Order Number 9229852

Inexact subgradient methods

Au, Kelly Thurston, Ph.D.

The University of Arizona, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

INEXACT SUBGRADIENT METHODS

by

Kelly Thurston Au

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF SYSTEMS AND INDUSTRIAL ENGINEERING
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

1 9 9 2

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Kelly Thurston Au

entitled INEXACT SUBGRADIENT METHODS

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Julia L. Higle — Date 4/7/92

Suvrajeet Sen — Date 4/7/92

Ferenc Szidarovszky — Date 4/7/92

Date

Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director Julia L. Higle, Suvrajeet Sen — Date 4/7/92

STATEMENT BY AUTHOR

ACKNOWLEDGMENTS

I would like to thank my committee members Julia Higle, Suvrajeet Sen and Ferenc Szidarovszky.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

In solving a mathematical program, the exact evaluation of the objective function and its subgradients can be computationally burdensome. For example, in a stochastic program, the objective function is typically defined through a multi-dimensional integration. Solution procedures for stochastic programs are usually based on functional approximation techniques, or statistical applications of subgradient methods. In this dissertation, we explore algorithms by combining functional approximation techniques with subgradient optimization methods. This class of algorithms is referred to as "inexact subgradient methods". First, we develop a basic inexact subgradient method and identify conditions under which this approach will lead to an optimal solution. We also offer an inexact subgradient algorithm by adaptively defining the steplengths via estimated bounds on the deviations from optimality. Second, we explore approaches in which functional approximation techniques can be combined with a primal-dual subgradient method. In these algorithms, the steplengths are defined via the primal and dual information. Hence suggestions to optimality can be reflected through the steplengths, as the iteration proceeds. We also incorporate space dilation operations, which stabilize the moving directions, within our basic inexact subgradient method. As an example of the applicability of these methods, we use statistically defined approximations, which are similar to those derived in Stochastic Decomposition, in some of our algorithms for the solutions of stochastic programs.

CHAPTER 1

INTRODUCTION

1.0 Introduction

In decision making processes related to economics, ecology, engineering, etc., one often gathers information on alternative actions, their associated consequences, and decision criteria to form mathematical models for identifying the 'best' decision . Many mathematical models can be formulated as mathematical programs or can be reduced to them. In general, a mathematical program is formulated as:

$$\begin{aligned} \text{Min} \quad & f(x) \\ s/t \quad & f_i(x) \leq 0, \quad i \in I \\ & x \in \mathcal{R}^n, \end{aligned} \tag{1.1}$$

where I is an index set. In a deterministic model, one assumes that precise information about the objective function f and the constraints f_i is available. For example, if it is a deterministic production planning problem, information about prices and future demands is presumed to be given. In practice, the demands may not be known with certainty at the time that the decision on the production level is made. In such cases, the assumptions typically associated with deterministic modelling techniques are inappropriate.

Uncertainties are invariably present in most engineering decision making problems, which can range from selecting the number of cashiers in a supermarket, to choosing a design for the structural scheme of a building, to deciding whether nuclear power is a viable energy source. Various approaches involving queueing models, decision tree models, reliability design, etc., have been developed to solve various stochastic problems. In a mathematical program, if some parameters are only known in a statistical sense, one can formulate it as a stochastic program.

Stochastic programs are applicable to a broad range of decision making problems. For example, in flood control (Prékopa and Szántai [1978]), the size of a reservoir is chosen so that flooding due to random stream inputs can be managed; in capacity planning (Murphy, Sen and Soyster [1982]), cost effective facilities are selected prior to the realization of random demands; and in financial planning (Kallberg, White and Ziemba [1982]), the portfolio of short term assets is adjusted in order to minimize the net cost of cash surpluses and deficits under uncertain transaction costs. Other test problems for stochastic programs can also be found in Ermoliev and Wets [1988].

In a stochastic program, the functions of interest are written as expected values of functions of random variables, so that

$$f(x) = \int_{\Omega} F(x,\omega) d\mathcal{P}(\omega)$$
$$f_i(x) = \int_{\Omega} F_i(x,\omega) d\mathcal{P}(\omega).$$

The random variable $\tilde{\omega}$ is defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and the functions

$$F: \mathcal{R}^n \times \Omega \rightarrow \mathcal{R}^1$$
$$F_i: \mathcal{R}^n \times \Omega \rightarrow \mathcal{R}^1$$

are presumed to be known. Other formulations, such as chance constraints and variance constraints, can be considered as special cases of this form (Wets and Ermoliev [1988]). In principle, any non-linear programming technique for the solution of (1.1) can be used for solving a stochastic optimization problem. However, due to their representations as expected values, it can be difficult to obtain explicit expressions for the functions f and f_i (analytically or numerically). Solution techniques for certain classes of stochastic optimization problems were first studied and developed by Beale [1955], Dantzig [1955], Tintner [1955] and Charnes and Cooper [1959].

1.1 Two Stage Stochastic Linear Program with Recourse

In designing procedures for solutions of stochastic programs, one can typically take advantage of commonly occurring problem structures. For the purpose of illustration, consider a production planning model, in which there are m production facilities and n markets. It is necessary to determine $\{x_i\}_{i=1}^m$, the amount of production at each location, prior to the realization of $\{\tilde{\omega}_j\}_{j=1}^n$, the demand in each market. Once the demand is realized, the amount of product transported from facility i to market j, y_{ij}, (at a cost of g_{ij} per unit) is determined. Letting c_i denote the per unit production costs at facility i, this problem can be formulated as follows:

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^{m} c_i x_i + E[h(x,\tilde{\omega})] \\ s/t \quad & x_i \geq 0 \qquad i = 1,\ldots,m \end{aligned}$$

where

$$\begin{aligned} h(x,\omega) = \text{Min} \quad & \sum_{i=1}^{m}\sum_{j=1}^{n} g_{ij} y_{ij} \\ s/t \quad & \sum_{i=1}^{m} y_{ij} \geq \omega_j \qquad j = 1,\ldots,n \\ & \sum_{j=1}^{n} y_{ij} \leq x_i \qquad i = 1,\ldots,m \\ & y_{ij} \geq 0. \end{aligned}$$

This formulation agrees with that of a two stage stochastic linear program with recourse (SLPR) which is typically stated as follows:

$$\begin{aligned} \text{Min} \quad & f(x) = cx + E[h(x,\tilde{\omega})] \\ s/t \quad & Ax \leq b \qquad\qquad (SLPR) \\ & x \geq 0 \end{aligned}$$

where

$$\begin{aligned} h(x,\omega) = \text{Min} \quad & gy \\ s/t \quad & Wy = \omega - Tx \qquad\qquad (S) \\ & y \geq 0, \end{aligned}$$

A, W, T and b are of fixed size $m_1 \times n_1$, $m_2 \times n_2$, $m_2 \times n_1$ and $m_1 \times 1$, respectively, $\tilde{\omega}$ is a random variable defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, and E denotes the expectation with respect to $\tilde{\omega}$. $E[h(x, \tilde{\omega})]$ is known as the recourse function and the problem (S) is referred to as a subproblem. In this dissertation, if x and y are two points, we use xy to denote the inner product. We use x^T to denote the transpose of x only if it is necessary.

As previously noted, the first stage decision x is made before any observation of the random variable $\tilde{\omega}$, is realized. Following its realization, a second stage or recourse decision which minimizes h is identified. The overall objective of the problem is the minimization of the expected value of the sum of the first and second stage costs. From the definition of the two stage SLPR, we note that the decision-making process is sequential in nature. Since the first stage decision is made without exact knowledge of ω, it is called a non-anticipative decision. On the other hand, the second decision allows the possibility of adapting to the combination of x and ω, and thus it is called an adaptive or anticipative decision (Ermoliev and Wets [1988]).

The objective function in $(SLPR)$ is convex, and under certain regularity conditions, it is continuous on $\mathbf{X} \equiv \{x \mid Ax \leq b, x \geq 0\}$ (Wets [1974]). On the surface, it would appear that one can use any convex programming method to solve SLPR. However, note that for each x, the evaluation of $f(x)$ requires the solution of a linear program

$$h(x, \omega) = \text{Min}\{gy \mid Wy = \omega - Tx, y \geq 0\},$$

for each possible outcome of $\tilde{\omega}$. For example, if $\tilde{\omega}$ is discrete with N outcomes, then at least N linear programs need to be solved implicitly for each evaluation of f. For continuous random variables, direct evaluation of $f(x)$ is typically out of the question.

Our research is aimed at the development of techniques for the solution of problems such as SLPR. Many such methods proceed by defining $\{f_k\}_{k=1}^{\infty}$, a sequence of approximations of f and a sequence of points, $\{x^k\}_{k=1}^{\infty}$, where

$$x^k \in \operatorname{argmin}\{f_k(x) \mid Ax \leq b, x \geq 0\} \tag{1.2}$$

(see, for example, Wets and Van Slyke [1969]). Under certain conditions, to be discussed in Chapter 2, one can show that $\{x^k\}_{k=1}^{\infty}$ accumulates at solutions to SLPR. These methods provide bounds for the objective function but typically require the discretization of continuous distributions. The finer the discretization, the heavier will be the computational load. Another class of methods for the solution of problems such as SLPR, based on methods of subgradient optimization, iterates from a point x^k using a direction d^k, a steplength s_k and the projection operator, $\mathrm{P_X}$ so that

$$x^{k+1} = \mathrm{P_X}(x^k - s_k d^k) \tag{1.3}$$

(see, for example, Ruszczynski [1987]). Typically, d^k estimates a subgradient of f at x^k via sampling, and its determination requires the solution of only a small number of subproblems. Consequently, convergence of $\{x^k\}_{k=1}^{\infty}$ to a solution is accomplished by imposing severe restrictions on the sequence of steplengths, $\{s_k\}_{k=1}^{\infty}$. This class of approaches requires less computational effort in each iteration, but the efficiency depends on the choice of steplengths and the accuracy of the subgradient estimates.

We will explore the combination of these two techniques so that the requirement in (1.2) can be relaxed via subgradient optimization methods in (1.3). In particular, we generate $\{f_k\}_{k=1}^{\infty}$, a sequence of approximations of f successively. In (1.3), we use the subgradient of f_k at x^k to determine the direction d^k. We also use f_k to provide estimated error bounds for the objective function, which will guide the selections of steplengths as well as stopping times of our algorithms.

1.2 Description of Dissertation

In this dissertation, we investigate methods by which subgradient methods can be combined with functional approximation techniques to solve problems such as two stage stochastic linear programs with recourse (SLPR).

In Chapter 2, we review literature on two stage SLPR and discuss its properties, common approximation schemes for the recourse function and various solution procedures. We also study subgradient methods, which are used as the backbone of the algorithms developed in this dissertation. Finally, we discuss the concept of space dilation and its application in subgradient methods.

In Chapter 3, we state the basic inexact subgradient algorithm and show that it inherits many of the asymptotic properties of a standard subgradient method. We also investigate inexact subgradient methods with adaptive steplengths and discuss their applications to the solution of two stage SLPR.

In Chapter 4, we use estimated primal-dual values to define the steplengths in inexact subgradient methods. Limiting results under both epi-convergence and epigraphical nesting of the approximations are presented.

In Chapter 5, we discuss the incorporation of space dilation operation within the basic inexact subgradient method and explore conditions under which these algorithms will lead to an optimal solution.

In Chapter 6, we describe the results of preliminary computational experimentation with the algorithms. Finally, we summarize the conclusions of the work and discuss possible future research.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

It is our goal to solve problems such as the two stage stochastic linear program with recourse (SLPR) by using objective function approximations within subgradient optimization procedures. In §2.1, we review the properties of SLPR. In §2.2, we discuss various approximation schemes used in the solution of SLPR. In §2.3, we study various algorithms for SLPR. In particular, in §2.4, we review the stochastic decomposition algorithm (Higle and Sen [1991a]). In §2.5, we discuss how one can solve a convex program by subgradient methods and how the direction can be perturbed by a space dilation operation. Our conclusions are presented in §2.6.

2.1 Two Stage Stochastic linear program with recourse (SLPR)

In this section, we present the properties of a stochastic linear program with recourse (SLPR) that are pertinent to our development in later chapters.

We restate a SLPR

$$\begin{aligned} \text{Min} \quad & f(x) = cx + E[h(x,\tilde{\omega})] \\ s/t \quad & Ax \leq b \\ & x \geq 0 \end{aligned} \qquad (SLPR)$$

where

$$\begin{aligned} h(x,\omega) = \text{Min} \quad & gy \\ s/t \quad & Wy = \omega - Tx \\ & y \geq 0, \end{aligned} \qquad (S)$$

A, W, T and b are fixed of size $m_1 \times n_1$, $m_2 \times n_2$, $m_2 \times n_1$ and $m_1 \times 1$, respectively.

Let

$$\mathbf{X}_1 = \{x \mid Ax \leq b, x \geq 0\}, \qquad \text{and} \qquad \mathbf{X}_2 = \{x \mid E[h(x,\tilde{\omega})] < \infty\}.$$

The feasible region, $\mathbf{X}$ can be described as $\mathbf{X} = \mathbf{X}_1 \cap \mathbf{X}_2$. If pos($W$)=$\{t \mid t = Wy, y \geq 0\}$, then $\mathbf{X}_2 = \{x \mid \tilde{\omega} - Tx \in \text{pos}(W), \text{ (wp1)}\}$ (Wets [1974]). It follows that in this case, the feasible region is a closed and convex set. If the matrix W is such that pos(W)= $\mathcal{R}^n$, we have $\mathbf{X} = \mathbf{X}_1$. A problem with this property is said to have complete recourse. In particular, when $W = I$ (or $-I$), the problem is said to have simple recourse, and in such cases, it is convenient to write

$$\begin{aligned} h(x,\omega) = \text{Min} \quad & g^+y^+ + g^-y^- \\ s/t \quad & y^+ - y^- = \omega - Tx \\ & y^+, y^- \geq 0. \end{aligned}$$

It follows that the subproblem can be written as:

$$\begin{aligned} h(x,\omega) = \text{Min} \quad & \sum_{i=1}^{n_2} (g_i^+ y_i^+ + g_i^- y_i^-) \\ s/t \quad & y_i^+ - y_i^- = \omega_i - (Tx)_i \qquad i = 1, \ldots, n_2, \\ & y_i^+, y_i^- \geq 0, \end{aligned}$$

and thus h is a separable function of the elements of $\tilde{\omega}$. This property substantially reduces the computational demands of the solution of SLPR. In particular, when $\tilde{\omega}$ is a discrete random variable, the separability property and the convex polyhedral objective allows the problem to be solved with the same efficiency as a linear program of n_1 variables and m_1+n_2 constraints (Wets [1983]). Furthermore, when the components $\tilde{\omega}_i$ are independent, separability of h implies that the expectation can be computed by single dimensional integration, rather than multi-dimensional integration as required in the general SLPR.

As noted in Wets [1974], h is a convex function. One can conclude that f is a convex function, as $E[h(x,\tilde{\omega})]$ is a convex combination of $\{h(x,\omega)\}_{\omega \in \Omega}$. Therefore, SLPR is a convex program. It is the representation of SLPR as a convex program that provides the basis for most of the solution procedures that have been derived.

In SLPR, let $\bar{x} \in \mathbf{X}$ be given. For each realization, $\omega \in \Omega$, let $\pi^*(\bar{x},\omega)$ solve the dual to the subproblem (S)

$$h(\bar{x},\omega) = \text{Max}\{\pi(\omega - T\bar{x}) \mid \pi W \leq g\}.$$

Hence.

$$\pi^*(\bar{x},\omega)(\omega - T\bar{x}) = h(\bar{x},\omega),$$

and since $\pi^*(\bar{x},\omega)W \leq g$, for each $\omega \in \Omega$, we have

$$\pi^*(\bar{x},\omega)(\omega - Tx) \leq h(x,\omega) \qquad \forall\, x \in \mathbf{X}.$$

Taking expectations with respect to ω on both sides,

$$\alpha + \beta\bar{x} \equiv E[\pi^*(\bar{x},\tilde{\omega})(\tilde{\omega} - T\bar{x})] = E[h(\bar{x},\tilde{\omega})], \tag{2.1a}$$

and

$$\alpha + \beta x \equiv E[\pi^*(\bar{x},\tilde{\omega})(\tilde{\omega} - Tx)] \leq E[h(x,\tilde{\omega})] \qquad \forall\, x \in \mathbf{X}. \tag{2.1b}$$

Hence, the affine function defined in (2.1) constitutes a supporting hyperplane

$$H = \{(x, z) \mid z - \beta x = \alpha\},$$

of the epigraph of the recourse function $E[h(x, \tilde{\omega})]$,

$$\{(x, z) \mid z \geq E[h(x, \tilde{\omega})],\ x \in \mathbf{X}\}.$$

In other words, a subgradient of the recourse function, $E[h(x, \tilde{\omega})]$, at $\bar{x}$ can be expressed as $-E[\pi^*(\bar{x}, \tilde{\omega})]T$.

The derivation of supporting hyperplanes provides conceptual guidelines for approximation techniques in many algorithms.

2.2 Approximation Schemes

In this section, we discuss two major classes of approximation techniques for the solution of SLPR, approximations of the objective function and the distribution of $\tilde{\omega}$ (Birge and Wets [1986]).

SLPR can be considered as a problem of the form

$$\begin{aligned} \text{Min} \quad & f(x) \\ s/t \quad & x \in \mathbf{X}, \end{aligned} \tag{2.2}$$

in which exact objective and subgradient evaluations are computationally burdensome. In such cases, it is common to derive a sequence of objective function approximations $\{f_k\}_{k=1}^{\infty}$ and solve

$$\begin{aligned} \text{Min} \quad & f_k(x) \\ s/t \quad & x \in \mathbf{X}. \end{aligned}$$

If $x^k \in \text{argmin}\{f_k(x) \mid x \in \mathbf{X}\}$, conditions need to be imposed on $\{f_k\}_{k=1}^{\infty}$ to ensure that $\{x^k\}_{k=1}^{\infty}$ accumulates at one or more optimal solutions. Within the framework of stochastic programming, the concept of epi-convergence has often been used to guide the manner in which $\{f_k\}_{k=1}^{\infty}$ is defined. We state the definition as follows (Kall [1986]).

Definition: *A sequence* $\{f_k\}_{k=1}^{\infty}$ *is said to epi-converge to* f, $(f_k \xrightarrow{\text{epi}} f)$ *if for every* x*, the following conditions are satisfied.*

i) $\underline{\lim}_{k\to\infty} f_k(x^k) \geq f(x)$, *whenever* $\{x^k\}_{k=1}^{\infty} \to x$;

ii) there exists $\{y^k\}_{k=1}^{\infty} \to x$ *such that* $\overline{\lim}_{k\to\infty} f_k(y^k) \leq f(x)$.

Geometrically, $f_k \xrightarrow{\text{epi}} f$ corresponds to the set convergence of the sequence of epigraphs

$$\text{epi}(f_k) = \{(x,z) \mid f_k(x) \leq z,\ x \in \mathbf{X}\},$$

to the epigraph of f,

$$\text{epi}(f) = \{(x,z) \mid f(x) \leq z,\ x \in \mathbf{X}\}.$$

Optimization algorithms that construct a sequence of epi-convergent functions usually rely on the following proposition.

Proposition 2.1. *(Attouch and Wets [1980].) If the sequence of functions* $\{f_k\}_{k=1}^{\infty}$ *epi-converges to* f *and* $x^k \in \text{argmin}\{f_k(x) \mid x \in \mathbf{X}\}$, *then every accumulation point of* $\{x^k\}_{k=1}^{\infty}$ *is an optimal solution to* (2.2).

We note that the epi-convergence of a sequence of functions constrains the limiting behavior of the functions everywhere on $\mathbf{X}$. Higle and Sen [1992] introduce a type of convergence which relaxes the requirements of epi-convergence and yields a sequence of points that accumulates at the solutions. As previously noted, $f_k \xrightarrow{\text{epi}} f$ implies that $\{\text{epi}(f_k)\}_{k=1}^{\infty}$ converges to $\text{epi}(f)$. In the following definition, $\text{epi}(f)$ is only required to be embedded in $\text{epi}(f_k)$ asymptotically.

Definition: *Let* $\underline{\lim}_{k\to\infty} \text{epi}(f_k) = \{(x,\alpha) \mid \alpha \geq \inf_{\{y^k\}_{k=1}^{\infty}\to x} \overline{\lim}_{k\to\infty} f_k(y^k)\}$. *A sequence of functions* $\{f_k\}_{k=1}^{\infty}$ *satisfies the epigraphical nesting condition if*

$$\text{epi}(f) \subseteq \underline{\lim}_{k\to\infty} \text{epi}(f_k).$$

We note that $\text{epi}(f) \subseteq \underline{\lim}_{k\to\infty} \text{epi}(f_k)$ is equivalent to

$$f(x) \geq \inf_{\{y^k\}_{k=1}^{\infty}\to x} \{\overline{\lim_{k\to\infty}} f_k(y^k)\} \quad \forall x.$$

Hence, asymptotically $\{f_k\}_{k=1}^{\infty}$ provides a lower bound on f. The following proposition from Higle and Sen [1992] indicates that the set convergence associated with epi-convergence of $\{f_k\}_{k=1}^{\infty}$ to f is stricter than is necessary to ensure the accumulation of the solutions $\{x^k\}_{k=1}^{\infty}$ to the solutions of (2.2).

Proposition 2.2. *(Higle and Sen [1992].) Let* $x^k \in \operatorname{argmin}\{f_k(x) \mid x \in \mathbf{X}\}$ *for all* k. *If*

$$\lim_{k \in \mathcal{K}} f_k(x^k) = f(\bar{x}) \quad \textit{and} \quad \operatorname{epi}(f) \subseteq \underline{\lim}_{k \in \mathcal{K}} \operatorname{epi}(f_k), \quad \textit{whenever} \quad \lim_{k \in \mathcal{K}} x^k = \bar{x},$$

then every accumulation point of $\{x^k\}_{k=1}^{\infty}$ *is optimal to (2.2).*

In the later chapters of this dissertation, we will require a slightly stronger form of epigraphical nesting. That is, we will require that

$$\overline{\lim_{k \in \mathcal{K}}} f_k(u^k) \leq f(x) \quad \text{whenever} \quad \{u^k\}_{k \in \mathcal{K}} \to x, \quad \forall\, x \in \mathbf{X}.$$

In addition to using functional approximations, one can approximate the probability measure $\mathcal{P}$ in SLPR by a sequence of measures $\{\mathcal{P}_k\}_{k=1}^{\infty}$. These measures might involve a discretization or partition of Ω in such a way that

$$\int_{\Omega} h(x,\omega) d\mathcal{P}_k(\omega) \to \int_{\Omega} h(x,\omega) d\mathcal{P}(\omega).$$

Algorithmically, it is convenient to define $\mathcal{P}_k$ from a partition of Ω that provides simultaneous upper and lower bounds on the objective value. That is, let Ω be partitioned as $\{\Omega_1, \ldots, \Omega_k\}$. Convexity of f ensures that one can obtain a lower bound on the objective using Jensen's inequality. Specifically, let $\bar{\omega}_i = E[\tilde{\omega} \mid \tilde{\omega} \in \Omega_i]$ and $p_i = P(\tilde{\omega} \in \Omega_i)$, for $i = 1, \ldots, k$. By Jensen's inequality,

$$E[h(x,\tilde{\omega}) \mid \tilde{\omega} \in \Omega_i] \geq h(x, \bar{\omega}_i)$$

for each i, and thus

$$E[h(x,\tilde{\omega})] \geq \sum_{i=1}^{k} p_i h(x, \bar{\omega}_i).$$

In deriving an upper bound on $E[h(x,\tilde{\omega})]$, note that a convex function is maximized at the extreme points of a compact set. We recall that $\tilde{\omega}$ is an n_2-dimensional random variable. Suppose further that the components of $\tilde{\omega}$ are independent and the support set Ω is compact. We specify the cells of the partition as

$$\Omega_i = \times_{j=1}^{n_2}[a_j^i, b_j^i], \qquad i = 1, \ldots, k.$$

It follows from Madansky [1959] that there exists a discrete random variable $\hat{\omega}_i$ with probability mass concentrated on the extreme points of Ω_i such that

$$E[(h,\tilde{\omega}) \mid \omega \in \Omega_i] \leq E[h(x,\hat{\omega}_i)],$$

for $i = 1, \ldots, k$. With $p_i = P(\omega \in \Omega_i)$, We have

$$E[h(x,\tilde{\omega})] \leq \sum_{i=1}^{N} p_i E[h(x,\hat{\omega}_i)].$$

An important feature of these bounds are their monotonicity; the finer the partition, the closer the bounds are to the expectation of h. Let

$$l_k = \text{Min}\{cx + \sum_{i=1}^{k} p_i h(x,\bar{\omega}_i) \mid Ax = b, x \geq 0\},$$

$$u_k = \text{Min}\{cx + \sum_{i=1}^{k} p_i h(x,\hat{\omega}_i) \mid Ax = b, x \geq 0\}.$$

The difference $u_k - l_k$ can be used as to provide a termination criterion as well as a measure of solution quality.

The works by Edmundson [1956] and Madansky [1959] are the basis for the upper bounds on the expectations of functions. Ben-Tal and Hochman [1972] refine the bounds by using additional information of the conditional expectations. Frauendorfer [1988] extends the results of upper bounds to multi-dimensional random variable with dependent components. While most of the upper bounds require bounded support set, Dulá [1989] introduces upper bounds on the expectations of functions defined over an unbounded set.

2.3 Solution methods for SLPR

In this section, we discuss various algorithms for the solutions of SLPR. Many such methods proceed by defining a sequence of approximations $\{f_k\}_{k=1}^{\infty}$, and a sequence of points $\{x^k\}_{k=1}^{\infty}$, where

$$x^k \in \text{argmin}\{f_k(x) \mid Ax \le b, x \ge 0\}.$$

In specifying f_k, it is common to use an outer linearization of f. This is naturally obtained by adding supporting hyperplanes (cutting planes, or cuts) of f iteratively. In SLPR, let $x^k \in \mathbf{X}$ and $\omega \in \Omega$ be given. Suppose that $\pi^*(x^k, \omega)$ solves the dual subproblem

$$h(x^k, \omega) = \text{Max}\{\pi(\omega - Tx^k) \mid \pi W \le g\}. \tag{2.3}$$

It follows from (2.1) that

$$\alpha_k + (c + \beta^k)x^k \equiv cx^k + E[\pi^*(x^k, \tilde{\omega})(\tilde{\omega} - Tx^k)] = cx^k + E[h(x^k, \tilde{\omega})], \tag{2.4a}$$

and

$$\alpha_k + (c + \beta^k)x \equiv cx + E[\pi^*(x^k, \tilde{\omega})(\tilde{\omega} - Tx)] \le cx + E[h(x, \tilde{\omega})], \tag{2.4b}$$

where α_k is a scalar and β^k is a point in $\mathcal{R}^{n_1}$. Hence, the affine function defined in (2.4) constitutes a supporting hyperplane of f at x^k, and is therefore a valid cutting plane.

