CENTER FOR
PARALLEL OPTIMIZATION

BARRIER DECOMPOSITION FOR THE
PARALLEL OPTIMIZATION OF
BLOCK-ANGULAR PROGRAMS

by

Gary Schultz

Computer Sciences Technical Report #1092

June 1992

BARRIER DECOMPOSITION FOR THE PARALLEL OPTIMIZATION OF BLOCK-ANGULAR PROGRAMS

By

Gary Lloyd Schultz

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN – MADISON

1991

Contents

Abstract

This thesis concerns methods for the computational solution of smooth, block-angular optimization problems that have very large numbers of variables. These problems include multicommodity network flows and are currently among the most challenging important problems in operations research and industrial mathematics.

After surveying relevant classical literature, we expand upon the barrier function theory of Fiacco and McCormick [1968]. These barrier functions allow us to generate a sequence of nonlinear approximating problems which have simpler constraints. Feasibility issues for the original problem are dealt with efficiently. Details regarding the accuracy of the approximation provided by the barrier problems are given for some barrier functions that involve a logarithm.

Next, we describe a new decomposition method for the approximating barrier problems that is amenable to parallel computation. This decomposition method does not require pseudo-convexity of the objective function, allowing the theory of this thesis to apply to local minimization problems with nonconvex objective functions and coupling constraints. Computational results obtained by combining the barrier approximation and the decomposition methods are given for the PDS problems—a class of real-world multicommodity flow problems from the U.S. Air Force Military Airlift Command. We solve PDS problems larger than are solved elsewhere in the literature, ranging to PDS-70 which has 382,311 flow variables.

Finally, our decomposition method is generalized to include the case of flow of information in an *asynchronous* computing environment. Convergence of the generalized method is proven, and we show that the method contains, as a concrete case, the restricted simplicial decomposition algorithm of Hearn *et al* [1987].

Acknowledgements

I must trace the success of this work back to my widowed Mother who wished and encouraged me to attend college. This is one wish of hers that I have carried out to the utmost.

The best thing that happened to me during graduate school was to marry my best friend, Janice. She deserves a good deal of credit for any success that I might enjoy.

There are a number of professors who have had an important positive influence on my career in graduate school. Foremost is my advisor, Bob Meyer, with whom I have truly enjoyed working. I thank Olvi Mangasarian and Michael Ferris for serving as readers on my thesis committee and Renato DeLeone and Stephen Robinson for rounding out my thesis committee. Thanks to my friend Arnold Neumaier for countless stimulating discussions while he visited Wisconsin in 1988–89. Thanks to Meyer, Mangasarian, DeLeone and Ferris for being receptive to my many queries. Thanks to Robinson for teaching the most enjoyable (and quite possibly the most difficult) courses that I took in graduate school. Thanks to John Strikwerda for all of the help in preparing for the numerical analysis screening exam. Finally, thanks to Howard Conner and Rod Smart for supporting my studies during my tenure at the Mathematics Computing Laboratory.

I appreciate the friendship and companionship of my two office mates, Rob Clark and Jonathan Yackel.

Most of all, I thank God, the Creator of the world, Jesus Christ, in whom is my hope, and the Holy Spirit, who guides me daily. It is only by His grace that my success is possible.

> Trust in the LORD with all your heart
>   and lean not on your own understanding;
> in all your ways acknowledge Him,
>   and He will make your paths straight.
>
> *Proverbs 3:5,6*

This research was partially funded by grants from the National Science Foundation (CCR-8709952, CCR-8907671 and DCR-8512862) and the Air Force Office of Scientific Research (AFOSR-86-0194 and AFOSR-89-0410).

List of Tables

List of Figures

Glossary of Symbols

The following is a list of notation that is used consistently throughout this thesis. The order corresponds roughly to the order in which the symbols are used in the text. The more fundamental notation introduced in §1.1 is not included in this glossary.

♠ Marks the end of a proof.

$\emptyset$ The empty set.

c The cost function.

K Number of blocks of variables.

$N = \sum_{k=1}^{K} N_{[k]}$ Number of variables in the optimization problem.

$\underline{b} \leq x \leq \overline{b}$ Simple bound constraints on x.

$Ax = a$ A block diagonal set of affine constraints.

$M = \sum_{k=1}^{K} M_{[k]}$ Number of rows of A.

$D(x) \leq d$ General coupling constraints.

$Dx \leq d$ Linear coupling constraints.

J Number of rows of D.

$x_{[k]}, A_{[k]}, \ldots$ The kth block of variables x, the portion of A corresponding to the kth block of variables, etc.

$\mathcal{B}$ The block constraints ($x \in \mathcal{B}$ iff $\underline{b} \leq x \leq \overline{b}$ and $Ax = a$).

$\mathcal{C}$ The coupling constraints ($x \in \mathcal{C}$ iff $D(x) \leq d$).

$\mathcal{L}$ The Lagrangian function.

p, q, r Lagrange multipliers or estimates of Lagrange multipliers.

v^* The objective function value at a particular local optimum.

X^* The set of local minima with the function value v^*.

$\rho(s) = \sum_{j=1}^{J} \rho_j(s_j)$ A barrier function.

τ A coefficient on the barrier function.

θ A shifted barrier.

f, f_τ An essentially smooth real valued function of N real variables, usually of the form $f_\tau(\cdot) = c(\cdot) + \tau\rho(\theta - D(\cdot))$.

$\mathcal{P}(\tau, \theta)$ The shifted barrier problem depending on τ and θ.

$\left\{x^{(i)}\right\}$ Sequence of (approximate) solutions to sequence of barrier problems.

$\left\{x^{(i,t)}\right\}_t = \left\{x^{(t)}\right\}$ For fixed i, a sequence tending to the solution of $\mathcal{P}\left(\tau^{(i)}, \theta^{(i)}\right)$.

$R = R(x) = \left\{ \; y \;\middle|\; \underline{R}(x) \leq y \leq \overline{R}(x) \; \right\}$ A decoupled resource allocation.

$g^{(t)}$ Approximation for $\nabla f\left(x^{(t)}\right)$ used in subproblems.

$G^{(t)}$ Positive semi-definite matrix of the quadratic form used in subproblems.

$\Xi^{(t)}$ A line segment feasible for the coordinator.

$y^{(t)}$ Search direction computed relative to $x^{(t)}$.

$Y^{(t)}$ Block diagonal matrix of search directions.

$w^{(t)}$ Vector of weights, one scalar for each column of $Y^{(t)}$.

$\widehat{w}^{(t)}$ Un-damped vector of weights; before a line search procedure.

$\Upsilon_{[k]}^{(t)}$ Set of indices of search directions in $Y_{[k]}^{(t)}$.

$v^{(t)} = \sum_{k=1}^{K} v_{[k]}^{(t)}$ Sizes of $\Upsilon_{[k]}^{(t)}$.

η A culling function.

$F^{(t)}$ Set of w feasible for master problem.

$W^{(t)}$ Set of w feasible for coordinator problem ($\subset F^{(t)}$).

Chapter 1

Notation, Problem Definition and Review of Classical Techniques

1.1 Notation

We shall let $\mathbb{R}$ denote the real numbers and $\mathbb{R}^N$ be the N-tuples (column vectors) of real numbers, both with the standard Euclidean topology. Similarly, $\mathbb{Z}$ will denote the standard ring of integers. The set of $M \times N$ matrices of real numbers is represented by $\mathbb{R}^{M\times N}$, with row vectors denoted by $\mathbb{R}^{1\times N}$ to distinguish them from column vectors. The transpose of a matrix or vector x is denoted by x^T. If x is a real vector, then the symbols x_+ and x_- represent the unique nonnegative vectors with the properties that $x = x_+ - x_-$ and $x_+^\mathsf{T} x_- = 0$ ($x_+ x_-^\mathsf{T} = 0$ if x is a row vector). If the symbol $\mathbb{R}$ is subscripted by a condition, it shall denote the subset of $\mathbb{R}$ satisfying the given condition. For example, $\mathbb{R}_{>0}$ is shorthand for the set of positive real numbers $\{\xi \in \mathbb{R} | \xi > 0\}$. We shall also use the standard notation $\mathbb{R}_+$ to mean $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_-$ to mean $\mathbb{R}_{\leq 0}$.

A vector or matrix that has each component equal to 0 (or 1) is represented by $\mathbf{0}$ (or $\mathbf{1}$). The identity matrix is denoted by $\mathbf{I}$. The sizes of each of these will be evident from the context. Often the sum of the components of $x \in \mathbb{R}^N$ will be written as $\mathbf{1}x$.

Ordinary subscripts will denote elements of a vector or matrix, so that x_n is the nth element of the (row or column) vector x and $A_{m,n}$ is the element of the matrix $A \in \mathbb{R}^{M \times N}$ in the mth row and the nth column. The mth row of the matrix A is denoted $A_{m\bullet}$ and the nth column of A is denoted by $A_{\bullet n}$. By analogy, if $A : \mathbb{R}^N \to \mathbb{R}^M$ is a general map, we denote the component functions (rows) of A by $A_{m\bullet} : \mathbb{R}^N \to \mathbb{R}$ $(m = 1, \ldots, M)$. This is consistent with identifying linear mappings from $\mathbb{R}^N$ to $\mathbb{R}^M$ with $M \times N$ matrices. The notation $x_{[k]}$, introduced in §1.3, denotes the kth *block* of variables, where the blocks are disjoint subsets of the variables.

Ordinary superscripts will denote raising to a power, so that ξ^2 is the square of ξ. Superscripts of the form $x^{(i)}$ denote the ith element of the sequence $\left\{x^{(i)}\right\}_{i=0}^{\infty}$.

If $\left\{x^{(i)}\right\}$ is a sequence, we let $\lim x^{(i)} = \lim_{i \to \infty} x^{(i)}$ denote its limit, if it exists. We often write $x^{(i)} \to \tilde{x}$ or $x^{(i)} \xrightarrow{i \to \infty} \tilde{x}$ when $\tilde{x} = \lim x^{(i)}$, or $x^{(i)} \downarrow \tilde{x}$ if $\tilde{x} = \lim x^{(i)}$ and $\tilde{x} \leq x^{(i)}$ for each i. We define the set of limit points:

$$\lim \text{pt}\, x^{(i)} = \lim \text{pt} \left\{x^{(i)}\right\} = \left\{x = \text{ limit of some subsequence of } \left\{x^{(i)}\right\}\right\}.$$

When using arguments which require generic subsequences of a given sequence $\left\{x^{(i)}\right\}$, we often **thin the sequence** so that $x^{(i)} \to \tilde{x}$. This merely means that we extract a subsequence that converges to $\tilde{x}$. This often simplifies the notation necessary for proofs requiring subsequences.

If ϕ is a function, $\nabla\phi$ is its gradient with respect to all of its variables. (The derivative of the univariate function ϕ is denoted ϕ'.) The symbol $\nabla_x\phi$ will be used to represent the gradient of ϕ with respect to only the variables named x, which will be clear from the context. If $\phi : \mathbb{R}^N \to \mathbb{R}^M$ and $x \in \mathbb{R}^N$, then $\nabla\phi(x) \in \mathbb{R}^{M \times N}$, i.e., the gradient of each component function is considered to be a row vector. The symbol ∂ is used for partial differentiation, differentiation of real functions (of a single variable) and to identify the dummy variable in an integration: e.g., $\int \phi(\xi)\partial\xi$.

We write $f \in O(g)$ (or $f(y) \in O(g(y))$) if

$$\limsup_{y \to 0} \left| \frac{f(y)}{g(y)} \right|$$

is finite. Therefore, the Taylor expansions of the function $f : \mathbb{R}^N \to \mathbb{R}$ give

$$f(x+y) - f(x) \in O(\|y\|)$$

if f is continuous at x,

$$f(x+y) - f(x) - \nabla f(x) y \in O\left(\|y\|^2\right)$$

if f is continuously differentiable at x and

$$f(x+y) - f(x) - \nabla f(x) y - y^\top \nabla^2 f(x) y \in O\left(\|y\|^3\right)$$

if f is twice continuously differentiable at x.

1.2 Nonlinear Programming

This section quotes some classical results about the general smooth nonlinear programming problem:

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad g(x) \leq \mathbf{0} \quad \text{and} \quad h(x) = \mathbf{0}, \tag{1.1}$$

where $f : \mathbb{R}^N \to \mathbb{R}$, $g : \mathbb{R}^N \to \mathbb{R}^J$ and $h : \mathbb{R}^N \to \mathbb{R}^M$ are continuously differentiable. Define the **Lagrangian** of (1.1) as

$$\mathcal{L}(x, u, v) = f(x) - ug(x) - vh(x),$$

where $u \in \mathbb{R}_-^{1 \times J}$ and $v \in \mathbb{R}^{1 \times M}$ are called **Lagrange multipliers** or simply **multipliers**. The derivative with respect to x is given by

$$\nabla_x \mathcal{L}(x, u, v) = \nabla f(x) - u \nabla g(x) - v \nabla h(x).$$

The following are called the **Karush-Kuhn-Tucker (KKT) conditions** [Karush 1939; Kuhn and Tucker 1951]:

$$g(x) \leq \mathbf{0} \qquad h(x) = \mathbf{0} \tag{1.2}$$

$$u \leq \mathbf{0} \qquad \nabla_x \mathcal{L}(x, u, v) = \mathbf{0} \tag{1.3}$$

$$ug(x) = 0. \tag{1.4}$$

A **KKT point** is either a tuple (x, u, v) that satisfies the KKT conditions, or a point x for which there exist (u, v) that together satisfy the KKT conditions. Condition (1.2) says that x must be feasible for (1.1) and is called the **primal feasibility** condition. Condition (1.3) is called the **dual feasibility** condition, because, as we shall see below, it is related to feasibility of a dual problem in the convex case. Condition (1.4) is called the **complementary slackness condition.** It is well known that, if x is a local minimizer of (1.1) and some *constraint qualification* holds at x, then there must exist multipliers p, q and r for which (x, u, v) is a KKT point. Many algorithms for solving (1.1) (including the ones developed in this thesis) take their motivation from the KKT conditions and are only guaranteed to converge to KKT points.

The following classical theorem [Fiacco and McCormick 1968, theorem 1] says that if no feasible descent direction emanates from the feasible point x, then x is a KKT point.

Theorem 1.1 *If x is feasible for (1.1) and*

$$\left\{ \begin{array}{c} \nabla h(x) z = \mathbf{0} \\ \forall \{j \textit{ such that } g_{j\bullet}(x) = 0\} \;\; \nabla g_{j\bullet}(x) z \leq 0 \end{array} \right\} \implies \nabla f(x) z \geq 0,$$

then x is a KKT point.

The theorem itself suggests a solution technique: "search along feasible directions until no such directions exist." The proof of the above theorem is based upon the cornerstone of inequality theory: Farkas' [1894] lemma. See Schrijver [1986] for an excellent treatment of Farkas' lemma, both theoretical and historical.

Given a point x—feasible for (1.1)—we say that the **Mangasarian–Fromovitz constraint qualification holds at** x [Mangasarian and Fromovitz 1967] if $\nabla h(x)$ has full row rank, and

$$\exists \left\{ z \in \mathbb{R}^N \right\} \quad \left\{ \begin{matrix} \nabla h(x) z = \mathbf{0} \\ g_{j\bullet}(x) = 0 \implies \nabla g_{j\bullet}(x)\, z < 0 \end{matrix} \right\}. \tag{1.5}$$

Similarly, we say that the **weakened Mangasarian–Fromovitz constraint qualification holds at** x if (1.5) holds regardless of the rank of $\nabla h(x)$. The following result of Gauvin [1977] characterizes the boundedness of the solutions of the KKT conditions.

Theorem 1.2 (Gauvin—Boundedness of multipliers) *Let x^* be a local minimum for (1.1). Then the solutions (u, v) to the KKT equations (1.2), (1.3) and (1.4) form a nonempty bounded set iff the Mangasarian–Fromovitz constraint qualification holds at x^*. Moreover, the set of multipliers u are bounded iff the weakened Mangasarian–Fromovitz constraint qualification holds at x^* and the set of multipliers v are bounded iff $\nabla h(x^*)$ has full row rank.*

In the event that f and g are convex, and h is affine $(h(x) = Ax - a)$, (1.1) is a **convex program.** There is a rich duality theory for convex programming which may be used in this case. Therefore, the **dual** of (1.1) is given by

$$\underset{x,u,v}{\text{maximize}} \quad \mathcal{L}(x, u, v) \quad \text{subject to} \quad \nabla_x \mathcal{L}(x, u, v) = \mathbf{0} \quad \text{and} \quad u \leq \mathbf{0}. \tag{1.6}$$

This is why (1.3) are called the dual feasibility conditions. When g is convex, the weakened Mangasarian–Fromovitz constraint qualification (1.5) is equivalent to the (global) condition of Slater [1950]:

$$\exists \left\{ x \in \mathbb{R}^N \right\} \quad Ax = a \;\; ; \;\; g(x) < \mathbf{0}. \tag{1.7}$$

Note that in the special case where (1.1) is a linear program:

$$\underset{x}{\text{minimize}} \quad fx \quad \text{subject to} \quad Dx \leq d \quad \text{and} \quad Ax = a,$$

x appears in the dual only in the objective function with a coefficient of $f - uD - vA$, which is equal to zero by the dual constraint. Therefore, the dual (1.6) simplifies to the usual linear programming dual

$$\begin{array}{ll} \underset{u,v}{\text{maximize}} & ud + va \\ \text{subject to} & uD + vA = f \quad \text{and} \quad u \leq \mathbf{0}, \end{array}$$

which is independent of x.

