \mid Wy = h^l - \chi^{(k)}] \quad .$$

Suppose π^l are the optimal dual multipliers of the above problem. Then

$$\Psi(\chi^{(k)}) = \sum_{l=1}^{L} f_l \psi(\chi^{(k)}, h^l)$$

$$g(\chi^{(k)}) = \sum_{l=1}^{L} f_l \pi^l \quad .$$

This can be carried out very efficiently using the dual simplex method coupled with techniques discussed by Wets in [73].

The method based upon outer linearization mentioned at the end of the previous section has been widely used to solve stochastic programs with recourse (see Van Slyke & Wets [68], Wets [73] and Birge [6]). This is a particular form of Benders' decomposition [3] and it is well known that approaches based upon Benders' decomposition can solve a wider class of ***nonlinear*** convex programs than approaches based upon the Dantzig–Wolfe decomposition, see, for example, Lasdon [39]). We shall not however discuss this approach in any detail here because it is already studied, in depth, in the references just cited.

3.2.2 Extensions

When $\Psi(\chi^k)$ and its subgradients are difficult to compute, the GLP approach continues to appear very promising but many open questions remain that center on convergence.

Two broad approaches can be distinguished:

(i) *Sampling*: Stochastic estimates of $\Psi(\chi)$ and its subgradient can be obtained by sampling the distribution. An approach that uses samples of fixed size and carries out the minimization of the Lagrangian subproblem (3.6b) using smoothing techniques is described by Nazareth [51]. Methods for minimizing noisy functions suggested recently by Atkinson et al. [1] would also be useful in this context. With a fixed level of noise, convergence proofs can rely upon the results of Poljak [58].

Another variant is to use samples of progressively increasing size tied to the progress of the algorithm and to solve the Lagrangian subproblem using stochastic quasi-gradient methods, see Ermoliev & Gaivoronski [18]. A particular algorithm (suggested jointly with A. Gaivoronski) is to replace $\Psi(\chi^{(k)})$ in (3.6) by some estimate $\Psi_s(\chi^{(k)})$ which is based upon a suitable sample size N. When no further progress is made, then this sample size is incremented by ΔN and the approximation

refined for all $\chi^{(k)}$ in the current basis. There are, of course, many further details that must be specified, but under appropriate assumptions convergence can be established.

(ii) *Approximate Distribution and Compute Bounds*: At issue here is how to simultaneously combine approximation and optimization. For example, Birge [7] assumes that converging approximations $\Psi_K^U(\chi)$ and $\Psi_K^L(\chi)$ are available for $K = 1,2,....$ and replaces $\Psi(\chi^{(k)})$ in (3.6a) by the upper bound $\Psi_k^U(\chi(k))$, $k = 1,...,K$. In the subproblem (3.6b), if

$$\Psi_{K+1}^U(\chi^{(K+1)}) + \pi^{(K)}\chi^{(K+1)} \geq \vartheta^{(K)}$$

then $\Psi_{K+1}^L(\chi^{(K+1)})$ is computed. If, further, the above inequality is satisfied using this lower bound in place of the upper bound, then $\chi^{(k+1)}$ is optimal. Otherwise the approximation is refined and the process continued. Approximation schemes for obtaining bounds rely on the properties of recourse problems, instead of purely on the distance between the given probability distribution and the approximating ones; this allows for sequential schemes that involve much fewer points as discussed by Kall & Stoyan [31] and Birge & Wets [8].

The interpretation of the optimal solution in Nazareth [51], suggests the possibility of an alternative approach to approximation by an increasingly large number of points. It is shown that if $\lambda_{k_j}^*$ and $\chi^{(k_j)}$, $j = 1,...,(m_2+1)$ give the optimal solution of (3.6a), then the problem (2.2) is equivalent to the associated discretized problem obtained by replacing the distribution of $h(w)$ by the distribution whose values are $\chi^{(k_j)}$, $j = 1,...,(m_2+1)$ with associated probabilities $\lambda_{k_j}^*$. Note that $\sum_{i=1}^{m_2+1} \lambda_{k_j}^* = 1$, $\lambda_{k_j}^* \geq 0$, so that these quantities do indeed define a probability distribution.

Let us conclude this section with a discussion of some other possibilities.

1. When the technology matrix is nonlinear, i.e. when T is replaced by a smooth nonlinear function, we have the possibility of a generalized programming algorithm where the *master program itself is nonlinear*. The question of convergence is open. Here an implementation based upon MINOS would be able to immediately draw upon the ability of this routine to solve programs with nonlinear constraints.

2. When some columns of T are stochastic, the transformation discussed at the end of Section 2 can also be used within the context of the GLP algorithm to keep the degree of nonlinearity low. This time inner approximation of $\Psi(\chi,\hat{x})$ would be

carried out over the convex hull of $(\chi^{(k)},\hat{x}^{(k)})$, $k = 1,\ldots,K$.