A classic representative of the functional approximation techniques is the L-shaped method presented by Van Slyke and Wets [1969]. This method, which is essentially motivated by cutting plane methods for deterministic problems, is an implementation of Benders' decomposition for SLPR (Benders [1962]). In each iteration, two types of cuts: feasibility and optimality cuts are generated and are added to a master program. In deriving the feasibility cuts, suppose that Ω has N finite outcomes and $p_i = P(\tilde{\omega} = \omega_i), i = 1, \cdots, N$. If x^k is infeasible (i.e. $x^k \notin X_2$, which is impossible for problems with complete recourse,) there exists $\omega^i \in \Omega$ such that $\omega^i - Tx^k \notin \text{pos}(W)$. Correspondingly, there exists σ such that

$$\sigma(\omega^i - Tx^k) > 0, \quad \text{and}$$
$$\sigma t \leq 0, \quad \forall t \in \text{pos(W)},$$

(Bazaraa and Shetty [1979]). Therefore, $\sigma(\omega^i - Tx) \leq 0$ becomes a condition for feasibility and is appended to the master program as a constraint. The optimality cuts are identical to those defined in (2.4), with (α_k, β^k) denoting the coefficients of the cut derived from x^k. If there are n_k feasibility cuts and m_k optimality cuts after k iterations, the master program is given by

$$\begin{aligned} \text{Min} \quad & cx + \eta \\ s/t \quad & Ax \leq b \\ & \sigma^t(\omega^{k(t)} - Tx) \leq 0 \qquad t = 1, \ldots, n_k \qquad k(t) \in \{1, \cdots, N\} \\ & \eta \geq \alpha_t + \beta^t x \qquad t = 1, \ldots, m_k \\ & x \geq 0. \end{aligned} \tag{2.5}$$

Suppose that (x^k, η_k) solves (2.5) and x^k is feasible with its optimality cut indexed by $m_k + 1$. If $\eta_k \geq \alpha_{m_k+1} + \beta^{m_k+1}x^k$, the process can terminate; otherwise an optimality cut $\{\alpha_{m_k+1}, \beta^{m_k+1}\}$ is appended to the master program (2.5). The approximation can be written as

$$f_k(x) = cx + \text{Max}\{\alpha_t + \beta^t x \mid t = 1, \ldots, m_k\}.$$

The drawback of this method is the amount of work required to solve the subprograms (2.3), especially when the number of outcomes is large. An extension of the L-shaped method is presented in Birge and Louveaux [1988] which is another way to implement Benders' decomposition. In each iteration, an optimality cut is added for each

$$h(x, \omega^i) = \text{Min}\{gy \mid Wy = \omega^i - Tx, \; y \geq 0\}, \quad i = 1, \ldots, N,$$

instead of adding a single optimality cut to $E[h(x,\tilde{\omega})]$. If there are n_k feasibility cuts and $m_{i(k)}$ optimality cuts for $1 \leq i \leq N$, after k iterations, the master program becomes

$$\begin{aligned}
\text{Min} \quad & cx + \sum_{i=1}^{N} \eta_i \\
s/t \quad & Ax \leq b \\
& \sigma^t(\omega^{k(t)} - Tx) \leq 0 \qquad t = 1, \ldots, n_k \qquad k(t) \in \{1, \cdots, N\} \\
& \eta_i \geq p_i \left(\alpha_{i(t)} + \beta^{i(t)} x\right) \qquad t = 1, \ldots, m_{i(k)} \qquad i = 1, \ldots, N \\
& x \geq 0.
\end{aligned}$$

Empirically, Birge and Louveaux [1988] note that although the multiple cuts in each iteration provide more information than a single cut, allowing termination of the algorithm with fewer iterations than the L-shaped method, the proliferation of cuts can become burdensome. Ruszczynski [1986] offers a similar decomposition type algorithm which also uses multiple cuts in defining f_k. The main differences are that a quadratic regularizing term is augmented to the master program and the number of cuts is less than $n_1 + 2N$. Other algorithms based on the L-shaped method include those that use upper and lower bounds on the recourse functions within a sample space partitioning scheme due to Frauendorfer and Kall [1986] and Frauendorfer [1988]. In the case of continuous random variables, the distributions are often discretized during the initialization of this class of methods.

Another class of techniques for SLPR, known as the stochastic quasigradient (SQG) methods, iterates from a point x^k using direction d^k and steplength s_k so that

$$x^{k+1} = \mathrm{P}_{\mathbf{X}}\{x^k - s_k d^k\}. \tag{2.6}$$

In each iteration, an observation of $\tilde{\omega}$ is used to estimate a subgradient of f at x^k. For example, it can be accomplished by randomly generating a collection of observations of $\tilde{\omega}$, $\{\omega^t\}_{i=1}^{N_k}$ and estimating a subgradient of f at x^k by

$$d^k = c - \frac{1}{N_k} \sum_{i=1}^{N_k} \{\pi^*(x^k, \omega^t) T\},$$

where $\pi^*(x^k, \omega^t)$ solves the dual subproblem $h(x^k, \omega^t)$. With $\{x^k\}_{k=1}^{\infty}$ defined in (2.6), the minimal requirement for the steplength to guaranteed convergence of $\{x^k\}_{k=1}^{\infty}$ to optimal solutions is

$$s_k \geq 0, \quad s_k \to 0, \quad \sum_{k=1}^{\infty} s_k = \infty, \quad \text{and} \quad \sum_{k=1}^{\infty} s_k^2 < \infty$$

(Urasiev [1989]). The source of the SQG algorithm can be traced back to Robbins and Monro [1951] which is motivated by finding the quantile of a distribution. Kiefer and Wolfowitz [1952] present the first method designed to use difference approximations for the gradients in stochastic minimization problems. The development of this class of methods and their application is discussed in Ermoliev [1983]. In Ruszczynski [1987], a sequence of stochastic quasigradients is aggregated to 'estimate' the subgradient at x^k, although the estimates are accurate only asymptotically. This aggregated estimate is then used as an input for a direction finding subproblem from which d^k is specified. Beyond our statement of SLPR, Ruszczynski [1987], finds stationary points when the objective function is weakly convex, and the constraints are convex and differentiable.

Although the convergence of the sequence to optimality is established and this method can handle continuous distributions without discretization, it is practically more important to design a process which terminates after a finite number of iterations within a neighborhood of a solution, x^*. A termination rule suggested by Pflug [1989] is to let ϵ and α respectively denote the size of a confidence region and the desired confidence level. The optimization process in (2.6) will stop at τ if

$$P_{x^1}\{\|x^\tau - x^*\| \leq \epsilon\} \geq 1 - \alpha, \quad \forall x^1 \in \mathbf{X},$$

where x^1 is the starting point. The derivation of a stopping time, τ is based on the sequence of *random* points $\{x^k\}_{k=1}^{\infty}$, while the asymptotic behavior of $\{x^k\}_{k=1}^{\infty}$ is controlled by the choice of steplengths. Hence, various steplength prescriptions will provide their pertaining stopping times τ. As a result, the success of this class of methods depends on the selection of steplengths.

A recent approach presented by Rockafellar and Wets [1991], called scenario aggregation, formulates a two stage (or in general a multi-stage) SLPR as a problem of identifying an optimal policy $(x^*(\omega), y^*(\omega))$ for all $\omega \in \Omega$. This algorithm is based on relaxing the nonanticipativity constraints, $x(\omega) = x$ for all $\omega \in \Omega$, via an augmented Lagrangian, and coordinating the decisions corresponding to the alternative scenarios via price adjustments. These adjustments are guided by a modification of the augmented Lagrangian approach. This method has also been extended to nonseparable problems by Robinson [1991].

Motivated by the L-shaped method, Infanger [1989] uses an "Importance sampling" procedure to obtain statistical estimates of the supporting hyperplanes while enjoying a reduction in the computational load associated with the solution of the subprogram. Another algorithm, stochastic decomposition (SD), proposed by Higle and Sen [1991a] also combines functional approximation with sampling, and is discussed in the following section.

2.4 Stochastic Decomposition

On occasion, we will suggest implementations of our algorithms that use functional approximations similar to those in the Stochastic Decomposition algorithm (SD) (Higle and Sen [1991a]). Like the L-shaped method, SD constructs a sequence of piecewise linear objective function approximations, $\{f_k\}_{k=1}^{\infty}$. Departing from deterministically motivated techniques, SD uses a random sample of $\tilde{\omega}$, and thus requires the previous cuts to be updated each time the sample size is increased. As a result, instead of providing supporting hyperplanes of the objective function, f, at certain points, f_k remains a statistically valid lower bound of f. That is, given the observations, $\{\omega^t\}_{t=1}^{k}$, one has that

$$f_k(x) \leq cx + \frac{1}{k}\sum_{t=1}^{k} h(x,\omega^t). \tag{2.7}$$

The specification of f_k is based on the dual to (S)

$$\begin{aligned} \text{Max} \quad & \pi(\omega - Tx) \\ s/t \quad & \pi W \leq g. \end{aligned}$$

Let V represent the set of extreme points of the dual feasible region of the subprogram and V_k be the set of extreme points identified through iteration k. For $t = 1, \ldots, k$, let

$$\pi_t^k \in \text{argmax}\{\pi(\omega^t - Tx^k) \mid \pi \in V_k\}.$$

Since $\pi_t^k \in V_k \subset V$, it follows that

$$\begin{aligned} \pi_t^k(\omega^t - Tx) &\leq \text{Max}\{\pi(\omega^t - Tx) \mid \pi \in V\} \\ &= h(x,\omega^t). \end{aligned} \tag{2.8}$$

Let

$$f_k(x) = cx + \text{Max}\{\alpha_t^k + \beta_t^k x \mid t = 1, \ldots, k\},$$

where the superscript on the cut coefficients corresponds the current iteration and the subscript refers to the iteration in which the cut was derived. In iteration k, cuts are derived from k observation of $\tilde{\omega}$, $\{\omega^t\}_{t=1}^k$. That is,

$$\alpha_k^k + \beta_k^k x \equiv \frac{1}{k}\sum_{t=1}^{k} \pi_t^k(\omega^t - Tx).$$

It follows from (2.8) that

$$\alpha_k^k + (\beta_k^k + c)x \leq cx + \frac{1}{k}\sum_{t=1}^{k} h(x,\omega^t).$$

Note that in their initial configuration, the cuts are based on unequal sample sizes. Thus, as a new observation is drawn, the cuts $\left\{\alpha_t^k, \beta_t^k\right\}_{t=1}^k$ are updated to $\left\{\alpha_t^{k+1}, \beta_t^{k+1}\right\}_{t=1}^k$ in a manner that preserves the inequality

$$\alpha_t^{k+1} + (\beta_t^{k+1} + c)x \leq cx + \frac{1}{k+1}\sum_{t=1}^{k+1} h(x,\omega^t).$$

In particular, assuming $h(x,\tilde{\omega}) \geq 0$ with probability one for all x, one can update as follows:

$$\left\{\alpha_t^{k+1}, \beta_t^{k+1}\right\} \quad \leftarrow \quad \frac{k}{k+1}\left\{\alpha_t^k, \beta_t^k\right\}, \quad t = 1, \cdots, k.$$

If h is not necessarily nonnegative, update procedures derived from dual feasibility of the subproblem are easily implemented, as described in Higle, Sen, and Yakowitz [1990]. We summarize the limiting results of the SD algorithm as follows.

Proposition 2.3. *(Higle and Sen [1991a]) Let $\{f_k\}_{k=1}^{\infty}$ denote the sequence of objective function approximations derived in SD and $\{x^k\}_{k=1}^{\infty}$ denote the sequence of iterates such that*

$$x^{k+1} \in \operatorname{argmin}\{f_k(x) \mid x \in X\}.$$

If

- *the feasible region associated with the dual of the recourse subproblem is a* *nonempty compact convex polyhedral set,*

· **X** *and* Ω *are nonempty compact sets under the Euclidean metric,*

then,

1) there exists a subsequence indexed by $\mathcal{K}^*$ *such that every accumulation point of* $\{x^k\}_{k\in\mathcal{K}^*}$ *is optimal,*

2) $\forall\, x \in \mathbf{X}$, $\overline{\lim}_{k\to\infty} f_k(u^k) \le f(x)$ *whenever* $\{u^k\} \to x$ *(wp1),*

3) $\{f_k(x^k)\}_{k=1}^{\infty} \to f(\bar{x})$ *whenever* $\{x^k\}_{k=1}^{\infty} \to \bar{x}$, *(wp1),*

4) there exists an index set $\bar{\mathcal{K}}$ *such that* $\lim_{k\in\bar{\mathcal{K}}} f_k(x^k) - f_{k-1}(x^k) = 0$.

The second property follows directly from (2.7). In the third property, the sequence of functional approximations $\{f_k\}_{k=1}^{\infty}$ is accurate asymptotically with respect to $\{x^k\}_{k=1}^{\infty}$. We note that this property holds for any sequence of points with respect to which the cuts are derived.

To terminate the algorithm, Higle and Sen [1991b] suggest that the optimality of x^k can be tested by using the generalized Kuhn-Tucker conditions. That is, they test whether there exist λ and μ such that

$$\begin{aligned} \lambda A + \mu &\in \partial f(\bar{x}^k) \\ \lambda(b - Ax) &= 0 \\ \mu^T \bar{x}^k &= 0 \\ \lambda \ge 0,\ \mu &\ge 0 \end{aligned} \tag{2.9}$$

(Clark [1983]). Solving a quadratic program from (2.9), one can obtain a point estimate of an upper bound on the objective value error $f(x^k) - f(x^*)$, where x^* is an optimal value. To assess the variability of the error bound estimators, the estimator is sampled multiple times using a method derived from the 'Bootstrap' procedure of Efron [1979]. The resulting empirical distribution is used to statistically verify optimality of the point x^k.

2.5 Subgradient Methods

Subgradient methods have been used to solve convex programs such as:

$$\begin{aligned} \text{Min} \quad & f(x) \\ s/t \quad & x \in \mathbf{X}, \end{aligned} \tag{2.10}$$

where the objective function is not necessarily differentiable. A subgradient method generates a sequence of points, $\{x^k\}_{k=1}^{\infty}$ as follows:

$$x^{k+1} = \mathrm{P}_{\mathbf{X}}(x^k - s_k d^k),$$

where s_k is the steplength and d^k is a subgradient of f at x^k. It can be shown that if

$$s_k \geq 0, \quad s_k \to 0 \quad \text{and} \quad \sum_{k=1}^{\infty} s_k = \infty,$$

int($\mathbf{X}$)$\neq \emptyset$ and $\{d^k\}_{k=1}^{\infty}$ is bounded, then every accumulation point of $\{x^k\}_{k=1}^{\infty}$ is an optimal solution to (2.10) (Shor [1985]).

Polyak [1969] considers a subgradient method in which

$$s_k = \lambda_k \frac{f(x^k) - \bar{f}}{\|d^k\|^2}, \quad 0 < \alpha \leq \lambda_k \leq \beta < 2,$$

and $\bar{f}$, the target value is an estimation of the optimal value, f^*. This method has been generalized by Allen et. al. [1987] who weaken the conditions imposed on $\{\lambda_k\}_{k=1}^{\infty}$ so that

$$s_k = \lambda_k \frac{f(x^k) - \bar{f}}{\|d^k\|^2}, \quad 0 < \lambda_k \leq \beta < 2, \quad \text{and} \quad \sum_{k=1}^{\infty} \lambda_k = \infty.$$

With this relaxation, the sequence $\{\lambda_k\}_{k=1}^{\infty}$ is allowed to converge to zero. Kim et. al. [1991] note that the target value is fixed throughout the process. They modify this kind of steplengths specification using adaptive target values so that

$$s_k = \lambda_k \frac{f(x^k) - \bar{f}_k}{\|d^k\|^2}, \quad 0 < \alpha \leq \lambda_k \leq \beta < 2,$$

$\bar{f}_k \leq f(x^k)$ and $\{\bar{f}_k\}_{k=1}^{\infty}$ is a monotone decreasing sequence bounded from below.

Based on a primal/dual relationship, Sen and Sherali [1986] provide steplengths for a pair of primal-dual iterates $\{x^k\}_{k=1}^{\infty}$ and $\{\pi^k\}_{k=1}^{\infty}$. Consider the problem

$$\begin{aligned} \text{Min} \quad & f(x) \\ s/t \quad & f_i(x) \leq 0 \qquad i = 1, \ldots, m \\ & x \in \mathbf{X} \subset \mathcal{R}^n, \end{aligned} \tag{2.11a}$$

where f and f_i, $i = 1, \ldots, m$, are convex functions with bounded subgradients on $\mathbf{X}$, a convex compact set. Letting $\pi = (\pi_1, \ldots, \pi_m)$, the lagrangian dual is given by solving

$$\begin{aligned} \text{Max} \quad & L(\pi) \\ s/t \quad & \pi_i \geq 0 \qquad i = 1, \ldots, m, \end{aligned} \tag{2.11b}$$

where

$$L(\pi) = \text{Min}\{f(x) + \sum_{i=1}^{m} \pi_i f_i(x) \mid x \in \mathbf{X}\}$$

(Bazaraa and Shetty [1979]). Note that optimality conditions ensure that $\sum_{i=1}^{m} \pi_i^k f_i(x^k) = 0$ whenever x^k and π^k are optimal to (2.11a) and (2.11b) respectively. Thus, a penalty function H such that

$$H(x^k) \geq f(x^k) + \sum_{i=1}^{m} \pi_i^k f_i(x^k) \geq L(\pi^k) \qquad \forall k,$$

with equalities holding throughout if and only if x^k and π^k are optimal to (2.11). is introduced.

Let η^k, δ^k respectively denote the subgradients of the penalty function H at x^k and the lagrangian function L at π^k. The steplength is

$$s_k = \lambda_k \frac{[H(x^k) - L(\pi^k)]}{\|\eta^k\|^2 + \|\delta^k\|^2},$$

where the coefficients $\{\lambda_k\}_{k=1}^{\infty}$ are subjected only to the condition, $0 < \lambda_k < 2$. Letting $\mathbf{\Pi} = \{\pi \mid \pi_i \geq 0, \ i = 1, \ldots, m\}$, the primal-dual subgradient algorithm proceeds according to

$$x^{k+1} = \mathrm{P}_{\mathbf{X}}(x^k - s^k \eta^k), \qquad \pi^{k+1} = \mathrm{P}_{\mathbf{\Pi}}(\pi^k + s^k \delta^k).$$

Under regularity conditions, the sequences $\{x^k\}_{k=1}^{\infty}$ and $\{\pi^k\}_{k=1}^{\infty}$ converge to an optimal point of (2.11), and $\{H(x^k)-L(\pi^k)\}_{k=1}^{\infty} \to 0$. Furthermore, the sequences $\{H(x^k)\}_{k=1}^{\infty}$ and $\{L(\pi^k)\}_{k=1}^{\infty}$ converge to the optimal value, f^*.

The function $H(x^k) - L(\pi^k)$ is referred to as the gap function, $G(x^k, \pi^k)$ (Sen and Sherali [1986]). It can be shown that it is a non-negative convex function defined on the set $\mathbf{Z} = \{(x, \pi) \mid x \in \mathbf{X}, \pi \geq 0\}$. Alternatively, with

$$s_k = \lambda_k \frac{G(x^k, \pi^k)}{\|d^k\|^2}, \quad d^k \in \partial G(x^k, \pi^k) \quad \text{and} \quad 0 < \lambda_k < 2,$$

the primal-dual subgradient method can be stated as

$$(x^{k+1}, \pi^{k+1}) = \mathrm{P}_{\mathbf{Z}}\{(x^k, \pi^k) - s_k d^k\}.$$

In this case, the target value suggested in Polyak's method is zero. Hence, no prior knowledge of the optimal value is required.

In primal-dual subgradient methods, the value of the gap function provides a natural basis for a termination criterion. However, the above procedure requires the exact evaluations of the objective, penalty or lagrangian functions, which will become computationally burdensome for the solution of SLPR. Within the framework of stochastic optimization, we recall from §2.3 that it is common to replace the objective function by an approximation and solve the corresponding approximated problem. In Chapter 3 and 4, we will develop algorithms by combining functional approximation techniques with subgradient optimization methods.

A subgradient method would become slow if the subgradient direction of the iterate is orthogonal to its direction to an optimal point. In keeping with the variable metric algorithm of nonlinear programming (e.g. Davidon [1959] and Broyden [1967]), Shor [1985] suggests using operators of space dilation in subgradient methods to change angles between the subgradient direction of the iterate and its direction to an optimal point. To define space dilation, let $\xi \in \mathcal{R}^n$ such that $\|\xi\| = 1$. For any point $x \in \mathcal{R}^n$, there exists a scalar $\gamma(\xi, x)$ and a point $y(\xi, x) \in \mathcal{R}^n$ such that

$$x = \gamma(\xi, x)\xi + y(\xi, x), \tag{2.12}$$

where

$$\xi y(\xi, x) = 0.$$

It follows that

$$\gamma(\xi, x) = x\xi \quad \text{and} \quad y(\xi, x) = x - (x\xi)\xi.$$

With (2.12), we can define space dilation as follows.

Definition: *(Shor [1985]) Let $\alpha \geq 0$ be a fixed real value and $\xi \in \mathcal{R}^n$ such that $\|\xi\| = 1$.*

$$R(\alpha, \xi)(x) = \alpha\gamma(\xi, x)\xi + y(\xi, x)$$

is called an operator of space dilation along direction ξ with coefficient α.

The notation, $R(\alpha, \xi)$, reflects the dependence of the operator on the fixed scalar α and point ξ. If $0 < \alpha < 1$, then the component of x parallel to ξ will be reduced. Let $\alpha \geq 0$ be a fixed real value and $\xi \in \mathcal{R}^n$ such that $\|\xi\| = 1$. The following standard properties are useful for our development.

Proposition 2.4. *(Shor [1985]) Let $\alpha \geq 0$ be a fixed real value and $\xi \in \mathcal{R}^n$ such that $\|\xi\| = 1$. The following statements are true.*

1. $R(\alpha, \xi)(x) = x + (\alpha - 1)(x\xi)\xi$.

2. The matrix representation of $R(\alpha, \xi)$ is $I + (\alpha - 1)\xi\xi^T$. Hence, it is a symmetric linear operator.

3. For $\alpha > 0$, $R(\alpha, \xi)^{-1} = R(\frac{1}{\alpha}, \xi)$.

We now proceed to discuss the application of space dilations in subgradient methods by considering an unconstrained convex program,

$$\begin{aligned} \text{Min} \quad & f(x) \\ s/t \quad & x \in \mathcal{R}^n. \end{aligned} \tag{2.13}$$

The following algorithm with space dilations along the previous subgradients for the solutions of (2.13) is presented in Shor [1985].

Subgradient Algorithm with Space Dilations

Step 0. $k \leftarrow 0$. x^1, $0 < \beta_1 < 1$, $B_1 = I$ are given.

Step 1. $k \leftarrow k + 1$. Let $\xi^k \in \partial f(x^k)$.

If $\xi^k = 0$, stop; otherwise $d^k = \frac{B_k^T(\xi^k)}{\|B_k^T(\xi^k)\|}$.

Step 2. Specify s_k.

Step 3. $x^{k+1} = x^k - s_k B_k d^k$.

Step 4. Specify $0 < \beta_k < 1$. $B_{k+1} = B_k R(\beta_k, d^k)$.

Repeat from Step 1.

We note that $\{R(\beta_k, d^k)\}_{k=1}^{\infty}$ are space dilation operators along d^k with coefficient, β_k. Since B_k is defined as a composition of $\{R(\beta_i, d^i)\}_{i=1}^{k-1}$, it follows from Proposition 2.4 that the transpose of B_k is also a composition of $\{R(\beta_i, d^i)\}_{i=1}^{k-1}$. Proposition 2.4 ensures the existence of B_k^{-1}. We can interpret Step 3 as a subgradient method under the transformation B_k^{-1}. To see this, let $u^{k+1} = B_k^{-1}(x^{k+1})$ and $u^k = B_k^{-1}(x^k)$. Step 3 becomes

$$u^{k+1} = u^k - s_k d^k.$$

It follows from the convexity of f that

$$f(x) \geq f(x^k) + \xi^k(x - x^k) \qquad \forall\, x \in \mathcal{R}^n.$$

Let $u = B_k^{-1}x$ and we have

$$f(B_k u) \geq f(B_k u^k) + \xi^k(B_k u - B_k u^k) \qquad \forall\, u \in \mathcal{R}^n.$$

Let $\phi_k(u) = f(B_k u)$ and it follows that

$$\phi_k(u) \geq \phi_k(u^k) + B_k^T(\xi^k)(u - u^k) \qquad \forall\, u \in \mathcal{R}^n.$$

Thus, $d^k = \frac{B_k^T(\xi^k)}{\|B_k^T(\xi^k)\|}$ is an normalized subgradient of ϕ_k at u^k, and the iterate in Step 3 is defined via a subgradient method. Also, B_k is a composition of space dilation operators along the subgradients of ϕ_k.

This class of algorithms is also referred to as generalized gradient methods with variable metric. That is, if we let $M_k = B_k B_k^T$, the basic iteration in Step 3 becomes

$$x^{k+1} = x^k - s_k \frac{M_k(\xi^k)}{\sqrt{\xi^k M_k \xi^k}},$$

and

$$M_{k+1} = M_k + (1-\beta_k^2)\frac{(M_k\xi^k)(M_k\xi^k)^T}{(\xi^k)^T M_k \xi^k}.$$

(Akgul [1984]). Nevertheless, each prescription of $\{\beta_k\}_{k=1}^{\infty}$ and $\{s_k\}_{k=1}^{\infty}$ will give a specific algorithm. For example, when $\beta_k = 1$ for all k, the above algorithm becomes a standard subgradient method.

With $\{\beta_k\}_{k=1}^{\infty}$ and $\{s_k\}_{k=1}^{\infty}$ specified in the following implementable fashion, the following proposition, from Shor [1985], can be used to construct an algorithm for (2.13).

Proposition 2.5. *Let $\{x^k\}_{k=1}^{\infty}$ denote the sequence of iterates generated by the algorithm with space dilations and x^* be an optimal solution to* (2.13). *Suppose the following assumptions are satisfied.*

1)f is a convex function,

2)there exists $M \geq 1$, such that $\xi(x-y) \leq M[f(x)-f(y)]$, $\xi \in \partial f(x)$ whenever, $h(\alpha) = f[(1-\alpha)x + \alpha y]$ is strictly decreasing in $\alpha \in [0,1]$,

3) $\lim_{\|x\|\to\infty} f(x) = +\infty$,

4) $\beta_k = \frac{M-1}{M+1}$, for all k,

5) $s_k = \frac{2M}{M+1}\frac{f(x^k)-\bar{f}}{\|B_k^t(\xi^k)\|}$ $\forall\, k$,

If, in addition, $\bar{f} \geq f(x^*)$, then the sequence $\{s_k\}_{k=1}^{\infty}$ is bounded and for any $\epsilon > 0$, there exists $\bar{k}$, such that $f(x^{\bar{k}}) < \bar{f} + \epsilon$; otherwise $\{s_k\}_{k=1}^{\infty}$ is unbounded.