1.3 Block Angular Programs

Suppose that the variables $x \in \mathbb{R}^N$ are partitioned into K (disjoint) blocks, so that $x_{[k]} \in \mathbb{R}^{N_{[k]}}$ with $N = \sum_k N_{[k]}$. Given two continuously differentiable functions $c : \mathbb{R}^N \to \mathbb{R}$ and $D : \mathbb{R}^N \to \mathbb{R}^M$, the linear maps $A_{[k]} \in \mathbb{R}^{M_{[k]} \times N_{[k]}}$ and the right-hand-sides $a_{[k]} \in \mathbb{R}^{M_{[k]}}$ for each k and the bounds $\underline{b} \in (\mathbb{R} \cup -\infty)^N$ and $\bar{b} \in (\mathbb{R} \cup +\infty)^N$. We shall be concerned with the **block angular programming problem** in the following form:

$$\left.\begin{array}{l} \text{Find a local minimizer } x^* \text{ of } c(x) \\ \text{subject to the } \textit{coupling constraints} \\ \qquad\qquad\qquad\qquad D(x) \leq d, \\ \text{the } \textit{block constraints} \\ \qquad\qquad A_{[k]} x_{[k]} = a_{[k]} \qquad \text{for } k = 1, \ldots, K, \\ \text{and the } \textit{bounds} \\ \qquad\qquad\qquad\qquad \underline{b} \leq x \leq \bar{b}. \end{array}\right\} \qquad (1.8)$$

Note that the bounds may be infinite, allowing for unbounded variables as well. We assume, without loss of generality, that $\underline{b} < \bar{b}$. The block constraints will often be abbreviated with the notation $Ax = a$ so that (1.8) takes the compact form

$$\underset{x}{\text{minimize}} \;\; c(x) \quad \text{subject to} \quad D(x) \leq d, \quad Ax = a, \quad \text{and} \quad \underline{b} \leq x \leq \bar{b}. \qquad (1.9)$$

We will also use the compact notation

$$\underset{\text{“coupling constraints”}}{\mathcal{C} := \{x \in \mathrm{I\!R}^N | D(x) \leq d\}} \quad \text{and} \quad \underset{\text{“block constraints”}}{\mathcal{B} := \{x \in \mathrm{I\!R}^N | Ax = a \text{ and } \underline{b} \leq x \leq \overline{b}\}} \tag{1.10}$$

throughout this thesis.

If D is a convex function, then the feasible region of (1.9) is convex, and if both c and D are convex, (1.9) is a convex program. If D is a linear map, then the feasible region for (1.9) is polyhedral, and if both c and D are linear, (1.9) is a linear program. When D is linear, the coefficient matrix takes on the traditional block-angular form [Lasdon 1970]:

$$\begin{pmatrix} \boxed{\quad A_{[1]} \quad} & & & \\ & \boxed{\quad A_{[2]} \quad} & & \\ & & \ddots & \\ & & & \boxed{\quad A_{[K]} \quad} \\ \multicolumn{4}{c}{\boxed{\qquad\qquad\qquad\qquad D \qquad\qquad\qquad\qquad}} \end{pmatrix} . \tag{1.11}$$

We shall be concerned with exploiting convexity or linearity when it exists. Moreover, there are applications where still more structure exists in the $A_{[k]}$ or D. We consider it fair to computationally exploit any structure in the problem, so long as it stems from considerations in the underlying real-world problem being modeled.

One special class of block angular linear program is the **multicommodity network flow problem**, a block-angular linear program where each $A_{[k]}$ is a node-arc-incidence matrix. (The classical network flow work is the book of Ford and Fulkerson [1962]. Another excellent book, which includes methods for solving multicommodity networks, is the book by Kennington and Helgason [1980].) In the case where the block constraints are network constraints, the decomposition methods developed below benefit from the excellent technology available for solving network flow problems.

Multicommodity networks typically model a a single network of infrastructure, shared by K distinct commodities. In this typical case, the networks are topologically the same for each commodity and the coupling constraints are **generalized upper bounding (GUB)** constraints;

$$\begin{array}{rclcl} A_{[1]} & = & A_{[k]} & & k = 2, \ldots, K \\ D_{[1]} & = & D_{[k]} & & k = 2, \ldots, K, \end{array}$$

where each $D_{\bullet n}$ is either a column of the identity matrix or zero.

There are many real scenarios in which block angular optimization problems occur. As an example of one, we shall consider a problem related to the Patient Distribution System (PDS) problems solved in §4. Consider the problem of evacuating patients from a number of battlefields in the Middle East to a collection of hospitals in the U.S. with transshipment points in between. The wounded are then classified into K categories depending on type and severity of injury. The interconnections between these points are also chosen to form a basic network (represented by the node-arc-incidence-matrix $A_{[k]}$) based on the available equipment and staff as well as geography. We have a decision variable for each injury type traveling each link in the network. For each type of injury we use the number of injured at each battlefield and the capacity of each hospital for dealing with that type of injury to form the divergence constraints $a_{[k]}$. Limits on the number of certain types of patients in particular aircraft (e.g., an aircraft has a limited number of respirators) may be modeled by the upper bounds $\overline{b}_{[k]}$ ($\underline{b}_{[k]} = \mathbf{0}$ in this instance). These constraints are all block separable so far. The coupling constraints $Dx \leq d$ are used to represent the capacity of aircraft traveling a link of the network; the sum of the passengers—regardless of injury type—in a particular aircraft at a particular time is bounded. Now we wish to find a feasible evacuation schedule that minimizes the time taken, presumably weighting the time for the more serious injuries more heavily. This is just one of the many scenarios in which block angular problems occur naturally.

Suppose the Lagrange multipliers on the constraints are denoted as follows:

$$\begin{array}{rcl@{\qquad\qquad}rcl} p \in \mathbb{R}_-^{1\times J} & : & D(x) \leq d & q \in \mathbb{R}^{1\times M} & : & Ax = a \\ r_+ \in \mathbb{R}_+^{1\times N} & : & \underline{b} \leq x & r_- \in \mathbb{R}_+^{1\times N} & : & x \leq \overline{b}. \end{array} \tag{1.12}$$

We write the Lagrangian of (1.8) (or of (1.9), if you like) as

$$\mathcal{L}(x,p,q,r) = c(x) + p(d - D(x)) + q(a - Ax) + r_+(\underline{b} - x) + r_-(x - \overline{b}) \;\;,\;\; p \leq \mathbf{0}.$$

Since c and D are continuously differentiable, the (KKT) equations for (1.8) are written

$$D(x) \leq d \qquad Ax = a \qquad \underline{b} \leq x \leq \overline{b} \tag{1.13}$$

$$p \leq \mathbf{0} \qquad r = \nabla c(x) - p\nabla D(x) - qA \tag{1.14}$$

$$p(d - D(x)) = 0 \tag{1.15}$$

$$r_+(\underline{b} - x) = 0 \qquad r_-(x - \overline{b}) = 0. \tag{1.16}$$

(These are specialized versions of the general conditions (1.2), (1.3) and (1.4).) Note that the multiplier r_n may be interpreted as a **reduced cost** for the primal variable x_n—which is to say that, for a KKT point, x_n has a positive (negative) reduced cost r_n only if x_n is at the corresponding lower (upper) bound. If c and D are convex, the dual of (1.8) is written as

$$\begin{array}{ll} \underset{x,p,q,r}{\text{maximize}} & c(x) + p(d - Dx) + q(a - Ax) + r_+(\underline{b} - x) + r_-(x - \overline{b}) \\ \text{subject to} & r = \nabla c(x) - p\nabla D(x) - qA \quad \text{and} \quad p \leq \mathbf{0}, \end{array} \tag{1.17}$$

which simplifies to

$$\begin{array}{ll} \underset{p,q,r}{\text{maximize}} & pd + qa + r_+\underline{b} - r_-\overline{b} \\ \text{subject to} & r = c - pD - qA \quad \text{and} \quad p \leq \mathbf{0} \end{array} \tag{1.18}$$

in the event that c and D are linear.

1.4 Dantzig-Wolfe Decomposition

Suppose that the block angular program (1.8) is a linear program. We will now survey a classical method for solving block-angular linear programs—the Dantzig-Wolfe decomposition method (or simply the decomposition method) [Dantzig and Wolfe 1960]. (For textbook coverage of Dantzig-Wolfe decomposition see Lasdon [1970] or—in the multicommodity network case—Kennington and Helgason [1980].) Suppose that the polyhedron $\{x \in \mathbb{R}^N | Ax = a; \underline{b} \leq x \leq \bar{b}\}$ has extreme points $\alpha_1, \ldots, \alpha_S$ and extreme rays $\beta_1, \ldots, \beta_T$. Then a classical result of polyhedral theory states that

$$\begin{aligned} \mathcal{B} &= \{ \ x \in \mathbb{R}^N \ \mid \ Ax = a \ ; \ \underline{b} \leq x \leq \bar{b} \ \} \\ &= \{ \ x \in \mathbb{R}^N \ \mid \ x = \sum_{i=1}^{S} \lambda_i \alpha_i + \sum_{i=1}^{T} \mu_i \beta_i \ ; \ \lambda \geq \mathbf{0} \ ; \ \mathbf{1}\lambda = 1 \ ; \ \mu \geq \mathbf{0} \ \}. \end{aligned} \tag{1.19}$$

It is well known that the number of extreme points and rays $S + T$ generally grows as an exponential function of the problem dimension N. Suppose we define $\tilde{c} \in \mathbb{R}^{1\times S}$, $\tilde{D} \in \mathbb{R}^{J\times S}$, $\bar{c} \in \mathbb{R}^{1\times T}$ and $\bar{D} \in \mathbb{R}^{J\times T}$ via

$$\begin{aligned} &\tilde{c}_i := c\alpha_i \quad \text{and} \quad \tilde{D}_{\bullet i} := D\alpha_i \quad \text{for } i = 1, \ldots, S, \text{ and} \\ &\bar{c}_i := c\beta_i \quad \text{and} \quad \bar{D}_{\bullet i} := D\beta_i \quad \text{for } i = 1, \ldots, T. \end{aligned}$$

Then the linear version of the original problem (1.9) is equivalent to the **master** linear program

$$\begin{aligned} &\underset{\lambda,\mu}{\text{minimize}} \ \ \tilde{c}\lambda + \bar{c}\mu \\ &\text{subject to} \ \ \tilde{D}\lambda + \bar{D}\mu \leq d \ ; \ \lambda \geq \mathbf{0} \ ; \ \mathbf{1}\lambda = 1 \ \text{ and } \ \mu \geq \mathbf{0} \end{aligned} \tag{1.20}$$

with fewer $(J + 1)$ rows, but very many $(S + T)$ columns and non-negativity constraints. If we introduce the artificial variable ξ with large positive cost Δ and the slack variables σ, the initial simplex tableau for (1.20) is written

$$\begin{array}{|c|c|c|c|c|}
\lambda & \mu & \sigma & \xi & \\
\hline
\tilde{D} & \bar{D} & \mathbf{I} & \mathbf{0} & d \\
\mathbf{1} & \mathbf{0} & \mathbf{0} & 1 & 1 \\
\hline
\tilde{c} & \bar{c} & \mathbf{0} & \Delta & 0 \\
\hline
 & & \uparrow & \uparrow & \\
 & & p & (\pi + \Delta) &
\end{array}.$$

At any time during the solution via the simplex method the values of the dual variables p and π may be read from the bottom row of the tableau as indicated under the columns for σ and ξ.

Because of the large number of columns of (1.20), the following **column generation** procedure is used to generate only a few (hopefully!) of the α and β. The reduced costs for (1.20) are given by

$$\begin{pmatrix} \tilde{c} & \bar{c} \end{pmatrix} - \begin{pmatrix} p & \pi \end{pmatrix} \begin{pmatrix} \tilde{D} & \bar{D} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} \tag{1.21}$$

where p and π may be read off of the simplex tableau as shown above. The usual steepest edge simplex pivot rule then picks the column with the minimum reduced cost to enter the basis. Recalling that $\tilde{c}$, $\tilde{D}$, $\bar{c}$ and $\bar{D}$ depend on the extreme points and rays, we find the next extreme point or ray by finding an extreme point or ray of

$$\underset{\gamma}{\text{minimize}} \quad (c - pD)\gamma \quad \text{subject to} \quad A\gamma = a \quad \text{and} \quad \underline{b} \leq \gamma \leq \bar{b}. \tag{1.22}$$

If the simplex method for (1.22) terminates with the extreme point γ^*, then we have computed a corresponding α_i in the above notation, while if it terminates with the extreme ray γ^*, then we have computed one of the columns denoted as β_i above. (This follows directly from (1.20) once we notice that the portion depending on π is a fixed constant.) Termination is determined when (1.21) is nonnegative, which is determined by the optimal value of the subproblems (1.22). Since there is a finite (albeit very large) number of columns of the master problem, finite termination of this method is easily deduced. Because of the form of the subproblems (1.22), the Dantzig-Wolfe decomposition method is called a **price directed** decomposition.

Another nice property of Dantzig-Wolfe decomposition is that lower bounds are easily obtained:

Theorem 1.3 (Lower Bounds) *Let v^* be the optimal objective function value for (1.8) (hence also for (1.20)). Suppose (λ, μ) is feasible for (1.20) with the simplex multipliers p and π from the tableau as shown above. If $z(p)$ is the optimal objective function of (1.22), then*

$$\tilde{c}\lambda + \bar{c}\mu + z(p) - \pi \leq v^* \leq \tilde{c}\lambda + \bar{c}\mu.$$

As stated above, the Dantzig-Wolfe decomposition method works for a general $A \in \mathrm{IR}^{M \times N}$. However, if the A is block diagonal—as it is in our formulation of a block-angular program—then each subproblem (1.22) decomposes into K independent linear programs. These may be solved in parallel on a parallel machine. Moreover, the extreme points $\{\alpha_i\}_{i=1}^S$ and rays $\{\beta_i\}_{i=1}^T$ of the polyhedron (1.19) are of the form

$$\{\alpha_i\}_{i=1}^S \cup \{\beta_i\}_{i=1}^T = \left(\{\alpha_{[1]_i}\}_{i=1}^{S_1} \cup \{\beta_{[1]_i}\}_{i=1}^{T_1}\right) \times \cdots \times \left(\{\alpha_{[K]_i}\}_{i=1}^{S_K} \cup \{\beta_{[K]_i}\}_{i=1}^{T_K}\right)$$

where the $\alpha_{[k]_i}, \beta_{[k]_i}$ are the extreme points and rays of the polyhedron defined only by the kth block of variables in (1.19). This suggests that all subproblems may not need to be solved in order to produce an entering column for the simplex method as applied to the master problem.

1.5 Simplicial Decomposition

Simplicial decomposition is a technique for nonlinear programming that was developed originally by von Hohenbalken [1977] and strongly portended by Holloway [1974]. Simplicial decomposition aims at solving the problem

$$\underset{x}{\text{minimize}} \;\; f(x) \quad \text{subject to} \quad x \in \mathcal{B},$$

where f is smooth and pseudo-convex and $\mathcal{B}$ is closed and convex. The theory of simplicial decomposition is based on the fundamental dimensionality result of convex analysis—Carathéodory's theorem. Restricted simplicial decomposition is a modification of simplicial decomposition that uses less memory.

Let $\alpha_1, \ldots, \alpha_s$ be some extreme points of $\mathcal{B}$ and $\beta_1, \ldots, \beta_t$ be some extreme rays of $\mathcal{B}$. If the smallest affine set containing all of the α_n and receding in the directions of all the β_n (i.e. the **affine hull**) has dimension $s + t - 1$, then we say that the α_n and the β_n are **affinely independent**. Affine independence implies that $s + t \leq N + 1$. In the case where the α_n and β_n are affinely independent, the set

$$\left\{ \; x = \sum_{n=1}^{s} \lambda_n \alpha_n + \sum_{n=1}^{t} \mu_n \beta_n \;\middle|\; \lambda \in \mathbb{R}_+^s \;\; ; \;\; \mathbf{1}\lambda = 1 \;\; ; \;\; \mu \in \mathbb{R}_+^t \; \right\}$$

is called the **generalized $s + t$-dimensional simplex** .

Theorem 1.4 (Carathéodory) *Let $\mathcal{B}$ be a closed convex subset of $\mathbb{R}^N$. Then for any $x \in \mathcal{B}$, there exist affinely independent extreme points $\alpha_1, \ldots, \alpha_s$ and extreme rays $\beta_1, \ldots, \beta_t$ of $\mathcal{B}$ such that x is in the generalized $s + t$-dimensional simplex generated by the α_n and the β_n.*

See Rockafellar [1970, §17] for a full discussion and proof of Carathéodory's theorem.

Simplicial decomposition is a *price directed* or *column generation* method analogous to Dantzig-Wolfe decomposition. A general iteration of the algorithm will now be described. Suppose that we have a set of affinely independent extreme points $\{\alpha_1, \ldots, \alpha_s\}$ and extreme directions $\{\beta_1, \ldots, \beta_t\}$ of $\mathcal{B}$ (so that $s+t \leq N+1$). Furthermore, assume that our current iterate x (approximately) solves the **master problem**

$$\begin{aligned} \underset{\lambda,\mu}{\text{minimize}} \;\; & f\left(\sum_{n=1}^{s} \lambda_n \alpha_n + \sum_{n=1}^{t} \mu_n \beta_n\right) \\ \text{subject to} \;\; & \lambda \geq \mathbf{0} \;\; ; \;\; \mathbf{1}\lambda = 1 \quad \text{and} \quad \mu \geq \mathbf{0}, \end{aligned} \tag{1.23}$$

in the sense that

$$x = \sum_{n=1}^{s} \bar{\lambda}_n \alpha_n + \sum_{n=1}^{t} \bar{\mu}_n \beta_n \tag{1.24}$$

for (approximately) optimal $\bar{\lambda}_n$ and $\bar{\beta}_n$. Moreover, suppose that each $\bar{\lambda}_n$ and $\bar{\mu}_n$ is strictly positive, since otherwise the corresponding α_n or β_n may be thrown away while maintaining (1.24). The gradient $\nabla f(x) \in \mathbb{R}^{1\times N}$ is now computed, and the subproblem

$$\underset{y}{\text{minimize}} \quad \nabla f(x)y \quad \text{subject to} \quad y \in \mathcal{B} \tag{1.25}$$

is solved to yield $\bar{y}$—either an extreme point or extreme ray of $\mathcal{B}$. (Solution of the subproblem is a finite procedure if $\mathcal{B}$ is a polyhedron.) This new column $\bar{y}$ is then added to the previously generated set of α_n and β_n and the master problem (1.23) is re-solved. Carathéodory's theorem allows us to always maintain $s + t \leq N + 1$, so that the number of columns that one need store at any iteration is bounded by $N + 1$. This procedure is repeated until a subproblem is solved which gives the optimal value of $\nabla f(x)\bar{y} = \nabla f(x)x$ (in this case theorem 1.1 applies).

In the case where $\mathcal{B}$ is a polyhedron, (1.25) is a linear program with a constraint set that does not depend on the current point x. Therefore, the simplex method may be used to solve (1.25), using the final linear programming basis from the previous iteration as the starting basis for the current iteration [von Hohenbalken 1977; Hearn *et al* 1987; Mulvey *et al* 1990]. In particular, Mulvey *et al* [1990] claim that starting with the old basis—as opposed to starting with an all artificial basis—results in a savings of about a factor of eight in the number of pivots used to compute a new extreme point. (Mulvey *et al* [1990] use simplicial decomposition specialized to generalized network problems, but their results may easily be generalized to work with any polyhedral $\mathcal{B}$.)