3. Generalized programming techniques appear to be useful for solving programs with probabilistic constraints, for example, of the form:

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & Ax = b \\ & \text{Prob}\,[\omega \mid Tx > h(\omega)] \geq \alpha \\ & x \geq 0 \end{array}$$

With the usual definition of tenders $Tx = \chi$ and under the appropriate assumptions on the distribution of $h(\cdot)$, we can express the above problem as:

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & Ax = b \\ & Tx - \chi = 0 \\ & g(\chi) \leq 0 \\ & x \geq 0 \end{array}$$

where $g(\chi) = \alpha - \text{Prob}\,[\omega \mid h(\omega) < \chi]$ is a nonlinear function which is log-concave for a wide variety of distribution functions, in which case the set $[\chi \mid g(\chi) \leq 0]$ is convex. In such a situation we can reformulate the constraint

$$g(\chi) \leq 0 \quad ,$$

as

$$\chi \in D = \{y \mid g(y) \leq 0\}$$

where

$$g(y) = \alpha - \int_{\zeta \leq \chi} p(\zeta)\mathrm{d}\zeta \quad .$$

Here $p(\cdot)$ denotes the density function of the random vector $h(\cdot)$. Assuming that we have already generated $\chi^1,\ldots,\chi^K$ in D such that

$$\{\chi = Tx \mid Ax = b\ ,\ x \geq 0\} \cap \text{co}\,\{\chi^1,\ldots,\chi^K\} \neq \phi \quad ,$$

we would be confronted at step K with the master problem:

$$\begin{array}{ll} \text{minimize} & cx \\ \text{subject to} & \sigma^K : Ax = b \\ & \pi^K : Tx - \sum_{i=1}^{K} \lambda_i \chi^i = 0 \\ & \vartheta^K : \sum \lambda_i = 1 \\ & \qquad x \geq 0,\ \lambda_i \geq 0\ ,\ i = 1,\ldots,K \end{array}$$

where $(\sigma^K, \pi^K, \vartheta^K)$ represent the dual variables associated with the optimal solution (χ^K, λ^K) of this master problem. The next tender χ^{K+1} is obtained by solving the Lagrangian subproblem, involving only χ:

$$\text{minimize } [\pi^K \chi \mid \chi \in D]$$

and this χ^{K+1} is introduced in the master problem unless

$$\pi^K \chi \geq \vartheta^K \quad ,$$

in which case x^K is an optimal solution of the master problem. To find

$$\chi^{K+1} \in \text{argmin } [\pi^K \chi \mid \int_{\zeta < \chi} p(\zeta) d\zeta \leq \alpha]$$

we consider the Lagrangian function

$$l(\chi, \beta) = \pi^K \chi + \beta(\int_{\zeta < \chi} p(\zeta) d\zeta - \alpha) \quad , \quad \beta \geq 0$$

and the dual problem

$$\text{maximize } h(\beta) \quad , \quad \beta \geq 0 \quad ,$$

where

$$h(\beta) = \inf [l(\chi, \beta) \mid \chi] \quad .$$

The function h is an l-dimensional concave function, its (generalized) derivative is a monotone increasing function, and, moreover, under strict log-concavity of the probability measure, its maximum is attained at a unique point. To search for the optimal β^K we can use a secant method for finding the zero of a monotone function. We have that for fixed β,

$$\chi(\beta) = \text{argmin } l(\chi, \beta)$$

is obtained by solving the following system of equations:

$$-\pi_i^K / \beta = \int_{\{\zeta_k < \chi_k \mid k \neq i\}} p(\zeta_1, \ldots, \zeta_{i-1}, \ldots, \chi_i, \zeta_{i+1}, \ldots, \zeta_{m_2}) d\zeta \quad , \quad i = 1, \ldots, m_2 \quad .$$

If p is simple enough, or if it does not depend on too many variables then this system can be solved by a quasi-Newton procedure that avoids multidimensional integration.

This application to chance-constrained stochastic linear programming is an open area and certainly deserves further investigation.

4. It is also worth pointing out that generalized programming methods have been recently applied to the study of problems with partially known distribution functions (incomplete information), see Ermoliev et al. [17] and Gaivoronski [22].

4. Variable Reduction (Partitioning) Methods

Methods in this group seek to restrict the search region to one defined by a subset of the variables and carry out one or more iterations of a gradient (or subgradient) based search procedure. The search region is then revised and the process continued. We can make a distinction between 'homogeneous' and 'global' methods (using the terminology of Lemarechal [42]). Homogeneous or active set methods, in the linearly constrained case, restrict the region of search to an *affine subspace* within which unconstrained minimization techniques can be used. We shall concentrate on the reduced gradient formulation of Murtagh & Saunders [49,50] as implemented in MINOS and seek extensions of an approach which has proved effective for large smooth problems. However the fact that extension is necessary, in contrast to the methods of the previous section, and the fact that there are theoretical issues of convergence that remain to be settled mean that such methods are still very much in the development stage.

Global methods treat all constraints *simultaneously* and define direction finding subproblems which usually involve minimization subject to inequality constraints (often just simple bound constraints). Convergence issues are more easily settled here. We shall consider some methods of this type.