As mentioned in Shor [1985], in many practical problems, the constant M can be chosen to be within the limits of 3 to 10. $\bar{f}$ is adjusted successively to ensure that there exists a subsequence of $\{f(x^k)\}_{k=1}^{\infty}$ approaching $f(x^*)$.

This class of methods corresponds to the conjugate gradient methods in smooth optimization in the sense that the previous gradients together with the current one are used to produce a new direction. This concept has been implemented by Ruszczynski and Syski [1983] to solve an unconstrained stochastic program. Here, the function f in (2.13) is assumed to be continuously differentiable and

$$f(x) = \int_{\Omega} F(x,\omega)d\mathcal{P}(\omega),$$

where $\tilde{\omega} \in (\Omega, \mathcal{A}, \mathcal{P})$. They generate a sequence of points as follows:

$$x^{k+1} = x^k - s_k d^k,$$

where

$$s_k \geq 0, \quad s_k \to 0 \quad \text{and} \quad \sum_k s_k = \infty.$$

The direction d^k is a convex combination of the current quasi-gradient obtained via sampling and the previous direction, thus imposing a stabilizing effect on the sequence $\{x^k\}_{k=1}^{\infty}$. We note that this is a modification of the SQG methods in which d^k is just the current quasi-gradient.

In Chapter 5, we will incorporate the space dilation methods into our algorithms to generate perturbed directions.

2.6 Conclusions

In this chapter, we have discussed the properties of stochastic linear program with recourse (SLPR), various approximation schemes and solution procedures. Additionally, we have reviewed some subgradient optimization methods and the application of space dilation operations within a subgradient method. In later chapters, we develop algorithms which combine objective function approximation techniques with subgradient methods to solve problems such as SLPR.

CHAPTER 3

INEXACT SUBGRADIENT METHODS

3.0 Introduction

In a mathematical program

$$\begin{aligned} \text{Min} \quad & f(x) \\ s/t \quad & x \in \mathbf{X}, \end{aligned} \tag{P}$$

where the objective function is non-differentiable, subgradient optimization methods generate a sequence of iterates $\{x^k\}_{k=1}^{\infty}$ so that

$$x^{k+1} = \mathrm{P}_{\mathbf{X}}(x^k - s_k d^k), \quad d^k \in \partial f(x^k).$$

$\mathrm{P}_{\mathbf{X}}$ is the projection operator onto the feasible set $\mathbf{X}$, s_k is the steplength and the direction, d^k is a subgradient of f at x^k. The standard requirement on the steplengths to ensure convergence are

$$s_k \geq 0, \quad s_k \rightarrow 0 \quad \text{and} \quad \sum_{k=1}^{\infty} s_k = \infty. \tag{3.1}$$

In some cases, the evaluations of the objective function and its subgradients may pose a computational burden. For example, in stochastic linear program with recourse (SLPR), the objective function involves an expectation with respect to a random variable, $\tilde{\omega}$, defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$. An application of subgradient algorithms to stochastic optimization has lead to stochastic quasigradient (SQG) methods described in §2.3. In SQG methods, the direction is a statistical estimate of the subgradient of f at x^k, (also known as a stochastic quasigradient of f at x^k). To ensure that this method can lead to optimal solutions, the typical guideline for the choice of steplengths is

$$s_k \geq 0, \; s_k \rightarrow 0, \; \sum_{k=1}^{\infty} s_k = \infty, \text{ and } \sum_{k=1}^{\infty} s_k^2 < \infty. \tag{3.2}$$

In this chapter, we will discuss subgradient methods that are based on a sequence of approximations of the objective function. We refer to this class of methods as inexact subgradient methods. Ideally, these functional approximations will be designed so that the subgradient evaluations are easily accomplished. Within our development, the sequence of functional approximations, $\{f_k\}_{k=1}^{\infty}$, is required to satisfy a special version of the epigraphical nesting property which is developed by Higle and Sen [1992], so that

$$\overline{\lim_{k\to\infty}} f_k(u^k) \leq f(x) \quad \text{whenever} \quad \{u^k\}_{k=1}^{\infty} \to x, \ \forall x \in \mathbf{X}.$$

Note that in this case, $\{f_k\}_{k=1}^{\infty}$ provides a lower bound on f, asymptotically.

This chapter is organized as follows. In §3.1, we develop the basic inexact subgradient methods and show that optimal solutions can be identified as in standard subgradient methods. In §3.2, we define the steplengths adaptively so that the conditions imposed on $\{s_k\}_{k=1}^{\infty}$ in (3.1) are relaxed. We apply this method to solve SLPR in §3.3. Our conclusions are presented in §3.4.

3.1 Inexact Subgradient Methods

In subgradient methods, when the function f is replaced by a sequence of approximations $\{f_k\}_{k=1}^{\infty}$, the basic method proceeds as follows:

Basic Inexact Subgradient Algorithm

Step 0. Initialize $k \leftarrow 0$. $x^1 \in \mathbf{X}$ is given.

Step 1. $k \leftarrow k+1$. Specify f_k, s_k.

Step 2. $d^k \in \partial f_k(x^k)$.

Step 3. $x^{k+1} = \mathrm{P}_{\mathbf{X}}\{x^k - s_k d^k\}$, and repeat from Step 1.

Note that Step 1 offers a great deal of flexibility on the choices of functional approximations and steplengths. Although our iterative process looks similar to a standard subgradient method, d^k in Step 2 is a subgradient of the approximation, f_k, at x^k. We include the following assumptions to ensure that the above algorithm will lead to an optimal solution. Throughout our development, x^* and f^* denote an optimal solution to (P) and the optimal value, respectively. Our asymptotic analysis is based on the following assumptions.

Assumptions

A1 f is continuous and $\mathbf{X}$ is a compact convex set.

A2 $\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below on $\mathbf{X}$.

A3 $\overline{\lim}_{k\to\infty} f_k(u^k) \leq f(x)$ whenever $\{u^k\}_{k=1}^{\infty} \to x, \quad \forall\, x \in \mathbf{X}$.

A4 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ whenever $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

A5 $\{d^k\}_{k=1}^{\infty}$ is a bounded sequence.

Brief comments regarding assumptions A1-A5 are in order. By assuming that f is continuous, we implicitly assume that $\mathbf{X}^*$, the set of optimal solutions to (P), is a closed set. A3 ensures that $\{f_k\}_{k=1}^{\infty}$ provides a lower bound on f asymptotically. In A4, the sequence of objective value approximations, $\{f_k(x^k)\}_{k=1}^{\infty}$ is required to be asymptotically accurate. This is only required as $\{x^k\}_{k=1}^{\infty}$ accumulates and is not required elsewhere on $\mathbf{X}$. Moreover, we note that the structure of the algorithm permits local improvement in the functional approximation near x^k, and thus this assumption is not unwieldy. A5 imposes a uniform bound on the subgradients defined in Step 2 of the basic algorithm. This assumption is easily satisfied if , for example, the functional approximations share a common Lipschitz constant.

In our asymptotic analysis of the inexact subgradient method, we will make use of the following index set:

$$\mathcal{K}(\epsilon) = \{k \mid f_k(x^k) \leq \operatorname*{Max}_{x^* \in \mathbf{X}^*} f_k(x^*) + \epsilon\}, \tag{3.3}$$

where $\epsilon > 0$. We will identify conditions under which the iterates, $\{x^k\}_{k=1}^{\infty}$ accumulate at the elements of $\mathbf{X}^*$, and thus we define

$$d(x, \mathbf{X}^*) = \operatorname*{Min}_{x^* \in \mathbf{X}^*} \|x - x^*\|^2. \tag{3.4}$$

Finally, we define

$$D(\epsilon) = \overline{\lim_{k \in \mathcal{K}(\epsilon)}}\, d(x^k, \mathbf{X}^*) \tag{3.5}$$

and note that $D(\epsilon)$ is well defined if, and only if, $\mathcal{K}(\epsilon)$ is an infinite set. Note that assumptions A1-A3 ensure that

$$\overline{\lim_{k \to \infty}} \operatorname*{Max}_{x^* \in \mathbf{X}^*} f_k(x^*) \leq f^*, \tag{3.6}$$

while A4 ensures that $\{f_k(x^k)\}_{k \in \mathcal{K}} \to f(\bar{x})$ whenever $\{x^k\}_{k \in \mathcal{K}} \to \bar{x}$. Thus, if $\bar{x}$ is an accumulation point of $\{x^k\}_{k \in \mathcal{K}(\epsilon)}$, (3.3) and (3.6) ensure that

$$f^* \leq f(\bar{x}) \leq f^* + \epsilon. \tag{3.7}$$

Moreover, it follows that $d(\bar{x}, \mathbf{X}^*)$ is at most $D(\epsilon)$. In the following lemma, we verify that $\lim_{\epsilon \to 0} D(\epsilon) = 0$.

Lemma 3.1. *Let $\{x^k\}_{k=1}^{\infty}$ be the sequence of iterates generated by the inexact subgradient algorithm, and suppose that assumptions A1-A4 hold.*

A1 *f is continuous and $\mathbf{X}$ is a compact convex set.*

A2 *$\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below on $\mathbf{X}$.*

A3 $\overline{\lim}_{k \to \infty} f_k(u^k) \leq f(x)$ *whenever* $\{u^k\}_{k=1}^{\infty} \to x, \quad \forall\, x \in \mathbf{X}$.

$\underline{A4}$ $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ *whenever* $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

If, in addition, there exists $\bar{x} \in \mathbf{X}^*$ *such that*

$$\lim_{k\to\infty} f_k(x^k) - f_k(\bar{x}) = 0 \tag{3.8}$$

then $\lim_{\epsilon\to 0} D(\epsilon) = 0$.

Proof. Let $\epsilon > 0$ be given, and note that since $\bar{x} \in \mathbf{X}^*$, (3.8) ensures that

$$f_k(x^k) \leq f_k(\bar{x}) + \epsilon \leq \operatorname*{Max}_{x^*\in\mathbf{X}^*} f_k(x^*) + \epsilon$$

on a subsequence of iterations, and thus $\mathcal{K}(\epsilon)$ is an infinite set. Let $\mathbf{X}(\epsilon)$ be the set of accumulation points of $\{x^k\}_{k\in\mathcal{K}(\epsilon)}$, and note that A1 ensures that $\mathbf{X}(\epsilon) \neq \emptyset$. As noted in the derivation of (3.7), A1-A4 ensure that

$$f^* \leq f(x) \leq f^* + \epsilon \qquad \forall\, x \in \mathbf{X}(\epsilon) \subseteq \mathbf{X}. \tag{3.9}$$

Again, A1 ensures that $\hat{\mathbf{X}} = \overline{\lim}_{\epsilon\to 0} \mathbf{X}(\epsilon) \neq \emptyset$, it follows from (3.9) and the continuity of f that $f(x) = f^*$, $\forall\, x \in \hat{\mathbf{X}}$. It follows from

$$d(x, \mathbf{X}^*) = 0 \iff x \in \mathbf{X}^* \iff f(x) = f^*$$

that $d(x, \mathbf{X}^*) = 0$ for all $x \in \hat{\mathbf{X}}$. Since $d(\cdot, \mathbf{X}^*)$ is a continuous function, we have that

$$\lim_{\epsilon\to 0} \overline{\lim_{k\in\mathcal{K}(\epsilon)}} d(x^k, \mathbf{X}^*) = 0,$$

and hence the result. ■

In the following result, we identify conditions under which every accumulation point of the sequence $\{x^k\}_{k=1}^{\infty}$ is an optimal solution to (P).

Lemma 3.2. *Let $\{x^k\}_{k=1}^{\infty}$ and $\{s_k\}_{k=1}^{\infty}$ denote the sequence of iterates and steplengths, respectively, associated with the inexact subgradient method. Suppose that assumptions A1-A5 hold.*

A1 f *is continuous and* $\mathbf{X}$ *is a compact convex set.*

A2 $\{f_k\}_{k=1}^{\infty}$ *are convex, continuous and uniformly bounded from below on* $\mathbf{X}$.

A3 $\overline{\lim}_{k\to\infty} f_k(u^k) \le f(x)$ *whenever* $\{u^k\}_{k=1}^{\infty} \to x, \quad \forall\, x \in \mathbf{X}$.

A4 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ *whenever* $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

A5 $\{d^k\}_{k=1}^{\infty}$ *is a bounded sequence.*

If, in addition, $s_k \ge 0$ *for all* k, $\lim_{k\to\infty} s_k = 0$, *and there exists* $\bar{x} \in \mathbf{X}^*$ *such that* $\underline{\lim}_{k\to\infty} f_k(x^k) - f_k(\bar{x}) = 0$, *then every accumulation point of* $\{x^k\}_{k=1}^{\infty}$ *is an optimal solution to* (P).

Proof. Let x^* denote an arbitrary point in $\mathbf{X}^*$. Note that A1 ensures that for all $x^* \in \mathbf{X}^*$, we have

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\le \|x^k - s_k d^k - x^*\|^2 \\ &= \|x^k - x^*\|^2 + s_k^2\|d^k\|^2 + 2s_k d^k(x^* - x^k) \\ &= \|x^k - x^*\|^2 + s_k\left(s_k\|d^k\|^2 + 2d^k(x^* - x^k)\right). \qquad (3.10)\end{aligned}$$

Since f_k is convex (Assumption A2) and $d^k \in \partial f_k(x^k)$, it follows that $f_k(x^*) \ge f_k(x^k) + d^k(x^* - x^k)$, and thus

$$\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 + s_k\left\{s_k\|d^k\|^2 + 2\left(f_k(x^*) - f_k(x^k)\right)\right\} \qquad (3.11)$$

for all $x^* \in \mathbf{X}^*$. As in the proof of Lemma 3.1, our hypotheses ensure that $\mathcal{K}(\epsilon)$ is an infinite index set for all $\epsilon > 0$, and consequently $D(\epsilon) = \overline{\lim}_{k\in\mathcal{K}(\epsilon)} d(x^k, X^*)$ is well defined. By hypothesis, $s_k \to 0$ and thus assumptions A1 and A5 ensure that there exists $N_\epsilon < \infty$ such that for every $k \ge N_\epsilon$

$$s_k\left\{s_k\|d^k\|^2 + 2d^k(x^* - x^k)\right\} \le \frac{\epsilon}{2} \quad , \quad s_k\|d^k\|^2 \le \epsilon. \qquad (3.12)$$

Moreover, we may assume without loss of generality that N_ϵ is chosen so that

$$d(x^k, \mathbf{X}^*) \le D(\epsilon) + \frac{\epsilon}{2} \quad \forall\, k \in \mathcal{K}(\epsilon),\ k \ge N_\epsilon. \qquad (3.13)$$

For the remainder of the proof, we let $k \geq N_\epsilon$, and consider two cases: 1) $k \in \mathcal{K}(\epsilon)$ and 2) $k \notin \mathcal{K}(\epsilon)$.

Case 1. ($k \in \mathcal{K}(\epsilon)$.) It follows from (3.4), (3.10), (3.12), and (3.13) that

$$\begin{aligned}&\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + \frac{\epsilon}{2}, \quad \forall x^* \in \mathbf{X}^* \\ &\Rightarrow d(x^{k+1}, \mathbf{X}^*) \leq d(x^k, \mathbf{X}^*) + \frac{\epsilon}{2} \\ &\Rightarrow d(x^{k+1}, \mathbf{X}^*) \leq D(\epsilon) + \epsilon. \end{aligned} \tag{3.14}$$

Case 2. ($k \notin \mathcal{K}(\epsilon)$.) Since $k \notin \mathcal{K}(\epsilon)$, we have $f_k(x^k) > f_k(x^*) + \epsilon$, for all $x^* \in \mathbf{X}^*$. Hence, it follows from (3.4), (3.11) and (3.12) that

$$\begin{aligned}&\|x^{k+1} - x^*\|^2 < \|x^k - x^*\|^2 + s_k \left\{s_k \|d^k\|^2 - 2\epsilon\right\} \quad \forall\, x^* \in \mathbf{X}^* \\ &\Rightarrow d(x^{k+1}, \mathbf{X}^*) \leq d(x^k, \mathbf{X}^*). \end{aligned} \tag{3.15}$$

Moreover, since $\mathcal{K}(\epsilon)$ is an infinite set, we see that whenever k is sufficiently large (i.e., whenever k exceeds $\text{Min}\{t \in \mathcal{K}(\epsilon) | t \geq N_\epsilon\}$), inequality (3.14) in combination with successive application of (3.15) yields

$$d(x^{k+1}, \mathbf{X}^*) \leq D(\epsilon) + \epsilon.$$

Thus, in both cases we see that whenever k is sufficiently large, $0 \leq d(x^k, \mathbf{X}^*) \leq D(\epsilon) + \epsilon$. Since the choice of ϵ was arbitrary, it follows from Lemma 3.1 that

$$\lim_{k \to \infty} d(x^k, \mathbf{X}^*) = 0,$$

and thus every accumulation point of $\{x^k\}_{k=1}^\infty$ is contained in $\mathbf{X}^*$. ■

As a result of Lemma 3.2, we see that whenever the algorithm can be shown to generate a subsequence on which (3.8) is satisfied, every accumulation point of $\{x^k\}_{k=1}^\infty$ is optimal to (P). As shown in the next theorem, this may be achieved by augmenting the conditions on the steplengths by requiring that $\sum_k s_k \to \infty$.

Theorem 3.3. *Let $\{x^k\}_{k=1}^{\infty}$ be the sequence of iterates identified by the inexact subgradient method and suppose that assumptions A1-A5 hold.*

A1 f *is continuous and* $\mathbf{X}$ *is a compact convex set.*

A2 $\{f_k\}_{k=1}^{\infty}$ *are convex, continuous and uniformly bounded from below on* $\mathbf{X}$.

A3 $\overline{\lim}_{k\to\infty} f_k(u^k) \le f(x) \quad$ *whenever* $\quad \{u^k\}_{k=1}^{\infty} \to x, \quad \forall\, x \in \mathbf{X}$.

A4 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x}) \quad$ *whenever* $\quad \{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

A5 $\{d^k\}_{k=1}^{\infty}$ *is a bounded sequence.*

If, in addition, the sequence of steplengths, $\{s_k\}_{k=1}^{\infty}$ *satisfy* $s_k \ge 0,\ s_k \to 0,\ \sum_k s_k \to \infty$, *then every accumulation point of* $\{x^k\}_{k=1}^{\infty}$ *is an optimal solution to* (P).

Proof. Let x^* be an optimal solution to (P). It follows from Lemma 3.2 that it is sufficient to show $\underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*) = 0$. As in (3.11), we have that

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \ge s_k \left\{-s_k\|d^k\|^2 + 2\left(f_k(x^k) - f_k(x^*)\right)\right\}.$$

Let $\gamma = \underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*)$, A1, A2 and A4 ensure that there exists $\mathcal{K}$ such that

$$\lim_{k\in\mathcal{K}} x^k = \hat{x} \in \mathbf{X}, \quad \lim_{k\in\mathcal{K}} f_k(x^k) = f(\hat{x}),$$

and

$$\underline{\lim}_{k\to\infty}\ f_k(x^k) - f_k(x^*) = \lim_{k\in\mathcal{K}} f_k(x^k) - f_k(x^*).$$

Since A3 ensures that $\overline{\lim}_{k\in\mathcal{K}} f_k(x^*) \le f^*$, it follows that

$$\gamma = \lim_{k\in\mathcal{K}} f_k(x^k) - f_k(x^*) \ge f(\hat{x}) - f^* \ge 0.$$

Suppose that $\gamma > 0$. Since $s_k \to 0$, A5 implies that there exists $N < \infty$ such that $-s_k\|d^k\|^2 + 2\left(f_k(x^k) - f_k(x^*)\right) \ge \gamma$ for all $k \ge N$. Thus,

$$\begin{aligned} &\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \ge \gamma s_k && \forall\, k \ge N \\ \Rightarrow\ &\|x^N - x^*\|^2 - \|x^{N+n} - x^*\|^2 \ge \gamma \sum_{k=N}^{N+n-1} s_k && \forall\, n \ge 0. \end{aligned}$$

But since $\sum_k s_k \to \infty$, this contradicts the compactness of $\mathbf{X}$. Thus, $\gamma = 0$, and the result now follows from Lemma 3.2. ■

As a result of Theorem 3.3, we see that inexact subgradient methods inherit the limiting behavior of standard subgradient algorithms. Additionally, it has the advantage of avoiding complicated objective and subgradient evaluations. However, in various practical subgradient algorithms (such as Held, Wolfe and Crowder [1974], Sen and Sherali [1986] and Allen et al [1987]), steplengths are defined adaptively. In the following section, we will adopt this form of presentation and suggest a method for the adaptive specification of steplengths that is similar in spirit to those of Allen el al [1987].

3.2 Inexact Subgradient Methods with Adaptive Steplengths

In this section, we use a sequence of convex objective function approximations to develop an inexact subgradient method with adaptively defined steplengths. This algorithm offers an extension of the result of Allen et. al. [1987]. Our method uses a sequence of real values $\{\bar{f}_k\}_{k=1}^{\infty}$ such that $\bar{f}_k \le f_k(x)$, for all $x \in \mathbf{X}$, and defines s_k adaptively using $\bar{f}_k$ as the iterative target value.

Inexact Subgradient Algorithm with Adaptive Steplengths

Step 0. Initialize $k \leftarrow 0$. $x^1 \in \mathbf{X}$ is given.

Step 1. $k \leftarrow k + 1$. Specify f_k, λ_k, $\bar{f}_k$.

Step 2. $d^k \in \partial f_k(x^k)$.

Step 3. If $d^k = 0$, then $s_k = 0$; otherwise

$s_k = \lambda_k \frac{f_k(x^k) - \bar{f}_k}{\|d^k\|^2}$, and $x^{k+1} = \mathrm{P}_{\mathbf{X}}\{x^k - s_k d^k\}$.

Repeat from Step 1.

Our primary focus is on Step 1, in which the functional approximation f_k, the steplength coefficient λ_k, and the 'target value' $\bar{f}_k$ are specified. We will establish conditions on the sequences under which one can be assured that the algorithm will lead to an optimal solution. In the following, x^* and f^* denote an optimal solution to (P) and the optimal value, respectively. We proceed as in

§3.1 with a similar set of assumptions, adapted for the new steplength definition as follows:

Assumptions

B1 f is continuous and $\mathbf{X}$ is a compact convex set.

B2 $\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below on $\mathbf{X}$.

B3 $\overline{\lim}_{k\to\infty} f_k(u^k) \leq f(x)$ whenever $\{u^k\}_{k\to\infty} \to x, \quad \forall\, x \in \mathbf{X}$.

B4 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ whenever $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

B5 $\{d^k\}_{k=1}^{\infty}$ is a bounded sequence.

B6 $\lambda_k \geq 0, \quad \{\lambda_k\}_k \to 0$ and $\sum_k \lambda_k = \infty$.

B7 $\{\bar{f}_k\}_{k=1}^{\infty}$ is bounded from below and $\bar{f}_k \leq f_k(x) \quad \forall\, x \in \mathbf{X}, \quad \forall\, k$.

B1-B5 are the same as A1-A5. B6 is similar to the steplength requirements in Theorem 3.3, although, we observe that it refers to the scaling factors, $\{\lambda_k\}_{k=1}^{\infty}$, instead of the steplengths, $\{s_k\}_{k=1}^{\infty}$. Since λ_k does not solely define the steplength, the conditions imposed on $\{s_k\}_{k=1}^{\infty}$ are relaxed. Assumption B7 ensures that

$$\underline{\lim}_{k\to\infty} f_k(x^k) - \bar{f}_k \geq \underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*),$$

while Assumptions B1-B4 ensure that there exists an index set $\mathcal{K}$ such that

$$\{x^k\}_{k\in\mathcal{K}} \to \bar{x}, \quad \text{and} \quad \underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*) = \lim_{k\in\mathcal{K}} f_k(x^k) - f_k(x^*) \geq f(\bar{x}) - f^*.$$

It follows that

$$\underline{\lim}_{k\to\infty} f_k(x^k) - \bar{f}_k \geq \underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*) \geq f(\bar{x}) - f^* \geq 0. \tag{3.16}$$

Hence, for large values of k, $f_k(x^k) - \bar{f}_k$ provides an estimate of a bound on the error associated with the iterate x^k, $f(x^k) - f^*$. As a result, if the estimated error bound is consistently small, the corresponding steplength will also become small.

We precede our proof of the optimality of all accumulation points with two lemmas. In the first lemma below, we identify a property of the sequence $\{\|x^k - x^*\|\}_{k=1}^{\infty}$. As before, x^* denotes an optimal solution to (P).

Lemma 3.4. *Let x^* be an optimal solution to (P), and $\{x^k\}$ denote the sequence of iterates identified by the inexact subgradient algorithm with adaptive steplengths. Suppose that assumptions B1, B2 and B7 hold.*

B1 *f is continuous and $\mathbf{X}$ is a compact convex set.*

B2 *$\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below $\mathbf{X}$.*

B7 *$\{\bar{f}_k\}_{k=1}^{\infty}$ is bounded from below and $\bar{f}_k \leq f_k(x) \quad \forall\, x \in \mathbf{X}, \quad \forall\, k$.*

If , in addition, there exists $N < \infty$ such that $0 < \lambda_k \leq \beta < 2$, for all $k > N$, then

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + s_k(2-\beta)\left\{f_k(x^*) - f_k(x^k) + \frac{\beta}{2-\beta}[f_k(x^*) - \bar{f}_k]\right\}$$

for all $k > N$.