A generalization of simplicial decomposition, called **restricted simplicial decomposition**, will be discussed in §5.3.1, where it will be considered as a special case of a new method developed in this thesis.

1.6 Resource Directed Decomposition

In opposition to the *price directed* decomposition methods of §1.4 and §1.5, are the *resource directed* decomposition methods. This dichotomy among primal decomposition methods is determined by the types of subproblems solved by each. Price directed decomposition methods are characterized by subproblems of the form (1.22) and (1.25), in which a modified linear functional is optimized subject to the original block constraints. Resource directed decomposition methods are then characterized by optimizing the original objective function subject to the original block constraints and some block-separable constraints approximating the coupling constraints $D(x) \leq d$. This approximation is interpreted as an allocation of scarce resources. (For a more in-depth coverage of resource directed decomposition see Lasdon [1970, chapter 9].)

Suppose we have a problem of the form

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \sum_{k=1}^{K} c_{[k]}\left(x_{[k]}\right) \\ \text{subject to} & \forall\{k=1,\ldots,K\} \;\; x_{[k]} \in \mathcal{B}_{[k]} \\ & \sum_{k=1}^{K} D_{[k]}\left(x_{[k]}\right) \leq d. \end{array} \tag{1.26}$$

We define the kth subproblem determined by the resource allocation $y_{[k]}$ and its optimal value function as

$$w_{[k]}\left(y_{[k]}\right) := \min\left\{ \;\; c_{[k]}\left(x_{[k]}\right) \;\middle|\; x_{[k]} \in \mathcal{B}_{[k]} \;\; \text{and} \;\; D_{[k]}\left(x_{[k]}\right) \leq y_{[k]} \;\; \right\}. \tag{1.27}$$

We define the set $V_{[k]}$ to be such that $y_{[k]} \in V_{[k]}$ iff the subproblem (1.27) is feasible for $y_{[k]}$. With our subproblems so defined, it is clear that (1.26) is equivalent to the **master** problem

$$\begin{array}{ll} \underset{y}{\text{minimize}} & w(y) := \sum_{k=1}^{K} w_{[k]}\left(y_{[k]}\right) \\ \text{subject to} & \forall\{k=1,\ldots,K\} \;\; y_{[k]} \in V_{[k]} \\ & \sum_{k=1}^{K} y_{[k]} \leq d. \end{array} \tag{1.28}$$

It is conventional to assign the value $+\infty$ to $w_{[k]}\left(y_{[k]}\right)$ when $y_{[k]} \notin V_{[k]}$. Therefore, the master need not explicitly contain the constraints $y_{[k]} \in V_{[k]}$. However, for a non-optimal y, any optimization method for solving (1.28) will need to find a better y, and this is typically possible only when $y_{[k]} \in V_{[k]}$. One way to avoid infeasibility of the subproblems is to modify the original problem by adding artificial variables z. Then (1.26) is replaced by

$$\begin{array}{ll} \underset{x,z}{\text{minimize}} & \sum_{k=1}^{K} \left[c_{[k]}\left(x_{[k]}\right) + \gamma \mathbf{1} z_{[k]}\right] \\ \text{subject to} & \forall\{k=1,\ldots,K\}\;\; x_{[k]} \in \mathcal{B}_{[k]} \\ & \forall\{k=1,\ldots,K\}\;\; z_{[k]} \geq \mathbf{0} \\ & \sum_{k=1}^{K} D_{[k]}\left(x_{[k]}\right) - z_{[k]} \leq d \end{array}$$

for γ sufficiently large. The subproblems arising from this formulation have constraints

$$x_{[k]} \in \mathcal{B}_{[k]} \quad \text{and} \quad D_{[k]}\left(x_{[k]}\right) - z_{[k]} \leq y_{[k]}.$$

These are feasible iff $\mathcal{B}_{[k]} \neq \emptyset$ (which is a necessary condition for feasibility of the original problem), because, given any $y_{[k]}$, the corresponding $z_{[k]}$ may be made arbitrarily large to compensate. Therefore, we may take $V_{[k]} = \mathbb{R}^{N_{[k]}}$ in (1.28).

Note that we have made no convexity assumptions so far in this section. The following theorem shows that, in the context of convex programming, w possesses certain convexity properties but still lacks differentiability.

Theorem 1.5 (Convexity and Non-smoothness of $w(\cdot)$) *Suppose c, D and $\mathcal{B}$ are convex, so that (1.26) is a convex program. Further suppose that c and D are continuously differentiable, $\mathcal{B}$ is compact and that (1.26) has a feasible solution. Then $w(\cdot)$ is a closed, proper convex function, but is, in general, not differentiable.*

We point out that if $D(x) \leq d$ are GUB constraints with each row having at most one non-zero per block, then the subproblems have only modified upper bounds. More generally, if D is linear and

$$\forall\{k=1,\ldots,K \text{ and } j=1,\ldots,J\}\;\; D_{[k]_{j\bullet}} \text{ has at most one non-zero,}$$

then the modified constraints of the subproblem $\left(D_{[k]}\left(x_{[k]}\right) \leq y_{[k]}\right)$ are modified *simple bounds* on the variables $x_{[k]}$. These types of constraints are handled efficiently by most optimization methods.

1.7 Barrier Function Methods

In this section we describe the classical barrier function methods first considered by Frisch [1954; 1955] and later analyzed in great detail by Fiacco and McCormick [1968]. This is important background because it is the theoretical backbone for the barrier function used later in the thesis. Occasionally we shall consider the particular barrier function based on the natural logarithm $\ln(\cdot)$, since it is the case most often used. It is also worth noting that the logarithmic barrier function has gotten increased attention ever since Gill *et al* [1986] showed that Karmarkar's [1984] polynomial time linear programming method is equivalent to a logarithmic barrier function method.

Consider a function

$$\rho \;:\; \mathbb{R}^J \to (\mathbb{R} \cup +\infty) \;:\; s \mapsto \begin{cases} \rho(s) := \sum_{j=1}^{J} \rho_j\left(s_j\right) & \text{if } s > \mathbf{0} \\ +\infty & \text{otherwise .} \end{cases} \tag{1.29}$$

Definition 1.6 *We say that ρ is a* **barrier function (for the positive orthant)** *if each ρ_j satisfies the following criteria:*

1. $\rho_j \;:\; \mathbb{R}_{>0} \to \mathbb{R}$ *is continuous*
2. $\lim_{\xi \downarrow 0} \rho_j(\xi) = +\infty$
3. ρ_j *is* **antitone**; *i.e.*, $\forall\{0 < \xi < \zeta\}\;\; \rho_j(\xi) \geq \rho_j(\zeta)$

We also wish to point out that ρ is L times continuously differentiable (or convex, or strictly convex) iff each ρ_j is L times continuously differentiable (or convex, or strictly convex, respectively).

The most common example is the **logarithmic barrier function**, where $\rho_j(\cdot) = -\ln(\cdot)$ for each j. We note in passing that Fiacco and McCormick [1968] consider functions lacking separability and antitonicity. Many convergence results still hold for these more general functions, but some useful results are complicated and the gain in generality does not seem to be used in practice.

If the function $D : \mathbb{R}^N \rightarrow \mathbb{R}^J$ (written componentwise as $D_{j\bullet} : \mathbb{R}^N \rightarrow \mathbb{R}$) gives the coupling constraints $D(x) \leq d$, we model these coupling constraints with the composite barrier function

$$\rho(d - D(x)) = \sum_{j=1}^{J} \rho_j \left(d_j - D_{j\bullet}(x)\right),$$

which, in the case of the logarithmic barrier function, reduces to

$$-\sum_{j=1}^{J} \ln \left(d_j - D_{j\bullet}(x)\right).$$

It is clear that this composite barrier function is finite-valued iff $D(x) < d$, thus $\rho\left(d - D(\cdot)\right)$ is the barrier function for the set $\mathrm{cl}\{x|D(x) < d\}$. Sometimes we shall call $\rho(\cdot)$ or $\rho\left(d - D(\cdot)\right)$ a "barrier function" and allow the context to resolve any ambiguities. In the case of the barrier function for the positive orthant $\rho(\cdot)$, we say that the **barrier** is the boundary of the positive orthant. On the other hand, when using the barrier function $\rho(d - D(\cdot))$, we say that the **barrier** is at the boundary of $\{x|D(x) < d\}$. This too will be clear from the context.

Suppose that we wish to

$$\underset{x}{\text{minimize}} \ \ c(x) \quad \text{subject to} \quad D(x) \leq d, \tag{1.30}$$

where there exists some point z with $D(z) < d$. The classical barrier function method would solve a sequence of unconstrained problems

$$\underset{x}{\text{minimize}} \ \ f_{\tau^{(i)}}(x) := c(x) + \tau^{(i)}\rho(d - D(x)) \tag{1.31}$$

(where $\tau^{(i)} \downarrow 0$ strictly monotonically) in the hopes that the computed sequence of minimizers $\{x^{(i)}\}$ converges to a minimizer of the constrained problem (1.30).

This may be done if the minimizing set is nice enough. We say that a set $X \subset Y$ is an **isolated (sub)set of** Y [Fiacco and McCormick 1968, page 46] if there is some closed set E such that $X \subset \operatorname{int} E$ and $\emptyset = (E \setminus X) \cap Y$. The following theorem [Fiacco and McCormick 1968, theorem 7] is a key to the local convergence theorem that will follow. It is stated in generic terms because we wish to quote it later on.

Theorem 1.7 (Existence of Compact Perturbation Set) *Suppose all functions f and g defining the generic nonlinear program*

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad g(x) \leq \mathbf{0}$$

are continuous. If a set X^* of local minima corresponding to a function value v^* is a nonempty, isolated, compact subset of

$$Y := \left\{ \ g(y) \leq \mathbf{0} \ \middle| \ f(y) = v^* \ \right\},$$

then there exists a compact set S such that $X^* \subset \operatorname{int} S$, and for any feasible point $y \in S \setminus X^*$, $f(y) > v^*$.

The assumption that X^* be isolated in Y assures that there cannot be feasible points of the same objective function value arbitrarily close to X^*. The following convergence result is classical [Fiacco and McCormick 1968, theorem 8]:

Theorem 1.8 (Convergence of Barrier Function Minimizers) *Suppose that*

1. *c and D are continuous,*

2. *there is some point z with $D(z) < d$,*

3. *the set of points X^* that are local minima of (1.30) corresponding to the value $v^* = c(X^*)$ is a nonempty compact set,*

4. *X^* is isolated in*

$$Y = \left\{ \ y \ \middle| \ D(y) \leq d \quad \textit{and} \quad c(y) = v^* \ \right\},$$

5. *at least one point in* X^* *is in the closure of* $\{x|D(x) < d\}$, *and*

6. $\tau^{(i)} \downarrow 0$ *monotonically.*

Then

1. *there exists a compact set* S *such that* $X^* \subset \operatorname{int} S$ *and for* $\tau^{(i)}$ *small enough, the unconstrained minima of* $f_{\tau^{(i)}}(\cdot)$ *in* $\{x|D(x) < d\} \cap \operatorname{int} S$ *exist and every limit point of any subsequence* $\left\{x^{(i)}\right\}$ *of such minimizers is in* X^*,

2. $\lim_{i \to \infty} c\left(x^{(i)}\right) = v^*$,

3. $\lim_{i \to \infty} \tau^{(i)} \rho \left(d - D(x^{(i)})\right) = 0$,

4. $\lim_{i \to \infty} f_{\tau^{(i)}}\left(x^{(i)}\right) = v^*$,

5. $\left\{c\left(x^{(i)}\right)\right\}$ *is a monotonically decreasing sequence, and*

6. $\left\{\rho\left(d - D(x^{(i)})\right)\right\}$ *is a monotonically increasing sequence.*

Assumptions 3 and 4 are made in order to be able to apply theorem 1.7. Assumption 5 is made to rule out possibilities where

$$\inf\left\{\ c(x)\ \middle|\ D(x) \leq d\ \right\} \ < \ \inf\left\{\ c(x)\ \middle|\ D(x) < d\ \right\}, \tag{1.32}$$

as was pointed out by Danskin [1967]. For an intuitive example where (1.32) holds, consider minimizing x_1 over the "tennis racket" shape in $\mathbb{R}^2$ given by

$$\left\{\ x \in \mathbb{R}^2\ \middle|\ -2 \leq x_1 \leq 2\ ,\quad -1.5 \leq x_2 \leq 1.5 \quad \text{and} \quad -T(x_1) \leq x_2 \leq T(x_1)\ \right\}$$

where

$$T(x_1) := \begin{cases} 0 & \text{if } x_1 < 0 \\ -2x_1^3 + 3x_1^2 & \text{if } x_1 \geq 0, \end{cases}$$

seen in figure 1. (Note that T is once continuously differentiable.) Here the minimum is at $(-2, 0)$, while the minimum over the closure of the interior occurs at the origin.

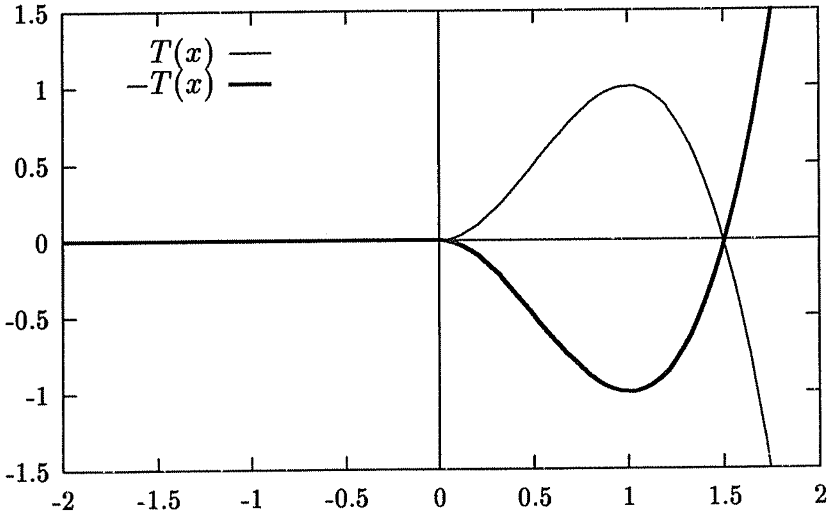


Figure 1: The "tennis racket" example.

We now wish to specialize our results to convex programs, where c and D are convex. The following lemma of Fiacco and McCormick [1968, lemma 11] shows that convex programs lead to convex barrier functions.

Lemma 1.9 (Convexity of Barrier Functions) *If ρ is a convex barrier function and D is convex, then the mapping*

$$x \mapsto \rho(d - D(x))$$

is convex. Furthermore, if c is also convex, then f_τ is convex for any $\tau > 0$.

If c and D are convex, the dual of (1.30) is

$$\begin{array}{ll} \underset{x,p}{\text{maximize}} & c(x) + p(d - D(x)) \\ \text{subject to} & \nabla c(x) - p\nabla D(x) = \mathbf{0} \quad \text{and} \quad p \leq \mathbf{0}. \end{array} \tag{1.33}$$

The results of the last theorem may be sharpened if one assumes convexity and boundedness of the feasible region [Fiacco and McCormick 1968, theorems 25 and 26].

Theorem 1.10 (Primal-Dual Convergence) *Suppose c, D and ρ are all convex functions and ρ is continuously differentiable. Further assume that the hypotheses of theorem 1.8 hold, and that the feasible set of (1.30) is bounded. Then each unconstrained problem (1.31) has a unique finite unconstrained minimizer $x^{(i)}$, every local minimizer is a global minimizer, and every limit point of the bounded sequence $\left\{x^{(i)}\right\}$ is a global minimizer of (1.30). Furthermore, the limit points of the sequence defined by*

$$p^{(i)} := \tau^{(i)}\nabla\rho\left(d - D\left(x^{(i)}\right)\right) = \tau^{(i)}\left(\ldots, \rho_j'\left(d_j - D_{j\bullet}\left(x^{(i)}\right)\right), \ldots\right)_{j=1}^{J} \in \mathbb{R}^{1\times J},$$

are optimal Lagrange multipliers of (1.30) and at least one limit point of $\left\{p^{(i)}\right\}$ exists.

The assumption of boundedness is made in order to bound the sequence $\left\{x^{(i)}\right\}$, thereby guaranteeing the existence of limit points. Fiacco and McCormick [1968] do this with a lemma (lemma 12 in their numbering scheme) that guarantees the boundedness of f_τ level sets. However, they make the following assumption:

> $\lim_{i\to\infty} \rho\left(s^{(i)}\right) = +\infty$ for every infinite sequence $\left\{s^{(i)}\right\}$ where $s^{(i)} > \mathbf{0}$
> and $\lim_{i\to\infty} s_j^{(i)} = 0$ for some j.

We consider this assumption too strong for our purposes because the logarithmic barrier function does not satisfy it: In two variables, if $s_1^{(i)} = 1/i \to 0$ and $s_2^{(i)} = \exp\left(i - \ln(1/i)\right)$, then

$$\rho\left(s^{(i)}\right) = -\ln\left(\frac{1}{i}\right) - i + \ln\left(\frac{1}{i}\right) = -i \to -\infty.$$

It is instructive to note that, since $x^{(i)}$ is an unconstrained minimizer of (1.31), the gradient

$$\nabla f_{\tau^{(i)}}\left(x^{(i)}\right) = \nabla c\left(x^{(i)}\right) + p^{(i)}\nabla D\left(x^{(i)}\right)$$

must vanish. But since $p^{(i)} \leq \mathbf{0}$ (by property 3 of barrier functions), this implies that the pair $\left(x^{(i)}, p^{(i)}\right)$ is feasible for the dual problem (1.33). Therefore we may bound the optimal solution value v^* from above by $c\left(x^{(i)}\right)$ (because $x^{(i)}$ is primal feasible) and from below by the dual function value

$$\begin{aligned} & c\left(x^{(i)}\right) + p^{(i)}\left(d - D\left(x^{(i)}\right)\right) \\ & = c\left(x^{(i)}\right) + \tau^{(i)} \sum_{j=1}^{J} \rho_j'\left(d_j - D_{j\bullet}\left(x^{(i)}\right)\right)\left(d_j - D_{j\bullet}\left(x^{(i)}\right)\right). \end{aligned} \tag{1.34}$$

In the special case of the logarithmic barrier function this yields a particularly nice ε-optimality result [Fiacco and McCormick 1968, page 102]:

Corollary 1.11 (Logarithmic Barrier ε-Optimality) *Suppose (1.30) is a convex program with a strictly interior feasible point. If $x^{(i)}$ is any minimizer of (1.31) then*

$$v^* \leq c\left(x^{(i)}\right) \leq v^* + \tau^{(i)} J.$$

Proof: The inequality $v^* \leq c\left(x^{(i)}\right)$ follows from the primal feasibility of $x^{(i)}$. If $\rho_j = -\ln$, we see that $\rho_j'(\xi)\xi = -1$. In this case the lower bound (1.34) for the optimal value v^* is $c\left(x^{(i)}\right) - \tau^{(i)} J$.