We also include here approaches where the partition of variables is more directly determined by the problem structure, in particular the grouping into linear and nonlinear variables.

Consider first the problem defined by

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \qquad (4.1)$$

where, initially, $f(x)$ is assumed to be smooth.

The variables at each cycle of the Murtagh and Saunders [49] reduced gradient method are partitioned into three groups, (x_B, x_S, x_N) representing m basic variables, s superbasic variables, and $nb = n - m - s$ non-basic variables

respectively. Non-basics are at their bound. A is partitioned as $[B|S|N]$ where B is an $m \times m$ nonsingular matrix, S is an $m \times s$ matrix, and N is an $m \times nb$ matrix. Let $g = \nabla f(x)$ be similarly partitioned as (g_B, g_S, g_N).

Each cycle of the method can be viewed as being roughly equivalent to:

(RG1) one or more iterations of a quasi-Newton method on an unconstrained optimization problem of dimension s determined by the *active* set $Ax = b$, $x_N = 0$. Here a reduced gradient is computed as

$$\mu = g_S - (g_B^T B^{-1})S = [-(B^{-1}S)^T \mid I_{s \times s} \mid 0]g = Z_S^T g \quad . \tag{4.2}$$

The columns of Z_S span the space in which the quasi-Newton search direction lies, and this is given by $p = -Z_S H Z_S^T g$ where H is an inverse Hessian approximation obtained by quasi-Newton update methods and defines the variable metric, e.g. $H = I$ gives the usual projected gradient direction. Along p a line search is usually performed. (Note that in actual computation H would not be computed. Instead we would work with approximations to the Hessian and solve systems of linear equations to compute the search direction p.)

(RG2) an iteration of the revised simplex method on a linear program of dimension $m \times nb$. Here components of the reduced gradient (Lagrange multipliers) corresponding to the nonbasic components are computed by

$$\lambda = g_N - (g_B^T B^{-1})N \tag{4.3}$$

$$\lambda = [-(B^{-1}N)^T \mid 0 \mid I_{nb \times nb}]g = Z_N^T g \tag{4.4}$$

This is completely analogous to the computation of μ in (4.2) above. The difference is in the way that λ is used, namely to revise the active set. In each case above prices π can be computed by $\pi = g_B^T B^{-1}$ and μ and λ computed as

$$\mu = g_S - \pi^T S \ , \ \lambda = g_N - \pi^T N \tag{4.5}$$

(It is worth noting that the ***convex simplex method*** is a special case of the above where (RG1) is omitted and (RG2) is replaced by a coordinate line search along a single coordinate direction in the reduced space given by $(Z_N)_k$, say, for which $\lambda_k < 0$. When there are nonlinear constraints present the above method can also be suitably generalized.)

In the *non-smooth* case we can proceed along three main directions:

1. Compute μ and λ in place of the above by

$$\mu = Z_S^T\{\text{argmin } [g^T(Z_S Z_S^T)g \mid g \in \partial f(x)]\} \tag{4.6}$$

$$\lambda = Z_N^T\{\text{argmin } [g^T(Z_N Z_N^T)g \mid g \in \partial f(x)]\}$$

where $\partial f(x)$ is the subdifferential of $f(x)$ at x. In effect we are computing steepest descent directions in the appropriate subspaces. Note that it is, in general, not correct to first compute a steepest descent direction $\bar{g}$ from

$$\bar{g} = \text{argmin } [g^T g \mid g \in \partial f(x)]$$

and then reduce $\bar{g}$ to give

$$\mu = Z_S^T \bar{g} \tag{4.7}$$

$$\lambda = Z_N^T \bar{g} \quad .$$

The reason for this is that the operations of minimization and projection are not interchangeable. However this approach does make it possible to restore use of the π vector and therefore yields useful heuristic methods, as we shall see in the next section. In order to ensure convergence, it is necessary to replace $\partial f(x)$ by $\partial_\varepsilon f(x)$ – the ε-subdifferential (except in special circumstances e.g. when $f(x)$ is polyhedral and line searches are exact). This is useful from a theoretical standpoint. However, from the point of view of computation it is usually impractical to use the subdifferential, let alone the epsilon-subdifferential (except again in rather special circumstances). One such instance is when the subdifferential is defined by a small set of vectors, say, $g_1,\ldots,g_N$. Then (4.6) leads to the problem:

$$\begin{aligned} \text{minimize} \quad & g^T Z_S H Z_S^T g \\ \text{subject to} \quad & g = \sum_{i=1}^{N} \lambda_i (Z_S^T g_i) \\ & \sum_{i=1}^{N} \lambda_i = 1 \\ & \lambda_i \geq 0 \quad . \end{aligned} \tag{4.8}$$

If g^* is its solution, then $\mu = Z_S^T g^*$, with a similar computation for Z_N^T. We also have $p = -Z_S H Z_S^T g^*$.