Proof: As in (3.10),

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + s_k\left\{s_k\|d^k\|^2 + 2d^k(x^* - x^k)\right\}.$$

If $\|d^k\| \neq 0$, then $s_k = \lambda_k \frac{f_k(x^k)-\bar{f}_k}{\|d^k\|^2}$. Thus, since $d^k \in \partial f_k(x^k)$ and f_k is convex, we have that

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + s_k\left\{\lambda_k[f_k(x^k) - \bar{f}_k] + 2d^k(x^* - x^k)\right\}\\ &\leq \|x^k - x^*\|^2 + s_k\left\{\lambda_k[f_k(x^k) - \bar{f}_k)] + 2[f_k(x^*) - f_k(x^k)]\right\}.\end{aligned}$$

Note that if $\|d^k\| = 0$, then $s_k = 0$ so that $x^{k+1} = x^k$ and the inequality remains valid. It follows from the hypothesis that for $\lambda_k \leq \beta < 2$,

$$\begin{aligned}\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + s_k\Big\{&\beta[f_k(x^k) - \bar{f}_k]\\ &+ (2-\beta)[f_k(x^*) - f_k(x^k)] + \beta[f_k(x^*) - f_k(x^k)]\Big\} \qquad \forall k > N\\ = \|x^k - x^*\|^2 + s_k(2-\beta)&\left\{f_k(x^*) - f_k(x^k) + \frac{\beta}{2-\beta}[f_k(x^*) - \bar{f}_k]\right\} \blacksquare\end{aligned}$$

In the following lemma, we identify a subsequential relationship between $\bar{f}_k$ and $f_k(x^k)$. Note that since $\bar{f}_k \leq f_k(x^*)$ for every $x^* \in \mathbf{X}^*$, assumptions B1-B4 together with Lemma 3.2 ensure that every accumulation point of $\{x^k\}_{k=1}^{\infty}$ is an optimal solution to (P) whenever $\underline{\lim}_{k\to\infty} f_k(x^k) - \bar{f}_k = 0$ (see inequality (3.16)). Thus, we will focus on the consequences of failing to satisfy this property.

Lemma 3.5. *Let x^* be an optimal solution to (P), and $\{x^k\}_{k=1}^{\infty}$ denote the sequence of iterates identified by the inexact subgradient algorithm with adaptive steplengths. Suppose that assumptions B1-B7 hold.*

B1 f *is continuous and* $\mathbf{X}$ *is a compact convex set.*

B2 $\{f_k\}_{k=1}^{\infty}$ *are convex, continuous and uniformly bounded from below on* $\mathbf{X}$.

B3 $\overline{\lim}_{k\to\infty} f_k(u^k) \leq f(x)$ *whenever* $\{u^k\}_{k\to\infty} \to x, \quad \forall\, x \in \mathbf{X}$.

B4 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ *whenever* $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

B5 $\{d^k\}_{k=1}^{\infty}$ *is a bounded sequence.*

B6 $\lambda_k \geq 0, \quad \{\lambda_k\}_k \to 0$ and $\sum_k \lambda_k = \infty$.

B7 $\{\bar{f}_k\}_{k=1}^{\infty}$ *is bounded from below and* $\bar{f}_k \leq f_k(x) \quad \forall\, x \in \mathbf{X}, \quad \forall\, k$.

If, in addition, $\underline{\lim}_{k\to\infty}\{f_k(x^k) - \bar{f}_k\} > 0$, *there exists* $N < \infty$ *such that for all* $k > N$, $\|d^k\| > 0$ *and* $0 < \lambda_k \leq \beta < 2$, *then for any* $\delta > 0$ *and* $N_1 > N$, *there exists* $k_1 > N_1$ *such that*

$$f_{k_1}(x^{k_1}) \leq f_{k_1}(x^*) + \frac{\beta}{2-\beta}\left(f_{k_1}(x^*) - \bar{f}_{k_1}\right) + \delta.$$

Proof: We proceed by contradiction. Given $\delta > 0$ and $N_1 > N$, suppose that

$$f_k(x^k) > f_k(x^*) + \frac{\beta}{2-\beta}\left(f_k(x^*) - \bar{f}_k\right) + \delta, \qquad \forall k > N_1.$$

From Lemma 3.4, we have

$$\|x^{k+1} - x^*\|^2 < \|x^k - x^*\|^2 - s_k(2-\beta)\delta \qquad \forall k > N_1. \tag{3.17}$$

Since $\|d^k\| > 0$, $s_k = \lambda_k \frac{f_k(x^k) - \bar{f}_k}{\|d^k\|^2}$, and thus (3.17) yields,

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 > \lambda_k(2-\beta)\delta\frac{f_k(x^k) - \bar{f}_k}{\|d^k\|^2}, \qquad \forall k > N_1.$$

Let $\gamma = \underline{\lim}_{k\to\infty} f_k(x^k) - \bar{f}_k$ and note that by hypothesis $\gamma > 0$. By assumption B5, there exists $\mu < \infty$ such that $\|d^k\|^2 \le \mu$ for all k, and thus there exists $\bar{N} \ge N_1$ such that

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 > \lambda_k \frac{(2-\beta)\delta\gamma}{2\mu} \qquad \forall\, k \ge \bar{N}$$

$$\Rightarrow \|x^{\bar{N}} - x^*\|^2 - \|x^{\bar{N}+n+1} - x^*\|^2 > \frac{(2-\beta)\delta\gamma}{2\mu} \sum_{k=\bar{N}}^{\bar{N}+n} \lambda_k \qquad \forall\, n \ge 0.$$

This yields a contradiction between B1 and B6. Thus, for any $\delta > 0$ and $N_1 > N$, there exists $k_1 > N_1$ such that

$$f_{k_1}(x^{k_1}) \le f_{k_1}(x^*) + \frac{\beta}{2-\beta}\left(f_{k_1}(x^*) - \bar{f}_{k_1}\right) + \delta. \qquad \blacksquare$$

With Lemma 3.5, we can now establish optimality of every accumulation point of $\{x^k\}_{k=1}^{\infty}$.

Theorem 3.6. *Let x^* be an optimal solution to (P), and $\{x^k\}_{k=1}^{\infty}$ denote the sequence of iterates identified by the inexact subgradient algorithm with adaptive steplengths. Suppose that B1-B7 hold.*

B1 f *is continuous and* $\mathbf{X}$ *is a compact convex set.*

B2 $\{f_k\}_{k=1}^{\infty}$ *are convex, continuous and uniformly bounded from below on* $\mathbf{X}$.

B3 $\overline{\lim}_{k\to\infty} f_k(u^k) \le f(x)$ *whenever* $\{u^k\}_{k\to\infty} \to x, \quad \forall\, x \in \mathbf{X}$.

B4 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ *whenever* $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

B5 $\{d^k\}_{k=1}^{\infty}$ *is a bounded sequence.*

B6 $\lambda_k \ge 0, \quad \{\lambda_k\}_k \to 0 \quad \text{and} \quad \sum_k \lambda_k = \infty$.

B7 $\{\bar{f}_k\}_{k=1}^{\infty}$ *is bounded from below and* $\bar{f}_k \le f_k(x) \quad \forall\, x \in \mathbf{X}, \quad \forall\, k$.

Then, every accumulation point of $\{x^k\}_{k=1}^{\infty}$ is optimal to (P).

Proof: Let $x^* \in \mathbf{X}^*$ be given, and note that since B1-B5 are the same as A1-A5, and our assumptions ensure that $s_k \ge 0, s_k \to 0$, it follows from Lemma 3.2 that it is sufficient to show that there exists $x^* \in \mathbf{X}^*$ such that

$$\underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*) = 0.$$

Lemma 3.5 identifies two cases, depending on $\underline{\lim}_{k\to\infty} \|d^k\|$. We shall consider these two cases separately.

Case 1. Suppose that $\underline{\lim}_{k\to\infty} \|d^k\| = 0$. Then, there exists an index set $\mathcal{K}$ such that $\lim_{k\in\mathcal{K}} \|d^k\| = 0$. Note that assumption B2 ensures that

$$f_k(x^k) + d^k(x - x^k) \leq f_k(x) \qquad \forall\, x \in \mathbf{X}.$$

Thus, since $\{d^k\}_{k\in\mathcal{K}} \to 0$,

$$\begin{aligned} 0 \leq \underline{\lim}_{k\to\infty}\, f_k(x^k) - f_k(x^*) &\leq \underline{\lim}_{k\in\mathcal{K}}\, f_k(x^k) - f_k(x^*) \\ &\leq \underline{\lim}_{k\in\mathcal{K}}\, d^k(x^k - x^*) = 0 \end{aligned} \tag{3.18}$$

where the first inequality follows from assumptions B1-B4 (see 3.16), and the last equality follows from assumption B1. The result now follows from Lemma 3.2.

Case 2. Now, suppose that $\underline{\lim}_{k\to\infty} \|d^k\| > 0$. As in (3.16), assumptions B1-B4 and B7 ensure that

$$\underline{\lim}_{k\to\infty}\, f_k(x^k) - \bar{f}_k \geq \underline{\lim}_{k\to\infty}\, f_k(x^k) - f_k(x^*) \geq 0.$$

If $\underline{\lim}_{k\to\infty} f_k(x^k) - \bar{f}_k = 0$, the result follows from Lemma 3.2. Thus, we suppose that $\underline{\lim}_{k\to\infty} f_k(x^k) - \bar{f}_k > 0$. We will use Lemma 3.5 to verify the hypotheses of Lemma 3.2. To begin, let $\rho \equiv \underline{\lim}_{k\to\infty} \bar{f}_k$, $f^* = f(x^*)$, and note that assumptions B3 and B7 ensure that $\rho \leq f^*$. In addition, for every $\epsilon > 0$ and $\beta \in (0, 2)$, it follows from B3 and B6 that there exists $N \leq \infty$ such that

$$\lambda_k \leq \beta \tag{3.19a}$$

$$f_k(x^*) \leq f^* + \epsilon \qquad \forall\, k \geq N \tag{3.19b}$$

$$\bar{f}_k \geq \rho - \epsilon \tag{3.19c}$$

Let $\epsilon_1 = \delta_1 = \frac{1}{6}$ and $\beta_1 = \frac{\delta_1}{f^* - \rho + \epsilon_1} \leq 1$. From (3.19), there exists $N_1 < \infty$ such that

$$\begin{aligned} &\lambda_k \leq \beta_1 \\ &f_k(x^*) \leq f^* + \epsilon_1 \qquad \forall\, k \geq N_1. \\ &\bar{f}_k \geq \rho - \epsilon_1. \end{aligned}$$

We note that we may assume without loss of generality that $\|d^k\| > 0$ for all $k > N_1$. Applying Lemma 3.5 with $\beta = \beta_1, \delta = \delta_1$, there exists $k_1 > N_1$ such that

$$f_{k_1}(x^{k_1}) \leq f_{k_1}(x^*) + \frac{\beta_1}{2-\beta_1}\left[f_{k_1}(x^*) - \bar{f}_{k_1}\right] + \delta_1.$$

Since $0 < \beta_1 \leq 1$ implying $\frac{\beta_1}{2-\beta_1} \leq \beta_1$ and $\beta_1 = \frac{\delta_1}{f^*-\rho+\epsilon_1} \leq 1$, it follows that

$$\begin{aligned} f_{k_1}(x^{k_1}) &\leq f_{k_1}(x^*) + \beta_1[f^* + \epsilon_1 - \rho + \epsilon_1] + \delta_1 \\ &\leq f_{k_1}(x^*) + \epsilon_1 + 2\delta_1 \\ &= f_{k_1}(x^*) + \frac{1}{2}. \end{aligned}$$

For $j = 2, 3, \cdots$, define $\epsilon_j = \delta_j = \frac{1}{3}(\frac{1}{2})^j$, and $\beta_j = \frac{\delta_j}{f^*-\rho+\epsilon_j} \leq 1$. Again, it follows from (3.19) that there exists $N_j > k_{j-1}$ such that

$$\begin{aligned} \lambda_k &\leq \beta_j \\ f_k(x^*) &\leq f^* + \epsilon_j \\ \bar{f}_k &\geq \rho - \epsilon_j \qquad \forall k \geq N_j. \end{aligned}$$

Applying Lemma 3.5 again with $\beta = \beta_j, \delta = \delta_j$, there exists $k_j > N_j$ such that

$$\begin{aligned} f_{k_j}(x^{k_j}) &\leq f_{k_j}(x^*) + \frac{\beta_j}{2-\beta_j}[f_{k_j}(x^*) - \bar{f}_{k_j}] + \delta_j \\ &\leq f_{k_j}(x^*) + \beta_j[f^* + \epsilon_j - \rho + \epsilon_j] + \delta_j \\ &\leq f_{k_j}(x^*) + \epsilon_j + 2\delta_j \\ &= f_{k_j}(x^*) + (\frac{1}{2})^j. \end{aligned}$$

Hence, there exists a subsequence $\{x^{k_j}\}_{j=1}^{\infty}$ such that

$$f_{k_j}(x^{k_j}) \leq f_{k_j}(x^*) + (\frac{1}{2})^j \qquad \forall j.$$

As in (3.18),

$$0 \leq \underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*) \leq \underline{\lim}_{j\to\infty} f_{k_j}(x^{k_j}) - f_{k_j}(x^*) = 0$$

and the result follows from Lemma 3.2. ■

The proof of Theorem 3.6 suggests that it may be possible to reduce the dependence of the algorithm on the externally controlled parameters, $\{\lambda_k\}_{k=1}^{\infty}$, if it is possible to ensure that $\underline{\lim}_{k\to\infty} \|d^k\| = 0$ or $\underline{\lim}_{k\to\infty} f_k(x^k) - \bar{f}_k$. Naturally, the former is difficult to ensure. With the latter, it is evident that consistency with the remaining assumptions would require that $\overline{\lim}_{k\to\infty} \bar{f}_k = f^*$. In §3.3, we investigate an inexact subgradient algorithm that can be used to solve stochastic linear programs with recourse, and show that this particular algorithm ensures $\overline{\lim}_{k\to\infty} \bar{f}_k = f^*$.

3.3 Application to the solution of Stochastic Linear Programs

In this section, we will discuss how the inexact subgradient method derived in §3.2 can be applied to solve a two stage stochastic linear program with recourse *(SLPR)*. The algorithm that we present here is an inexact subgradient algorithm based on the sequence of objective function approximations constructed by a stochastic decomposition (SD) algorithm (Higle and Sen [1991a]). Assuming that $(SLPR)$ satisfies the complete recourse property so that f is continuous (Wets [1982]) on $\mathbf{X}$, the sequence of SD approximations will be shown to ensure that the hypotheses in Theorem 3.6 are satisfied. In addition, we will show that $\overline{\lim}_{k\to\infty} \bar{f}_k = f^*$.

<u>An Inexact Subgradient Algorithm with SD Approximations (IXSSD)</u>

<u>Step 0.</u> $V_0 \leftarrow \emptyset,\ x^1, y^1 \in \mathbf{X},\ k \leftarrow 0$.

<u>Step 1.</u> $k \leftarrow k+1$. Randomly generate an observation of $\tilde{\omega}$, ω^k. For $u = x, y$, let $\pi(u) \in \text{argmax}\{\pi(\omega^k - Tu^k) \mid \pi W \le g\}$, and $V_k \leftarrow V_{k-1} \cup \{\pi(u) \mid u = x, y\}$.

a) For $t = 1, \cdots, k$, $u = x, y$, let $\pi(u)_t^k \in \text{argmax}\{\pi(\omega^t - Tu^k) \mid \pi \in V_k\}$,

$\alpha(u)_k^k + \beta(u)_k^k x = \frac{1}{k}\sum_{t=1}^{k} \pi(u)_t^k(\omega^t - Tx)$,

$\alpha(u)_t^k \leftarrow \frac{k-1}{k}\alpha(u)_t^{k-1} + \frac{1}{k}\pi(u)_k^k\omega^k,\quad \beta(u)_t^k \leftarrow \frac{k-1}{k}\beta(u)_t^{k-1} - \frac{1}{k}\pi(u)_k^k T.$

$f_k(x) = cx + \text{Max}\{\alpha(u)_t^k + \beta(u)_t^k x \mid u = x, y;\ t = 1, \cdots, k\}$.

b) $y^{k+1} \in \text{argmin}\{f_k(x) \mid x \in \mathbf{X}\}$.

$\bar{f}_k = f_k(y^{k+1})$.

c) Specify λ_k.

<u>Step 2.</u> $d^k \in \partial f_k(x^k)$.

<u>Step 3.</u> If $d^k = 0$, then $s_k = 0$, otherwise let

$s_k = \lambda_k \frac{f_k(x^k) - \bar{f}_k}{\|d^k\|^2}$, and $x^{k+1} = \mathrm{P}_{\mathbf{X}}(x^k - s_k d^k)$.

Repeat from Step 1.

Note that f_k is defined by two sequences of cuts associated with $\{x^k\}_{k=1}^{\infty}$ and $\{y^k\}_{k=1}^{\infty}$, respectively. As in §2.4, the superscript, k, of $\alpha(u)_t^k$ and $\beta(u)_t^k$ represents the current iteration, and the subscript, t, denotes the iteration at which the cuts were generated. At iteration k, for $t < k$, $\{\alpha(u), \beta(u)\}_t^{k-1}$ are updated and denoted by $\{\alpha(u), \beta(u)\}_t^k$.

The asymptotic optimality of the sequence of iterates $\{x^k\}$ is easily established by combining the results of Higle and Sen [1991a] (summarized in Proposition 2.3), with Theorem 3.6.

Proposition 3.7. *Let $\{x^k\}_{k=1}^{\infty}$ denote the sequence of iterates identified in the inexact subgradient algorithm with SD approximations. In SLPR, if*

- $\mathbf{X}$ *is a non-empty compact and convex set,*
- $\{\pi \mid \pi W \leq g\}$ *is a non-empty compact convex polyhedral set,*
- Ω *is a non-empty compact set,*
- $\lambda_k \geq 0$, $\{\lambda_k\} \to 0$, and $\sum_k \lambda_k \to \infty$.

then with probability one every accumulation point of $\{x^k\}_{k=1}^{\infty}$ is optimal to $(SLPR)$. Furthermore, $\overline{\lim}_{k\to\infty} \bar{f}_k = f^*$ (wp1).

Proof: By Theorem 3.6, it is sufficient to show that

<u>B1</u> f is continuous and $\mathbf{X}$ is a compact convex set.

<u>B2</u> $\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below on $\mathbf{X}$.

<u>B3</u> $\overline{\lim}_{k\to\infty} f_k(u^k) \leq f(x)$ whenever $\{u^k\}_{k\to\infty} \to x$, $\forall\, x \in \mathbf{X}$.

<u>B4</u> $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ whenever $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

<u>B5</u> $\{d^k\}_{k=1}^{\infty}$ is a bounded sequence.

<u>B6</u> $\lambda_k \geq 0$, $\{\lambda_k\}_k \to 0$ and $\sum_k \lambda_k = \infty$.

<u>B7</u> $\{\bar{f}_k\}_{k=1}^{\infty}$ is bounded from below and $\bar{f}_k \leq f_k(x)$ $\forall\, x \in \mathbf{X}$, $\forall\, k$.

are satisfied with probability one. B1 and B6 are satisfied by hypotheses. B2 and

B7 are satisfied easily from the definitions in Step 1a and 1b. The subgradient d^k can be expressed as a convex combination of $\{c - \pi T\}$, and $\pi \in \{\pi \mid \pi W \leq g\}$ is bounded by hypothesis. Thus, B5 is satisfied. B3 and B4 follow from Proposition 2.3. Hence, it follows from Theorem 3.6 that with probability one every accumulation point of $\{x^k\}_{k=1}^{\infty}$ is optimal to $(SLPR)$. In addition, it follows from Proposition 2.3 that

$$\{f_k(y^k)\}_{k \in \mathcal{K}} \to f(\bar{y}) \quad \text{whenever} \quad \{y^k\}_{k \in \mathcal{K}} \to \bar{y} \quad \text{(wp1)},$$

and there exists an index set $\bar{\mathcal{K}}$ such that

$$\{f_k(y^{k+1}) - f_k(y^k)\}_{k \in \bar{\mathcal{K}}} \to 0 \quad \text{(wp1)}.$$

In combination, there exists an index set $\mathcal{K} \subseteq \bar{\mathcal{K}}$ such that

$$\{y^k\}_{k \in \mathcal{K}} \to \bar{y} \quad \text{and} \quad f^* \leq f(\bar{y}) = \lim_{k \in \mathcal{K}} f_k(y^k) = \lim_{k \in \mathcal{K}} f_k(y^{k+1}) = \lim_{k \in \mathcal{K}} \bar{f}_k \quad \text{(wp1)}.$$

On the other hand, B3 and B7 ensure that $\overline{\lim}_{k \to \infty} \bar{f}_k \leq f^*$ (wp1). Thus, $\overline{\lim}_{k \to \infty} \bar{f}_k = f^*$ (wp1), and the result follows. ■

We note that since $\overline{\lim}_{k \to \infty} \bar{f}_k = f^*$, the optimality result of this algorithm does not depend on the external inputs $\{\lambda_k\}$ (see *Case 2* of Theorem 3.6), and thus it is possible to suggest termination test based on the appearance of optimality (i.e., when $f_k(x^k) - \bar{f}_k$ is sufficiently small).

This inexact subgradient algorithm incorporates statistically defined functional approximations within subgradient methods. As a solution procedure for SLPR, it proceeds similarly to stochastic quasigradient methods, but the requirements imposed on $\{s_k\}_{k=1}^{\infty}$ are more relaxed than (3.2).

3.4 Conclusion

In this chapter, we have developed algorithms that are applicable to optimization problems in which the objective function may be difficult to evaluate. Under mild conditions, inexact subgradient algorithms inherit the same properties as subgradient methods. Furthermore, with adaptively defined steplengths, convergence results can be established with relaxed conditions imposed on the steplengths. In stochastic optimization, it can be interpreted as another statistical application of subgradient methods with mild steplength requirements.

CHAPTER 4

INEXACT PRIMAL-DUAL SUBGRADIENT METHODS

4.0 Introduction

We begin with the program stated in Chapter 3,

$$\begin{aligned} \text{Min} \quad & f(x) \\ s/t \quad & x \in \mathbf{X}. \end{aligned} \tag{P}$$

The inexact subgradient method generates a sequence of iterates through

$$x^{k+1} = \mathrm{P}_{\mathbf{X}}\{x^k - s_k d^k\},$$

where $\mathrm{P}_{\mathbf{X}}$ denotes the projection onto $\mathbf{X}$. If the feasible set, $\mathbf{X}$, is simply defined (e.g., as a box or a ball in $\mathcal{R}^n$,), then this projection is easily undertaken. However, the projection onto a more complicated region can be computationally demanding. In this chapter, we will revise our problem statement as follows.

$$\begin{aligned} f^* = \quad \text{Min} \quad & f(x) \\ s/t \quad & Ax \leq b \\ & x \in \mathbf{X}. \end{aligned} \tag{Q}$$

In this statement, $\mathbf{X}$ is a non-empty compact convex set onto which a projection is easily undertaken. We develop algorithms for (Q) by adapting the primal-dual subgradient algorithm presented in Sen and Sherali [1986].

The primal-dual subgradient algorithm (see Sen and Sherali [1986]), employs a Lagrangian function along with a penalty function to generate a sequence of primal and dual iterates which converges to a pair of primal and dual optimal solutions of (Q). The penalty function provides an upper bound on the Lagrangian function. Ideally, the penalty and Lagrangian functions would agree at optimal solutions. Consequently, the difference between these functions is referred to as the gap function, and is used as a guideline for the choice of steplengths. The iterate is defined via the projection performed onto the relaxed feasible set $\mathbf{X}$. In our algorithms, the gap function is replaced by a gap approximation. Hence, we refer to these algorithms as inexact primal-dual subgradient methods.

In §4.1, we offer the notation associated with the methodology presented in this chapter and develop the inexact primal-dual subgradient method. In §4.2, we verify that the algorithm leads to an optimal solution when the functional approximations epi-converge to f. In §4.3, we assume that the functional approximations satisfy an epigraphical nesting property discussed in §2.2, and study the limiting behavior of the gap approximations. In §4.4, we use the asymptotic results of §4.3 to develop the inexact primal-dual subgradient method with objective function approximations similar to those constructed in the stochastic decomposition algorithm. Finally, our conclusions are presented in §4.5.

4.1 Notation and Algorithm

In the problem,

$$\begin{aligned} f^* = \quad & \text{Min } f(x) \\ s/t \quad & Ax \leq b \\ & x \in \mathbf{X}, \end{aligned} \qquad (Q)$$

we assume that $\mathbf{X}$ is a non-empty compact convex set onto which a projection is easily undertaken, and A is an $m \times n$ matrix. Let a_i represent the row vector of A for $i = 1, \cdots, m$, and $b = (b_1, \cdots, b_m)$. Alternatively, we can write the constraints $Ax \leq b$ as $a_i x \leq b_i$, $i = 1, \cdots, m$. The constraints $Ax \leq b$ will be relaxed in a Lagrangian fashion through the following function definition:

$$L(\pi) = \text{Min } \{f(x) + \pi(Ax - b) \mid x \in \mathbf{X}\}.$$

Let

$$v^* = \underset{\pi \geq 0}{\text{Max}} \underset{x \in \mathbf{X}}{\text{Min}} f(x) + \pi(Ax - b) \qquad (D)$$

denote the Lagrangian value. From duality theory, $v^* \leq f^*$ always holds. We assume that $v^* = f^*$. This assumption can be easily accomplished if (Q) is a convex program, and a regularity condition (e.g., the Slater constraint qualification) is satisfied. We also assume that $(Q)/(D)$ has a finite optimal solution, which is denoted as (x^*, π^*), so that $f(x^*) = L(\pi^*)$. By adapting the primal-dual subgradient method, we will develop algorithms in which the function f is approximated by a sequence of convex functions, $\{f_k\}_{k=1}^{\infty}$ defined on $\mathbf{X}$ and explore conditions under which an optimal solution can be identified. The following notation is used throughout this chapter.

Notation

· $\epsilon = (\epsilon_1, \cdots, \epsilon_m)$ with $\epsilon_i > 0$ for each i.
· $\mathbf{Z} = \{\ (x, \pi) \mid x \in \mathbf{X},\ \pi \geq 0\ \}$ and $z \in \mathbf{Z}$.
· $H_k(x, \pi) = f_k(x) + \sum_{i=1}^{m} (\pi_i + \epsilon_i) \operatorname{Max}\{a_i x - b_i, 0\}$.
· $l_k(x, \pi) = f_k(x) + \pi(Ax - b)$.
· $L_k(\pi) = \operatorname{Min}\{l_k(x, \pi) \mid x \in \mathbf{X}\}$.
· $G_k(x, \pi) = H_k(x, \pi) - L_k(\pi)$.
· $H(x, \pi) = f(x) + \sum_{i=1}^{m} (\pi_i + \epsilon_i) \operatorname{Max}\{a_i x - b_i, 0\}$.
· $l(x, \pi) = f(x) + \pi(Ax - b)$.
· $L(\pi) = \operatorname{Min}\{l(x, \pi) \mid x \in \mathbf{X}\}$.
· $G(x, \pi) = H(x, \pi) - L(\pi)$.