♠

This is one of the most useful mathematical properties of the logarithmic barrier function. For the sake of contrast, consider the barrier function based upon $\rho_j(\cdot) = (\cdot)^{-\ell}$ for some $\ell > 0$. In this case we get $\rho_j'(\xi)\xi = -\ell\xi^{-\ell}$ in the proof above. Then the lower bound (1.34) on v^* depends on the values $d_j - D_{j\bullet}\left(x^{(i)}\right)$. (Small components of this residual contribute large negative amounts to the lower bound (1.34).)

Chapter 2

Interior Point Methods for Block-Angular Problems

In this chapter we describe a scheme that allows us to deal with the block constraints of the block angular problem (1.8) explicitly and the coupling constraints implicitly via a sequence of modified objective functions. One might call this kind of a method a "partial interior point methods" due to the property that the sequence of points generated will be interior to the region defined by *some, but not all* of the inequality constraints. The discussion that follows here is analogous to that of §1.7, which in turn is based on the work of Fiacco and McCormick [1968]. The motivation for the Sequential Unconstrained Minimization Techniques of Fiacco and McCormick was the fact that the technology for solving smooth, unconstrained minimization problems was available and robust. We take our motivation from the observation that solving K smaller linear programs is very easy relative to solving one linear program, K times larger. Moreover, in the multicommodity flow problem, the block constrained problems have only network constraints, making them even easier to solve.

Recall the definition (1.10):

$$\underset{\text{``coupling''}}{\mathcal{C} := \{x \in \mathbb{R}^N | D(x) \le d\}} \quad \text{and} \quad \underset{\text{``block-structured''}}{\mathcal{B} := \{x \in \mathbb{R}^N | Ax = a \text{ and } \underline{b} \le x \le \overline{b}\}}$$

where the A has block diagonal structure. We assume that a point $x^{(0)} \in \mathcal{B}$ is easily computable. This is the case if the problem

$$\underset{x}{\text{minimize}} \ \ \tilde{c}x \quad \text{subject to} \quad x \in \mathcal{B} \tag{2.1}$$

is easily solvable for some $\tilde{c} \in \mathrm{I\!R}^{1\times N}$. This problem may be (trivially) decomposed into the K linear programs

$$\underset{x_{[k]}}{\text{minimize}} \ \ \tilde{c}_{[k]}x_{[k]} \quad \text{subject to} \quad x_{[k]} \in \mathcal{B}_{[k]} \quad \text{for} \quad k = 1, \ldots, K,$$

where

$$\mathcal{B}_{[k]} := \left\{ \ x_{[k]} \ \middle| \ A_{[k]}x_{[k]} = a_{[k]} \quad \text{and} \quad \underline{b}_{[k]} \leq x_{[k]} \leq \bar{b}_{[k]} \ \right\}.$$

We note that if (2.1) is not easily solvable, then the original problem (1.8) is also probably intractable. If (2.1) is infeasible, then $\mathcal{B} = \emptyset$ and the original problem (1.8) is also infeasible.

In subsequent iterations, the algorithm solves additional subproblems with block constraints. The solutions of the subproblems are then *coordinated* into a new approximation to the solution using a smaller (typically K-dimensional) search. During this coordination problem, information about $\mathcal{C}$ is introduced into the objective function by using a barrier function.

§2.1 will deal with methods for computing feasible points and §2.2 will develop the convergence theory for minimizers of barrier problems. §2.3 then summarizes this interior point method.

2.1 Shifting Barriers to Obtain Feasibility

Our goal is to replace the coupling region $\mathcal{C}$ with the barrier function

$$x \mapsto \rho(d - D(x)).$$

However, simply beginning with an arbitrary starting point $x^{(0)} \in \mathcal{B}$ does not work in general because $x^{(0)}$ is most likely not in $\operatorname{dom} \rho(d - D(\cdot))$. The problem here is that the barrier function provides no derivative information useful for finding a feasible point. The solution that we propose is to modify the original problem by replacing the right-hand-side of the coupling term d with the parameter $\theta \in \mathbb{R}^J$ ($\theta \geq d$). Therefore, the **shifted barrier problem** with which we are concerned is

$$\mathcal{P}(\tau, \theta) \quad := \quad \left\{ \underset{x}{\text{minimize}} \;\; f_\tau(x, \theta) := c(x) + \tau\rho(\theta - D(x)) \quad \text{subject to} \quad x \in \mathcal{B} \right\}, \tag{2.2}$$

i.e., $\mathcal{P}(\tau, \theta)$ represents the optimization problem with parameters $\tau \in \mathbb{R}_{>0}$ and $\theta \in \mathbb{R}^J$. The function $\rho(\theta - D(\cdot))$ is a barrier function designed to model the constraints $D(x) < \theta$. Thus, for $\theta = d$, we have $\operatorname{dom}\rho(\theta - D(\cdot)) = \operatorname{int}\mathcal{C}$. Allowing $\theta \neq d$ has the property of "shifting" the barrier, hence the name.

Once an initial point $x^{(0)} \in \mathcal{B}$ has been found by solving the relaxed problem (2.1), the parameters $\theta^{(1)}$ and $\tau^{(1)}$ are chosen so that $\tau^{(1)} > 0$, $\theta^{(1)} > D(x^{(0)})$ and $\theta^{(1)} \geq d$. This will have the effect of making $x^{(0)} \in \mathcal{B}$ an interior point of the domain of $c(\cdot) + \tau^{(1)}\rho\left(\theta^{(1)} - D(\cdot)\right)$ and the region $\{x | D(x) \leq \theta\}$ a relaxation of $\{x | D(x) \leq d\}$. We then compute $x^{(1)}$ by approximately minimizing the barrier problem $\mathcal{P}\left(\tau^{(1)}, \theta^{(1)}\right)$. In general, if $D(x^{(i)}) < d$, we have produced a feasible point and we may set $\theta^{(i+1)} = d$. If not, then we choose $\theta^{(i+1)}$ as described below while maintaining $\tau^{(i+1)} = \tau^{(i)}$, then set $i \leftarrow i + 1$ and do the process again. We will prove that if a point $x \in \mathcal{B} \cap \operatorname{int}\mathcal{C}$ exists, then *such a point is generated in a finite number of iterations*, under appropriate assumptions given below.

Suppose first that successive values of $\theta^{(i)}$ are chosen so that

$$D\left(x^{(i)}\right) < \theta^{(i+1)} \leq \theta^{(i)} \qquad \text{and} \qquad \theta^{(i+1)} \geq d. \tag{2.3}$$

This implies that $\theta^{(\infty)} := \lim_{i \to \infty} \theta^{(i)}$ is well defined. We make one more assumption on our choice of the $\theta^{(i)}$:

$$\text{Either} \quad \theta^{(\infty)} = d \quad \text{or} \quad \exists \{j\} \text{ such that } \liminf_{i} \left(\theta_j^{(i)} - D_{j\bullet}\left(x^{(i)}\right)\right) = 0. \tag{2.4}$$

2.1.1 Properties of Shifted Barriers

We first explore the behavior of barrier functions near the barrier.

Theorem 2.1 *Consider a single component j of the coupling constraints. Suppose the sequence $\left\{\xi^{(i)}\right\} \subset \mathbb{R}_{>0}$ and that ρ_j is L times continuously differentiable ($L \geq 0$). Then*

$$\left\{\xi^{(i)}\right\} \text{ is bounded away from } 0 \iff \left\{\rho_j\left(\xi^{(i)}\right)\right\} \text{ is bounded above.}$$

Also, if $\inf_i \left\{\xi^{(i)}\right\} = 0$, *then*

$$\sup_i \partial^\ell \rho_j \left(\left\{\xi^{(i)}\right\}\right) = (-1)^\ell \infty \qquad \text{for } \ell = 0, \ldots, L.$$

Moreover, if $\left\{\xi^{(i)}\right\}$ *is bounded, then for each* $\ell = 0, \ldots, L$

$$\left\{\xi^{(i)}\right\} \text{ is bounded away from } 0 \iff \left\{\partial^\ell \rho_j\left(\xi^{(i)}\right)\right\} \text{ is bounded.}$$

Proof: Define

$$\xi^{\text{inf}} := \inf_i \xi^{(i)} \qquad \text{and} \qquad \xi^{\text{sup}} := \sup_i \xi^{(i)}.$$

If $\xi^{\text{inf}} > 0$, then properties 3 and 1 of definition 1.6 imply that

$$\sup_i \rho_j\left(\xi^{(i)}\right) = \rho_j\left(\xi^{\text{inf}}\right) < +\infty.$$

Conversely, if $\left\{\rho_j\left(\xi^{(i)}\right)\right\}$ is bounded from above, then properties 2 and 1 of definition 1.6 show that $\xi^{\text{inf}} > 0$.

Suppose that, for some integer ℓ between 1 and L,

$$\limsup_i (-1)^\ell \partial^\ell \rho_j\left(\xi^{(i)}\right) < \infty. \tag{2.5}$$

Using the fundamental theorem of calculus:

$$\partial^{\ell-1}\rho_j\left(\xi^{(i)}\right) = \int_1^{\xi^{(i)}} \partial^\ell \rho_j(\zeta)\,\partial\zeta + \kappa = -\int_{\xi^{(i)}}^1 \partial^\ell \rho_j(\zeta)\,\partial\zeta + \kappa,$$

for some constant κ. Proceeding by induction, we see that (2.5) implies (2.5) with ℓ replaced by $\ell - 1$. However, the first part of the theorem shows that if $\xi^{\text{inf}} = 0$, then

$$\limsup_i (-1)^0 \partial^0 \rho_j \left(\xi^{(i)}\right) = \infty,$$

contradicting the assumption (2.5) for $\ell = 0$, and hence for any $\ell = 0, \ldots, L$.

Now suppose that $\left\{\xi^{(i)}\right\}$ is bounded. We already know that $\xi^{\text{inf}} = 0$ implies that $\partial^\ell \rho_j \left(\xi^{(i)}\right)$ is unbounded, so we consider the case where $\xi^{\text{inf}} > 0$. Then, for all $\ell = 0, \ldots, L$, $\partial^\ell \rho_j$ is continuous on the compact interval $\left[\xi^{\text{inf}}..\xi^{\text{sup}}\right]$. Therefore, the image $\partial^\ell \rho_j \left[\xi^{\text{inf}}..\xi^{\text{sup}}\right]$ is compact, hence bounded.

♠

This componentwise result translates directly into a statement about the full barrier function.

Corollary 2.2 (Divergence at Barriers) *Suppose that the sequences* $\left\{x^{(i)}\right\} \subset \mathrm{IR}^N$ *and* $\left\{\theta^{(i)}\right\} \subset \mathrm{IR}^J$ *satisfy* $D\left(x^{(i)}\right) < \theta_j^{(i)}$ *for each* $i = 0, 1, \ldots,$ *and that* ρ *is* L *times continuously differentiable. Then*

$$\textit{each component of } \left\{\theta^{(i)} - D\left(x^{(i)}\right)\right\} \textit{ is bounded away from } 0$$
$$\Updownarrow$$
$$\left\{\rho\left(\theta^{(i)} - D\left(x^{(i)}\right)\right)\right\} \textit{ is bounded from above.}$$

Moreover, if some component of $\left\{\theta^{(i)} - D\left(x^{(i)}\right)\right\}$ *is not bounded away from* 0, *then*

$$\left\{\left\|\nabla^\ell \rho\left(\theta^{(i)} - D\left(x^{(i)}\right)\right)\right\|\right\} \textit{ is unbounded} \qquad \textit{for } \ell = 0, \ldots, L. \tag{2.6}$$

The norm in the preceding corollary depends on the dimension ℓ of the tensor. These may be built up as one builds a matrix norm from a vector norm, from any norm that yields the standard topology in finite-dimensional Euclidean space.

Proof: The proof follows easily from the componentwise statement of the theorem.

♠

Corollary 2.3 (Boundedness Away from Barriers) *Suppose the hypotheses of corollary 2.2. If the sequence* $\left\{\theta^{(i)} - D\left(x^{(i)}\right)\right\}$ *is bounded, then the converse of corollary 2.2 holds, namely (2.6) implies that each component of* $\left\{\theta^{(i)} - D\left(x^{(i)}\right)\right\}$ *is bounded away from* 0.

Proof: Again, the proof follows from the componentwise theorem.

♠

We are now motivated to study properties of the composite barrier functions ϕ and ψ defined by

$$x \overset{\phi}{\mapsto} \rho(\theta - D(x)) \qquad \text{and} \qquad \theta \overset{\psi}{\mapsto} \rho(\theta - D(x)). \tag{2.7}$$

To do this, we review the following concepts. A function ϕ is said to be **essentially smooth** [Rockafellar 1970, §26] if $\operatorname{dom}\phi \neq \emptyset$, ϕ is differentiable on $\operatorname{dom}\phi$, and $\lim\left\|\nabla\phi(x^{(i)})\right\| = +\infty$ for all sequences $\{x^{(i)}\} \subset \operatorname{dom}\phi$ converging to a point $\bar{x} \in \operatorname{bdy}\operatorname{dom}\phi$.

Theorem 2.4 *Let* ϕ *and* ψ *be given by (2.7) where* ρ *is a smooth barrier function in the sense of definition 1.6. Then* ϕ *is an essentially smooth function with* $\operatorname{dom}\phi = \{x|D(x) < \theta\}$, *and* ψ *is an essentially smooth, antitone function with* $\operatorname{dom}\psi = \{\theta|D(x) < \theta\}$.

Proof: The statement about the domains of the functions clearly hold. ψ is smooth on $\operatorname{dom}\psi$ and, since D is smooth, ϕ is smooth on $\operatorname{dom}\phi$. Corollary 2.2 then shows that both ϕ and ψ are essentially smooth. Property 3 of definition 1.6 shows that ψ is antitone.

♠

Theorem 2.5 *Suppose the hypotheses of theorem 2.4 hold. If* ρ *is (strictly) convex, then so is* ψ. *If both* D *and* ρ *are (strictly) convex, then so is* ϕ.

Proof: The result for ψ is trivial. The result for ϕ depends on property 3 of definition 1.6 and may be found in Fiacco and McCormick [1968, lemma 11] or Rockafellar [1970, theorem 5.1].

♠

We shall denote the restriction of a mapping ϕ to a set S by $\phi|_S$ and if S is a subspace of T, the quotient space of S in T is denoted by T/S. Where y is a point and S is a set, we use $y+S$ to denote $\{y+s|s \in S\}$.

Theorem 2.6 *Suppose the hypotheses of theorem 2.4 hold. Also suppose that ρ is (strictly) convex and D is a linear map. Let $y \in \mathbb{R}^N$ be fixed and let $S := y + \ker D$ and $S^{\div} := y + (\mathbb{R}^N / \ker D) = y + \operatorname{range} D^{\mathsf{T}}$, so that S and $S^{\div}$ are translated subspaces. Then $\phi|_S$ is constant and $\phi|_{S^{\div}}$ is (strictly) convex.*

Proof: By the definition of kernel, any points in S are mapped to the same value, proving that $\phi|_S$ is constant. For $0 < \lambda < 1$,

$$\begin{aligned} \rho\left(\theta - D\left[(1-\lambda)x + \lambda y\right]\right) &= \rho\left((1-\lambda)\left[\theta - Dx\right] + \lambda\left[\theta - Dy\right]\right) \\ &\leq (1-\lambda)\rho\left(\theta - Dx\right) + \lambda\rho\left(\theta - Dy\right). \end{aligned}$$

If ρ is *strictly* convex, then the inequality is strict in the event that $Dx \neq Dy$. But the definition of $S^{\div}$ guarantees that no two points $x, y \in S^{\div}$ are mapped to the same point by D.

♠

We now write out formulae for the first two gradients of ϕ and ψ. Suppose D and ρ are continuously differentiable and $D(x) < \theta$. Then

$$\begin{aligned} \nabla\psi(\theta) &= \nabla_\theta \rho(\theta - D(x)) = \nabla\rho(\theta - D(x)) \text{ and} && (2.8) \\ \nabla\phi(x) &= \nabla_x \rho(\theta - D(x)) = \nabla\rho(\theta - D(x))\nabla D(x). && (2.9) \end{aligned}$$

If also D and ρ are twice continuously differentiable, then the Hessian matrices are given by

$$\begin{aligned} \nabla^2\psi(\theta) &= \nabla_{\theta\theta}\rho(\theta - D(x)) = \nabla^2\rho(\theta - D(x)) \text{ and} && (2.10) \\ \nabla^2\phi(x) &= \nabla_{xx}\rho(\theta - D(x)) \\ &= (\nabla D(x))^{\mathsf{T}}\nabla^2\rho(\theta - D(x))\nabla D(x) && (2.11) \\ &\quad + \nabla\rho(\theta - D(x)) * \nabla^2 D(x), \end{aligned}$$

where the tensor product "$*$" multiplies the $1 \times J$ vector by the $J \times N \times N$ tensor forming an $N \times N$ matrix. Therefore, (2.11) is written componentwise as

$$\frac{\partial^2 \phi}{\partial x_n \partial x_\ell}(x) = \sum_{j=1}^{J} \left[\rho_j''(\theta - D(x)) \frac{\partial D_{j\bullet}}{\partial x_n}(x) \frac{\partial D_{j\bullet}}{\partial x_\ell}(x) + \rho_j'(\theta - D(x)) \frac{\partial^2 D_{j\bullet}}{\partial x_n \partial x_\ell}(x) \right].$$

2.1.2 Computing a Feasible Point

Theorem 2.7 (Finite Feasibility) *Suppose $c(\cdot)$ is bounded from below on $\mathcal{B}$. Let $\theta^{(i+1)}$ be chosen to satisfy (2.3) and (2.4). If $x^{(i)} \in \mathcal{B}$ be computed so that $\sup_i f_\tau\left(x^{(i)}, \theta^{(i)}\right)$ is finite, then a point $z \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$ exists and will be found in a finite number of steps. On the other hand, if no such z exists, then $f_\tau\left(x^{(i)}, \theta^{(i)}\right) \to +\infty$.*

Proof: By corollary 2.2 and because $c\left(x^{(i)}\right)$ is bounded from below, there is some γ such that $\forall\{i,j\}\ \theta_j^{(i)} - D_{j\bullet}\left(x^{(i)}\right) \geq \gamma > 0$. Therefore, for some finite $\hat{\imath}$, $\forall\{j\}\ d_j - D_{j\bullet}\left(x^{(\hat{\imath})}\right) \geq \gamma/2 > 0$, and a point with $D\left(x^{(\hat{\imath})}\right) < d$ has been found in a finite number of iterations. If no such z exists, then (2.3) and (2.4) imply that $\theta_j^{(i)} - D_{j\bullet}\left(x^{(i)}\right) \to 0$ for at least one $j \in \{1, \ldots, J\}$. Corollary 2.2 then shows that $f_\tau\left(x^{(i)}, \theta^{(i)}\right) \to +\infty$.