2. Utilize bundle methods in which the subdifferential is replaced by an approximation composed from subgradients obtained at a number of prior iterations. For

the unconstrained case algorithms are given by Lemarechal [40,41] and an implementable version is given by Kiwiel [35]. An extension of [40] to handle linear constraints in the reduced gradient setting is given by Lemarechal et al. [45]. However, as the authors point out theoretical issues of convergence remain to be settled in the latter case.

3. Utilize non-monotonic methods (see, for example, Shor [66]) which require only a single subgradient at each iteration. In effect non-monotonic iterations will be carried out in subspaces (see RG1 and RG2 above) determined by Z_S and Z_N, using reduced subgradients $Z_S^T g$ and $Z_N^T g$. Again convergence issues remain open.

Line searches suitable for use in the above cases (1) and (2) are given by Mifflin [48] and Lemarechal [43].

The reduced gradient method as formulated above benefits from additional structure in objective and constraints, in particular the partition between variables that occur linearly and variables that occur nonlinearly. We shall see instances of this in the discussion of recourse problems. In particular, it is easy to show that when $f(x)$ is replaced by $cx + \Psi(\chi)$, an optimal solution exists for which the number of superbasics does not exceed the number of nonlinear variables χ.

Instead of obtaining an active set from $x_N = 0$, another approach which gives a 'global' method is to reduce the gradient or subgradient only through the equality constraints $Ax = b$ (these are always active) and define reduced problems to find the search direction involving bound constraints on the x_N variables. This is discussed in Bihain [5]. (See also Strodiot et al. [67].)

Reduced gradient methods, as discussed above, benefit from the partition of the problem into linear and nonlinear variables, but they do not *explicitly* utilize it. It is however possible to take more immediate advantage of this partition. Possible approaches are given, for example, by Rosen [62] and by Ermoliev [15]. Consider the problem

$$\begin{array}{ll} \text{minimize} & cx + F(y) \\ \text{subject to} & Ax + By = b \\ & x \geq 0, y \geq 0 \end{array}$$

If the nonlinear variables y are fixed at certain values we obtain a simpler problem, in this case a linear program (which may have further structure, for example, when A is block-diagonal). The optimal dual multipliers π^* of this linear program (assumed feasible), can then be used to define a reduced subproblem, for example,

$F(y) - (\pi^*)^T By$, $y \geq 0$. This is then solved to revise the current values of y, for example, by computing a reduced subgradient by $g - \pi^* B$, $g \in F(y)$ and carrying out (nonmonotonic) iterations in the positive orthant of the y variables (see Ermoliev [15]). An alternative approach is given by Rosen [62].

4.1 Applications to Recourse Problems

Since the number of nonlinear variables χ in (2.2) is usually small relative to the number of linear variables, the reduced gradient approach outlined above is a natural choice. When $\Psi(\chi)$ is smooth (and the gradient is computable) the reduced gradient method can be used directly. In the form of the convex simplex method, which is a special case of the reduced gradient method, it has been suggested for the simple recourse problem by Wets [69] and Ziemba [77]. Wets [71] extends the convex simplex method to solve problems with simple recourse when the objective is nonsmooth.

For C1 problems $\partial\Psi_i(\chi_i) = [v_i^-, v_i^+]$ (see Nazareth & Wets [56]). The computation of μ and λ in (4.6) thus requires that we solve ***bound constrained quadratic programs***. We can utilize structure in the basis matrix in defining these quadratic programs. Since the χ variables are unrestricted, they can be assumed to be always in the basis. A basis matrix will thus have the form

$$B = \left[\begin{array}{c|c} D & 0 \\ \hline E & I \end{array}\right] \tag{4.9a}$$

and its inverse (never, of course, computed directly) will therefore be given by

$$B^{-1} = \left[\begin{array}{c|c} D^{-1} & 0 \\ \hline -ED^{-1} & I \end{array}\right] . \tag{4.9b}$$

Let $g_B = (c_B, g_\chi)$ where c_B are coefficients of the objective row corresponding to the x variables in the basis and g_χ is a subgradient of $\Psi(\chi)$ at the current value of χ. Also, since superbasics and non-basics are always drawn from c, we shall use c_S and c_N in place of g_S and g_N. Thus we define $g = (c_B, g_\chi, c_S, c_N)$. The quadratic programs (4.6) then takes the form

$$\begin{array}{ll} \text{minimize} & g^T Z_S Z_S^T g \\ \text{subject to} & v_i^- \leq (g_\chi)_i \leq v_i^+ , \quad i = 1, \ldots, m_2 \end{array} \tag{4.10}$$

where g is defined above, $Z_S^T = (-(B^{-1}S)^T \mid I_{s \times s} \mid 0)$ with B defined by (4.9a). Note that usually g_χ will have relatively few components. A similar bound

constrained quadratic program can be defined for Z_N^T. Both can be solved very efficiently using a routine like QPSOL, see [25]. The above approach also requires a line search and an efficient one based upon a specialized version of generalized upper bounding, is given in Nazareth [54]. An implementation could thus be based upon MINOS, QPSOL and this line search.