Compared with the objective functions f_k and f, the functions $H_k(x, \pi)$ and $H(x, \pi)$ contain additional terms which will contribute penalties if the constraints $Ax \leq b$ are violated. Thus, they are referred to as penalty functions. Furthermore, for any $(x, \pi) \in \mathbf{Z}$,

$$\begin{aligned}
H_k(x, \pi) &= f_k(x) + \sum_{i=1}^{m} (\pi_i + \epsilon_i) \operatorname{Max}\{a_i x - b_i, 0\} \\
&\geq f_k(x) + \sum_{i=1}^{m} \pi_i (a_i x - b_i) \\
&= f_k(x) + \pi(Ax - b) \\
&\geq \operatorname{Min}\{f_k(x) + \pi(Ax - b) \mid x \in \mathbf{X}\} \\
&= L_k(\pi).
\end{aligned}$$

Therefore,

$$H_k(x,\pi) \geq f_k(x) + \pi(Ax-b) \geq L_k(\pi), \qquad \forall\,(x,\pi) \in \mathbf{Z}.$$

Similarly, we have

$$H(x,\pi) \geq f(x) + \pi(Ax-b) \geq L(\pi), \qquad \forall\,(x,\pi) \in \mathbf{Z}. \tag{4.1}$$

It follows from the definitions of the gap approximation G_k and the gap function G that

$$G_k(x,\pi) \geq 0 \quad \text{and} \quad G(x,\pi) \geq 0, \qquad \forall\,(x,\pi) \in \mathbf{Z}.$$

Since f_k is convex, H_k is convex and L_k is concave. Hence, G_k is a non-negative convex function. We state the following lemma which is used as a guideline for the identification of optimal solutions for $(Q)/(D)$.

Lemma 4.1. *A point $(\bar{x},\bar{\pi}) \in \mathbf{Z}$ is optimal to $(Q)/(D)$ if and only if $G(\bar{x},\bar{\pi}) = 0$.*

Proof. If $(\bar{x},\bar{\pi})$ is optimal to $(Q)/(D)$, it follows that $f(\bar{x}) = L(\bar{\pi})$. Since $(\bar{x},\bar{\pi})$ is feasible, it follows from the definition of H that $H(\bar{x},\bar{\pi}) = f(\bar{x})$. Thus,

$$G(\bar{x},\bar{\pi}) = H(\bar{x},\bar{\pi}) - L(\bar{\pi}) = 0.$$

Conversely, if $G(\bar{x},\bar{\pi}) = 0$, then $H(\bar{x},\bar{\pi}) = L(\bar{\pi})$, and thus from (4.1), we have that

$$\begin{aligned} H(\bar{x},\bar{\pi}) &= f(\bar{x}) + \sum_{i=1}^{m}(\bar{\pi}_i + \epsilon_i)\,\mathrm{Max}\{a_i\bar{x} - b_i, 0\} \\ &= f(\bar{x}) + \sum_{i=1}^{m}\bar{\pi}_i(a_i\bar{x} - b_i) \\ &= L(\bar{\pi}). \end{aligned}$$

It follows that

$$\sum_{i=1}^{m}(\bar{\pi}_i + \epsilon_i)\,\mathrm{Max}\{a_i\bar{x} - b_i, 0\} = \sum_{i=1}^{m}\bar{\pi}_i(a_i\bar{x} - b_i).$$

Since for all i, $\pi_i \geq 0$ and $\epsilon_i > 0$, it follows that

$$a_i\bar{x} - b_i \leq 0 \quad \text{and} \quad \bar{\pi}_i(a_i\bar{x} - b_i) = 0, \qquad \forall i = 1, \cdots, m.$$

Hence, $f(\bar{x}) = L(\bar{\pi})$ and $(\bar{x}, \bar{\pi})$ is optimal. ∎

Consequently, our methods will be aimed at locating a point $(\bar{x}, \bar{\pi}) \in \mathbf{Z}$ such that $G(\bar{x}, \bar{\pi}) = 0$. We will explore various conditions and algorithmic requirements under which this can be accomplished.

The following algorithm is an adaptation of the primal-dual subgradient method presented in Sen and Sherali [1986]. In our presentation, we allow for the use of approximations of the penalty and Lagrangian values in each iteration, thus leading to an inexact primal-dual subgradient method.

Inexact Primal-Dual Subgradient Algorithm

Step 0. $k \leftarrow 0$. (x^1, π^1), $\epsilon_i > 0$ are given.

Step 1. $k \leftarrow k + 1$. Specify G_k.

Step 2. Let $d^k \in \partial G_k(x^k, \pi^k)$.

If $d^k = 0$, let $s_k = 0$; otherwise let $s_k = \lambda_k \frac{G_k(x^k, \pi^k)}{\|d^k\|^2}$.

Step 3. $(x^{k+1}, \pi^{k+1}) = \mathrm{P}_{\mathbf{Z}}\left\{(x^k, \pi^k) - s_k d^k\right\}$

Repeat from Step 1.

As in chapter 3, the iterate (x^{k+1}, π^{k+1}) is defined via the current iterate, (x^k, π^k), the steplength, s_k, and a subgradient, d^k. Note that if

$$\xi^k \in \partial f_k(x^k), \quad \text{and} \quad y^k \in \operatorname{argmin}\{\ l_k(x, \pi^k) \mid x \in \mathbf{X}\ \},$$

then

$$d^k = \left(\xi^k + \sum_{i: a_i x^k - b_i > 0} (\pi_i^k + \epsilon_i) a_i, \quad -(Ay^k - b)\right)$$

is a subgradient of G_k at (x^k, π^k). Recall from Lemma 4.1 that an iterate (x^k, π^k) is optimal if and only if $G(x^k, \pi^k) = 0$. The choice of the steplength in Step 2 suggests that the sequence of values $\{G_k(x^k, \pi^k)\}_{k=1}^{\infty}$ has a zero target without any prior knowledge of the optimal value of (Q).

In §4.2 and §4.3, we will establish conditions on the sequence of objective function approximations, $\{f_k\}_{k=1}^{\infty}$ which ensure that the inexact primal-dual subgradient method leads to optimal solutions.

4.2 Asymptotic Results with Epi-convergent Approximations

In this section, we consider a sequence of approximations, $\{f_k\}$ that epi-converges to f, study the limiting properties of the gap approximations, G_k, and discuss how the inexact primal-dual subgradient method leads to an optimal solution. We precede our asymptotic analysis with the following assumptions.

Assumptions

C1 f is continuous and $\mathbf{X}$ is a compact convex set.

C2 $\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below on $\mathbf{X}$.

C3 $f_k \xrightarrow{\text{epi}} f$.

C4 $\{f_k(x^k)\}_{k\in\mathcal{K}} \rightarrow f(\bar{x}) \quad$ whenever $\quad \{x^k\}_{k\in\mathcal{K}} \rightarrow \bar{x}$.

C5 $\{d^k\}_{k=1}^{\infty}$ is a bounded sequence.

C6 $0 < \beta \leq \lambda_k \leq 1, \ \forall\ k$.

We note that C1, C2, C4 and C5 are the same as A1, A2, A4 and A5. In C3, we assume that the sequence of function approximations $\{f_k\}$ epi-converges to f. Recall that

$$l_k(x,\pi) = f_k(x) + \pi(Ax - b).$$

Thus, $f_k \xrightarrow{\text{epi}} f$ implies that $l_k(\cdot,\bar{\pi}^k) \xrightarrow{\text{epi}} l(\cdot,\bar{\pi})$ whenever $\bar{\pi}^k \rightarrow \bar{\pi}$. This observation will be exploited in Lemma 4.3, where we investigate the asymptotic relationship between $\{L_k\}_{k=1}^{\infty}$ and L. First, we state a proposition which will be useful in the lemma.

Proposition 4.2. *Let a and $\{a_k\}_{k=1}^{\infty}$ respectively denote a real value and an infinite sequence of bounded real numbers. If for every index set $\mathcal{K}$, there exists a further index set $\mathcal{K}' \subseteq \mathcal{K}$ such that $\{a_k\}_{k\in\mathcal{K}'} \rightarrow a$, then*

$$\lim_{k\rightarrow\infty} a_k = a.$$

Proof. Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be the index sets such that

$$\underline{\lim}_{k\to\infty} a_k = \lim_{k\in\mathcal{K}_1} a_k \quad \text{and} \quad \overline{\lim}_{k\to\infty} a_k = \lim_{k\in\mathcal{K}_2} a_k.$$

Without loss of generality, by hypothesis,

$$\lim_{k\in\mathcal{K}_1} a_k = \lim_{k\in\mathcal{K}_2} a_k = a.$$

Hence, the result follows. ■

With this proposition, we can now establish the following result.

Lemma 4.3. *Let $\{\bar{\pi}^k\}_{k=1}^{\infty}$ denote a sequence of non-negative bounded points in $\mathcal{R}^m$. Suppose that assumptions C1 and C3 hold.*

C1 f is continuous and $\mathbf{X}$ is a compact convex set.

C3 $f_k \overset{\text{epi}}{\to} f$.

Then

$$\{L_k(\bar{\pi}^k)\}_{k=1}^{\infty} \to L(\bar{\pi}) \quad \text{whenever} \quad \{\bar{\pi}^k\}_{k=1}^{\infty} \to \bar{\pi}.$$

Proof. Let $\{\bar{\pi}^k\}_{k=1}^{\infty} \to \bar{\pi}$,

$$\bar{y}^k \in \operatorname{argmin}\{l_k(x,\bar{\pi}^k) | x \in \mathbf{X}\}$$

and $\mathcal{K}$ denote any index set. The compactness of $\mathbf{X}$ (C1) ensures that there exists a further index set $\mathcal{K}' \subset \mathcal{K}$ such that $\{\bar{y}^k\}_{k\in\mathcal{K}'} \to \bar{y}$. Since $l_k(\cdot,\bar{\pi}^k) \overset{\text{epi}}{\to} l(\cdot,\bar{\pi})$, it follows from Kall [1986] that

$$\{l_k(\bar{y}^k,\bar{\pi}^k)\}_{k\in\mathcal{K}'} \to l(\bar{y},\bar{\pi}) \quad \text{and} \quad \bar{y} \in \operatorname{argmin}\{l(x,\bar{\pi}) \mid x \in \mathbf{X}\}.$$

Since

$$l_k(\bar{y}^k,\bar{\pi}^k) = L_k(\bar{\pi}^k), \quad \text{and} \quad l(\bar{y},\bar{\pi}) = L(\bar{\pi}),$$

it follows that $\{L_k(\bar{\pi}^k)\}_{k\in\mathcal{K}'} \to L(\bar{\pi})$. Thus, from Proposition 4.2,

$$\{L_k(\bar{\pi}^k)\}_{k=1}^{\infty} \to L(\bar{\pi}).$$ ■

With Lemma 4.3, we can also establish an asymptotic relationship between the functions $\{G_k\}$ and G.

Lemma 4.4. *Let $\{(\bar{x}^k, \bar{\pi}^k)\}_{k=1}^{\infty}$ denote a given sequence of bounded points in* **Z**. *Suppose that assumptions C1 and C3 hold.*

C1 f *is continuous and* **X** *is a compact convex set.*

C3 $f_k \xrightarrow{\text{epi}} f$.

Then,

$$\lim_{k\to\infty} G_k(\bar{x}^k, \bar{\pi}^k) = G(\bar{x}, \bar{\pi})$$

whenever

$$\{(\bar{x}^k, \bar{\pi}^k)\}_{k=1}^{\infty} \to (\bar{x}, \bar{\pi}), \text{ and } \{f_k(\bar{x}^k)\}_{k=1}^{\infty} \to f(\bar{x}).$$

Proof. Suppose that $\{(\bar{x}^k, \bar{\pi}^k)\}_{k=1}^{\infty} \subseteq \mathbf{Z}$ such that

$$\{(\bar{x}^k, \bar{\pi}^k)\}_{k=1}^{\infty} \to (\bar{x}, \bar{\pi}) \quad \text{and} \quad \{f_k(\bar{x}^k)\}_{k=1}^{\infty} \to f(\bar{x}).$$

Since $f_k \xrightarrow{\text{epi}} f$, it follows from Lemma 4.3 that $\lim_{k\to\infty} L_k(\bar{\pi}^k) = L(\bar{\pi})$. Hence,

$$\begin{aligned}
\lim_{k\to\infty} G_k(\bar{x}^k, \bar{\pi}^k) &= \lim_{k\to\infty} f_k(\bar{x}^k) + \sum_{i=1}^{m} (\bar{\pi}_i^k + \epsilon_i)\operatorname{Max}\{a_i\bar{x}^k - b_i, 0\} - L_k(\bar{\pi}^k) \\
&= f(\bar{x}) + \sum_{i=1}^{m} (\bar{\pi}_i + \epsilon_i)\operatorname{Max}\{a_i\bar{x} - b_i, 0\} - L(\bar{\pi}) \\
&= H(\bar{x}, \bar{\pi}) - L(\bar{\pi}) \\
&= G(\bar{x}, \bar{\pi}).
\end{aligned}$$

■

Thus, epi-convergence ensures that the gap approximations, $\{G_k\}_{k=1}^{\infty}$, approach the gap function G asymptotically. In the following lemma, we can establish the existence of a subsequence on which $\{G_k(x^k, \pi^k)\}_{k=1}^{\infty}$ converges to zero.

Lemma 4.5. *Let $\{(x^k, \pi^k)\}_{k=1}^{\infty}$ denote the sequence of iterates generated by the inexact primal-dual subgradient method. Suppose that assumptions C1-C3, C5 and C6 hold.*

C1 f *is continuous and* **X** *is a compact convex set.*

C2 $\{f_k\}_{k=1}^{\infty}$ *are convex, continuous and uniformly bounded from below on* **X**.

C3 $f_k \stackrel{\text{epi}}{\to} f$.

C5 $\{d^k\}_{k=1}^{\infty}$ *is a bounded sequence.*

C6 $0 < \beta \leq \lambda_k \leq 1, \; \forall \; k$.

If, in addition, $\{\pi^k\}_{k=1}^{\infty}$ *is bounded, then there exists a subsequence indexed by* $\mathcal{K}^*$ *such that* $\lim_{k \in \mathcal{K}^*} G_k(x^k, \pi^k) = 0$.

Proof. Let (x^*, π^*) denote an optimal solution to $(Q)/(D)$. Since $f_k \stackrel{\text{epi}}{\to} f$, we assume without loss of generality that there exists a sequence of bounded points $\{(\hat{x}^k, \hat{\pi}^k)\}_{k=1}^{\infty}$ in $\mathbf{Z}$, such that

$$\{(\hat{x}^k, \hat{\pi}^k)\}_{k=1}^{\infty} \to (x^*, \pi^*) \quad \text{and} \quad \lim_{k \to \infty} f_k(\hat{x}^k) = f(x^*).$$

It follows from Lemma 4.4 and Lemma 4.1 that

$$\lim_{k \to \infty} G(\hat{x}^k, \hat{\pi}^k) = G(x^*, \pi^*) = 0.$$

To ease our presentation, let $\hat{z}^k = (\hat{x}^k, \hat{\pi}^k)$ and $z^k = (x^k, \pi^k)$, the iterate generated by the inexact primal-dual subgradient method. We consider two cases according to the limiting behavior of $\{\|d^k\|\}_{k=1}^{\infty}$.

Case 1. Suppose that $\underline{\lim}_{k=1}^{\infty} \|d^k\| = 0$. Then, there exists a subsequence indexed by $\mathcal{K}$ such that $\lim_{k \in \mathcal{K}} \|d^k\| = 0$. Since G_k is a convex function and $d^k \in \partial G_k(z^k)$, we have that

$$G_k(z^k) + d^k(\hat{z}^k - z^k) \leq G_k(\hat{z}^k), \qquad \forall \; k.$$

Note that $\lim_{k \in \mathcal{K}} \|d^k\| = 0$ ensures that

$$0 \leq \underline{\lim}_{k \in \mathcal{K}} G_k(z^k) \leq \underline{\lim}_{k \in \mathcal{K}} G_k(\hat{z}^k).$$

Since $\lim_{k \to \infty} G_k(\hat{z}^k) = 0$, it follows that

$$\underline{\lim}_{k \to \infty} G_k(z^k) = 0.$$

Case 2. Suppose that there exists an integer $N_1 < \infty$ such that $\|d^k\| > 0$ for all $k > N_1$. We will prove $\underline{\lim}_{k \to \infty} G_k(z^k) = 0$ by contradiction. Our hypotheses ensure that $\underline{\lim}_{k\to\infty} G_k(z^k) < \infty$. Suppose that $\gamma = \underline{\lim}_{k\to\infty} G_k(z^k) > 0$. Note that

$$\begin{aligned}\|z^{k+1} - \hat{z}^k\|^2 &\le \|z^k - s_k d^k - \hat{z}^k\|^2 \\ &= \|z^k - \hat{z}^k\|^2 + s_k^2\|d^k\|^2 + 2s_k d^k(\hat{z}^k - z^k).\end{aligned}$$

Since $s_k = \lambda_k \frac{G_k(z^k)}{\|d^k\|^2}$, G_k is convex, and $\lambda_k \le 1$ (C6), we have

$$\begin{aligned}\|z^{k+1} - \hat{z}^k\|^2 &\le \|z^k - \hat{z}^k\|^2 + s_k\lambda_k G_k(z^k) + 2s_k\left\{G_k(\hat{z}^k) - G_k(z^k)\right\} \\ &\le \|z^k - \hat{z}^k\|^2 + s_k\left\{2G_k(\hat{z}^k) - G_k(z^k)\right\} \quad \forall\, k > N_1.\end{aligned}$$

Since $\lim_{k\to\infty} G_k(\hat{z}^k) = 0$, and $\gamma > 0$, there exists N_2, $N_1 < N_2 < \infty$ such that for all $k > N_2$,

$$G_k(z^k) > \frac{\gamma}{2} \quad \text{and} \quad G_k(\hat{z}^k) < \frac{\gamma}{8}.$$

Therefore,

$$\|z^{k+1} - \hat{z}^k\|^2 \le \|z^k - \hat{z}^k\|^2 - s_k\frac{\gamma}{4} \qquad \forall k > N_2.$$

By hypothesis, there exists $\mu < \infty$ such that $\|d^k\|^2 < \mu$ for all k. It follows from C6 that $s_k > \frac{\beta\gamma}{2\mu}$ for all $k > N_2$ and

$$\|z^{k+1} - \hat{z}^k\|^2 < \|z^k - \hat{z}^k\|^2 - \frac{\gamma^2\beta}{8\mu} \qquad \forall k > N_2. \tag{4.2}$$

This implies that

$$\|z^{k+1} - \hat{z}^k\| < \|z^k - \hat{z}^k\| \le \|z^k - \hat{z}^{k-1}\| + \|\hat{z}^{k-1} - \hat{z}^k\| \qquad \forall\, k > N_2.$$

Taking limit infimums on both sides, we have

$$\begin{aligned}\underline{\lim}_{k\to\infty} \|z^{k+1} - \hat{z}^k\| \le \underline{\lim}_{k\to\infty} \|z^k - \hat{z}^k\| &\le \underline{\lim}_{k\to\infty} \|z^k - \hat{z}^{k-1}\| + \|\hat{z}^{k-1} - \hat{z}^k\| \\ &= \underline{\lim}_{k\to\infty} \|z^k - \hat{z}^{k-1}\|,\end{aligned}$$

since $\lim_{k\to\infty} \hat{z}^k = z^*$, and thus $\lim_{k\to\infty} \|\hat{z}^{k-1} - \hat{z}^k\| = 0$. It follows that

$$\underline{\lim}_{k\to\infty} \|z^{k+1} - \hat{z}^k\| = \underline{\lim}_{k\to\infty} \|z^k - \hat{z}^k\|.$$

On the other hand, taking limit infimums on both sides of (4.2), we have that

$$0 = \varliminf_{k\to\infty} \|z^{k+1} - \hat{z}^k\|^2 - \varliminf_{k\to\infty} \|z^k - \hat{z}^k\|^2 \le -\frac{\gamma^2\beta}{8\mu}$$

which contradicts $\gamma = \varliminf_{k\to\infty} G_k(z^k) > 0$. Hence, $\varliminf_{k\to\infty} G_k(z^k) = 0$ and the result follows. ■

With Lemmas 4.1 and 4.4, we can establish the existence a subsequence, $\{(x^k, \pi^k)\}_{k\in\mathcal{K}^*}$, such that every accumulation point of this subsequence is optimal to $(Q)/(D)$.

Theorem 4.6. *Let $\{(x^k, \pi^k)\}_{k=1}^{\infty}$ denote the sequence of iterates generated by the inexact primal-dual subgradient method. Suppose that assumptions C1-C6 hold.*

C1 *f is continuous and $\mathbf{X}$ is a compact convex set.*

C2 *$\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below on $\mathbf{X}$.*

C3 *$f_k \stackrel{\text{epi}}{\to} f$.*

C4 *$\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ whenever $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.*

C5 *$\{d^k\}_{k=1}^{\infty}$ is a bounded sequence.*

C6 *$0 < \beta \le \lambda_k \le 1$, $\forall\, k$.*

If, in addition, $\{\pi^k\}_{k=1}^{\infty}$ is bounded, then there exists an index set $\mathcal{K}^*$ such that every accumulation point of $\{(x^k, \pi^k)\}_{k\in\mathcal{K}^*}$ is optimal to $(Q)/(D)$.

Proof. From Lemma 4.5, there exists an index set $\mathcal{K}^*$, such that

$$\lim_{k\in\mathcal{K}^*} G(x^k, \pi^k) = 0.$$

If $\mathcal{K} \subseteq \mathcal{K}^*$ such that $\{(x^k, \pi^k)\}_{k\in\mathcal{K}} \to (\bar{x}, \bar{\pi}) \in \mathbf{Z}$, it follows from Lemma 4.4 that

$$0 = \lim_{k\in\mathcal{K}} G_k(x^k, \pi^k) = G(\bar{x}, \bar{\pi}).$$

Hence, the result follows from Lemma 4.1. ■

In the next section, we replace the assumption of epi-convergence with an assumption of epigraphical nesting, under which we study the limiting behavior of the gap approximations.

4.3 Asymptotic Results with Epigraphically Nested Approximations

In this section, we investigate the inexact primal-dual subgradient method under conditions that are less stringent than those imposed by the theory of epi-convergence. We start with the following assumptions.

Assumptions

D1 f is continuous and $\mathbf{X}$ is a compact convex set.

D2 $\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below on $\mathbf{X}$.

D3 $\overline{\lim}_{k\in\mathcal{K}} f_k(u^k) \leq f(x)$ whenever $\{u^k\}_{k\in\mathcal{K}} \to x$, $\forall\, x \in \mathbf{X}$.

D4 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ whenever $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

D5 $\{d^k\}_{k=1}^{\infty}$ is a bounded sequence.

D6 $\lambda_k \geq 0$, $\{\lambda_k\}_k \to 0$, and $\sum_k \lambda_k = \infty$.

We note that the assumptions are the same as C1-C6 except D3 and D6. D6 together with the remaining assumptions ensure that $s_k \geq 0$, $s_k \to 0$, (a requirement in Lemma 3.2). In D4, we assume that the objective function approximations, $\{f_k\}_{k=1}^{\infty}$ satisfy a special form of epigraphical nesting property. We note that epigraphical nesting provides a substantial relaxation of the requirements of epi-convergence. Nonetheless, a result similar to Lemma 4.3, which is critical to proving Theorem 4.6, can be established with this weaker requirement.

Lemma 4.7. *Let $\{\bar{\pi}^k\}_{k=1}^{\infty}$ denote any non-negative sequence of points in $\mathcal{R}^m$, $\bar{y}^k \in \operatorname{argmin}\{l_k(x, \bar{\pi}^k) \mid x \in \mathbf{X}\}$. Suppose that assumptions D1 and D3 hold.*

D1 f is continuous and $\mathbf{X}$ is a compact convex set.

D3 $\overline{\lim}_{k\in\mathcal{K}} f_k(u^k) \leq f(x)$ whenever $\{u^k\}_{k\in\mathcal{K}} \to x$, $\forall\, x \in \mathbf{X}$.

If, in addition,

$$\{f_k(\bar{y}^k)\}_{k\in\mathcal{K}} \to f(\bar{y}) \quad \text{whenever} \quad \{\bar{y}^k\}_{k\in\mathcal{K}} \to \bar{y},$$

then

$$\{L_k(\bar{\pi}^k)\}_{k\in\mathcal{K}} \to L(\bar{\pi}) \quad \text{whenever} \quad \{\bar{\pi}^k\}_{k\in\mathcal{K}} \to \bar{\pi}.$$

Proof. Suppose that there exists an index set, $\mathcal{K}$ such that $\{\bar{\pi}^k\}_{k\in\mathcal{K}} \to \bar{\pi}$ and $\{\bar{y}^k\}_{k\in\mathcal{K}} \to \bar{y}$. Compactness of $\mathbf{X}$ ensures that $\bar{y} \in \mathbf{X}$. It follows from hypotheses that

$$\begin{aligned}\lim_{k\in\mathcal{K}} l_k(\bar{y}^k, \bar{\pi}^k) &= \lim_{k\in\mathcal{K}} \{f_k(\bar{y}^k) + \bar{\pi}^k(A\bar{y}^k - b)\} \\ &= f(\bar{y}) + \bar{\pi}(A\bar{y} - b) \\ &= l(\bar{y}, \bar{\pi}).\end{aligned}$$

Additionally, let x be any point in $\mathbf{X}$ and $\{u^k\}_{k\in\mathcal{K}} \to x$. It follows from D3 that

$$\begin{aligned}\overline{\lim_{k\in\mathcal{K}}}\, l_k(u^k, \bar{\pi}^k) &= \overline{\lim_{k\in\mathcal{K}}} \{f_k(u^k) + \bar{\pi}^k(Au^k - b)\} \\ &\le f(x) + \bar{\pi}(Ax - b) \\ &= l(x, \bar{\pi}).\end{aligned}$$

This implies that $\mathrm{epi}\{l(\cdot, \bar{\pi})\} \subseteq \underline{\lim}_{k\in\mathcal{K}}\, \mathrm{epi}\{l_k(\cdot, \bar{\pi}^k)\}$. Since

$$\bar{y}^k \in \mathrm{argmin}\{l_k(x, \pi^k) \mid x \in \mathbf{X}\}, \quad \text{and} \quad \{l_k(\bar{y}^k, \pi^k)\}_{k\in\mathcal{K}} \to l(\bar{y}, \bar{\pi}),$$

it follows that $\bar{y} \in \mathrm{argmin}\{l(x, \bar{\pi}) \mid x \in \mathbf{X}\}$ (See Higle and Sen [1992], Theorem 5). Therefore,

$$\lim_{k\in\mathcal{K}} L_k(\bar{\pi}^k) = L(\bar{\pi}).$$ ∎

The following lemma reflects the limiting behavior of the gap approximations, as in Lemma 4.4.

Lemma 4.8. *Let $\{(\bar{x}^k, \bar{\pi}^k)\}_{k=1}^{\infty}$ denote any sequence of points in $\mathbf{Z}$, and*

$$\bar{y}^k \in \mathrm{argmin}\{\ l_k(x, \bar{\pi}^k) \mid x \in \mathbf{X}\ \}, \quad \forall\ k.$$

Suppose that assumptions D1 and D3 hold.