♠

The above theorem does not give us a *computationally feasible* test for assuring that a point $z \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$ will be found, in the event that such a z exists. Under the assumption of convexity, however, such a test is possible:

Corollary 2.8 (Computable Finite Feasibility) *Suppose the barrier problem $\mathcal{P}(\tau, \theta)$ is* convex *for any $\tau > 0$ and θ. Let the optimal value function be defined as*

$$f_\tau^*(\theta) := \min_{x \in \mathcal{B}} f_\tau(x, \theta).$$

Further suppose that there exists some $z \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$. If the $x^{(i)}$ are chosen so that

$$f_\tau\left(x^{(i)}, \theta^{(i)}\right) \leq f_\tau^*\left(\theta^{(i)}\right) + \beta \tag{2.12}$$

for a constant $\beta > 0$, such a z will be computed in a finite number of steps. Moreover, we may actually compute each $x^{(i)}$ in a finite number of steps.

Proof: Assume such a z exists. Then

$$\forall\{i\} \;\; f_\tau^*\left(\theta^{(i)}\right) \leq f_\tau\left(z, \theta^{(i)}\right) \leq f_\tau\left(z, d\right) < +\infty$$

where the second inequality follows from antitonicity in θ (theorem 2.4). Therefore, $\sup_i f_\tau\left(x^{(i)}, \theta^{(i)}\right)$ is finite.

The final remark is clearly true, for any convergent primal-dual sequence will have its duality gap to to zero. We may consider the convex program solved whenever this gap becomes less than or equal to β.

♠

Assume from now on that the $x^{(i)}$ are computed so that $D\left(x^{(i)}\right) < \theta^{(i)}$. We will now give a particular method for computing θ. This is a very simple, intuitive scheme designed to satisfy the abstract conditions (2.3) and (2.4). After computing $x^{(0)} \in \mathcal{B}$, set

$$\theta_j^{(1)} \leftarrow \begin{cases} d_j & \text{if } D_{j\bullet}\left(x^{(0)}\right) < d_j \\ D_{j\bullet}\left(x^{(0)}\right) + \Theta & \text{if } D_{j\bullet}\left(x^{(0)}\right) \geq d_j \end{cases} \tag{2.13}$$

where $\Theta > 0$ is a constant. In general, after computing $x^{(i)}$, set

$$\theta_j^{(i+1)} \leftarrow \begin{cases} d_j & \text{if } D_{j\bullet}\left(x^{(i)}\right) < d_j \\ \lambda_\theta D_{j\bullet}\left(x^{(i)}\right) + (1-\lambda_\theta)\theta_j^{(i)} & \text{if } D_{j\bullet}\left(x^{(i)}\right) \geq d_j \end{cases} \tag{2.14}$$

where $\lambda_\theta \in (0, 1)$ is a constant.

Theorem 2.9 *Suppose the $x^{(i)}$ are computed so that $D\left(x^{(i)}\right) < \theta^{(i)}$. If the θ are computed by the rules (2.13) and (2.14), then (2.3) and (2.4) are satisfied.*

Proof: We show the result for each component j. If $D_{j\bullet}\left(x^{(i)}\right) \geq d_j$, then $\theta_j^{(i+1)} = \lambda_\theta D_{j\bullet}\left(x^{(i)}\right) + (1-\lambda_\theta)\theta_j^{(i)}$. Condition (2.3) follows from this since

$D_{j\bullet}\left(x^{(i)}\right) < \theta_j^{(i)}$ and $\lambda_\theta \in (0,1)$. If $D_{j\bullet}\left(x^{(i)}\right) < d_j$, then $\theta_j^{(\hat{\imath})} = d_j$ for all $\hat{\imath} > i$ and again condition (2.3) follows. If, for some finite i, $D_{j\bullet}\left(x^{(i)}\right) < d_j$, then condition (2.4) follows. In the other case, $(\forall\{i\}\ D_{j\bullet}\left(x^{(i)}\right) \geq d_j)$ the limit

$$\theta_j^{(\infty)} := \lim_{i\to\infty} \theta_j^{(i)}$$

is well defined since the sequence is monotonic and bounded. Then (2.14) shows that for any $s_j^\infty \in \lim \mathrm{pt}\{D_{j\bullet}\left(x^{(i)}\right)\}$ we have

$$\theta_j^{(\infty)} = \lambda_\theta s_j^\infty + (1-\lambda_\theta)\theta_j^{(\infty)}.$$

Therefore $s_j^\infty = \theta_j^{(\infty)}$, which shows that (2.4) holds.

♠

2.2 Convergence of Barrier Minimizers

In this section we assume that $\theta = d$ and that we have found some point in $\mathcal{B} \cap \operatorname{int} \mathcal{C}$. This means that the barriers are at fixed locations. Since θ is constant, we shall drop the dependence of f_τ on θ and write $f_\tau(x)$ for the objective function of $\mathcal{P}(\tau, d)$ in (2.2). We will study the convergence of (local) minimizers of the barrier problem $\mathcal{P}(\tau, d)$ (2.2) as $\tau \downarrow 0$. These results will mimic the results of §1.7.

2.2.1 Convergence Without Convexity

If c and D are not convex, the optimization problem (1.8) may have local optima that are not global. In general, we are concerned only with local optimality. In §2.2.2 we shall add the assumption of convexity.

Theorem 2.10 (Convergence of Barrier Function Minimizers) *Suppose that*

1. *c and D are continuous,*

2. *$\mathcal{B}$ is closed, $\mathcal{C} = \{x|D(x) \leq d\}$ and $\mathcal{B} \cap \operatorname{int} C$ is nonempty,*

3. *the set of points X^* that are local minima of (1.8) corresponding to the value $v^* = c(X^*)$ is a nonempty compact set,*

4. *X^* is isolated in*

$$Y = \left\{ \; y \in \mathcal{B} \;\middle|\; D(y) \leq d \quad \textit{and} \quad c(y) = v^* \; \right\},$$

5. *at least one point in X^* is in the closure of $\operatorname{int} \mathcal{C}$, and*

6. *$\tau^{(i)} \downarrow 0$ monotonically.*

Then

1. *there exists a compact set S such that $X^* \subset \operatorname{int} S$ and for $\tau^{(i)}$ small enough, the minima of $\mathcal{P}\left(\tau^{(i)}, d\right)$ in $\mathcal{B} \cap \operatorname{int} \mathcal{C} \cap \operatorname{int} S$ exist and every limit point of any subsequence $\left\{x^{(i)}\right\}$ of such minimizers is in X^*,*

2. $\lim_{i \to \infty} c\left(x^{(i)}\right) = v^*$,

3. $\lim_{i \to \infty} \tau^{(i)} \rho\left(d - D\left(x^{(i)}\right)\right) = 0$,

4. $\lim_{i \to \infty} f_{\tau^{(i)}}\left(x^{(i)}\right) = v^*$,

5. *$\left\{c\left(x^{(i)}\right)\right\}$ is a monotonically decreasing sequence, and*

6. *$\left\{\rho\left(d - D\left(x^{(i)}\right)\right)\right\}$ is a monotonically increasing sequence.*

Proof: This proof is a modification of Fiacco and McCormick's [1968] proof of theorem 1.8 above.

By theorem 1.7 there is a set S such that $X^* \subset \operatorname{int} S$ and $c(y) > v^*$ for all $y \in (\mathcal{B} \cap \operatorname{int} \mathcal{C} \cap S) \setminus X^*$. Define $x^{(i)} \in \mathcal{B} \cap \operatorname{int} \mathcal{C} \cap S$ so that

$$f_{\tau^{(i)}}\left(x^{(i)}\right) = \min\{f_{\tau^{(i)}}(x) \,|x \in \mathcal{B} \cap \operatorname{int} \mathcal{C} \cap S\},$$

i.e., $x^{(i)}$ is a local (localized by S) minimizer for $\mathcal{P}\left(\tau^{(i)}, d\right)$. There is the technical point here that such a minimizer exists. This is easily taken care of by noting that S is compact and $f_{\tau^{(i)}}(x) \xrightarrow[x\to \mathrm{bdy}\,\mathcal{C}]{} +\infty$ for $x \in \mathrm{int}\,\mathcal{C}$ [Fiacco and McCormick 1968, corollary 8]. Let y be one of the limit points of the uniformly bounded sequence $\left\{x^{(i)}\right\}$, so that $y \in \mathcal{B} \cap \mathcal{C} \cap S$. We will now show that $y \in X^*$ by assuming that $y \notin X^*$ and deriving a contradiction. By the basic property of S, $y \notin X^*$ implies that $c(y) > v^*$. By assumption 5 there is a $z \in \mathcal{B} \cap \mathrm{int}\,\mathcal{C} \cap S$ where $v^* < c(z) < c(y)$. But then

$$\liminf_{i\to\infty} f_{\tau^{(i)}}\left(x^{(i)}\right) \geq c(y) > c(z) = \lim_{i\to\infty} f_{\tau^{(i)}}(z).$$

This contradicts the assumption that $x^{(i)}$ solves $\mathcal{P}\left(\tau^{(i)}, d\right)$ locally for large i. Therefore $y \in X^* \subset \mathrm{int}\, S$ and, for i large enough, $x^{(i)} \in \mathcal{B} \cap \mathrm{int}\,\mathcal{C} \cap \mathrm{int}\, S$. Therefore conclusion 1 is shown. Then conclusions 2, 3 and 4 follow directly.

Let $c^{(i)}$ denote $c\left(x^{(i)}\right)$, and $\rho^{(i)}$ denote $\rho\left(d - D\left(x^{(i)}\right)\right)$. Then, because each $x^{(i)}$ is a *global* minimizer of $f_{\tau^{(i)}}(x)$ subject to the constraints $x \in \mathcal{B} \cap S$, we have

$$\begin{aligned} c^{(i)} + \tau^{(i)}\rho^{(i)} &\leq c^{(i+1)} + \tau^{(i)}\rho^{(i+1)} \qquad \text{and} \\ c^{(i+1)} + \tau^{(i+1)}\rho^{(i+1)} &\leq c^{(i)} + \tau^{(i+1)}\rho^{(i)}. \end{aligned} \tag{2.15}$$

Adding $\tau^{(i+1)}/\tau^{(i)}$ times the first inequality to the second causes all of the ρ terms to drop out, and leaves one with

$$\left[1 - \frac{\tau^{(i+1)}}{\tau^{(i)}}\right] c^{(i+1)} \leq \left[1 - \frac{\tau^{(i+1)}}{\tau^{(i)}}\right] c^{(i)}.$$

Assumption 6 then verifies condition 5, and condition 6 follows from condition 5 and (2.15).

♠

Theorem 2.11 (Convergence of Multipliers) *Suppose*

1. *the hypotheses of theorem 2.10 hold,*

2. *c and D are continuously differentiable functions,*

3. *ρ is a continuously differentiable barrier function,*

4. *$\left\{x^{(i)}\right\}$ is a bounded sequence of minimizers as guaranteed in the conclusion of theorem 2.10,*

5. *$q^{(i)}$ and $r^{(i)}$ are optimal Lagrange multipliers for $\mathcal{P}\left(\tau^{(i)}, d\right)$ associated with $x^{(i)}$, and*

$$p^{(i)} := \tau^{(i)} \nabla \rho \left(d - D\left(x^{(i)}\right)\right) \in \mathbb{R}^{1 \times J}. \tag{2.16}$$

Then

1. *the finite limit points of $\left\{\left(p^{(i)}, q^{(i)}, r^{(i)}\right)\right\}$ are optimal Lagrange multipliers of the original problem (1.8) (see (1.12)),*

2. *if the Mangasarian–Fromovitz constraint qualification (cf. 1.2) holds at the limit point x^*, then the corresponding subsequence of multipliers $\left\{\left(p^{(i)}, q^{(i)}, r^{(i)}\right)\right\}$ is bounded, and*

3. *if the weakened Mangasarian–Fromovitz constraint qualification (cf. 1.2) holds at the limit point x^*, then the corresponding subsequences of $\left\{p^{(i)}\right\}$ and $\left\{r^{(i)}\right\}$ are bounded.*

Proof: Since $x^{(i)}$ is feasible for $\mathcal{P}\left(\tau^{(i)}, d\right)$, it must satisfy (1.13). Antitonicity of ρ shows that $p^{(i)} \leq \mathbf{0}$. Since $q^{(i)}$ and $r^{(i)}$ are optimal Lagrange multipliers to $\mathcal{P}\left(\tau^{(i)}, d\right)$ we know that

$$\begin{aligned}
r^{(i)} &= \nabla f_{\tau^{(i)}}\left(x^{(i)}\right) - q^{(i)} A \\
&= \nabla c\left(x^{(i)}\right) - p^{(i)} \nabla D\left(x^{(i)}\right) - q^{(i)} A, \\
0 &= r_{+}^{(i)}\left(\underline{b} - x^{(i)}\right) \text{ and} \\
0 &= r_{-}^{(i)}\left(x^{(i)} - \overline{b}\right).
\end{aligned}$$

Therefore, conditions (1.14) and (1.16) are satisfied. If a finite limit point exists, we may thin the sequence so that it converges to this limit point. We are assuming that $\tau^{(i)} \downarrow 0$, so that theorem 2.1 shows

$$\lim_i p_j^{(i)} > 0 \implies \lim_i d_j - D_{j\bullet}\left(x^{(i)}\right) = 0$$

and

$$\lim_i p_j^{(i)} = 0 \quad \Longleftarrow \quad \lim_i d_j - D_{j\bullet}\left(x^{(i)}\right) > 0.$$

Therefore, any finite limit point of the multiplier sequence is an optimal multiplier tuple for the original problem. Conclusions 2 and 3 follow directly from theorem 1.2 (Gauvin's theorem).

♠

2.2.2 Convergence For Convex Programs

We now assume that c and D are convex and try to sharpen the results of §2.2.1. The first result is that the level sets of the barrier problem are bounded for any barrier function that becomes ever more "flat" far from the barrier. The second result is that the Lagrange multipliers are automatically bounded.

As pointed out by the example on page 22, the classical assumptions made on barrier functions to guarantee boundedness of level sets is too strong. We will see below that the weakest property that can be used, in some sense, is the property to be defined below. We call a barrier function ρ **eventually flat** if all component functions ρ_j satisfy

$$\lim_{\xi \to \infty} \left[\rho_j(\xi) - \rho_j(\xi + 1)\right] = 0. \tag{2.17}$$

(Note that the antitonicity of ρ implies that the sequence is non-negative.) If $S \subset \mathbb{R}^J$, we say that ρ **is eventually flat on** S if (2.17) holds for each j for which $s_j^{(i)} \to \infty$ for some sequence $\left\{s^{(i)}\right\} \subset S \cap \mathbb{R}^J_{>0}$. Without loss of generality, we may use $S \cap \mathbb{R}^J_{>0}$ in place of S; and we will use either convention in what follows. We make the following observations which follow directly from the definition:

- ρ is eventually flat iff ρ is eventually flat on $\mathbb{R}^J$
- *any* ρ is eventually flat on a bounded set
- if ρ is eventually flat on $S \subset \mathbb{R}^J$, then ρ is eventually flat on the smallest Cartesian product $S_1 \times \cdots \times S_J \subset \mathbb{R}^J$ containing S

Eventually flat functions may diverge to $-\infty$, e.g., $\rho_j = -\ln$. However, the following theorem guarantees that they must do so at a *sublinear* rate. This result will be used subsequently to bound solutions to $\mathcal{P}(\tau, \theta)$ for all $\tau > 0$.

Theorem 2.12 *If ρ is eventually flat and antitone on $S \subset \mathbb{R}^J_{>0}$ and $\left\{s^{(i)}\right\} \subset S$ satisfies $\left\|s^{(i)}\right\| \to \infty$, then*

$$\forall \{i\} \;\; \rho\left(s^{(i)}\right) \leq \alpha \left\|s^{(i)}\right\| + \beta$$

implies that $\alpha \geq 0$.