It is possible to avoid the use of quadratic programming by using a heuristic technique in which a steepest-descent direction is first computed as the solution of the expression preceeding (4.7). This is given by:

$$\begin{aligned} &\text{minimize} && g_\chi^T g_\chi \\ &\text{subject to} && v_i^- \le (g_\chi)_i \le v_i^+ , \quad i=1,\ldots,m_2 \end{aligned} \tag{4.11}$$

The solution $\bar{g}_\chi$ is given explicitly by:

$$(\bar{g}_\chi)_i = \begin{cases} v_i^- \text{ if } v_i^- > 0 \\ 0 \text{ if } 0 \in [v_i^-, v_i^+] \\ v_i^+ \text{ if } v_i^+ < 0 \end{cases} . \tag{4.12}$$

Projected quantities $Z_S^T\bar{g}$ and $Z_N^T\bar{g}$ can then be computed with $\bar{g}$ defined analogously to g (just before expression (4.10)). This and use of the line search in Nazareth [54] suggests a very convenient ***heuristic*** extension of MINOS. Even the construction of a specialized line search can be avoided by utilizing line search methods designed for smooth problems (again heuristic in this context) as discussed by Lemarechal [44].

For C2 problems, computing μ and λ by (4.6) again requires that we solve the following special structured quadratic program (Nazareth & Wets [56]):

$$\text{find } g \in R^{m_2} \text{ such that } \|g\|_M^2 \text{ is minimized}$$

such that

$$g_\chi = \sum_{l=1}^{L} f_l \pi^l \text{ and } \pi^l W \le q , \quad \pi^l (h^l - \chi) \ge \Psi(\chi, h^l) , \quad l = 1,\ldots,L$$

where h^l and f^l define the probability distribution of the scenarios, as in Section 3.2.1. M defines the metric and for different choices, the objective takes the form $g^T g$ (or equivalently, in this case, $g_\chi^T g_\chi$), $g^T Z_S Z_S^T g$ or $g^T Z_N Z_N^T g$. Again special purpose techniques can be devised to solve such problems. It is however often impractical to consider use of the above steepest descent approach because only

$\Psi(\chi)$ and a subgradient are available. In this case an algorithm would have to be designed around bundle techniques or non-monotonic optimization as discussed in Section 4, items (2) and (3) (after expression (4.8)), using reduced subgradients given by $Z_S^T g$ and $Z_N^T g$, with g and other quantities defined as in the paragraph preceding (4.10). In this case an implementation could be based upon a routine for minimizing nonsmooth functions, see Bihain [5].

In the above methods the χ variables would normally always be in the basis, since they have no bounds on their value. This means that there are always some variables in the basis which correspond to the nonsmooth part of the objective function. An alternative approach is to try and restore a more simple pricing strategy by keeping the χ variables always *superbasic* and define a basis only in the x variables. The alternating method of Qi [61] is an attempt in this direction although it is not implementable in the form given in [61]. Other methods along these lines are given by Birge [6]. However, the numerical results given by Birge [6] show that the approach may not be as promising as the method based upon outer linearization (the so-called L-shaped method) mentioned at the end of Section 3.2.1.

4.2 Extensions

As with generalized linear programming, we think that much can be done by extending the above approach, when $\Psi(\chi)$ and its subgradient are hard to compute, but there are many open questions. As in Section 3.2.2, two broad approaches can be followed:

(i) *Sampling*: Potentially the most valuable approach seems to be an *alternating* method in which one would carry out iterations in the χ space and combine them in some suitable way with subgradient (or stochastic quasi-gradient) iterations in the x space (along the lines suggested by Ermoliev [15]). It is also possible to consider 'homogeneous' or active set methods which extend the reduced gradient approach and interleave iterations involving two projection operators into the space defined by superbasic and non-basic variables respectively.

(ii) *Approximate Distribution and Compute Bounds*: For a discussion of this approach see Birge & Wets [8].

5. Lagrange Multiplier Methods

We conclude this chapter with very brief mention of methods which have recently achieved much popularity for smooth and non-smooth optimization and are thus likely to lead to useful methods for solving recourse problems. Bertsekas [4] and Powell [59] give comprehensive reviews in the smooth case. Lemarechal [42] explains connections with minimax optimization and other methods of non-smooth optimization.

A distinguishing feature of methods in this category is that they combine cutting plane techniques with use of a quadratic penalty term in the computation of search directions and that they often treat the constraints 'globally', again in the sense of Lemarechal [42]. For an example of the use of a (parameterized) quadratic penalty term in unconstrained minimization see the proximal point method of Rockafellar [63]; in smooth nonlinear programming, see Wilson [75] and in nonsmooth optimization, see Pschenichnyi & Danilin [60].

Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \\ & x \geq 0 \quad . \end{array} \tag{5.1}$$

The search direction finding problem then takes the form:

$$\begin{array}{ll} \text{minimize} & v + (1/2)d^{T}Bd \\ \text{subject to} & v \geq -\alpha_i + g_i^{T}d \quad , \; i \in I \\ & Ax = b \\ & x \geq 0 \end{array} \tag{5.2}$$

where I denotes an index set and g_i, $i \in I$ a set of subgradients of $f(x)$. α_i is a scalar. If $B = 0$, I has only one element and $f(x)$ is smooth (so that g_i corresponds to a gradient), note the connection with the method of Frank & Wolfe [21] (see also Section 3.1). When $B = I$, the identity matrix, we have the method suggested by Pschenichnyi & Danilin, see [60].

By dualizing (5.2) it is easy to establish ties with steepest descent methods determined by bundles of subgradients in the appropriate reduced space together with the appropriate definition of a metric (see (4.8) and also Han [27,28], Lemarechal [42], Kiwiel [35] and Demyanov & Vasiliev [12]). Recently Kiwiel [36] has suggested a method which further exploits the structure in (4.8) and has also considered extensions of methods under consideration in this section when there is uncertainty in the value of the function.

Finally, for application of ideas underlying Lagrange multiplier methods to stochastic programs with recourse, see Rockafellar & Wets [64], Merkovsky, Dempster & Gunn [47].

References

[1] E.N. Atkinson, B.W. Brown, J.E. Dennis, J.R. Thompson, "Minimization of noisy functions", presented at SIAM Conference on Numerical Optimization, Boulder, Colorado, June 1984.

[2] E.M.L. Beale, Private communication.

[3] J.F. Benders, "Partitioning procedures for solving mixed variables programming problems", *Numerische Mathematik* **4**(1962) 238–252.

[4] D.P. Bertsekas, "Constrained Optimization and Lagrange Multiplier Methods", Academic Press, New York (1982).

[5] A. Bihain, "Numerical and algorithmic contributions to the constrained optimization of some classes of non-differentiable functions", Ph.D. dissertation, Facultes Universitaires Notre-Dame De La Paix, Belgium (1984).

[6] J. Birge, "Decomposition and partitioning methods for multistage stochastic linear programs", Tech. Report 82–6, Dept. of IE & OR, University of Michigan (1982).

[7] J. Birge, "Using sequential approximations in the L-shaped and generalized programming algorithms for stochastic linear programs", Tech. Report 83–12, Dept. of IE & OR, University of Michigan (1983).

[8] J. Birge and R. J-B Wets, "Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse", WP-83-111, IIASA, Laxenburg, Austria (1983) and in A. Prekopa and R.J-B. Wets eds., *Mathematical Programming Study*, to appear.

[9] J. Bisschop and A. Meeraus, "Matrix augmentation and partitioning in the updating of the basis inverse", *Mathematical Programming* **19**(1977) 7–15.

[10] H. Cleef, "A solution procedure for the two-stage stochastic program with simple recourse", *Z. Operations Research* **35**(1981) 1–13.

[11] G.B. Dantzig, *Linear Programming and Extensions*, Princeton University Press (1980).

[12] V.F. Demyanov and L.V. Vasiliev, "Nondifferentiable Optimization", (in Russian), Nauka (1981).

[13] B.C. Eaves and W. Zangwill, "Generalized cutting plane algorithms", *SIAM J. Control* **9**(1971) 529–542.

[14] J. Edwards, J. Birge, A. King and L. Nazareth, "A standard input format for computer codes which solve stochastic programs with recourse and a library of utilities to simplify its use", WP-85-03, IIASA, Laxenburg, Austria (1984). (Also this volume, see [19].)

[15] Yu. Ermoliev, "Methods of nondifferentiable and stochastic optimization and their applications", in: E. Nurminski, ed., Progress in Nondifferentiable Optimization, CP-82-S8, IIASA, Laxenburg, Austria (1982).

[16] Yu. Ermoliev, "Stochastic quasi-gradient methods and their application in systems optimization", *Stochastics* **8**(1983).

[17] Yu. Ermoliev, A. Gaivoronski and C. Nedeva, "Stochastic optimization problems with incomplete information on distribution functions", WP-83-113, IIASA, Laxenburg, Austria (1983).

[18] Yu. Ermoliev and A. Gaivoronski, "Stochastic quasigradient methods and their implementation", WP-84-55, IIASA, Laxenburg, Austria (1984).

[19] Yu. Ermoliev and R. J-B Wets, eds., *Numerical Techniques for Stochastic Optimization Problems*, Springer-Verlag, (to appear).

[20] R. Fletcher, *Practical Methods of Optimization, Vol. 2: Constrained Optimization*, J. Wiley, Chichester (1981).

[21] M. Frank and P. Wolfe, "An algorithm for quadratic programming", *Naval Research Logistics Quarterly*, III(1956) 95–110.

[22] A. Gaivoronski, "Optimization of functionals which depend on distribution functions: 1. Nonlinear functional and linear constraints", WP-83-114, IIASA, Laxenburg, Austria (1983).

[23] P.E. Gill, W. Murray and M.H. Wright, *Practical Optimization*, Academic Press, London and New York (1981).