<u>D1</u> f is continuous and $\mathbf{X}$ is a compact convex set.

<u>D3</u> $\overline{\lim}_{k\in\mathcal{K}} f_k(u^k) \le f(x)$ whenever $\{u^k\}_{k\in\mathcal{K}} \to x$, $\forall\ x \in \mathbf{X}$.

If, in addition,

$$\{f_k(\bar{x}^k)\}_{k\in\mathcal{K}} \to f(\bar{x}) \quad \text{whenever} \quad \{\bar{x}^k\}_{k\in\mathcal{K}} \to \bar{x},$$

$$\{f_k(\bar{y}^k)\}_{k\in\mathcal{K}} \to f(\bar{y}) \quad \text{whenever} \quad \{\bar{y}^k\}_{k\in\mathcal{K}} \to \bar{y},$$

then

$$\{G_k(\bar{x}^k, \bar{\pi}^k)\}_{k\in\mathcal{K}} \to G(\bar{x}, \bar{\pi}) \quad \text{whenever} \quad \{(\bar{x}^k, \bar{\pi}^k)\}_{k\in\mathcal{K}} \to (\bar{x}, \bar{\pi}).$$

Proof. Let $\{(\bar{x}^k, \bar{\pi}^k)\}_{k\in\mathcal{K}} \to (\bar{x}, \bar{\pi})$. It follows from Lemma 4.7 that

$$\{L_k(\bar{\pi}^k)\}_{k\in\mathcal{K}} \to L(\bar{\pi}).$$

It follows from our hypothesis that

$$\begin{aligned}\lim_{k\in\mathcal{K}} G_k(\bar{x}^k, \bar{\pi}^k) &= \lim_{k\in\mathcal{K}} \left\{ f_k(\bar{x}^k) + \sum_i (\bar{\pi}_i^k + \epsilon_i) \operatorname{Max}(a_i \bar{x}^k - b_i, 0) - L_k(\bar{\pi}^k) \right\} \\ &= f(\bar{x}) + \sum_i (\bar{\pi}_i + \epsilon_i) \operatorname{Max}(a_i \bar{x} - b_i, 0) - L(\bar{\pi}) \\ &= G(\bar{x}, \bar{\pi}).\end{aligned}$$

■

In the above lemma, if $\underline{\lim}_{k\to\infty} G_k(\bar{x}^k, \bar{\pi}^k) = 0$, it follows from Lemma 4.1 that $(\bar{x}, \bar{\pi})$ is optimal to $(Q)/(D)$. Next, we investigate the asymptotic behavior of $\{G_k\}_{k=1}^{\infty}$ at certain points of $\mathbf{Z}$.

Lemma 4.9. *Let $\hat{x}^k \in \operatorname{argmin}\{f_k(x) \mid Ax \le b, x \in \mathbf{X}\}$, and $\hat{\pi}^k$ denote the dual multiplier of $A\hat{x}^k \le b$ associated with the problem of minimizing f_k. Suppose that assumptions D1 and D3 hold.*

D1 *f is continuous and $\mathbf{X}$ is a compact convex set.*

D3 $\overline{\lim}_{k\in\mathcal{K}} f_k(u^k) \le f(x) \quad$ whenever $\quad \{u^k\}_{k\in\mathcal{K}} \to x, \ \forall\, x \in \mathbf{X}$.

If, in addition,

$$\{f_k(\hat{x}^k)\}_{k\in\mathcal{K}} \to f(\hat{x}) \quad \text{whenever} \quad \{\hat{x}^k\}_{k\in\mathcal{K}} \to \hat{x},$$

then

$$\lim_{k\in\mathcal{K}} G_k(\hat{x}, \hat{\pi}) = 0 \quad \text{whenever} \quad \{\hat{x}^k, \hat{\pi}^k\}_{k\in\mathcal{K}} \to (\hat{x}, \hat{\pi}).$$

Proof. Let $\{(\hat{x}^k, \hat{\pi}^k)\}_{k\in\mathcal{K}} \to (\hat{x}, \hat{\pi}) \in \mathbf{Z}$. Since

$$\hat{x}^k \in \operatorname{argmin}\{f_k(x) \mid Ax \leq b,\ x \in \mathbf{X}\}$$

and $\hat{\pi}^k$ is the corresponding dual multiplier, it follows that

$$f_k(\hat{x}^k) = f_k(\hat{x}^k) + \hat{\pi}^k(A\hat{x}^k - b) = L_k(\hat{\pi}^k).$$

Thus, $G_k(\hat{x}^k, \hat{\pi}^k) = 0$, and

$$\hat{x}^k \in \operatorname{argmin}\{l_k(x, \hat{\pi}^k) \mid x \in \mathbf{X}\} \qquad \forall\, k. \tag{4.3}$$

It follows from Lemma 4.8,

$$G(\hat{x}, \hat{\pi}) = 0. \tag{4.4}$$

By definition,

$$f_k(\hat{x}^k) \leq f_k(\hat{x}) \quad \forall\, k. \tag{4.5}$$

It follows from D3 that

$$\overline{\lim_{k\in\mathcal{K}}}\, f_k(\hat{x}) \leq f(\hat{x}). \tag{4.6}$$

Our hypothesis ensures that

$$f(\hat{x}) = \lim_{k\in\mathcal{K}} f_k(\hat{x}^k). \tag{4.7}$$

It follows from (4.5) to (4.7) that

$$f(\hat{x}) = \lim_{k\in\mathcal{K}} f_k(\hat{x}^k) \leq \underline{\lim}_{k\in\mathcal{K}}\, f_k(\hat{x}) \leq \overline{\lim_{k\in\mathcal{K}}}\, f_k(\hat{x}) \leq f(\hat{x}).$$

As a result,

$$\lim_{k\in\mathcal{K}} f_k(\hat{x}) = f(\hat{x}). \tag{4.8}$$

Next, let

$$\hat{y}^k \in \operatorname{argmin}\{l_k(x, \hat{\pi}) \mid x \in \mathbf{X}\}.$$

Without loss of generality, suppose that $\{\hat{y}^k\}_{k\in\mathcal{K}} \to \hat{y}$. Since $L_k(\hat{\pi}) = l_k(\hat{y}^k, \hat{\pi})$, it follows that

$$\begin{aligned}\underline{\lim}_{k\in\mathcal{K}} L_k(\hat{\pi}) &= \underline{\lim}_{k\in\mathcal{K}} f_k(\hat{y}^k) + \hat{\pi}(A\hat{y}^k - b) \\ &= \underline{\lim}_{k\in\mathcal{K}} f_k(\hat{y}^k) + \hat{\pi}^k(A\hat{y}^k - b).\end{aligned}$$

It follows from (4.3) that, $\hat{x}^k \in \operatorname{argmin}\{\ l_k(x, \hat{\pi}^k) \mid x \in \mathbf{X}\}$. It follows that

$$\begin{aligned}\underline{\lim}_{k\in\mathcal{K}} L_k(\hat{\pi}) &\geq \underline{\lim}_{k\in\mathcal{K}} f_k(\hat{x}^k) + \hat{\pi}^k(A\hat{x}^k - b) \\ &= \underline{\lim}_{k\in\mathcal{K}} L_k(\hat{\pi}^k).\end{aligned}$$

Since $\{\hat{\pi}^k\}_{k\in\mathcal{K}} \to \hat{\pi}$, Lemma 4.7 ensures that

$$\underline{\lim}_{k\in\mathcal{K}} L_k(\hat{\pi}) \geq \lim_{k\in\mathcal{K}} L_k(\hat{\pi}^k) = L(\hat{\pi}). \tag{4.9}$$

Since

$$0 \leq G_k(\hat{x}, \hat{\pi}) = f_k(\hat{x}) + \sum_i (\hat{\pi}_i + \epsilon_i)\,\mathrm{Max}\{a_i\hat{x} - b_i, 0\} - L_k(\hat{\pi}),$$

it follows from (4.8) and (4.9) that

$$\begin{aligned}0 \leq \underline{\lim}_{k\in\mathcal{K}} G_k(\hat{x}, \hat{\pi}) \leq \overline{\lim}_{k\in\mathcal{K}} G_k(\hat{x}, \hat{\pi}) &\leq f(\hat{x}) + \sum_i (\hat{\pi}_i + \epsilon_i)\,\mathrm{Max}\{a_i\hat{x} - b_i, 0\} - L(\hat{\pi}) \\ &= G(\hat{x}, \hat{\pi}).\end{aligned}$$

From (4.4), $G(\hat{x}, \hat{\pi}) = 0$. Hence, $\lim_{k\in\mathcal{K}} G_k(\hat{x}, \hat{\pi}) = 0$. ∎

Our goal is to identify conditions such that $\underline{\lim}_{k\to\infty} G_k(x^k, \pi^k) = 0$, where $\{(x^k, \pi^k)\}_{k=1}^{\infty}$ is the sequence of iterates generated by the inexact primal-dual subgradient method. In the following lemma, we investigate the consequences that arise when this condition fails to be met.

Lemma 4.10. *Let $\{(x^k, \pi^k)\}_{k=1}^{\infty}$ denote the sequence of iterates generated in the inexact primal-dual subgradient method. Suppose that assumptions D1, D2, D5 and D6 hold.*

D1 f *is continuous and* $\mathbf{X}$ *is a compact convex set.*

D2 $\{f_k\}_{k=1}^{\infty}$ *are convex, continuous and uniformly bounded from below on* $\mathbf{X}$.

D5 $\{d^k\}_{k=1}^{\infty}$ *is a bounded sequence.*

D6 $\lambda_k \geq 0, \quad \{\lambda_k\}_k \to 0, \quad \text{and} \quad \sum_k \lambda_k = \infty.$

If, in addition, $\{(x^k, \pi^k)\}_{k=1}^{\infty}$ *is bounded, there exists* $N < \infty$ *such that for all* $k > N$, $\|d^k\| > 0$, *and* $\underline{\lim}_{k\to\infty} G_k(x^k, \pi^k) = \gamma > 0$, *then for any finite point* $(\hat{x}, \hat{\pi})$ *in* $\mathbf{Z}$, *there exists an index set* $\{t_n\}_{n=1}^{\infty}$, $t_n < t_{n+1}$, $\forall n$ *such that*

$$0 \leq G_{t_n}(x^{t_n}, \pi^{t_n}) < 2G_{t_n}(\hat{x}, \hat{\pi}) + \frac{1}{n} \qquad \forall\, n > 0.$$

Proof. To ease our presentation, let $z^k = (x^k, \pi^k)$ and $\hat{z} = (\hat{x}, \hat{\pi})$. It is sufficient to show that $\forall\ \delta_1 > 0$, and $\forall\ N_1 > N$, there exists $k > N_1$, such that

$$0 \leq G_k(z^k) < 2G_k(\hat{z}) + \delta_1.$$

We proceed by contradiction. Suppose that for some $\delta_1 > 0$, $N_1 > N$, we have

$$G_k(z^k) \geq 2G_k(\hat{z}) + \delta_1 \qquad \forall\ k > N_1.$$

Note that

$$\begin{aligned}\|z^{k+1} - \hat{z}\|^2 &\leq \|z^k - s_k d^k - \hat{z}\|^2 \\ &= \|z^k - \hat{z}\|^2 + s_k^2\|d^k\|^2 + 2s_k d^k(\hat{z} - z^k).\end{aligned}$$

By definition, $s_k = \lambda_k \frac{G_k(z^k)}{\|d^k\|^2}$, and G_k is convex. Thus,

$$\|z^{k+1} - \hat{z}\|^2 \leq \|z^k - \hat{z}\|^2 + s_k\lambda_k G_k(z^k) + 2s_k\left\{G_k(\hat{z}) - G_k(z^k)\right\}.$$

Since $\lambda_k \to 0$, we assume without loss of generality that for all $k > N_1$, $\lambda_k \leq 1$, and thus

$$\begin{aligned}\|z^{k+1} - \hat{z}\|^2 &\leq \|z^k - \hat{z}\|^2 + s_k\left\{2G_k(\hat{z}) - G_k(z^k)\right\} \\ &\leq \|z^k - \hat{z}\|^2 + s_k(-\delta_1) \qquad\qquad \forall k > N_1.\end{aligned}$$

This implies that

$$\|z^k - \hat{z}\|^2 - \|z^{k+1} - \hat{z}\|^2 \geq s_k\delta_1 \qquad \forall k > N_1.$$

Since $\{d^k\}_{k=1}^{\infty}$ is bounded, there exists $\mu > 0$ such that $\|d^k\|^2 < \mu$ for all k. By hypothesis, there exists N_2, $N_1 < N_2 < \infty$ such that $G_k(z^k) > \frac{\gamma}{2}$ for all $k \geq N_2$. Thus,

$$\|z^k - \hat{z}\|^2 - \|z^{k+1} - \hat{z}\|^2 \geq \lambda_k \frac{\gamma}{2\mu}\delta_1 \qquad \forall k \geq N_2$$

$$\Rightarrow \sum_{i=0}^{n} \|z^{N_2+i} - \hat{z}\|^2 - \|z^{N_2+i+1} - \hat{z}\|^2 \geq \frac{\gamma}{2\mu}\delta_1 \sum_{i=0}^{n} \lambda_{N_2+i} \qquad \forall\, n \geq 0.$$

This yields a contradiction between the boundedness of $\{z^k\}_{k=1}^{\infty}$ and D6. ■

In the following theorem, we identify conditions under which a subsequence of the iterates, $\{(x^k, \pi^k)\}_{k=1}^{\infty}$ generated by the inexact primal-dual subgradient method may lead to an optimal solution of $(Q)/(D)$.

Theorem 4.11. *Let $\{(x^k, \pi^k)\}_{k=1}^{\infty}$ denote the sequence of iterates generated by the primal-dual subgradient method,*

$$y^k \in \operatorname{argmin}\{l_k(x, \pi^k) \mid x \in \mathbf{X}\}, \quad \hat{x}^k \in \operatorname{argmin}\{f_k(x) \mid Ax \leq b,\ x \in \mathbf{X}\},$$

and $\hat{\pi}^k$ denote the dual multiplier of $A\hat{x}^k \leq b$, in the problem of minimizing f_k. Suppose that D1-D3, D4'(see below), D5, D6 hold.

D1 *f is continuous and $\mathbf{X}$ is a compact convex set.*

D2 *$\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below on $\mathbf{X}$.*

D3 $\overline{\lim}_{k\in\mathcal{K}} f_k(u^k) \leq f(x)$ whenever $\{u^k\}_{k\in\mathcal{K}} \to x,\ \forall\, x \in \mathbf{X}$.

D4' *For* $(u^k, u) = (x^k, \bar{x}), (y^k, \bar{y}), (\hat{x}^k, \hat{x})$,

$$\{f_k(u^k)\}_{k\in\mathcal{K}} \to f(u) \quad \text{whenever} \quad \{u^k\}_{k\in\mathcal{K}} \to u.$$

D5 *$\{d^k\}_{k=1}^{\infty}$ is a bounded sequence.*

D6 $\lambda_k \geq 0, \quad \{\lambda_k\}_k \to 0, \quad \text{and} \quad \sum_k \lambda_k = \infty$.

If, in addition, there exists an unique $(x^*, \pi^*) \in \mathbf{Z}$ such that $G(x^*, \pi^*) = 0$,

$\{(x^k,\pi^k)\}_{k=1}^{\infty}$ and $\{(\hat{x}^k,\hat{\pi}^k)\}_{k=1}^{\infty}$ are bounded, then there exists a subsequence indexed by $\mathcal{K}^*$ such that $\{(x^k,\pi^k)\}_{k\in\mathcal{K}^*} \to (x^*,\pi^*)$, the optimal solution of $(Q)/(D)$. Furthermore, if there exists $N < \infty$ such that $Ax^k \le b$, $\forall\ k \ge N$, then $\{x^k\}_{k=1}^{\infty} \to x^*$.

Proof. We first show that $\underline{\lim}_{k\to\infty} G_k(x^k,\pi^k) = 0$ by considering two cases according to the limiting behavior of the subgradient approximations $\{d^k\}_{k=1}^{\infty}$.

<u>Case 1.</u> Suppose that $\underline{\lim}_{k\to\infty} \|d^k\| = 0$. Then, there exists an index set $\mathcal{K}$ such that $\lim_{k\in\mathcal{K}} \|d^k\| = 0$. It follows from the convexity of G_k, and the boundedness of $\{(x^k,\pi^k)\}$ and $\{(\hat{x}^k,\hat{\pi}^k)\}$ that

$$\begin{aligned} &G_k(x^k,\pi^k) + d^k\left[(\hat{x}^k,\hat{\pi}^k) - (x^k,\pi^k)\right] \le G_k(\hat{x}^k,\hat{\pi}^k) \\ &\Rightarrow 0 \le \lim_{k\in\mathcal{K}} G_k(x^k,\pi^k) \le \lim_{k\in\mathcal{K}} G_k(\hat{x}^k,\hat{\pi}^k). \end{aligned}$$

By definition, $G_k(\hat{x}^k,\hat{\pi}^k) = 0$, and thus

$$\underline{\lim}_{k\to\infty} G_k(x^k,\pi^k) = 0.$$

<u>Case 2.</u> Suppose that $\underline{\lim}_{k\to\infty} \|d^k\| > 0$. Then, there exists $N < \infty$ such that $\|d^k\| > 0$, $\forall k > N$. We will show that $\underline{\lim}_{k\to\infty} G_k(x^k,\pi^k) = 0$ by contradiction. Suppose that $\underline{\lim}_{k\to\infty} G_k(x^k,\pi^k) > 0$. Let (x^*,π^*) denote *the* optimal solution to $(Q)/(D)$. Lemma 4.10 ensures that there exists an index set $\mathcal{K}$ such that

$$0 < \underline{\lim}_{k\to\infty} G_k(x^k,\pi^k) \le \underline{\lim}_{k\in\mathcal{K}} G_k(x^k,\pi^k) \le \underline{\lim}_{k\in\mathcal{K}} 2G_k(x^*,\pi^*) \qquad (4.11)$$

On the other hand, since $G_k(\hat{x}^k,\hat{\pi}^k) = 0$, Lemma 4.8 ensures that every accumulation point of $\{(\hat{x}^k,\hat{\pi}^k)\}_{k=1}^{\infty}$ is optimal, and thus it follows from our assumption of a unique optimal solution that $\{(\hat{x}^k,\hat{\pi}^k)\}_{k\in\mathcal{K}} \to (x^*,\pi^*)$. Hence applying Lemma 4.9, we have $\underline{\lim}_{k\in\mathcal{K}} G_k(x^*,\pi^*) = 0$ which contradicts (4.11).

Hence, in both cases $\underline{\lim}_{k\to\infty} G_k(x^k,\pi^k) = 0$, and it follows from Lemma 4.1 and 4.8 that there exists a index set $\mathcal{K}^*$ such that every accumulation point of $\{(x^k,\pi^k)\}_{k\in\mathcal{K}^*}$ is optimal to $(Q)/(D)$. Again, our assumption of a unique solution ensures that $\{(x^k,\pi^k)\}_{k\in\mathcal{K}^*} \to (x^*,\pi^*)$.

In addition, if there exists N such that $Ax^k \leq b, \; \forall\, k \geq N$, we have that

$$x^{k+1} = \mathrm{P}_{\mathbf{X}'}(x^k - s_k \xi^k), \qquad \forall\, k \geq N,$$

where $\mathbf{X}' = \{x \in \mathbf{X} \mid Ax \leq b\}$ and $\xi^k \in \partial f_k(x^k)$. Thus, for all $k > N$, the generation of the primal iterate, x^k, coincides with the basic inexact subgradient algorithm in §3.1. We verify the hypotheses of Lemma 3.2 in the following. Assumptions A1-A5 are satisfied, $s_k \geq 0$, and $s_k \to 0$. Since f_k is convex, we have that

$$f_k(x^k) - f_k(x^*) \leq \xi^k(x^k - x^*), \qquad \forall\, k > N.$$

D3 ensures that $\overline{\lim}_{k \in \mathcal{K}^*} f_k(x^*) \leq f^*$, the optimal value of (Q), while D4' ensures that $\lim_{k \in \mathcal{K}^*} f_k(x^k) = f(x^*) = f^*$. It follows that

$$0 \leq \underline{\lim}_{k \in \mathcal{K}^*} f_k(x^k) - f_k(x^*) \leq \underline{\lim}_{k \in \mathcal{K}^*} \xi^k(x^k - x^*) = 0.$$

The last equality is ensured by assumption D5. It follows from Lemma 3.2 that every accumulation point of $\{x^k\}_{k=1}^{\infty}$ is optimal. Our hypothesis ensures the $\{x^k\}_{k=1}^{\infty} \to x^*$. ■

In the next section, we apply the inexact primal-dual subgradient method to solve a stochastic linear program with recourse.

4.4 Application to the Solution of Stochastic Linear Programs

In this section, we will discuss how the inexact primal-dual subgradient method can be used to develop a statistically motivated method for the solution of a two stage stochastic linear program with recourse (SLPR). As in Chapter 3, we assume that SLPR satisfies the complete recourse property, and thus f is continuous. A sequence of objective function approximations, $\{f_k\}_{k=1}^{\infty}$ is specified using the types of approximations derived within a stochastic decomposition (SD) algorithm (see Higle and Sen [1991a]).

Inexact Primal-Dual Subgradient Method
with SD type of Approximations (IPDSD)

Step 0. Initialize $k \leftarrow 0,\ n \leftarrow 0,\ \{\delta_n\}_{n=1}^{\infty} \rightarrow 0,\ V_1 = \phi,\ (x^1, \pi^1)$, ν_1 and f_1.

Step 1. $k \leftarrow k+1$. Let

$$y^k \in \operatorname{argmin}\{l_k(x, \pi^k) \mid x \in \mathbf{X}\}, \quad \text{and} \quad \hat{x}^k \in \operatorname{argmin}\{f_k(x) \mid Ax \leq b,\ x \in \mathbf{X}\}.$$

Step 2. Specify $(\bar{x}, \bar{\pi}),\ \nu_{k+1}(\cdot)$.

Let $d^k \in \partial G_k(x^k, \pi^k)$. If $d^k = 0$, let $s_k = 0$; otherwise let
$s_k = \frac{G_k(x^k, \pi^k)}{\|d^k\|^2}$, $\quad (\bar{x}, \bar{\pi}) = \mathrm{P}_{\mathbf{Z}}\left\{(x^k, \pi^k) - s_k d^k\right\}$.

a) Generate ω^{k+1}.

b) For $u = y^k, \hat{x}^k$, solve $\lambda(u, \omega^{k+1}) \in \operatorname{argmax}\left\{\ \lambda(\omega^{k+1} - Tu) \mid \lambda W \leq g\ \right\}$.

Let $\bar{V} \leftarrow V_k \cup \{\lambda(y^k, \omega^{k+1}),\ \lambda(\hat{x}^k, \omega^{k+1})\}$.

c) For $u = y^k, \hat{x}^k$, $t = 1, \ldots, k+1$, solve

$$\lambda(u)_t^{k+1} \in \operatorname{argmax}\left\{\lambda(\omega^t - Tu) \mid \lambda \in \bar{V}\right\}.$$

d) For $u = y^k, \hat{x}^k$, let $\alpha(u)_{k+1}^{k+1} + \beta(u)_{k+1}^{k+1} x \equiv \frac{1}{k+1} \sum_{t=1}^{k+1} \lambda(u)_t^{k+1}(\omega^t - Tx)$,

e) $\nu_{k+1}(x) = cx + \operatorname{Max}\{\alpha(u)_{k+1}^{k+1} + \beta(u)_{k+1}^{k+1} x \mid u = y^k, \hat{x}^k\}$.

Step 3. Specify (x^{k+1}, π^{k+1}).

If $|\nu_{k+1}(y^k) - f_k(y^k)| < \delta_n$ and $|\nu_{k+1}(\hat{x}^k) - f_k(\hat{x}^k)| < \delta_n$, then

$n \leftarrow n+1,\ k_n \leftarrow k \quad$ and $\quad (x^{k+1}, \pi^{k+1}) \leftarrow (\bar{x}, \bar{\pi})$;
otherwise $(x^{k+1}, \pi^{k+1}) \leftarrow (x^k, \pi^k)$.

Step 4. Specify f_{k+1}.

a) Solve $\lambda(x^{k+1}, \omega^{k+1}) \in \text{argmax}\left\{ \lambda(\omega^{k+1} - Tx^{k+1}) \mid \lambda W \leq g \right\}$.
Let $V_{k+1} \leftarrow \bar{V} \cup \{\lambda(x^{k+1}, \omega^{k+1})\}$.

b) For $t = 1, \ldots, k+1$, solve $\lambda(x^{k+1})_t^{k+1} \in \text{argmax}\left\{\lambda(\omega^t - Tx^{k+1}) \mid \lambda \in V_{k+1}\right\}$.

c) Let $\alpha(x^{k+1})_{k+1}^{k+1} + \beta(x^{k+1})_{k+1}^{k+1} x \equiv \frac{1}{k+1} \sum_{t=1}^{k+1} \lambda(x^{k+1})_t^{k+1}(\omega^t - Tx)$.

d) For $u = x^{k+1}, y^k, \hat{x}^k$, and $t = 1, \ldots, k$, let
$\alpha(u)_t^{k+1} \leftarrow \frac{k}{k+1}\alpha(u)_t^k + \frac{1}{k+1}\lambda(u)_{k+1}^{k+1}\omega^{k+1}, \quad \beta(u)_t^{k+1} \leftarrow \frac{k}{k+1}\beta(u)_t^k - \frac{1}{k+1}\lambda(u)_{k+1}^{k+1}T.$

e) $f_{k+1}(x) = cx + \text{Max}\{\alpha(u)_t^{k+1} + \beta(u)_t^{k+1}x \mid u = x^{k+1}, y^k, \hat{x}^k, t = 1, \ldots, k+1\}$.

Repeat from Step 1.

The objective function approximation f_{k+1} is defined as a convex piecewise linear function. The superscript of $\{\alpha(u), \beta(u)\}$ represents the index of the next iteration, the subscript denotes the iteration at which the cuts are first generated, and the notation u represents the point $(\hat{x}^k, y^k, x^{k+1})$ at which the cut is derived. In Step 2b and 4a, dual subproblems are solved exactly at the sample point ω^{k+1} and $\hat{x}^k, y^k, x^{k+1}$ to update the collection of dual extreme points. In Step 2c and 4b, under the collection, the subproblems are 'solved' at these points and $\{\omega^t\}_{t=1}^{k+1}$. In Step 2d and 4c, the average of the aggregated objectives are used to define the cuts $\{\alpha, \beta\}_{k+1}^{k+1}$. As a result, the cuts will be more accurate at these points as the sample size of $\{\omega^t\}_{t=1}^{k+1}$ increases. Since the cuts $\{\alpha, \beta\}_{k+1}^{k+1}$ at y^k and $\hat{x}^k$ are used to define x^{k+1} in the test of Step 3, we derive the cut at x^{k+1} in Step 4a-4c instead of Step 2b-2d.