Proof: By the equivalence of norms on $\mathbb{R}^J$ [Ortega and Rheinboldt 1970, theorem 2.2.1], we may choose the *uniform norm* $\|\cdot\|_\infty$. We define $\mathcal{J}^0$ to be the set of j for which $\left\{s_j^{(i)}\right\}$ is bounded and $\mathcal{J}^\infty$ to be the j for which $\left\{s_j^{(i)}\right\}$ is unbounded.

$$\begin{aligned}
\forall \{i\} \;\; \alpha \left\|s^{(i)}\right\|_\infty + \beta - \rho(\mathbf{1}) &\geq \rho\left(s^{(i)}\right) - \rho(\mathbf{1}) = \sum_{j=1}^{J} \rho_j\left(s_j^{(i)}\right) - \rho_j(1) \\
&\geq \sum_{j=1}^{J} \rho_j\left(\left\lfloor s_j^{(i)} \right\rfloor + 1\right) - \rho_j(1) \\
&= \sum_{j=1}^{J} \sum_{\ell=1}^{\left\lfloor s_j^{(i)} \right\rfloor} \rho_j(\ell+1) - \rho_j(\ell) .
\end{aligned}$$

For brevity, we define

$$\xi_j^{(\ell)} := \rho_j(\ell) - \rho_j(\ell+1)$$

and note that eventual flatness guarantees that $\lim_{\ell\to\infty} \xi_j^{(\ell)} = 0$ for any $j \in \mathcal{J}^\infty$. Without loss of generality $s^{(i)} \subset \mathbb{R}^J_{>0}$ and $\left\|s^{(i)}\right\|_\infty \to \infty$, so that we may divide by $\left\|s^{(i)}\right\|_\infty$, yielding

$$\forall \{i\} \;\; \alpha \;\; \geq \;\; \frac{1}{\left\|s^{(i)}\right\|_\infty} \left[-\sum_{j\in\mathcal{J}^\infty} \sum_{\ell=1}^{\left\lfloor s_j^{(i)} \right\rfloor} \xi_j^{(\ell)} \underbrace{- \sum_{j\in\mathcal{J}^0} \sum_{\ell=1}^{\left\lfloor s_j^{(i)} \right\rfloor} \xi_j^{(\ell)} + \rho(\mathbf{1}) - \beta}_{\text{bounded}} \right] .$$

Therefore, to show that $\alpha \geq 0$ it suffices to show that

$$\forall \{j \in \mathcal{J}^\infty\} \quad \liminf_{i\to\infty} \sum_{\ell=1}^{\lfloor s_j^{(i)} \rfloor} \frac{\xi_j^{(\ell)}}{\left\| s^{(i)} \right\|_\infty} = 0. \tag{2.18}$$

Antitonicity of ρ shows that, $\forall\{\ell, j\}$ $\xi_j^{(\ell)} \geq 0$, and so

$$0 \leq \sum_{\ell=1}^{\lfloor s_j^{(i)} \rfloor} \frac{\xi_j^{(\ell)}}{\left\| s^{(i)} \right\|_\infty} \leq \sum_{\ell=1}^{\lfloor s_j^{(i)} \rfloor} \frac{\xi_j^{(\ell)}}{\left\lfloor s_j^{(i)} \right\rfloor}.$$

For any $\varepsilon > 0$, pick L so that

$$\forall\{\ell > L\} \ \xi_j^{(\ell)} < \frac{\varepsilon}{2},$$

and

$$0 \leq \sum_{\ell=L+1}^{\lfloor s_j^{(i)} \rfloor} \frac{\xi_j^{(\ell)}}{\left\lfloor s_j^{(i)} \right\rfloor} < \frac{\varepsilon\left(\left\lfloor s_j^{(i)} \right\rfloor - L\right)}{2\left\lfloor s_j^{(i)} \right\rfloor} = \frac{\varepsilon}{2} - \frac{\varepsilon L}{2\left\lfloor s_j^{(i)} \right\rfloor} < \frac{\varepsilon}{2}.$$

Now, by the definition of $\mathcal{J}^\infty$, we may pick i sufficiently large so that

$$0 \leq \sum_{\ell=1}^{L} \frac{\xi_j^{(\ell)}}{\left\lfloor s_j^{(i)} \right\rfloor} < \frac{\varepsilon}{2}.$$

Then the ith term of the sequence of partial sums in (2.18) is non-negative and $< \varepsilon$.

♠

The following theorem is used to guarantee that minimizers of $\mathcal{P}(\tau, d)$ exist (assuming boundedness of X^*) no matter what τ is. This is in contrast to the case without convexity where the τ must be sufficiently small for minimizers to exist. The result is the same as lemma 12 of Fiacco and McCormick [1968]. As was pointed out earlier, the assumptions made by Fiacco and McCormick are too strong because they exclude the log barrier function.

Theorem 2.13 (Boundedness of f_τ Contours) *Suppose c and D are continuous and convex, the solution set for (1.8) (denoted as X^*) is bounded (hence compact) and ρ is a barrier function that is eventually flat on the image set $d - D(\mathcal{B})$. Then, for any $\tau > 0$ and any $z \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$,*

$$L_\tau(z) = \left\{ \; x \in \mathcal{B} \;\middle|\; f_\tau(x) \leq f_\tau(z) \; \right\}$$

is a compact subset of $\mathcal{B}$. Moreover, if ρ is also convex, then so is $L_\tau(z)$.

Proof: $\tau > 0$ and any $z \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$ imply that $f_\tau(z) < \infty$. Closedness follows from continuity of the problem functions and convexity follows from convexity of $\mathcal{B}$ and lemma 1.9. Therefore, it suffices to show that $L = L_\tau(z)$ is bounded. To do this, take any sequence $\left\{y^{(i)}\right\} \subset L \subset \mathcal{B}$ and suppose that $\limsup_{i\to\infty} \left\|y^{(i)}\right\| = \infty$ and we shall derive, at length, a contradiction.

We first prove that:

$$\exists\{\alpha_c > 0 \text{ and } \beta_c\}\forall\{i\} \;\; c\left(y^{(i)}\right) \geq \alpha_c \left\|y^{(i)}\right\| + \beta_c. \tag{2.19}$$

By theorem 1.7, there exists some compact set $S \in \mathbb{R}^N$ such that $X^* \subset \operatorname{int} S$ and for each $x \in (\mathcal{B} \cap \mathcal{C} \cap S) \setminus X^*$, $c(x) > v^* := c(X^*)$. We now let

$$\begin{aligned} u^* &:= \inf\left\{ \; c(x) \;\middle|\; x \in \operatorname{bdy} S \; \right\} \text{ and} \\ w^* &:= \sup\left\{ \; \|z - x\| \;\middle|\; z \in \operatorname{bdy} S \text{ and } x \in X^* \; \right\}. \end{aligned}$$

Clearly, $u^* > v^*$, $w^* > 0$ and both are finite. Let $z^{(i)} := y^{(i)} - x$ for some $x \in X^*$, and let $\lambda^{(i)} \geq 0$ be defined so that $x + \lambda^{(i)} z^{(i)} \in \operatorname{bdy} S$ so that $\inf_i \lambda^{(i)} > 0$. Since $x + \lambda^{(i)} z^{(i)} \in \operatorname{bdy} S$, we have

$$\forall\{i\} \;\; w^* \geq \left\|x + \lambda^{(i)} z^{(i)} - x\right\| = \lambda^{(i)} \left\|z^{(i)}\right\| \geq \lambda^{(i)} \left[\left\|y^{(i)}\right\| - \|x\|\right].$$

Note that, since $\left\|z^{(i)}\right\| \to \infty$, it suffices to consider indices i for which $\lambda^{(i)} \leq 1$. Convexity of c now shows that

$$c\left(x + \lambda^{(i)} z^{(i)}\right) \leq \lambda^{(i)} c\left(x + z^{(i)}\right) + \left(1 - \lambda^{(i)}\right) c(x),$$

from which we have

$$\begin{aligned} c\left(y^{(i)}\right) = c\left(x + z^{(i)}\right) &\geq \frac{c\left(x + \lambda^{(i)} z^{(i)}\right) - c(x)}{\lambda^{(i)}} + c(x) \geq \frac{u^* - v^*}{\lambda^{(i)}} + v^* \\ &\geq \frac{u^* - v^*}{w^*}\left\|y^{(i)}\right\| - \frac{u^* - v^*}{w^*}\|x\| + v^*, \end{aligned}$$

finishing the proof of the claim (2.19).

Now suppose that $\left\{f_\tau\left(y^{(i)}\right)\right\}$ remains bounded from above while $y^{(i)} \rightarrow \infty$. Then

$$f_\tau\left(y^{(i)}\right) = c\left(y^{(i)}\right) + \tau\rho\left(d - D\left(y^{(i)}\right)\right) \geq \alpha_c \left\|y^{(i)}\right\| + \beta_c + \tau\rho\left(d - D\left(y^{(i)}\right)\right)$$

implies that

$$\exists\{\alpha_\rho < 0 \text{ and } \beta_\rho\}\forall\{i\} \;\; \rho\left(d - D\left(y^{(i)}\right)\right) \leq \alpha_\rho \left\|y^{(i)}\right\| + \beta_\rho.$$

Let $s^{(i)} := d - D\left(y^{(i)}\right) \subset d - D(L) \subset d - D(\mathcal{B})$ and let x be any point. Then, since D is convex, a finite subgradient $G \in \mathrm{I\!R}^{J \times N}$ exists which depends only on x and D:

$$D\left(y^{(i)}\right) \geq D(x) + G\left(y^{(i)} - x\right).$$

Therefore,

$$\left\|s^{(i)}\right\| = \left\|d - D\left(y^{(i)}\right)\right\| \leq \|d - D(x)\| + \|Gx\| + \|G\|\left\|y^{(i)}\right\|,$$

and, since $\alpha_\rho < 0$,

$$\exists\{\alpha'_\rho < 0 \text{ and } \beta'_\rho\}\forall\{i\} \;\; \rho\left(s^{(i)}\right) \leq \alpha'_\rho \left\|s^{(i)}\right\| + \beta'_\rho.$$

But theorem 2.12 shows that $\alpha'_\rho \geq 0$, a contradiction.

♠

The assumption of eventual flatness is, in fact, the weakest assumption possible, as the following theorem shows.

Theorem 2.14 *Suppose ρ is a barrier function that is* not *eventually flat. Then there exists some $z \in \mathrm{I\!R}^N$, $\tau > 0$ and some problem of the form (1.8) for which*

- *c and D are continuous and convex,*
- *the optimal solution set X^* of (1.8) is compact, and*
- *$z \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$,*

for which $L_\tau(z)$ is unbounded.

Proof: We may choose a sequence $\left\{s_1^{(i)}\right\}$ such that

$$\lim_{i \to \infty} \rho_1\left(s_1^{(i)}\right) - \rho_1\left(s_1^{(i)}+1\right) = \sigma > 0.$$

Without loss of generality, suppose

$$\forall\{i\}\ \ 0 \leq s_1^{(i+1)} - s_1^{(i)} \leq s_1^{(i+2)} - s_1^{(i+1)}.$$

Then the piecewise linear map from $\mathbb{R}_+$ to $\mathbb{R}$ defined from $i \mapsto -s_1^{(i)}$ for all i and $0 \mapsto 0$ is uniquely determined, continuous, convex and has range $\mathbb{R}_-$. This map is our choice for $D_{1\bullet}$. For all $j \neq 1$, set $\forall\{i\}\ \ s_j^{(i)} = 1$. Let $-D_{j\bullet}$ be the jth row of the identity map for $j \neq 1$ and $d = \mathbf{0}$ so that $J = N$, $\mathcal{C} = \mathbb{R}_+^J$ and $x^{(i)}$ is given by $x_j^{(i)} = s_j^{(i)} = 1$ for all $j \neq 1$ and $x_1^{(i)} = i$. We let $\mathcal{B} = \mathbb{R}^N$ and $c(\cdot) = \mathbf{1}\cdot$ so that the optimal solution set is the singleton $\mathbf{0}$, which is compact. Now we see that

$$f_\tau\left(x^{(i)}\right) = J - 1 + i + \tau \sum_{j \neq 1} \rho_j(1) + \tau \rho_1\left(s_1^{(i)}\right)$$

so that

$$\begin{aligned} f_\tau\left(x^{(i)}\right) - f_\tau\left(x^{(i+1)}\right) &= -1 + \tau\left[\rho_1\left(s_1^{(i)}\right) - \rho_1\left(s_1^{(i+1)}\right)\right] \\ &\geq -1 + \tau\left[\rho_1\left(s_1^{(i)}\right) - \rho_1\left(s_1^{(i)}+1\right)\right] \\ &\xrightarrow[i \to \infty]{} -1 + \tau\sigma. \end{aligned}$$

Therefore, if $z = x^{(1)}$ and $\tau > 1/\sigma$, the tail of the unbounded sequence $\left\{x^{(i)}\right\}$ is contained in $L_\tau(z)$, because $f_\tau\left(x^{(i)}\right) \to -\infty$.

♠

We now write out an expression for the dual of $\mathcal{P}\left(\tau^{(i)}, \theta\right)$ from (2.2):

$$\begin{array}{ll} \underset{x,q,r}{\text{maximize}} & f_\tau(x) + q(a - Ax) + r_+(\underline{b} - x) + r_-(x - \overline{b}) \\ \text{subject to} & r = \nabla f_\tau(x) - qA. \end{array} \tag{2.20}$$

Theorem 2.15 (Primal-Dual Convergence) *Suppose that c and D are convex and the hypotheses of theorem 2.11 are satisfied. Then the weakened Mangasarian–Fromovitz constraint qualification is satisfied by any limit point of the sequence of minimizers $\left\{x^{(i)}\right\}$. Therefore, the limit points of the bounded sequences $\left\{p^{(i)}\right\}$ and $\left\{r^{(i)}\right\}$ are optimal Lagrange multipliers (dual variables) for (1.8). Moreover, if A has full row rank, then the limit points of the bounded sequence $\left\{q^{(i)}\right\}$ are optimal Lagrange multipliers for (1.8).*

Proof: We are assuming that some $z \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$ exists. Since $\underline{b} < \overline{b}$, we may choose a point x arbitrarily close to z such that $x \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$ and $\underline{b} < x < \overline{b}$. But this is the Slater constraint qualification, which is equivalent to the weakened Mangasarian–Fromovitz constraint qualification under convexity (cf. (1.7)).

♠

2.2.3 Barrier-Linear Functions

We shall take a slight digression to introduce a technique for constructing a barrier function that agrees with a linear function for all sufficiently large arguments. Specifically, we wish to construct a barrier function ρ such that

$$\rho(u) = hu \quad \text{for } u \geq t,$$

where $h \in \mathbb{R}_-^{1\times J}$ and $t \in \mathbb{R}_{>0}^J$ are given. Such function ρ is called a **barrier-linear function** , and, in the event that the slope $h = \mathbf{0}$, we call ρ a **barrier-zero function**. Suppose we begin with a barrier function $\tilde{\rho}$ that satisfies the following assumption:

Assumption 2.16 *$\tilde{\rho}$ is a barrier function in the sense of definition 1.6. Moreover, each component function $\tilde{\rho}_j$ is twice continuously differentiable and $\tilde{\rho}_j''(t_j) \leq \tilde{\rho}_j''(\xi)$ whenever $0 < \xi \leq t_j$. If ρ_j'' is antitone, then the above inequality holds for all t_j.*

The inequality of second derivatives may seem at first to be restrictive. However, the analysis of theorem 2.1 shows that ρ_j'' should go to $+\infty$ near 0. Therefore, we think that this assumption is not restrictive, and the end of this subsection will show that all of the standard barrier functions exhibit this behavior for any t_j.

We shall do the construction componentwise by perturbing the component barrier functions with quadratic terms and then insisting on two continuous derivatives. Let

$$\rho_j(\xi) := \begin{cases} \tilde{\rho}_j(\xi) + \frac{\alpha_j}{2}\xi^2 + \beta_j\xi + \gamma_j & \text{if } \xi < t_j \\ h_j\xi & \text{if } \xi \geq t_j. \end{cases} \tag{2.21}$$

The following conditions are needed for ρ_j to be twice continuously differentiable:

$$\left.\begin{array}{lllllllll} \tilde{\rho}_j(t_j) & + & \frac{\alpha_j}{2}t_j^2 & + & \beta_j t_j & + & \gamma_j & = & h_j t_j \quad (\rho_j \text{ continuous at } t_j) \\ \tilde{\rho}_j'(t_j) & + & \alpha_j t_j & + & \beta_j & & & = & h_j \quad (\rho_j' \text{ continuous at } t_j) \\ \tilde{\rho}_j''(t_j) & + & \alpha_j & & & & & = & 0 \quad (\rho_j'' \text{ continuous at } t_j). \end{array}\right\} \tag{2.22}$$

This linear system in α_j, β_j and γ_j clearly has a unique solution, easily computable, for any $\tilde{\rho}_j$, h_j and t_j. The following theorem describes the important properties of the barrier-linear function ρ.

Theorem 2.17 *Suppose $\tilde{\rho}$ satisfies assumption 2.16 and that ρ is given by (2.21) and the coefficients α, β and γ solve (2.22). Then ρ is a twice continuously differentiable, convex barrier function in the sense of definition 1.6.*

Proof: We prove the theorem componentwise. ρ_j is twice continuously differentiable at t_j by (2.22), and it is clearly twice continuously differentiable on the rest of $\mathbb{R}_{>0}$ by definition. When $\xi \geq t_j$, $\rho_j''(\xi) = 0$, and when $\xi < t_j$ we use (2.21), (2.22) and assumption 2.16 to write

$$\rho_j''(\xi) = \tilde{\rho}_j''(\xi) + \alpha_j = \tilde{\rho}_j''(\xi) - \tilde{\rho}_j''(t_j) \geq 0,$$

so that ρ_j is convex. Clearly, conditions 1 and 2 of definition 1.6 are satisfied, so that we only need to verify the antitonicity condition 3. This is easily verified by noting that $\rho_j'(\xi) = h_j \leq 0$ for all $\xi \geq t_j$ and that ρ_j' is monotone, because ρ_j is convex. Therefore, $\rho_j(\xi) \leq h_j \leq 0$ for all $\xi > 0$, proving antitonicity.

♠

We conclude this section by noting that the most common barrier functions satisfy assumption 2.16. From our experience, the most common barrier functions are those constructed from

$$\tilde{\rho}_j(\cdot) = -\ln(\cdot) \qquad \text{and} \qquad \bar{\rho}_j(\cdot) = (\cdot)^{-\ell} \quad \text{for } \ell > 0.$$

(Note that if another base is used in the logarithm, the resulting function is a scaled version of the natural logarithm, because $\log_a(\xi) = \log_a(b)\log_b(\xi)$ for any two bases a and b.) It is clear that definition 1.6 is satisfied and that both are twice continuously differentiable. The final property would follow for any $t_j > 0$ if $\tilde{\rho}_j''$ and $\bar{\rho}_j''$ were antitone. This indeed is the case, as is apparent from

$$\tilde{\rho}_j''(\xi) = \frac{1}{\xi^2} \qquad \text{and} \qquad \bar{\rho}_j''(\xi) = \ell(\ell+1)(\xi)^{-(\ell+2)} \quad \text{for } \ell > 0.$$

Therefore, any of these typical barrier functions—or any non-negative combination of them—may be used to construct a barrier-linear function.

2.2.4 ε-Optimality for Special Barriers

We shall show that, for a particular class of barrier functions, there is a quantitative relationship between the size of the barrier coefficient τ and the error in the KKT conditions of the minimizer of $\mathcal{P}(\tau, d)$. Moreover, for convex problems, this yields a bound on the objective function value. We then derive an ε-optimality result which depends on the natural logarithm in much the same way as corollary 1.11. This may also be done for the barrier-zero function constructed from the logarithm.