[24] P.E. Gill, W. Murray, M.A. Saunders, and M.H. Wright, "Sparse matrix methods in optimization", Tech. Report SOL-82-17, Systems Optimization Lab., Dept. of Operations Research, Stanford University (1983).

[25] P.E. Gill, W. Murray, M.A. Saunders and M.H. Wright, "User's guide for SOL/QPSOL: A FORTRAN package for quadratic programming", SOL 83-7, Systems Optimization Laboratory, Stanford University (1983).

[26] R.E. Griffith and R.A.Stewart, "A nonlinear programming technique for the optimization of continuous processing systems", *Management Science* **7**(1961) 379–392.

[27] S.P. Han, "Superlinearly convergent variable metric algorithms for general nonlinear programming problems", *Mathematical Programming* **11**(1976) 263–282.

[28] S.P. Han, "Variable metric method for minimizing a class of nondifferentiable functions", *Mathematical Programming* **20**(1981) 1–12.

[29] IBM Document No. H20-0476-2, "Mathematical Programming System/360 Version 2, Linear and Separable Programming – User's Manual", IBM Corportion, New York.

[30] P. Kall, "Computational methods for solving two-stage stochastic linear programs", *Z. Agnew.-Math. Phys.* **30**(1979) 261–271.

[31] P. Kall and D. Stoyan, "Solving stochastic programming problems with recourse including error bounds", *Math. Operations Forsch. Statist. Ser. Optimization* **13**(1982) 431–447.

[32] J. Kallberg and M. Kusy, "Code Instruction for S.L.P.R., a stochastic linear program with simple recourse", Tech. Report, University of British Columbia (1976).

[33] J. Kallberg and W. Ziemba, "An extended Frank–Wolfe algorithm with application to portfolio selection", in: P. Kall and A. Prekopa, eds., Springer-Verlag, Berlin (1981).

[34] J.E. Kelley, "The cutting plane method for solving convex programs", *J. Soc. Ind. Appl. Math.* **8**(1960) 703–712.

[35] K.C. Kiwiel, "An aggregate subgradient method for nonsmooth convex minimization", *Mathematical Programming* **27**(1983) 320–341.

[36] K.C. Kiwiel, "A descent algorithm for large-scale linearly constrained convex nonsmooth minimization", CP-84-15, IIASA, Laxenburg, Austria (1984).

[37] E. Komaroni, "A dual method for probabilistic constrained problems", A. Prekopa and R.J-B. Wets, eds.,*Mathematical Programming Study*, to appear.

[38] M. Kusy and W. Ziemba, "A bank asset and liability management model", CP-83-59, IIASA, Laxenburg, Austria (1983).

[39] L.S. Lasdon, "Optimization Theory for Large Scale Systems", Macmillan, London (1970).

[40] C. Lemarechal, "Nonsmooth Optimization and Descent Methods", RR-78-4, IIASA, Laxenburg, Austria (1977).

[41] C. Lemarechal, "Bundle methods in nonsmooth optimization", in: C. Lemarechal and R. Mifflin, eds., Nonsmooth Optimization, IIASA Proceedings Series, Vol. 3, 79–102. Laxenburg, Austria (1977).

[42] C. Lemarechal, "Nonlinear programming and nonsmooth minimization: A unification", Report No.332, INRIA, France (1978).

[43] C. Lemarechal, "A view of line searches", in: A. Auslender, W. Otelli, and J. Stoer, eds., *Optimization and Optimal Control*. Lecture Notes in Control and Information Science 30, Sprinter Verlag, Berlin (1981).

[44] C. Lemarechal, "Numerical experiments in nonsmooth optimization", in: E.A. Nurminski, ed., *Progress in Nondifferentiable Optimization*, CP-82-S8, 61–84, IIASA, Laxenburg, Austria (1982).

[45] C. Lemarechal, J.J Strodiot and A. Bihain, "On a bundle algorithm for nonsmooth optimization", in: O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds., *Nonlinear Programming*, Academic Press, New York (1981).

[46] K. Madsen and H. Schjaer-Jacobsen, "Linearly constrained minimax optimization", *Mathematical Programming* **14**(1978) 208–223.

[47] R. Merkovsky, M.A.H. Dempster and E.A. Gunn, "Some Lagrangian approaches to stochastic programming with recourse", Presented at 12^{th} International Symposium on Mathematical Programming, Boston, Mass., August 1985.

[48] R. Mifflin, "Stationarity and superlinear convergence of an algorithm for univariate locally Lipschitz constrained minimization", *Mathematical Programming* **28**(1984) 50–71.

[49] B. Murtagh and M. Saunders, "Large-Scale linearly constrained optimization", *Mathematical Programming* **14**(1978) 41–72.

[50] B. A. Murtagh and M.A. Saunders, "MINOS 5.0 User's Guide", Report No. SOL 83–20, Systems Optimization Laboratory, Stanford University (1983).

[51] J.L. Nazareth, "Algorithms based upon generalized linear programming for stochastic programs with recourse", in: F. Archetti, ed., *Proceedings of IFIP International Workshop on Stochastic Programming: Algorithms and Applications*. Springer-Verlag, (to appear).