In iteration k, we perform the test

$$|\nu_{k+1}(y^k) - f_k(y^k)| < \delta_n \quad \text{and} \quad |\nu_{k+1}(\hat{x}^k) - f_k(\hat{x}^k)| < \delta_n,$$

in Step 3. If the test is satisfied, we let $k_n = k$, and thus

$$(x^{k_n+1}, \pi^{k_n+1}) = \text{P}_{\text{Z}}\left\{(x^{k_n}, \pi^{k_n}) - s_{k_n} d^{k_n}\right\}.$$

As the process continues, the iterates will remain unchanged until the index k_{n+1} is identified in Step 3. Thus,

$$(x^{k_n+1}, \pi^{k_n+1}) = (x^{k_n+2}, \pi^{k_n+2}) = \cdots = (x^{k_{n+1}}, \pi^{k_{n+1}}).$$

As a result,

$$(x^{k_{n+1}}, \pi^{k_{n+1}}) = \mathrm{P}_{\mathbf{Z}}\left\{(x^{k_n}, \pi^{k_n}) - s_{k_n} d^{k_n}\right\}.$$

This method extracts a subsequence using the test in Step 3. We caution that the identification of the subsequence indexed by $\{k_n\}_{n=1}^{\infty}$ depends on the input parameters $\{\delta_n\}_{n=1}^{\infty}$. Within this subsequence, we will ensure that the conditions in Theorem 4.11 are satisfied in the following proposition using results from Higle and Sen [1991a] (summarized in Proposition 2.3).

Proposition 4.12. *In SLPR, if*

· *the feasible region of the recourse dual subproblem is a compact convex polyhedral set,*

· $\mathbf{X}$ *is a compact convex set and* Ω *is a compact,*

· $\lambda_k \geq 0, \quad \{\lambda_k\}_k \to 0, \quad$ and $\quad \sum_k \lambda_k = \infty.$

· $\{(\hat{x}^{k_n}, \hat{\pi}^{k_n})\}_{n=1}^{\infty}$ *and* $\{(x^{k_n}, \pi^{k_n})\}_{n=1}^{\infty}$ *are bounded, where* $\hat{\pi}^k$ *is the dual multiplier of* $A\hat{x}^k \leq b$ *associated with the problem of minimizing* f_k*, all* k*,*

· *there is a unique* (x^*, π^*) *such that* $G(x^*, \pi^*) = 0$*,*

then with probability one, there exists an index set $\mathcal{N}^*$ *such that*

$$\{(x^{k_n}, \pi^{k_n})\}_{n \in \mathcal{N}^*} \to (x^*, \pi^*),$$

the optimal solution of the primal and dual problem of SLPR. Furthermore, $\{x^{k_n}\}_{n=1}^{\infty} \to x^*$ *(wp1).*

Proof: Let $\{(x^k, \pi^k)\}_{k=1}^{\infty}$ denote the sequence of iterates generated by the inexact primal-dual subgradient method with SD type of approximations,

$$y^k \in \operatorname{argmin}\{l_k(x, \pi^k) \mid x \in \mathbf{X}\}, \quad \hat{x}^k \in \operatorname{argmin}\{f_k(x) \mid Ax \leq b,\ x \in \mathbf{X}\}.$$

It is sufficient to show that with probability one, D1-D3, D4', D5, and D6 of Theorem 4.11 are satisfied within the subsequence identified in Step 3.

D1 f is continuous and $\mathbf{X}$ is a compact convex set.

D2 $\{f_k\}_{k=1}^{\infty}$ are convex, continuous and uniformly bounded from below on $\mathbf{X}$.

D3 $\overline{\lim}_{k\in\mathcal{K}} f_k(u^k) \le f(x)$ whenever $\{u^k\}_{k\in\mathcal{K}} \to x$, $\forall\, x \in \mathbf{X}$.

D4' For $(u^k, u) = (x^k, \bar{x}), (y^k, \bar{y}), (\hat{x}^k, \hat{x})$,

$$\{f_k(u^k)\}_{k\in\mathcal{K}} \to f(u) \quad \text{whenever} \quad \{u^k\}_{k\in\mathcal{K}} \to u,$$

D5 $\{d^k\}_{k=1}^{\infty}$ is a bounded sequence.

D6 $\lambda_k \ge 0$, $\{\lambda_k\}_k \to 0$, and $\sum_k \lambda_k = \infty$.

From our hypotheses, D1 and D6 are satisfied. D2 is satisfied since f_k is a piecewise linear convex function. D3 follows from Proposition 2.3. It follows from Theorem 2 in Higle and Sen [1991a] that

$$\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x}) \quad \text{whenever} \quad \{x^k\}_{k\in\mathcal{K}} \to \bar{x}, \quad \text{wp1},$$

$$\{f_{k+1}(y^k)\}_{k\in\mathcal{K}} \to f(\bar{y}) \quad \text{whenever} \quad \{y^k\}_{k\in\mathcal{K}} \to \bar{y}, \quad \text{wp1}, \tag{4.12a}$$

$$\{f_{k+1}(\hat{x}^k)\}_{k\in\mathcal{K}} \to f(\hat{x}) \quad \text{whenever} \quad \{\hat{x}^k\}_{k\in\mathcal{K}} \to \hat{x}, \quad \text{wp1}. \tag{4.12b}$$

As a simple extension of Theorem 3 in Higle and Sen [1991a] (see Proposition 2.3), there exists an index set $\bar{\mathcal{K}}$ such that

$$\{f_{k+1}(y^k) - f_k(y^k)\}_{k\in\bar{\mathcal{K}}} \to 0 \quad \text{and} \quad \{f_{k+1}(\hat{x}^k) - f_k(\hat{x}^k)\}_{k\in\bar{\mathcal{K}}} \to 0, \quad \text{wp1}.$$

By hypothesis, for k sufficiently large, $f_{k+1}(y^k) = \nu_{k+1}(y^k)$ and $f_{k+1}(\hat{x}^k) = \nu_{k+1}(\hat{x}^k)$. It follows that there exists $N < \infty$ such that the index set $\{k_n\}_{n>N} \subseteq \bar{\mathcal{K}}$, and thus

$$\lim_{n\to\infty} \{f_{k_n+1}(y^{k_n}) - f_{k_n}(y^{k_n})\} = 0, \quad \text{wp1},$$

and

$$\lim_{n\to\infty} \{f_{k_n+1}(\hat{x}^{k_n}) - f_{k_n}(\hat{x}^{k_n})\} = 0, \quad \text{wp1}.$$

It follows from (4.12) that

$$\{f_{k_n}(y^{k_n})\}_{n\in\mathcal{N}} \to f(\bar{y}) \quad \text{whenever} \quad \{y^{k_n}\}_{n\in\mathcal{N}} \to \bar{y}, \quad \text{wp1},$$

$$\{f_{k_n}(\hat{x}^{k_n})\}_{n\in\mathcal{N}} \to f(\hat{x}) \quad \text{whenever} \quad \{\hat{x}^{k_n}\}_{n\in\mathcal{N}} \to \hat{x}, \quad \text{wp1}.$$

Hence, with probability one, D4' is satisfied on the subsequence index by $\{k_n\}_{n=1}^{\infty}$. Finally, let ξ^k denote a subgradient of f_k at x^k. Since ξ^k is a convex combination of $c - \lambda T$, where λ belongs to the dual feasible region of the recourse subproblem, our hypothesis ensure that $\{\xi^k\}$ is bounded. It follows from

$$d^k = \left(\xi^k + \sum_{i:a_i x^k - b_i > 0} (\pi_i^k + \epsilon_i) a_i, \quad -(Ay^k - b) \right),$$

and our hypotheses that $\{d^k\}_{k=1}^{\infty}$ is bounded, and thus D5 is satisfied. The optimality result now follows from Theorem 4.11. ■

We identify conditions to ensure that inexact primal-dual subgradient method with SD type of approximations will lead to a solution of a SLPR.

4.5 Conclusions

In this chapter, we established asymptotic optimality for the inexact primal-dual subgradient method under the assumption of both epi-convergence and a form of epigraphical nesting. Using the functional approximations similar to those derived in stochastic decomposition, an inexact primal-dual subgradient method is derived to solve a two stage stochastic linear program with recourse. Note that H_k and L_k provide upper and lower bounds for the optimal value of f_k. One may find that the optimal value f^* (if known) lies outside the range $[H_k, L_k]$, especially in the early stages of the process. Nevertheless, the gap approximation G_k estimates the gap, G and defines the steplength s_k. If $G_k(x^k, \pi^k)$ is close to zero, we expect G to be small, and thus G_k can be used as a criterion to suggest optimality.

CHAPTER 5

SPACE DILATION IN INEXACT SUBGRADIENT METHODS

5.0 Introduction

In chapters 3 and 4, the improvements in the choice of steplengths cannot significantly accelerate convergence if at each iteration the subgradient direction is almost orthogonal to the direction towards the minimum (optimal direction). In this chapter, we will adapt the method of space dilation for subgradient optimization presented in Shor [1985] (see §2.5), which reduces at subsequent iterations the components of the subgradient parallel to the previous subgradients. In §5.1. we construct a basic iterative process with space dilation and establish conditions under which exact subgradient evaluations within the process can lead to optimality. In §5.2, we explore the situation when the exact subgradient is replaced by a subgradient of an objective function approximation, taken from an epigraphical nesting property. In other words, we will incorporate space dilations within inexact subgradient methods. In §5.3, we present our conclusions.

5.1 Basic Iterative Process

We start with a basic iterative process with space dilation and provide an asymptotic analysis. Recall from §2.5 that if $\beta \geq 0$ is a scalar and $d \in \mathcal{R}^n$, then $R(\beta, d)$ is a space dilation operator along the direction d with coefficient β.

Basic Iterative Process

Step 0. $k \leftarrow 0$. x^1, $\epsilon_1, \epsilon_2 > 0$, $0 < \epsilon_1 \leq \beta_1 \leq \epsilon_2 < 1, B_1 = I$ are given.

Step 1. $k \leftarrow k + 1$.

Step 2. Specify ξ^k.

If $\xi^k = 0$, let $s_k = 0$ and $d^k = 0$;

otherwise specify s_k and let $d^k = \frac{B_k^T(\xi^k)}{\|B_k^T(\xi^k)\|}$.

Step 3. $x^{k+1} = x^k - s_k B_k d^k$.

Step 4. Specify $0 < \epsilon_1 \leq \beta_k \leq \epsilon_2 < 1$. $B_{k+1} = B_k R(\beta_k, d^k)$.

Repeat from Step 1.

In the above process, we note that $\{\xi^k\}_{k=1}^{\infty}$ is a sequence of unspecified points. If ξ^k is non-zero, the iterate is defined via

$$x^{k+1} = x^k - s_k B_k d^k,$$

where s_k is a steplength and d^k is a perturbed direction. With additional requirements, we establish the following limiting result.

Lemma 5.1. *Let x^* denote any point, and $\{x^k\}_{k=1}^{\infty}$ be the sequence of iterates generated by the basic iterative process. If*

- $\{B_k^{-1}(x^k - x^*)\}_{k=1}^{\infty}$ *and* $\{x^k\}_{k=1}^{\infty}$ *are bounded,*
- $s_k \geq 0, \quad \{s_k\}_k \to 0,$
- *there exists $N < \infty$ such that $\|\xi^k\| > 0$ for all $k \geq N$,*
- $\underline{\lim}_{k \to \infty} d^k \left(B_k^{-1}(x^k - x^*)\right) \geq 0,$

then

$$\underline{\lim}_{k \to \infty} d^k \left(B_k^{-1}(x^k - x^*)\right) = 0.$$

Proof. To ease our presentation, let $z^k = B_k^{-1}(x^k - x^*)$ and $\alpha_k = \frac{1}{\beta_k} > 1$. It follows from Proposition 2.4 that the inverse of the space dilation operator $R(\beta_k, d^k)$ is $R(\alpha_k, d^k)$. In Step 4, $B_{k+1} = B_k R(\beta_k, d^k)$ implies that

$$B_{k+1}^{-1} = R(\alpha_k, d^k) B_k^{-1}.$$

Thus,

$$\begin{aligned}
\|z^{k+1}\|^2 &= \|B_{k+1}^{-1}(x^{k+1} - x^*)\|^2 \\
&= \|R(\alpha_k, d^k) B_k^{-1}(x^k - s_k B_k d^k - x^*)\|^2 \qquad \forall\, k \geq N \\
&= \|R(\alpha_k, d^k)(z^k - s_k d^k)\|^2.
\end{aligned}$$

For all $x \in \mathcal{R}^n$, $R(\alpha_k, d^k)(x) = x + (\alpha_k - 1)(x d^k) d^k$ and $\|d^k\| = 1$ for all k, and thus it follows that

$$\begin{aligned}
\|z^{k+1}\|^2 &= \|z^k - s_k d^k + (\alpha_k - 1)\left((z^k - s_k d^k) d^k\right) d^k\|^2 \qquad \forall\, k \geq N \\
&= \|z^k - s_k d^k + (\alpha_k - 1)(z^k d^k) d^k - (\alpha_k - 1) s_k d^k\|^2 \\
&= \|z^k + (\alpha_k - 1)(z^k d^k) d^k - \alpha_k s_k d^k\|^2.
\end{aligned}$$

Expanding the right hand side, we have

$$\begin{aligned}
\|z^{k+1}\|^2 &= \|z^k\|^2 + \alpha_k^2 s_k^2 + (\alpha_k - 1)^2 (z^k d^k)^2 + 2(\alpha_k - 1)(z^k d^k)^2 \\
&\quad - 2\alpha_k s_k (z^k d^k) - 2\alpha_k(\alpha_k - 1) s_k (z^k d^k) \\
&= \|z^k\|^2 + \alpha_k^2 s_k^2 + (\alpha_k^2 - 1)(z^k d^k)^2 - 2\alpha_k^2 s_k (z^k d^k) \qquad \forall\, k \geq N
\end{aligned}$$

Since $\alpha_k^2 s_k^2 \geq 0$, it follows that

$$\begin{aligned}
&\|z^{k+1}\|^2 \geq \|z^k\|^2 + (\alpha_k^2 - 1)(z^k d^k)^2 - 2\alpha_k^2 s_k (z^k d^k) \\
\Rightarrow\ &\|z^{k+1}\|^2 - \|z^k\|^2 \geq (\alpha_k^2 - 1)(z^k d^k) \left\{ (z^k d^k) - 2\frac{\alpha_k^2}{\alpha_k^2 - 1} s_k \right\} \qquad \forall\, k \geq N
\end{aligned} \tag{5.2}$$

It follows from $0 < \epsilon_1 \leq \beta_k \leq \epsilon_2 < 1$ in Step 4, and $\alpha_k = \frac{1}{\beta_k}$ that

$$\frac{\alpha_k^2}{\alpha_k^2 - 1} \leq \frac{\epsilon_2^2}{\epsilon_1^2 (1 - \epsilon_2^2)}.$$

We will show that $\eta = \underline{\lim}_{k\to\infty}(z^k d^k) = 0$ by contradiction. By hypothesis, $\eta \geq 0$. Suppose that $\eta > 0$. Since $\{s_k\}_k \to 0$, there exists N_1, $N < N_1 < \infty$ such that

$$(z^k d^k) > \frac{\eta}{2}, \quad \text{and} \quad s_k < \frac{\epsilon_1^2(1-\epsilon_2^2)}{\epsilon_2^2}(\frac{\eta}{8}) \qquad \forall\, k \geq N_1.$$

With $\alpha_k^2 \geq \frac{1}{\epsilon_2^2} > 1$, (5.2) becomes

$$\begin{aligned} \|z^{k+1}\|^2 - \|z^k\|^2 &\geq (\alpha_k^2 - 1)(\frac{\eta}{2})\{\frac{\eta}{2} - \frac{\eta}{4}\} \\ &\geq (\frac{1}{\epsilon_2^2} - 1)\frac{\eta^2}{8} \qquad \forall\, k \geq N_1. \end{aligned}$$

It follows that

$$\begin{aligned} &\sum_{k=N_1}^{N_1+n} \|z^{k+1}\|^2 - \|z^k\|^2 \geq \sum_{k=N_1}^{N_1+n} (\frac{1}{\epsilon_2^2} - 1)\frac{\eta^2}{8} \\ \Rightarrow\ &\|z^{N_1+n+1}\|^2 - \|z^{N_1}\|^2 \geq (n+1)(\frac{1}{\epsilon_2^2} - 1)\frac{\eta^2}{8} \qquad \forall\, n \geq 0. \end{aligned}$$

This yields a conflict with the boundedness of $\{z^k\}_{k=1}^{\infty}$. ■

Next, we consider an unconstrained problem,

$$\begin{aligned} \text{Min} \quad & f(x) \\ s/t \quad & x \in \mathcal{R}^n, \end{aligned} \tag{U}$$

of which the optimal value exists and is finite. In the following result, We apply the basic iterative process to solve (U) by defining the ξ^k in Step 2 as a subgradient of f at x^k.

Theorem 5.2. *Let x^* denote an optimal solution of (U), and $\{x^k\}_{k=1}^{\infty}$ be the sequence of iterates generated by the basic iterative process. If*

- *$\{\|B_k^T(\xi^k)\|\}_{k=1}^{\infty}$, $\{B_k^{-1}(x^k - x^*)\}_{k=1}^{\infty}$ and $\{x^k\}_{k=1}^{\infty}$ are bounded,*
- *$s_k \geq 0$, $\{s_k\}_k \to 0$,*
- *f is convex, bounded and $\xi^k \in \partial f(x^k)\ \forall\, k$,*

then there exists either $\bar{k} < \infty$ such that $x^{\bar{k}}$ is optimal to (U), or a subsequence indexed by $\mathcal{K}^*$ such that

$$\lim_{k\in\mathcal{K}^*} f(x^k) = f(x^*).$$

Proof. We consider two cases in our analysis.

<u>Case 1.</u> Suppose that there exists $\bar{k} < \infty$ such that $\xi^{\bar{k}} = 0$. Since f is convex, it follows that

$$f(x) \geq f(x^{\bar{k}}) + \xi^{\bar{k}}(x^{\bar{k}} - x) \qquad \forall\, x,$$

and thus $x^{\bar{k}}$ is optimal to (U).

<u>Case 2.</u> Suppose that $\xi^k \neq 0$ for all k. It follows from Step 2 that $d^k = \frac{1}{\|B_k^T(\xi^k)\|} B_k^T(\xi^k)$, for all k, and thus,

$$\begin{aligned} d^k\left(B_k^{-1}(x^k - x^*)\right) &= \frac{1}{\|B_k^T(\xi^k)\|} B_k^T(\xi^k)\left(B_k^{-1}(x^k - x^*)\right) \\ &= \frac{1}{\|B_k^T(\xi^k)\|}\xi^k(x^k - x^*) \qquad \forall\, k. \end{aligned}$$

Since f is convex and x^* is an optimal solution, it follows that

$$d^k\left(B_k^{-1}(x^k - x^*)\right) \geq \frac{1}{\|B_k^T(\xi^k)\|}\left[f(x^k) - f(x^*)\right] \geq 0 \qquad \forall\, k.$$

The boundedness of $\{\|B_k^T(\xi^k)\|\}_{k=1}^{\infty}$ implies that there exists μ, $0 < \mu < \infty$ such that $\|B_k^T(\xi^k)\| < \mu$ for all k. Hence,

$$\begin{aligned} & d^k\left(B_k^{-1}(x^k - x^*)\right) \geq \frac{1}{\mu}\left[f(x^k) - f(x^*)\right] \geq 0 \\ \Rightarrow\ & \underline{\lim}_{k\to\infty} d^k\left(B_k^{-1}(x^k - x^*)\right) \geq \underline{\lim}_{k\to\infty} \frac{1}{\mu}\left[f(x^k) - f(x^*)\right] \geq 0 \end{aligned}$$

It follows from Lemma 5.1 that

$$\begin{aligned} & 0 = \underline{\lim}_{k\to\infty} d^k B_k^{-1}(x^k - x^*) \geq \underline{\lim}_{k\to\infty} \frac{1}{\mu}\left[f(x^k) - f(x^*)\right] \geq 0 \\ \Rightarrow\ & \underline{\lim}_{k\to\infty} f(x^k) - f(x^*). \end{aligned}$$

Hence, there exists an index set $\mathcal{K}^*$ such that

$$\lim_{k\in\mathcal{K}^*} f(x^k) - f(x^*) = 0. \qquad \blacksquare$$

The algorithm suggested in Theorem 5.2 uses exact subgradients of f. We note that it belongs to the family of subgradient methods with space dilations described in §2.5. In the next section, we will use a subgradient of an objective approximation to define the direction in the basic iterative process and establish a similar limiting result.

5.2 Inexact Subgradient Methods with Space Dilations

In this section, We consider cases in which the function and subgradient evaluations of f become complicated. We generate a sequence of objective function approximations, $\{f_k\}_{k=1}^{\infty}$, and use their subgradients to define the directions in the basic iterative process, thus leading to inexact subgradient algorithm with space dilations.

We precede our analysis with the following assumptions associated with the basic iterative process.

Assumptions

E1 $\{\|B_k^T(\xi^k)\|\}_{k=1}^{\infty}$, $\{B_k^{-1}(x^k - x^*)\}_{k=1}^{\infty}$ and $\{x^k\}_{k=1}^{\infty}$ are bounded.

E2 $s_k \geq 0, \quad \{s_k\}_k \to 0$.

E3 $\{f_k\}_{k=1}^{\infty}$ are convex, uniformly bounded and $\xi^k \in \partial f_k(x^k)$.

E4 $\overline{\lim}_{k\to\infty} f_k(x) \leq f(x) \qquad \forall\, x \in \mathcal{R}^n$.

E5 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x}) \quad$ whenever $\quad \{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

E1-E2 appeared in the hypotheses of Theorem 5.2. In E4, we impose a type of epigraphical nesting property on $\{f_k\}_{k=1}^{\infty}$. In E5, We also assume that $\{f_k\}_{k=1}^{\infty}$ is asymptotically accurate as the iterates accumulate. Under assumptions E3-E5, we establish the following asymptotic results on $\{f_k\}_{k=1}^{\infty}$.

Lemma 5.3. *Let x^* be an optimal solution of (U), and $\{x^k\}_{k=1}^{\infty}$ denote the sequence of iterates generated by the basic iterative process. Suppose that assumptions E3-E5 hold.*

E3 $\{f_k\}_{k=1}^{\infty}$ *are convex, uniformly bounded and* $\xi^k \in \partial f_k(x^k)$.

E4 $\overline{\lim}_{k\to\infty} f_k(x) \leq f(x) \qquad \forall\, x \in \mathcal{R}^n$.

E5 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x}) \quad$ *whenever* $\quad \{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

If, in addition, $\{x^k\}_{k=1}^{\infty}$ *is bounded, then*

$$\underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*) \geq 0.$$

Proof. Without loss of generality, let $\mathcal{K}$ be an index set such that

$$\{x^k\}_{k\in\mathcal{K}} \to \bar{x} \quad \text{and} \quad \underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*) = \lim_{k\in\mathcal{K}} f_k(x^k) - f_k(x^*).$$

It follows from E4-E5 that

$$\begin{aligned} \underline{\lim}_{k\to\infty} f_k(x^k) - f_k(x^*) &= f(\bar{x}) - \lim_{k\in\mathcal{K}} f_k(x^*) \\ &\geq f(\bar{x}) - \overline{\lim}_{k\to\infty} f_k(x^*) \\ &\geq f(\bar{x}) - f(x^*) \geq 0. \end{aligned}$$

■

With Lemma 5.3, we show that the algorithm leads to an optimal solution in the following theorem.

Theorem 5.4. *Let x^* be an optimal solution of (U), and $\{x^k\}_{k=1}^{\infty}$ denote the sequence of the iterates generated by the basic iterative process. Suppose that assumptions E1-E5 hold.*

E1 $\{\|B_k^T(\xi^k)\|\}_{k=1}^{\infty}$, $\{B_k^{-1}(x^k - x^*)\}_{k=1}^{\infty}$ *and* $\{x^k\}_{k=1}^{\infty}$ *are bounded.*

E2 $s_k \geq 0, \quad \{s_k\} \to 0.$

E3 $\{f_k\}_{k=1}^{\infty}$ *are convex, uniformly bounded and* $\xi^k \in \partial f_k(x^k)$.

E4 $\overline{\lim}_{k\to\infty} f_k(x) \leq f(x) \qquad \forall\, x \in \mathcal{R}^n.$

E5 $\{f_k(x^k)\}_{k\in\mathcal{K}} \to f(\bar{x})$ *whenever* $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$.

If, in addition, $\{x^k\}_{k=1}^{\infty}$ is bounded, then there exists an index set $\mathcal{K}^*$ such that every accumulation point of $\{x^k\}_{k\in\mathcal{K}^*}$ is optimal to (U).

Proof. We consider two cases according to the limiting behavior of $\{\xi^k\}_{k=1}^{\infty}$.

Case 1. Suppose that there exists an index set $\mathcal{K}^*$ such that $\lim_{k\in\mathcal{K}^*} \|\xi^k\| = 0$. Let $\mathcal{K} \subset \mathcal{K}^*$ such that $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$. Since f_k is convex, it follows that

$$f_k(x^k) + \xi^k(x - x^k) \leq f_k(x) \qquad \forall\, x \in \mathcal{R}^n.$$

Since $\{\xi^k\}_{k\in\mathcal{K}} \to 0$, assumptions E4-E5 ensure that

$$f(\bar{x}) = \lim_{k\in\mathcal{K}} f_k(x^k) \leq \overline{\lim}_{k\in\mathcal{K}} f_k(x) \leq \overline{\lim}_{k\to\infty} f_k(x) \leq f(x) \qquad \forall\, x \in \mathcal{R}^n$$

and the result follows.