Throughout this subsection, assume that $\tau^{(i)} > 0$ and that $x^{(i)}$ solves $\mathcal{P}\left(\tau^{(i)}, d\right)$. We also assume that $q^{(i)}$ and $r^{(i)}$ are optimal Lagrange multipliers for $\mathcal{P}\left(\tau^{(i)}, d\right)$, and that

$$p^{(i)} := \tau^{(i)} \nabla \rho \left(d - D\left(x^{(i)}\right)\right) \in \mathbb{R}^{1\times J}.$$

We already saw (in the proof of theorem 2.11) that all of the KKT conditions except (1.15) are satisfied by $\left(x^{(i)}, p^{(i)}, q^{(i)}, r^{(i)}\right)$. Therefore, the error in the KKT conditions occurs in only one single equation:

$$\varepsilon^{(i)} := p^{(i)}\left(d - D\left(x^{(i)}\right)\right) = \tau^{(i)} \sum_{j=1}^{J} \rho_j'\left(d_j - D_{j\bullet}\left(x^{(i)}\right)\right)\left(d_j - D_{j\bullet}\left(x^{(i)}\right)\right).$$

Therefore, the error in the KKT conditions may be bounded if ρ is chosen in such a way that $\rho_j'(\xi)\xi$ may be bounded *a priori* for all $\xi > 0$. We shall show that the logarithmic barrier function and the logarithmic barrier-zero function have this property.

Theorem 2.18 (Logarithmic Barrier KKT Error) *If ρ is the logarithmic barrier function (i.e., $\rho_j(\cdot) = -\ln(\cdot)$ for each component j), then*

$$\varepsilon^{(i)} = -\tau^{(i)} J.$$

Proof: Since $\rho_j = -\ln$, we see that $\rho_j'(\xi)\xi = -1$ for all $\xi > 0$.

♠

We now consider the logarithmic barrier-zero function that is zero for all $s \geq t$ (where $t > \mathbf{0}$ is a fixed vector). In the notation used in §2.2.3, we have $\tilde{\rho}_j(\cdot) = -\ln(\cdot)$, and $h = \mathbf{0}$. Therefore, (2.22) gives

$$\alpha_j = -\frac{1}{t_j^2} \qquad \beta_j = 0 \qquad \text{and} \qquad \gamma_j = \ln(t_j) + \frac{1}{2},$$

and so

$$\rho_j(\xi) = \begin{cases} -\ln(\xi) - \dfrac{\xi^2}{2t_j^2} + \ln(t_j) + \dfrac{1}{2} & \text{if } \xi < t_j \\ 0 & \text{if } \xi \geq t_j \end{cases}$$

for each j. The following ε-optimality result is available:

Theorem 2.19 (Logarithmic Barrier-Zero KKT Error) *If ρ is the logarithmic barrier-zero function (cf. §2.2.3) that is zero for all arguments $\geq t$ and J^* is the number of components j for which $d_j - D_{j\bullet}\left(x^{(i)}\right) < t_j$, then*

$$-2\tau^{(i)}J^* < \varepsilon^{(i)} < -\tau^{(i)}J^*.$$

Proof: We shall use the shorthand $s^{(i)} = d - D\left(x^{(i)}\right)$ to clean up the notation. Now consider each component j separately. If $s_j^{(i)} \geq t_j$, then $\rho_j'\left(s_j^{(i)}\right) = 0$ so that the contribution of this component j to the sum defining $\varepsilon^{(i)}$ is 0. If, on the other hand, $s_j^{(i)} < t_j$, then

$$\rho_j'\left(s_j^{(i)}\right)s_j^{(i)} = -1 - \frac{\left(s_j^{(i)}\right)^2}{t_j^2} \in (-2.. -1),$$

and this component contributes between -2 and -1 times $\tau^{(i)}$ to the sum.

♠

We now suppose that (1.8) is convex and v^* is the optimal objective function value.

Theorem 2.20 (Logarithmic Barrier Objective Error) *Suppose c and D are convex and that ρ is the logarithmic barrier function. Then*

$$v^* \leq c\left(x^{(i)}\right) \leq v^* + \tau^{(i)}J.$$

Proof: The inequality $v^* \leq c\left(x^{(i)}\right)$ follows from the primal feasibility of $x^{(i)}$. We already saw above that (1.14) is satisfied, so that $\left(x^{(i)}, p^{(i)}, q^{(i)}, r^{(i)}\right)$ is feasible for the dual problem (1.17). Therefore, a lower bound on v^* is provided by the dual objective function value

$$c\left(x^{(i)}\right) + \underbrace{p^{(i)}\left(d - D\left(x^{(i)}\right)\right)}_{\varepsilon^{(i)}} + q^{(i)}\underbrace{\left(a - Ax^{(i)}\right)}_{\mathbf{0}} + \underbrace{r_+^{(i)}\left(\underline{b} - x^{(i)}\right)}_{0} + \underbrace{r_-^{(i)}\left(x^{(i)} - \overline{b}\right)}_{0},$$

where the zero terms are a consequence of primal feasibility and complementary slackness. But now the result follows easily from theorem 2.18.

♠

A similar result holds for the logarithmic barrier-zero function.

Theorem 2.21 (Logarithmic Barrier-Zero Objective Error) *Suppose c and D are convex and that ρ is the logarithmic barrier-zero function (cf. §2.2.3) that is zero for all arguments $\geq t$. If J^* is the number of components j for which $d_j - D_{j\bullet}\left(x^{(i)}\right) < t_j$, then*

$$v^* \leq c\left(x^{(i)}\right) < v^* + 2\tau^{(i)}J^*.$$

Proof: Again $v^* \leq c\left(x^{(i)}\right)$ by primal feasibility and the dual objective function provides a lower bound on $c\left(x^{(i)}\right)$. The result then follows from theorem 2.19.

♠

2.3 The Three Phase Method

The method we have been developing fits naturally into a three phase framework. The RELAXED PHASE finds a point $x^{(0)} \in \mathcal{B}$. (If no such point exists, (1.8) is infeasible.) This is accomplished by throwing out the constraints $\mathcal{C}$ and minimizing some linear functional over $\mathcal{B}$. This may be done in a finite number of steps since this is nothing more than K linear programs, which should be solved concurrently on a parallel machine. The name of the RELAXED PHASE comes from the relaxation of the constraints $\mathcal{C}$.

In the event that a point in $\mathcal{B}$ was found in the RELAXED PHASE, the FEASIBILITY PHASE uses techniques developed in §2.1 to generate a feasible point in

$\mathcal{B} \cap \operatorname{int} \mathcal{C}$. We shall let $\hat{\imath}$ be the first iteration for which $x^{(\hat{\imath})} \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$, while not ruling out the possibility that $\hat{\imath} = \infty$. If c, D and ρ are convex, then so is $\mathcal{P}(\tau, \theta)$ for any τ and θ, in which case corollary 2.8 shows that a point in $\mathcal{B} \cap \operatorname{int} \mathcal{C}$ may be computed (assuming such a point exists). In the absence of convexity, theorem 2.7 still allows us to compute feasible points in practice, when the region $\mathcal{B} \cap \operatorname{int} \mathcal{C}$ is "large." This is done by choosing a finite upper bound for the sequence of objective functions $\left\{ f_\tau \left(x^{(i)}, \theta^{(i)} \right) \right\}$, and attempting to maintain the bound. Of course, if $\mathcal{B} \cap \operatorname{int} \mathcal{C} = \emptyset$, then this will not be possible (theorem 2.7 shows that $f_\tau \left(x^{(i)}, \theta^{(i)} \right) \to +\infty$ for any sequence $\left\{ \theta^{(i)} \right\}$ that satisfies (2.3) and (2.4)). We choose $\tau^{(1)} = \tau^{(i)}$ and $\theta^{(i)}$ by the rules (2.13) and (2.14) for all $i \leq \hat{\imath}$.

In the event that the FEASIBILITY PHASE is successful, the REFINE PHASE will iteratively approximate an optimal solution of (1.8). (For each iteration $i > \hat{\imath}$, $\theta^{(i)} = d$.) Since we cannot compute exact solutions to the nonlinear $\mathcal{P}(\tau, d)$, we shall not let the sequence $\left\{ \tau^{(i)} \right\}$ converge to 0, but we rather let

$$\tau^{(i+1)} = \max \left\{ \ \tau_{\text{inf}} \ , \ \ \lambda_\tau \tau^{(i)} \ \right\}, \tag{2.23}$$

where $\tau^{(1)}$ and τ_{inf} are positive constants and $\lambda_\tau \in (0..1)$ is a constant. For i sufficiently large, therefore, the problem $\mathcal{P}\left(\tau_{\text{inf}}, d\right)$ remains constant. If the $x^{(i)}$ are chosen as one iteration of a convergent method for the nonlinear program $\mathcal{P}\left(\tau^{(i)}, \theta^{(i)}\right)$, then the limit points of the REFINE PHASE will be solutions of $\mathcal{P}\left(\tau_{\text{inf}}, d\right)$. One may use any of the ε-optimality results in §2.2.4 to choose τ_{inf} suitably small, provided, of course, that the logarithmic barrier or the logarithmic barrier-zero functions are used to determine ρ. Moreover, one may compute the Lagrange multipliers and check the KKT equations or the objective function bounds.

Figure 2 gives the three phase algorithm in detail. (Note that $\hat{\imath}$ is defined only to relate to the above discussion and is not actually used in the algorithm.) For the time being, we shall let the statement:

"Generate $x^{(i)}$ as an *approximate* solution of $\mathcal{P}\left(\tau^{(i)}, \theta^{(i)}\right)$"

remain vague. The next chapter will give methods for computing the particular $x^{(i)}$.

```
    Assume that the parameters
        $\Theta > 0$        $\lambda_\theta \in (0..1)$
        $\tau^{(1)} > 0$        $\tau_{\text{inf}} > 0$        $\lambda_\tau \in (0..1)$
    are given and that we wish to solve (1.8)
RELAXED PHASE
    $i \leftarrow 0$
    Compute $x^{(0)}$ as the solution of the "relaxed" problem (2.1)
    If we determine that $\mathcal{B} = \emptyset$ then quit
    Set $\theta^1$ as in (2.13)
    If $x^{(0)} \in \operatorname{int} \mathcal{C}$
      Then go to the REFINE PHASE
      Otherwise go to the FEASIBILITY PHASE
FEASIBILITY PHASE
    $i \leftarrow i + 1$
    Generate $x^{(i)}$ as an *approximate* solution of $\mathcal{P}\left(\tau^{(i)}, \theta^{(i)}\right)$
    Set $\theta^{(i+1)}$ as in (2.14)
    $\tau^{(i+1)} \leftarrow \tau^{(i)}$
    If $x^{(i)} \in \mathcal{B} \cap \operatorname{int} \mathcal{C}$
      Then $\hat{\imath} \leftarrow i$ and go to the REFINE PHASE
      Otherwise repeat the FEASIBILITY PHASE
REFINE PHASE
    $i \leftarrow i + 1$
    Set $\theta^{(i)} \leftarrow d$
    Generate $x^{(i)}$ as an *approximate* solution of $\mathcal{P}\left(\tau^{(i)}, \theta^{(i)}\right)$
    $\tau^{(i+1)} \leftarrow \max\left\{ \tau_{\text{inf}} \ , \ \lambda_\tau \tau^{(i)} \right\}$ (cf. (2.23))
    Repeat the REFINE PHASE until $x^{(i)}$ is "suitable"
```

Figure 2: The three phase method.

Chapter 3

Decomposition of $\mathcal{P}(\tau, \theta)$

The previous chapter generates a sequence of block-constrained nonlinear barrier problems $\left\{\mathcal{P}\left(\tau^{(i)}, \theta^{(i)}\right)\right\}$ for which we must compute approximate minimizers $\left\{x^{(i)}\right\}$. However, we are only concerned with approximating a local minimizer of the original problem (1.8). Therefore, we may interpret "approximate minimizer of $\mathcal{P}\left(\tau^{(i)}, \theta^{(i)}\right)$" quite loosely, so long as the three phase method of figure 2 computes a point that is suitable in practice. Suppose that we have an approximate solution of $\mathcal{P}\left(\tau^{(i-1)}, \theta^{(i-1)}\right)$, namely $x^{(i-1)}$, from the last iteration. The method of this chapter will produce a sequence $\left\{x^{(i,t)}\right\}_t$, with $x^{(i,0)} := x^{(i-1)}$, the limit points of which are solutions to $\mathcal{P}\left(\tau^{(i)}, \theta^{(i)}\right)$. At some point the sequence is truncated and the last point is taken to be the "approximate solution of $\mathcal{P}\left(\tau^{(i)}, \theta^{(i)}\right)$," i.e., $x^{(i)} := x^{(i,T(i))}$ for some $T(i)$. For the remainder of this chapter, the outer iteration i is irrelevant and we shall drop dependence on i, so that we are generating a sequence $\left\{x^{(t)}\right\}$, the limit points of which are solutions of $\mathcal{P}(\tau, \theta)$.

For convenience, we restate the definition of the barrier problem from (2.2) with the objective function f:

$$\mathcal{P}(\tau, \theta) \; := \; \left\{ \begin{array}{ll} \underset{x}{\text{minimize}} & f(x) := c(x) + \tau\rho(\theta - D(x)) \\ \text{subject to} & x \in \mathcal{B} \end{array} \right\}, \tag{3.1}$$

where $f : \mathcal{B} \to \mathbb{R} \cup +\infty$ is essentially smooth (by theorem 2.4) and the block constraints $\mathcal{B}$ are given by (1.10), i.e.,

$$\mathcal{B} := \left\{ \; x \in \mathbb{R}^N \;\middle|\; \forall\{k = 1, \ldots, K\} \; A_{[k]}x_{[k]} = a_{[k]} \;\text{ and }\; \underline{b} \leq x \leq \overline{b} \; \right\}.$$

Moreover, we are assuming that we know some $x^{(0)} \in \mathcal{B} \cap \operatorname{dom} f$. The KKT conditions for (3.1) are written as

$$Ax = a \qquad \underline{b} \leq x \leq \overline{b} \tag{3.2}$$

$$r = \nabla f(x) - qA \tag{3.3}$$

$$r_{+}(\underline{b} - x) = 0 \qquad r_{-}(x - \overline{b}) = 0. \tag{3.4}$$

We point out that the (restricted) simplicial decomposition of §1.5 may be used on this problem if f is pseudo-convex. This will surely be the case if c is pseudo-convex and D and ρ are convex (cf. lemma 1.9). Another difficulty with any method that solves a nonlinear coordinator problem is that it must be able to handle the case where $\mathcal{B} \not\subset \operatorname{dom} f$.

We propose a method for obtaining approximate solutions of $\mathcal{P}(\tau, \theta)$ that consists of approximating the objective function to allow the *computation of search directions separately for each block*, and then solving an coordinator problem to produce a new approximation to the solution. It uses a linear or convex quadratic approximation of the objective function along with a simple trust region to produce descent directions. The coordinator takes advantage of the block structure of the constraints by using a multi-dimensional search rather than a line search. Analyzing this decomposition method in terms of descent allows us to avoid the assumption of pseudo-convexity.

3.1 Concurrent Block Search Directions

This section will show how to design subproblems that will form the basis for our decomposition method. A key point is that the subproblems are linear programs with block separable feasible region $\mathcal{B} = \mathcal{B}_{[1]} \times \cdots \times \mathcal{B}_{[K]}$. Each subproblem is, therefore, K independent linear programming subproblems. Since solution of subproblems is typically the computational task that takes most of the time, we may expect parallel computing to become an important tool for the solution of such problems. In fact, much recent work in decomposition methods [Medhi 1987; Ho *et al* 1988; Lee *et al* 1989; Pinar and Zenios 1990] is motivated by parallel computing technology.

Suppose we are given a current point $x^{(t)} \in \mathcal{B} \cap \operatorname{dom} f$. Then the linearization of $\mathcal{P}(\tau, \theta)$ around $x^{(t)}$ is

$$\underset{y}{\text{minimize}} \ \ \nabla f\left(x^{(t)}\right) y \quad \text{subject to} \quad Ay = a \quad \text{and} \quad \underline{b} \leq y \leq \overline{b}. \tag{3.5}$$

The form of this subproblem shows that we shall are working with a price directed decomposition. That this is exactly the form of subproblems in Dantzig-Wolfe decomposition (1.22) and (restricted) simplicial decomposition (1.25).

3.1.1 Resource Allocation Techniques

From a computational point of view, subproblems of the form (3.5) are not made much more difficult if the bounds are changed. In particular, if f has finite poles (i.e., $f(x) = +\infty$ for some finite x), we would like the solution to the subproblem to stay closer to $\operatorname{dom} f$.

Definition 3.1 *A* **decoupled resource allocation for** $\mathcal{B}$ *is an interval-valued map*

$$R \ : \ \mathcal{B} \to \left\{\textit{intervals in } \mathrm{I\!R}^N\right\} \ : \ x \mapsto \left\{ \ z \in \mathrm{I\!R}^N \ \middle| \ \underline{R}(x) \leq z \leq \overline{R}(x) \ \right\}$$

that also satisfies the following criteria:

1. $\forall\{x \in \mathcal{B}\} \ \ \underline{b} \leq \underline{R}(x) \leq x \leq \overline{R}(x) \leq \overline{b}$,

2. *for any* bounded *sequence* $\left\{x^{(t)}\right\} \subset \mathcal{B}$

$$\bigcup_{t=0}^{\infty} R\left(x^{(t)}\right) \ \ \textit{is bounded,}$$

and

3. for any $z \in \mathcal{B}$ and any bounded *sequence* $\left\{x^{(t)}\right\} \subset \mathcal{B}$ *with*

$$\alpha^{(t)} := \max\left\{ \;\alpha \;\middle|\; 0 \le \alpha \le 1 \quad \textit{and} \quad x^{(t)} + \alpha\left(z - x^{(t)}\right) \in R\left(x^{(t)}\right) \;\right\},$$

$\liminf_{t\to\infty} \alpha^{(t)} > 0$.