[52] J.L. Nazareth, "Variants on Dantzig–Wolfe decomposition with applications to multistage problems", WP-83-61, IIASA, Laxenburg, Austria (1983).

[53] J.L. Nazareth, "Hierarchical implementation of optimization methods", in: P. Boggs, R. Byrd and B. Schnabel, eds., *Numerical Optimization, 1984*, SIAM, Philadelphia, 199-210 (1985).

[54] J.L. Nazareth, "An efficient algorithm for minimizing a multivariate polyhedral function along a line", in R.W. Cottle, ed., *Mathematical Programming Study*, (to appear).

[55] J.L. Nazareth, "Design and implementation of a stochastic programming optimizer with recourse and tenders", IIASA WP-85-63 Also this volume, see [19].)

[56] J.L. Nazareth and R. J-B Wets, "Algorithms for stochastic programs: the case of non-stochastic tenders", A. Prekopa and R. J-B Wets, eds., *Mathematical Programming Study*, (to appear).

[57] S.C. Parikh, Lecture notes on stochastic programming, unpublished, University of California, Berkeley (1968).

[58] B. Poljak, "Nonlinear programming methods in the presence of noise", *Mathematical Programming* **14**(1978) 87–97.

[59] M.J.D. Powell, "Algorithms for nonlinear constraints that use Lagrangian functions", *Mathematical Programming* **14**(1978) 224–248.

[60] B.N. Pshenichnyi and Yu. Danilin, "Methodes Numeriques dans les Problemes d'extremum", Editions de Moscou, Paris (1977).

[61] L. Qi, "An alternating method to solve stochastic programming with simple recourse", Tech. Report No.515, Computer Science Department, University of Wisconsin (1983).

[62] J.B. Rosen, "Convex partition programming", in: R.L. Graves and P. Wolfe, eds., *Recent Advances in Mathematical Programming*, 159–176, McGraw-Hill, New York (1963).

[63] R.T. Rockafellar, "Monotone operators and the proximal point algorithm", *SIAM Journal on Control and Optimization* **14**(1976) 877–898.

[64] R.T. Rockafellar and R. J-B Wets, "A lagrangian finite generation technique for solving linear–quadratic problems in stochastic programming", WP-84-25, Laxenburg, Austria (1984).

[65] J.F. Shapiro, "Mathematical programming: structures and algorithms", John Wiley, New York (1979).

[66] N.Z. Shor, "Generalized gradient methods of nondifferentiable optimization employing space dilation operators", in: A. Bachem, M. Groetschel and B. Korte, eds., *Mathematical Programming: The State of the Art*, 501–529. Springer Verlag, Berlin (1983).

[67] J.J. Strodiot, V.H. Nguyen and N. Heukmes, "Eps-optimal solutions in nondifferentiable convex programming and some related questions", *Mathematical Programming* **25**(1983) 307–328.

[68] R. Van Slyke and R. J-B Wets, "L-shaped linear program with applications to optimal control and stochastic linear programs", *SIAM Journal on Applied Mathematics* **17**(1969) 638–663.

[69] R. Wets, "Programming under uncertainty: the complete problem", *Z. Wahrsch. verw. Gebiete* **4**(1966) 316–339.

[70] R. Wets, "Solving stochastic programs with simple recourse II", Proceedings of Johns Hopkins Symposium on Systems and Information, 1–6 (1975).

[71] R. Wets, "A statistical approach to the solution of stochastic programs with (convex) simple recourse", in: A. Wierzbicki, ed., *Generalized Lagrangians in Systems and Economic Theory*, IIASA Proceedings Series, Pergamon Press, Oxford (1983).

[72] R. Wets, "Solving stochastic programs with simple recourse", *Stochastics* **10**(1983) 219–242.

[73] R. Wets, "Stochastic programming: solution techniques and approximation schemes", in: A. Bachem, M. Groetschel and B. Korte, eds., *Mathematical Programming: The State-of-the-Art*, 566–603, Springer-Verlag, Berlin (1983).

[74] A.C. Williams, "Approximation formulas for stochastic linear programming", *SIAM Journal on Applied Mathematics* **14**(1966) 668–677.

[75] R.B. Wilson, "A simplicial algorithm for concave programming", Ph.D. Thesis, Graduate School of Business Administration, Harvard University (1963).

[76] W.I. Zangwill, *Nonlinear Programming: A Unified Approach*, Prentice-Hall (1963).

[77] W.T. Ziemba, "Computational algorithms for convex stochastic programs with simple recourse", *Operations Research* **18**(1970) 415–431.

[78] W.T. Ziemba, "Solving nonlinear programming problems with stochastic objective functions", *Journal of Financial and Quantitative Analysis* VII(1972), 1809–1827.

[79] W.T. Ziemba, "Stochastic programs with simple recourse", in: P. Hammer and G. Zoutendijk, eds., *Mathematical Programming in Theory and Practice*, 213–274, North-Holland, Amsterdam (1974).