Case 2: Suppose that there exists $N < \infty$ such that $\|\xi^k\| > 0$ for all $k \geq N$. It follows from Step 2 that $d^k = \frac{1}{\|B_k^T(\xi^k)\|} B_k^T(\xi^k)$, for all $k \geq N$. Thus,

$$d^k\left(B_k^{-1}(x^k - x^*)\right) = \frac{1}{\|B_k^T(\xi^k)\|} B_k^T(\xi^k)\left(B_k^{-1}(x^k - x^*)\right)$$
$$= \frac{1}{\|B_k^T(\xi^k)\|} \xi^k(x^k - x^*) \qquad \forall\, k \geq N.$$

Since f_k is convex, it follows that

$$d^k\left(B_k^{-1}(x^k - x^*)\right) \geq \frac{1}{\|B_k^T(\xi^k)\|}\left[f_k(x^k) - f_k(x^*)\right] \qquad \forall\, k \geq N.$$

The boundedness of $\{\|B_k^t(\xi^k)\|\}_{k=1}^{\infty}$ implies that there exists μ, $0 < \mu < \infty$ such that $\|B_k^T(\xi^k)\| < \mu$ for all k. Thus

$$d^k\left(B_k^{-1}(x^k - x^*)\right) \geq \frac{1}{\mu}\left[f_k(x^k) - f_k(x^*)\right] \qquad \forall\, k \geq N.$$

It follows from Lemma 5.3 that

$$\varliminf_{k\to\infty} d^k\left(B_k^{-1}(x^k - x^*)\right) \geq \frac{1}{\mu} \varliminf_{k\to\infty} f_k(x^k) - f_k(x^*) \geq 0.$$

From Lemma 5.1, we have that

$$\varliminf_{k\to\infty} d^k\left(B_k^{-1}(x^k - x^*)\right) = 0,$$

and thus,

$$\varliminf_{k\to\infty} f_k(x^k) - f_k(x^*) = 0.$$

It follows that there exists an index set $\mathcal{K}^*$ such that

$$\lim_{k\in\mathcal{K}^*} f_k(x^k) - f_k(x^*) = 0.$$

If $\mathcal{K} \subseteq \mathcal{K}^*$ and $\{x^k\}_{k\in\mathcal{K}} \to \bar{x}$, it follows from assumptions E4-E5 that

$$f(x^*) \leq f(\bar{x}) = \lim_{k\in\mathcal{K}} f_k(x^k) = \lim_{k\in\mathcal{K}} f_k(x^*) \leq \varlimsup_{k\to\infty} f_k(x^*) \leq f(x^*).$$

Hence, the result follows. ■

Inexact subgradient methods with space dilations inherit the limiting optimality of the algorithm with exact subgradient evaluations presented in §5.1.

5.3 Conclusions

In this chapter, we use space dilations to develop a basic iterative process. We identify conditions under which exact and inexact subgradient evaluations within the basic iterative process can lead to an optimal solution of (U).

CHAPTER 6

EXPERIMENTAL RESULTS AND CONCLUSIONS

6.0 Introduction

In this chapter, we discuss the results of our initial computational experimentation of our algorithms, IXSSD (§3.3) and IPDSD (§4.4), in the solution of two stage stochastic linear programs with recourse. In §6.1, we discuss our implementational concerns. In §6.2, we briefly describe the test problems. In §6.3, we discuss the computational results. In §6.4, we offer conclusions for this dissertation.

6.1 Implementational Concerns

The algorithms, IXSSD and IPDSD, are implemented using the C programming language on Sun Sparc station 1+, and CPLEX linear optimizer to solve the associated linear programs. The projection operation in IXSSD is implemented by ZQPCVX algorithm written in Fortran (Powell [1989]). For each test problem (See §6.2), we use 30 independent replications with the same input parameters but different initial seeds for random numbers which are generated by Paul Sanchez's RNG routine. The input parameters are simply chosen in a convenient fashion. In IXSSD, we let the step coefficient, $\lambda_k = \frac{1}{k}$, $\forall\, k$, while in IPDSD, we let $\lambda_k = \frac{10.5}{k}$. It follows that the hypotheses on $\{\lambda_k\}_{k=1}^{\infty}$ in Proposition 3.7 and 4.12 are satisfied. In addition, in IPDSD, we recall that $\{\delta_k\}_{k=1}^{\infty}$ is used to decide whether a new point is generated via the projection operation and ϵ is a parameter for the definition of H_k. We let $\delta_k = \frac{100}{k}$ and $\epsilon_i = 0.01$, $1 \le i \le n$. The multiples, 10.5 in λ_k and 100 in δ_k, are heuristically chosen in specifying λ_k and δ_k so that larger steps and projection operations are allowed in earlier stages.

When running the algorithms, termination is not permitted until at least 30 iterations are completed and no more than 400 iterations are allowed. The bound $f_k(x^k) - f_k$ in IXSSD, and the gap $G_k(x^k, \pi^k)$ in IPDSD, are point estimates of the error bound, $f(x^k) - f^*$, and thus provide preliminary criteria in our termination procedure.

6.1.1 Termination Procedure

There are two major tests in our termination procedure.

1st Test. In each iteration, we first test if bounds, $f_k(x^k) - \bar{f}_k$ in IXSSD and $G_k(x^k, \pi^k)$ IPDSD, are sufficiently small, that is

$$\frac{f_k(x^k) - \bar{f}_k}{|f_k(x^k)|} \leq \text{tolerance} \quad \text{and} \quad \frac{G_k(x^k, \pi^k)}{|H_k(x^k, \pi^k)|} \leq \text{tolerance},$$

tolerance $= 0.05$. If this test is satisfied, we will proceed to the second test; otherwise we will continue the iterative processes.

2nd Test. In the second test, we obtain the variability of $f_k(x^k) - \bar{f}_k$ using a statistical procedure derived by Higle and Sen [1991b]. In particular, we generate multiple observations of the bound, $f_k(x^k) - \bar{f}_k$ using the bootstrap procedure (Efron [1979]) which involves resampling the observations $\{\omega^t\}_{t=1}^k$ (with replacement).

In the k^{th} iteration, we let $\{\alpha_j^k(u), \beta_j^k(u)\}_{j \in \mathcal{J}}$ be the collection of cuts 'tight' at x^k. We recall that j is the iteration at which the cut was first derived using j observations of $\tilde{\omega}$, and u indicates the point with which the cut is associated. A cut is referred to as 'tight' at x^k if

$$|\alpha_j^k(u) + \beta_j^k(u)x^k - f_k(x^k)| < \text{cut tolerance} = 10^{-6}.$$

After resampling the k observations of $\tilde{\omega}$, M samples, denoted by $\{S_m\}_{m=1}^M$, are obtained. For each m, we evaluate

$$\hat{\alpha}_{jm}^k(u) + \hat{\beta}_{jm}^k(u)x \equiv \frac{1}{k} \sum_{t \in S_m} \pi_t^k(u)(\omega^t - Tx) \quad j \leq k, \ t \leq j.$$

Letting

$$\hat{f}_m(x) = cx + \text{Max}\{\hat{\alpha}_{jm}^k(u) + \hat{\beta}_{jm}^k(u)x \mid j \in \mathcal{J}, \ \forall \, u\},$$

and

$$\hat{f}_m^* = \text{Min}\{\hat{f}_m(x) \mid x \in \mathbf{X}\},$$

X is the feasible set, the collection $\{\hat{f}_m(x^k) - \hat{f}_m^*\}_{m=1}^M$ provides M estimates of $f(x^k) - f^*$. If a sufficient fraction of these estimates are close to zero, we infer that $f(x^k) - f^*$ is acceptably small and terminate the algorithm. We use $M = 30$ as the bootstrap sample size and terminate the algorithm as soon as 90% of the estimated error bounds are smaller than the bootstrap tolerance (= 0.05). Of course, if both tests are not satisfied within 400 iterations, we will terminate the process as well.

6.1.2 Errors in Estimation

Upon termination, we consider two types of error. Letting K denote the terminal iteration, we note that $f_K(x^K)$ is a statistical estimate of $f(x^K)$, and thus is subject to error. Furthermore, x^K may not be optimal, so that $f(x^K)$ and f^* may differ. Specifically, the errors are defined as follows:

Error in Obj. Estimate = $|\frac{f_K(x^K) - f(x^K)}{f(x^K)}|$

Deviation from Opt. = $|\frac{f(x^K) - f^*}{f^*}|$.

We use the Evaluator routine written by R. Odio [1990] for the evaluation of $f(x^k)$. This is done by using the sample mean of a large number of iid observations of $h(x^k, \tilde{\omega})$. The size of the sample is chosen so that the width of the 95% confidence interval for $E[h(x^k, \tilde{\omega})]$ is within 0.2% of the sample mean. The value f^* can be obtained by the L-shaped method of Van Slyke and Wets [1969].

6.2 Test Problems

To test and evaluate the performance of both algorithms, we run several test problems, SCRS8, SCAGR7, PGP2 and CEP1. SCRS8 and SCAGR7 are two stage versions of the multistage problems described in Ho and Loute [1981] and Birge [1985]. The descriptions of PGP2 and CEP1 and their associated data are given by Higle, sen and Yakowitz [1990].

We briefly describe the problems as follows:

SCRS8 is a dynamic energy model for the transition from fossil fuel to renewable energy resources. The model uses estimates of the remaining quantities of domestic oil and gas resources as well as technical and environmental feasibility of new methods for synthetic fuel production.

SCAGR7 is a dairy farm expansion planning model, used to maintain a profit maximizing livestock mix by determining crop acreage, feed purchases, and newborn cattle disposition.

PGP2 is a power generation planning model used to determine the capacities of various types of generators required to ensure that demand for electrical power is met. Total expansion cost in this problem is constrained by a fixed budget.

CEP1 is a capacity planning model for manufacturing plant in which several parts are produced on several machines.

In table 6.1 below, we summarize characteristics of the test problems. The table reports dimensions of the first and second stage problems and the dimension of the random components.

	SCRS8	SCAGR7	PGP2	CEP1
1st Stage				
Rows	29	16	3	10
Columns	37	20	4	8
2nd Stage				
Rows	29	39	8	8
Columns	38	102	23	15
# of r.v.'s	3	3	3	3
# of possible outcomes	8	8	729	343

Table 6.1: Test Problem Characteristics

Each problem satisfies the requirements:

· the feasible region of the recourse dual subproblem is a compact convex polyhedral set.

· X is a compact convex set and Ω is a compact set.

We note that in Proposition 3.3, all of the assumptions are satisfied, while in Proposition 4.12, assumptions such as boundedness of sequences of points cannot be verified.

6.3 Computational Results

In this section, we discuss some preliminary computational results of the algorithms, IXSSD and IPDSD, for each test problem presented in §6.2.

6.3.1 Runs of IXSSD

In IXSSD, we use the bound ratio, $\frac{f_k(x^k)-\bar{f}_k}{|f_k(x^k)|}$, for the first test of the termination procedure. We recall the definitions of the errors we evaluated:
Error in Obj. Estimate = $|\frac{f_K(x^K)-f(x^K)}{f(x^K)}|$
Deviation from Opt. = $|\frac{f(x^K)-f^*}{f^*}|$.
For each test problem, we run the algorithm 30 times. The averaged terminal iteration, bound ratio, the errors as well as their standard deviations (in parentheses) are summarized in the following table.

	Terminal iteration	Bound ratio	Error in obj. est.	Deviation from opt.
SCRS8	30 (0.0)	0.0 (0.0)	0.0135 (0.010536)	0.017344 (0.020199)
*SCAGR7	60.3 (69.644023)	0.00065 (0.0)	0.000063 (0.0)	0.000667 (0.0)
PGP2	34.87 (9.383496)	0.009807 (0.000616)	0.045663 (0.050428)	0.053424 (0.06109)
*CEP1	193.23 (184.22038)	0.05213 (0.031305)	0.102626 (0.087966)	0.081498 (0.102299)

Table 6.2: Performance Characteristics for IXSSD

0.0: the value is smaller that 10^{-6}.
*: some runs are terminated at the 400th iteration.

In SCRS8, we note that the average terminal iteration is 30 with standard deviation 0, while in CEP1, terminal iteration is 193.23 with standard deviation 184.22038. Among the test problems, we note that SCAGR7 has the smallest error in objective estimate and deviation from optimality. We note that the bound ratios in SCRS8, SCAGR7 and PGP2, are smaller than the deviation from optimality, although we verified in Chapter 3 that these bounds would overestimated the deviation from optimality in the long run. This situation is not surprising, since there is also error in the estimated objective values within our maximum number of iterations.

For each problem, the upper limit on the number of iterations is 400. All 30 in SCRS8 and PGP2 terminate in less than 400 iterations. A run in SCAGR7 and 13 runs in CEP1 require more than 400 iterations. In SCAGR7, we note that the the deviation from optimality of the outlier is not significant. In CEP1, the bound ratio, error in objective estimate and deviation from optimality are the highest among the test problems. A major reason for such performances is that there are sampled outcomes associated with high costs in the second stage. Such situation also occurs in the runs of IPDSD for CEP1.

6.3.2 Runs of IPDSD

We recall that in the algorithm of IPDSD, some constraints can be relaxed by including them in the gap approximations. In our computations, we relax all the constraint except $x \geq 0$. In order to perform an projection operation onto a compact set, we estimate an upper bound on the feasible points by solving $u = \text{Max}\{x_i \mid Ax \leq b,\ x_i \geq 0,\ i = 1, \cdots, n\}$. The projection operation is then performed onto the box, $\{x \in \mathcal{R}^n \mid 0 \leq x_i \leq u, i = 1, \cdots, n\}$. We also recall that the projection operation is only performed within the subsequence indexed by $\{k_n\}_{n=1}^{\infty}$, which are extracted in Step 3 of IPDSD.

As in §6.3.1, we run each problem 30 times. In addition to the averaged terminal iteration, gap ratio, $\frac{G_k(x^k, \pi^k)}{|H_k(x^k, \pi^k)|}$, error in objective estimate and deviation from optimality, we also offer the terminal index n of the subsequence indexed $\{k_n\}_{n=1}^{\infty}$. Their standard deviations are given in parentheses.

	Terminal iteration	Terminal index (n)	Gap ratio	Error in obj. est.	Deviation from opt.
SCRS8	30 (0.0)	30 (0.0)	0.0 (0.0)	0.014248 (0.011045)	0.017754 (0.021517)
SCAGR7	30 (0.0)	0.0 (0.0)	0.00071 (0.0)	0.000059 (0.0)	0.000671 (0.0)
*PGP2	338.57 (139.71836)	56.57 (17.213948)	0.039677 (0.018520)	0.03076 (0.051962)	0.034857 (0.059321)
*CEP1	400 (0.0)	0.03 (0.173205)	0.353951 (0.269193)	0.054646 (0.035986)	0.049133 (0.049346)

Table 6.3: Performance Characteristics for IPDSD

0.0: the value is smaller that 10^{-6}.

*: some runs are terminated at the 400th iteration.

This table can be interpreted in a manner similar to that of Table 6.2. In SCRS8, since the terminal index is the same as the terminal iteration, each iterate is specified via the subgradient method during the 30 iterations. In SCAGR7, since the terminal index is zero, no iterate is defined via the subgradient method and thus the initial iterate is also the final iterate.

All 30 runs in SCRS8 and SCAGR7 terminate in less than 400 iterations. 25 runs in PGP2 and all runs in CEP1 require more than 400 iterations. While SCRS8 and SCAGR7 offer gap ratios smaller than the deviations from optimality, PGP2 and CEP1 overestimate the deviations. In PGP2, although the gap ratio is smaller than the tolerance (0.05), the second test in the termination procedure (bootstrap routine) fails to identify those outcomes associated with low costs in the second stage. In CEP1, the big gap ratio is due to consistently poor lower bounds given by Lagrangian functions. In comparison with IXSSD for PGP2 and CEP1, this algorithm offers smaller errors in objective estimates and deviations from optimality.

We note that there is no projection operation in SCAGR7 and only one out of 30 in CEP1. These indicate that the limiting optimality established in Proposition 4.12 cannot be verified in finite time for some test problems, with certain chosen input parameters. Since this method requires more input parameters and assumptions than IXSSD, we expect slower performance in general although it offers better objective estimates (see CEP1).

6.4 Conclusions

In this dissertation, we combine subgradient optimization methods with functional approximation techniques, which we refer to as inexact subgradient methods. Our work is motivated by mathematical program in which the evaluation of the objective function and its subgradients is computationally demanding, such as the two stage stochastic program with recourse (SLPR).

Most solution procedures for SLPR generate a sequence of iterates which leads to an optimal solution under suitable conditions. One common kind of procedure, based on functional approximations, generates a sequence of functional approximations, $\{f_k\}_{k=1}^{\infty}$, successively, and define a solution of this function as the next iterate. Another kind of procedure, the gradient methods, uses a quasi-gradient of the objective function at the current iterate to define the direction. With a chosen steplength, it moves to the next iterate. We have combined these techniques using a subgradient of the functional approximation, f_k, to define the direction of our gradient procedure.

In Chapter 3, we investigate inexact subgradient methods when the steplengths satisfy the standard requirements in subgradient methods,

$$s_k \geq 0, \quad \{s_k\}_k \to 0 \quad \text{and} \quad \sum_k s_k = \infty,$$

and the functional approximations satisfy an epigraphical nesting property. In addition, when the steplengths are defined adaptively by the gap between the objective approximation at the current iterate and a lower bound, we identify conditions under which every accumulation point of the iterates is optimal.

In Chapter 4, we use estimated primal-dual values to generate steplengths in inexact subgradient methods. Two kinds of guidelines, epi-convergence and epigraphical nesting property, are used to define the objective function approximations. We refer to these algorithms as inexact primal-dual subgradient methods. We also identify conditions so that the iterates generated in these algorithms may lead to an optimal solution.

In Chapter 5, we discuss how the space dilation operation can be incorporated into a basic inexact subgradient method to solve unconstrained problems. The functional approximations here satisfy an epigraphical nesting property. The steplengths satisfy the conditions, $s_k \geq 0,\ \{s_k\}_k \rightarrow 0$.

The algorithms of the inexact subgradient method with SD type of approximations (IXSSD) and the inexact primal-dual subgradient method with SD type of approximations (IPDSD) are used to solve some SLPR test problems. Our preliminary computational results are summarized in Chapter 6. Note that in these implementations, since our functional approximations are defined statistically, IXSSD and IPDSD can be classified as SQG methods. Departing from most SQG algorithms, we offer relaxed steplengths requirements, statistical termination procedures, and guidelines in designing the functional approximations that satisfy the imposed assumptions.

A brief discussion of possible future research is offered as follows.

· Two kinds of guidelines for functional approximations are used. It is possible to use an alternative guideline and study its practical implementation in inexact subgradient methods.

· When space dilations are considered, different adaptive steplength specifications can be investigated. It is also important to extend these methods to solve constrained problems. Space dilation is a technique that combines current and previous information about the subgradients. We can also adapt the techniques of conjugate gradient methods by aggregating the previous and current subgradients in a fashion similar to the algorithm derived by Ruszczynski and Syski [1983].

· In our implementation, we note that the preliminary results presented are limited within our maximum number of iterations. We may either increase the limit or allow interactive runs. The experiments can be further developed so that the input parameters can be defined interactively or adaptively. We adopt the bootstrap procedure of resampling $\tilde{\omega}$ in our termination procedure. It is possible to try alternative termination criteria and test their efficiency.

REFERENCES

Allen, E., R. Helgason, J. Kennington and B. Shetty [1987], "A generalization of Polyak's result for subgradient optimization," *Mathematical Programming*, 37, pp. 309–317.

Attouch, H. and R. J.-B. Wets [1980], "Approximation and convergence in nonlinear optimization" in *Nonlinear Programming 4*, 0. Mangasarian, R.R. Meyer and S.M. Robinson (eds.), Academic Press, N.Y., pp. 367–394.

Bazaraa, M.S. and C.M. Shetty [1979], *Nonlinear Programming: Theory and Applications,* John Wiley and Sons, N.Y.

Beale, E.M. [1955], "On minimizing a convex function subject to linear inequalities," *J. Royal Stat. Soc.*, 17B, pp. 173–184.

Benders, J.F. [1962], "Partitioning Procedures for solving mixed-variables programming problems," *Numerische Mathematik*, 4, pp. 238–252.

Ben-Tal, A. and E. Hochman [1972], "More bounds on the expectation of a convex function of a random variable," *Journal of Applied Probability*, 9, pp. 803–812.

Birge, J.R. [1985], "Decomposition and partitioning methods for two stage stochastic linear program," *Operations Research*, 33, pp. 989–1007.

Birge, J.R. and F.V. Louveaux [1988], "A multicut algorithm for two stage linear programs," *European Journal of Operational Research*, 34, pp. 384–392.

Birge, J.R. and R. J.-B. Wets [1986], "Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse," *Mathematical Programming Study*, 27, pp. 54–102.

Broyden C.G. [1967], "Quasi-Newton methods and their application to function minimization," *Math. Comp.*, 21, pp. 368–381.

Charnes, A. and W.W. Cooper [1959], "Chance-constrained programming," *Management Science*, 5. pp. 73–79.

Clark, F.H. [1983], *Optimization and Nonsmooth Analysis*, John Wiley and Sons, N.Y.

Davidon, W.C. [1959], "Variation algorithm for minimization," *Computer Journal*, 10, pp. 406–410.

Dantzig, G.B. [1955], "Linear programming under uncertainty," *Management Science*, 1, pp. 197–206.

Dulá, J.H. [1989], "An upper bound on the expectation of simplicial functions of multivariate random variables," Working paper, Southern Methodist University, Dallas, TX.

Edmundson, H.P. [1959],"Bounds on the expectation of a convex function of a random variable," *The RAND Corporation Paper,*, 982.

Efron, B. [1979], "Bootstrap Methods: another look at the jackknife," *The Annals of Statistics,* 7, pp. 1–26.

Ermoliev, Y. and R.J.-B. Wets, [1988], "Stochastic programming, an introduction," in *Numerical Techniques for Stochastic Optimization,* Y. Ermoliev and R.J.-B. Wets (eds.), Springer-Verlag, Berlin, pp. 1–32.

Frauendorfer, K. [1988], "Solving SLP recourse problems with arbitrary multivariate distributions — the dependent case," *Mathematics of Operations Research*, 13, no. 3, pp. 377–394.

Frauendorfer, K. and P. Kall, [1986], " A solution method for SLP recourse problems with arbitrary distributions — the independent case," *Problems of Control and Information Theory.*

Held, M., P. Wolfe, and H.P. Crowder, [1974], "Validation of subgradient optimization," *Mathematical Programming*, 6, pp. 62–88.

Higle, J.L. and S. Sen [1991a], "Stochastic Decomposition: an algorithm for stage linear programs with recourse", *Mathematics of Operations Research*, 16, pp. 650-669.

Higle, J.L. and S. Sen [1991b], "Statistical verification of optimality conditions", *Annals of Operations Research*, 30, pp. 215-240.

Higle, J.L. and S. Sen [1992], "On the convergence of algorithms with applications to stochastic and nondifferentiable optimization," *Mathematics of Operations Research*, 17, pp. 112–131.

Higle, J.L., S. Sen and D. Yakowitz [1990], "Finite master programs in stochastic decomposition', working paper, SIE Dept, University of Arizona, Tucson, AZ 85721.

Ho, J.K. and E. Loute [1981], "A set of staircase linear programming test problems," *Mathematical Programming*, 20, pp. 245–251.

Infanger, G. [1989], "Monte Carlo (importance) sampling within a Bender's decomposition for stochastic linear programs," Tech. Report SOL 89-13, Department of Operations Research, Stanford University, Stanford, CA.

Kall, P. [1986], "Approximation to optimization problems: an elementary review," *Mathematics of Operations Research,* 11, pp. 9–18.

Kallberg, J.G., R.W. White and W.T. Ziemba [1982], "Short-term financial planning under uncertainty". *Management Science*, 28, pp. 670–682.

Kiefer, J. and J. Wolfowitz [1952], "Stochastic estimation of the maximum of a regression function," *Annals of Math. Stat.*, 23, pp. 462–466.

Kim, S., H. Ahn, and S. Cho, [1991], "Variable target value subgradient method," *Mathematical Programming*, 49, pp. 359-369.

Madansky, A. [1959], "Bounds on the expectation of a convex function of a multivariate random variable," *Annals of Math. Stat.*, 30, pp. 743–746.

Murphy, F.H., S. Sen and A.L. Soyster [1982], "Electric utility capacity expansion planning with uncertain load forecasts", *AIIE Transactions*, 14, pp. 52–59.

Odio, R. [1990], "Implementation of the conditional stochastic decomposition algorithm," master report, Sie Dept, University of Arizona, Tucson, AZ 85721.

Pflug, G.C. [1988], "Stepsize rules, stopping times and their implementation in stochastic quasi-gradient algorithms," in *Numerical Techniques for Stochastic Optimization,* Y. Ermoliev and R.J.-B. Wets (eds.), Springer-Verlag, Berlin, pp. 353-372.

Prékopa, A. and T. Szañtai [1978], "Flood control reservoir system design using stochastic programming," *Mathematical Programming Study,* 9, North-Holland, pp. 183–151.

Polyak, B.T. [1969], "Minimization of unsmooth functions," *USSR Computational Mathematics and Mathematical Physics,* 9, pp. 14–29.

Powell, M.J.D. [1989], "ZQPCVX Algorithm", Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England.

Rockafellar, R.T. and R. J.-B. Wets, [1991], "Scenarios and policy aggregation in optimization under uncertainty", *Mathematics of Operations Research*, 16, pp. 119–147.

Robbins, H. and S. Monroe, [1951], "On a stochastic approximation technique," *Annals of Math. Stat.*, 22, pp. 400–407.

Robinson S.M. [1991], "Extended scenario analysis", *Annals of Operations Research*, 31, pp. 385–397.

Ruszczynski, A. and W. Syski, [1983], "Stochastic approximation method with gradient averaging for unconstrained problems," *IEEE Transactions on Automatic Control,* 78, pp. 1097–1105.

Ruszczynski, A. [1987], "A linearization method for nonsmooth stochastic programming problems," *Mathematics of Operations Research*, 12, pp. 32–49.

Ruszczynski, A. [1986], "A regularized decomposition method for minimizing a sum of polyhedral functions," *Mathematical Programming*, 35, pp. 309–333.

Sen, S. and H.D. Sherali, [1986], "A class of convergent primal-dual subgradient algorithms for decomposable convex program," *Mathematical Programming*, 35, pp. 279–297.

Shor, N. Z. [1985], *Minimization methods for non differentiable functions,* Springer-Verlag, Berlin.

Tintner, G. [1955], "Stochastic linear programming with applications to agricultural economics," in *Proc. Second Symposium in linear programming*, H. A. Antosiewicz (ed.), Washington, pp. 197–207.

Van Slyke, R. and R.J.-B. Wets [1969], "L-Shaped linear programs with application to optimal control and stochastic programming," *SIAM J. on Applied Mathematics*, 17, pp. 638–663.

Wets, R.J.-B. [1974], "Stochastic programs with fixed recourse: the equivalent deterministic program," *SIAM Review,* 16, pp. 309–339.

Wets, R.J.-B. [1982], "Stochastic programming: solution techniques and approximation schemes," *Mathematical Programming: The State-of-the-Art*, A. Bachem, M. Grotshel and B. Korte (eds.), Springer-Verlag, Berlin, pp. 566–603.

Wets, R.J.-B. [1983], "Solving stochastic programs with simple recourse," *Stochastics*, 10, pp. 219–242.