If a decoupled resource allocation R is given, the subproblems that we solve are the linear programs

$$\begin{array}{ll} \text{minimize}_y & \nabla f\left(x^{(t)}\right) y \\ \text{subject to} & Ay = a \quad \text{and} \quad \underline{R}\left(x^{(t)}\right) \le y \le \overline{R}\left(x^{(t)}\right). \end{array} \tag{3.6}$$

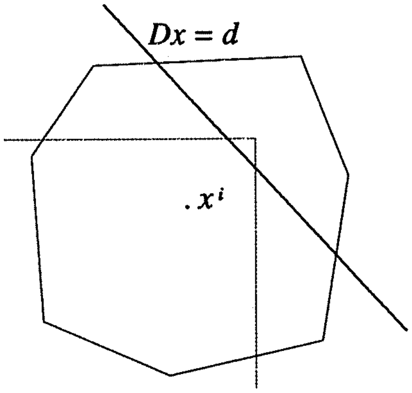
For computational ease we limit ourselves to *decoupled* resource allocations, i.e., interval valued R. Condition 1 guarantees that the feasible region of (3.6) is a subset of $\mathcal{B}$. The subproblems (3.6) are feasible ($x^{(t)}$ is a feasible point) and bounded (by condition 2). Condition 3 is used to guarantee that the boundary of $R\left(x^{(i)}\right)$ cannot constrain $x^{(i)}$ from moving toward to any other feasible point z. This condition is very important, because the convergence proof of the decomposition method is based only upon the price directed ideas; this condition is included so that the price directed proof continues to work. Therefore, decoupled resource allocations are heuristic in nature and designed to achieve a practical improvement in the computational efficiency of the method without altering the theoretical convergence properties.

Now suppose that the function f is given by $f(\cdot) = c(\cdot) + \tau\rho(\theta - D(\cdot))$ from $\mathcal{P}(\tau, \theta)$. Then $x^{(t)} + \delta y \in \operatorname{dom} f$ iff

$$\mathbf{0} < \theta - D\left(x^{(t)} + \delta y\right) \approx \theta - D\left(x^{(t)}\right) - \nabla D\left(x^{(t)}\right)\delta y.$$

(If D is convex, the "$\approx$" may be replaced by a "$\le$".) Suppose that only one component—the nth component—of δy is different from 0. Then we would want, in some sense,

$$\nabla D(x^{(t)})_{\bullet n}\delta y_n \le \theta - D(x^{(t)}).$$



$\mathcal{B}$ is represented by the solid polygon.

Figure 3: A typical decoupled resource allocation.

Suppose that we wish to relax this constraint by multiplying the right-hand-side by the scaling factor $\omega > 0$. Further suppose that we want none of these constraints active to within some constant $\Omega > 0$ at $\delta y = \mathbf{0}$. We can obtain these by defining

$$R\left(x^{(t)}\right) = \left\{ \; x^{(t)} + \delta y \; \middle| \; \begin{array}{c} \underline{b} \leq x^{(t)} + \delta y \leq \bar{b} \\ -\kappa \mathbf{1} \leq \delta y \leq \kappa \mathbf{1} \\ \forall\{n = 1, \ldots, N\} \;\; \nabla D(x^{(t)})_{\bullet n} \delta y_n \leq \Psi \end{array} \right\}, \tag{3.7}$$

for

$$\Psi = \max\left\{ \; \Omega \mathbf{1} \; , \; \omega\left(\theta - D\left(x^{(t)}\right)\right) \; \right\} > \mathbf{0}$$

and κ a, positive real number. In this case the values of $\underline{R}_n$ and $\overline{R}_n$ are then computed by the following "ratio test":

$$\begin{aligned} \underline{R}_n(x) &= \max\Bigg[\; \underline{b}_n \; , \; x_n - \kappa \; , \\ & \qquad x_n + \max\left\{ \; \frac{\max\left\{\Omega, \omega\left(\theta_j - D_{j\bullet}(x)\right)\right\}}{\nabla D(x)_{j,n}} \; \middle| \; \nabla D(x)_{j,n} < 0 \; \right\} \; \Bigg] \\ \overline{R}_n(x) &= \min\Bigg[\; \bar{b}_n \; , \; x_n + \kappa \; , \\ & \qquad x_n + \min\left\{ \; \frac{\max\left\{\Omega, \omega\left(\theta_j - D_{j\bullet}(x)\right)\right\}}{\nabla D(x)_{j,n}} \; \middle| \; \nabla D(x)_{j,n} > 0 \; \right\} \; \Bigg]. \end{aligned}$$

Note that, if $\underline{b}$ and $\overline{b}$ are both finite, κ may be chosen large enough to be effectively ignored. In this case, if D is monotonic (e.g., if D is a non-negative matrix), then only the upper bounds will be modified, which is to say that $\underline{R} = \underline{b}$. We now verify that the definition in (3.7) produces a decoupled resource allocation.

Theorem 3.2 *Suppose R is defined by (3.7) where Ω, ω and κ are positive real numbers. Then R is a decoupled resource allocation for $\mathcal{B}$ in the sense of definition 3.1.*

Proof: Condition 1 is true because $\Omega > 0$ and condition 2 follows from the finiteness of κ. Suppose that $z \in \mathcal{B}$, $\left\{x^{(t)}\right\} \subset \mathcal{B}$ and $\alpha^{(t)}$ is defined as in condition 3 of definition 3.1. We suppose that $\liminf_t \alpha^{(t)} = 0$ and will show that this implies that the sequence $\left\{x^{(t)}\right\}$ is unbounded, which will finish the proof.

We may thin the sequence so that $\lim_{t\to\infty} \alpha^{(t)} = 0$, $\forall\{t\}\ \ \alpha^{(t)} < 1$ and there is some particular n such that

$$\forall\{t\}\ \ x_n^{(t)} + \alpha^{(t)}\left(z_n - x_n^{(t)}\right) = R_n^?\left(x^{(t)}\right),$$

where $R_n^?$ is either $\underline{R}_n$ or $\overline{R}_n$. Then $\alpha^{(t)} < 1$ implies that $z_n - x_n^{(t)} \neq 0$, so that

$$\alpha^{(t)} = \frac{R_n^?\left(x^{(t)}\right) - x_n^{(t)}}{z_n - x_n^{(t)}} = \frac{\left|R_n^?\left(x^{(t)}\right) - x_n^{(t)}\right|}{\left|z_n - x_n^{(t)}\right|} \to 0. \tag{3.8}$$

In the case that $R_n^?\left(x^{(t)}\right)$ is equal to $\underline{b}_n$ or $\overline{b}_n$, we know that

$$\left|z_n - x_n^{(t)}\right| \leq \left|R_n^?\left(x^{(t)}\right) - x_n^{(t)}\right|$$

because $z \in \mathcal{B}$. But then $\alpha^{(t)} \geq 1$, so that $R_n^?\left(x^{(t)}\right)$ can never take on the values $\underline{b}_n$ or $\overline{b}_n$ for our chosen subsequence. If $R_n^?\left(x^{(t)}\right) = x_n^{(t)} \pm \kappa$, then the numerator of (3.8) becomes κ, in which case $\left\{x^{(t)}\right\}$ must be unbounded. If $R_n^? = \underline{R}_n$, then

$$R_n^?\left(x^{(t)}\right) - x_n^{(t)} \leq \max\left\{ \frac{\Omega}{\nabla D\left(x^{(t)}\right)_{j,n}} \;\middle|\; \nabla D\left(x^{(t)}\right)_{j,n} < 0 \right\} < 0,$$

while if $R^?_n = \overline{R}_n$, then

$$R^?_n\left(x^{(t)}\right) - x^{(t)}_n \geq \max\left\{ \frac{\Omega}{\nabla D\left(x^{(t)}\right)_{j,n}} \;\middle|\; \nabla D\left(x^{(t)}\right)_{j,n} > 0 \right\} > 0.$$

If $\left\{x^{(t)}_n\right\}$ is bounded, then so is $\left\{\nabla D\left(x^{(t)}\right)\right\}$ (D is continuously differentiable on $\mathbb{R}^N$), so that the numerator of (3.8) is bounded away from 0. This then implies that the denominator of (3.8) is unbounded, so that $\left\{x^{(t)}_n\right\}$ must be at once bounded and unbounded. This contradiction shows that $\left\{x^{(t)}_n\right\}$ cannot be bounded.

♠

For such a decoupled resource allocation, $R(x^{(t)}) \not\subset \operatorname{dom} f$ even when $x^{(t)} \in \operatorname{dom} f$. But constraints of this form in the subproblems tend to limit extreme violation of this "domain constraint." Enforcement of "domain constraints" and convergence of the overall process is ensured by the coordinator, discussed below.

3.1.2 Convex Quadratic Subproblems

We will, in fact prove that a more general subproblem may be used. Suppose that we solve the subproblem

$$\begin{array}{ll} \underset{y}{\text{minimize}} & g^{(t)}\left(y - x^{(t)}\right) + \left(y - x^{(t)}\right)^{\mathsf{T}} G^{(t)}\left(y - x^{(t)}\right) \\ \text{subject to} & Ay = a \quad \text{and} \quad \underline{R}\left(x^{(t)}\right) \leq y \leq \overline{R}\left(x^{(t)}\right), \end{array} \tag{3.9}$$

where the approximate gradients $g^{(t)}$ satisfy

$$\lim_{t\to\infty} \left\| g^{(t)} - \nabla f\left(x^{(t)}\right) \right\| = 0 \tag{3.10}$$

and the symmetric matrices $G^{(t)}$ satisfy

$$\exists\{\beta > 0\}\forall\{t, x\} \;\; 0 \leq x^{\mathsf{T}} G^{(t)} x \leq \beta \left\|x\right\|^2 . \tag{3.11}$$

Note that the subproblems (3.6) are of the form (3.9) with $g^{(t)} = \nabla f\left(x^{(t)}\right)$ and $G^{(t)} = \mathbf{0}$. (3.10) means that the gradient of the subproblem's objective function at the current point tends to the gradient of the original problem at the current point, while (3.11) requires that the $G^{(t)}$ are symmetric, positive semidefinite with bounded eigenvalues.

There are two reasons why it is reasonable to generalize the subproblems allowed from (3.6) to (3.9). First, inexact gradient values $g^{(t)}$ mean that an inexact solution of (3.6) will still produce convergence. Second, in a massively parallel context, the methods for solving problems of the form (3.9) are most effective when $G^{(t)}$ are positive definite diagonal matrices [Zenios and Lasken 1988; Bertsekas and Tsitsiklis 1989, §5.5 and §6.6; Tseng 1990] with mild condition numbers [Hager and Hearn 1990].

The following theorem says that any subproblem sequence defined by such a quadratic approximation and a decoupled resource allocation allows the coordinator (cf. §3.2) to guarantee convergence.

Theorem 3.3 (Sufficiency of Search Directions) *Suppose that $\left\{x^{(t)}\right\} \subset \mathcal{B} \cap \operatorname{dom} f$ has $x^{(t)} \to \tilde{x} \in \mathcal{B} \cap \operatorname{dom} f$, that $y^{(t)}$ are solutions to (3.9) for some decoupled resource allocation R and that (3.10) and (3.11) hold. If there exists some $\tilde{y} \in$* lim pt $\left\{y^{(t)}\right\}$ *with*

$$\nabla f(\tilde{x})(\tilde{y}-\tilde{x}) \geq 0, \tag{3.12}$$

then

$$\forall\{z \in \mathcal{B}\} \quad \nabla f(\tilde{x})(z-\tilde{x}) \geq 0,$$

which implies, by theorem (1.1), that $\tilde{x}$ is a KKT point.

Proof: Without loss of generality, we may thin the sequence so that $y^{(t)} \to \tilde{y}$. Then (3.10), (3.11) and (3.12) imply that

$$\begin{gathered}\forall\{\varepsilon>0\}\exists\{t_0>0\}\forall\{t>t_0\} \\ g^{(t)}\left(y^{(t)}-x^{(t)}\right)+\left(y^{(t)}-x^{(t)}\right)G^{(t)}\left(y^{(t)}-x^{(t)}\right)>-\varepsilon.\end{gathered} \tag{3.13}$$

Now if there is some $z \in \mathcal{B}$ with

$$\nabla f(\tilde{x})(z-\tilde{x}) = -\zeta < 0,$$

we shall contradict (3.13). Without loss of generality, we may further thin the sequence so that $G^{(t)} \longrightarrow \tilde{G}$ (the sequence is bounded by (3.11) and the finite dimensional space of $N \times N$ real matrices is complete). We also define $\alpha^{(t)}$ as the maximal step sizes in condition 3 of definition 3.1 and $\tilde{\alpha} := \liminf_t \alpha^{(t)} > 0$. We may then pick a $\lambda \in (0..1)$ (independent of t) such that

$$\lambda\tilde{\alpha}\nabla f(\tilde{x})(z-\tilde{x}) + \lambda^2\tilde{\alpha}^2(z-\tilde{x})^{\mathsf{T}}\tilde{G}(z-\tilde{x}) < -\frac{\lambda\zeta\tilde{\alpha}}{2} < 0.$$

But then, for the subsequence in question,

$$\begin{array}{l}\exists\{t_1 > 0\}\forall\{t > t_1\}\\ g^{(t)}\left[\lambda\alpha^t\left(z - x^{(t)}\right)\right] + \left[\lambda\alpha^t\left(z - x^{(t)}\right)\right]^{\mathsf{T}} G^{(t)}\left[\lambda\alpha^t\left(z - x^{(t)}\right)\right] < -\frac{\lambda\zeta\tilde{\alpha}}{4} < 0,\end{array}$$

and the quantity in the square brackets gives

$$\lambda\alpha^t\left(z - x^{(t)}\right) = \underbrace{\lambda\alpha^t\left(z - x^{(t)}\right) + x^{(t)}}_{v^{(t)}} - x^{(t)}.$$

Note that $v^{(t)}$ is feasible for (3.9) (by the definition of $\alpha^{(t)}$ and $\lambda \in (0..1)$) and the objective function values of (3.9) at $v^{(t)}$ are bounded from above by $-\lambda\zeta\tilde{\alpha}/4 < 0$ for t (in the subsequence) sufficiently large. This contradicts (3.13) because $y^{(t)}$ is the minimizer of (3.9).

♠

3.2 The Master Problem and Coordination

The line search procedure used in classical nonlinear programming algorithms is a method for turning suitable descent directions into suitable descent, guaranteeing convergence. However, the block structure of our problem makes a multidimensional search quite natural. We will determine a master problem for this decomposition method and use it to motivate a multi-dimensional search that we will call the **coordinator problem**. This coordinator has the property that it may be solved in a finite number of steps and its solution at each iteration will allow us to apply theorem 3.3.

Let $y^{(t)}$ be any solution of the subproblem (3.9), which typically is in the form (3.6). The definition of R shows that $y^{(t)} \in \mathcal{B}$, and that one would *expect* to have

$$\forall \{k = 1, \ldots, K\} \quad \nabla f\left(x^{(t)}\right)_{[k]} \left(y^{(t)}_{[k]} - x^{(t)}_{[k]}\right) < 0.$$

That is to say that $y^{(t)} - x^{(t)}$ is expected to be a *descent direction* for $\mathcal{P}(\tau, \theta)$ at $x^{(t)}$. Let

$$Y^{(t)} := \begin{pmatrix} y^{(t)}_{[1]} - x^{(t)}_{[1]} & & \\ & \ddots & \\ & & y^{(t)}_{[K]} - x^{(t)}_{[K]} \end{pmatrix}, \tag{3.14}$$

so that each column of $Y^{(t)}$ is a search direction that is nonzero in only one block. Furthermore, note that $AY^{(t)} = \mathbf{0}$ for all t, because $A_{[k]} y^{(t)}_{[k]} = A_{[k]} x^{(t)}_{[k]} = a_{[k]}$ for all k and t. This property enables us to ignore the affine constraints $Ax = a$ in constructing the master problem. Therefore, the **master problem** for this decomposition method is

$$\underset{w}{\text{minimize}} \ \ f\left(x^{(t)} + Y^{(t)} w\right) \quad \text{subject to} \quad \underline{b} \leq x^{(t)} + Y^{(t)} w \leq \bar{b}. \tag{3.15}$$

The following proposition shows that the constraints of 3.15 are quite simple.

Proposition 3.4 *The constraints of (3.15) are simple bounds computable by a ratio test.*

Proof: The columns of $Y^{(t)}$ are complementary, so that the linear inequality constraints on w have no coupling, i.e., they are simple bounds. The lower bounds are given by

$$\underline{w}^{(t)}_k = \max \left\{ \max_{\{n | Y^{(t)}_{n,k} > 0\}} \frac{\left(\underline{b} - x^{(t)}\right)_n}{Y^{(t)}_{n,k}} \ , \ \max_{\{n | Y^{(t)}_{n,k} < 0\}} \frac{\left(\bar{b} - x^{(t)}\right)_n}{Y^{(t)}_{n,k}} \right\}$$

and the upper bounds by

$$\overline{w}_k^{(t)} = \min\left\{ \min_{\{n|Y^{(t)}_{n,k}>0\}} \frac{\left(\overline{b}-x^{(t)}\right)_n}{Y^{(t)}_{n,k}} \quad , \quad \min_{\{n|Y^{(t)}_{n,k}<0\}} \frac{\left(\underline{b}-x^{(t)}\right)_n}{Y^{(t)}_{n,k}} \right\}.$$

♠

If $w_*^{(t)}$ solves the master problem (3.15), then we could solve the master problem and take $x^{(t+1)} = x^{(t)} + Y^{(t)} w_*^{(t)}$ for the next iterate. However, the nonlinearity of f makes exact solution of (3.15) impossible. Therefore, we will consider coordinator problems that are finitely satisfiable and yet retain the convergence properties. We will not attempt to prove any finite termination results analogous to those of simplicial decomposition, because the finite termination results of simplicial decomposition assume the exact solution of the master problems (3.15).

3.2.1 A Coordination Method for Convex Problems

In this section, we consider the case when f is convex. This implies that (3.15) is also convex for all t. Rather than solving (3.15) exactly, we will specify in this section and §3.2.2 sets of approximate optimality conditions that we wish the result of the coordinator to satisfy. Define

$$\Xi^{(t)} := \left\{ \quad x = (1-\lambda)x^{(t)} + \lambda y^{(t)} \;\middle|\; 0 \leq \lambda \leq 1 \quad \right\},$$

to be the line segment from the current point $x^{(t)}$ to the subproblem solution $y^{(t)}$. Note that any $x \in \Xi^{(t)}$ may be written as $x = x^{(t)} + Y^{(t)} w$ for some $w = \lambda \mathbf{1}$ feasible for (3.15). Our coordinator problem computes $x^{(t+1)} = x^{(t)} + Y^{(t)} w^{(t)}$ that satisfies the conditions (3.16), (3.17) and (3.18) as follows:

$$\forall\{t\} \;\; f\left(x^{(t+1)}\right) - f\left(x^{(t)}\right) \leq \mu \left[\min_{x\in\Xi^{(t)}} f(x) - f\left(x^{(t)}\right)\right] \qquad (3.16)$$

for a constant $\mu \in (0..1)$ independent of t

$$\forall\{t\} \;\; \left\|\left[\nabla f\left(x^{(t+1)}\right) - u^{(t)}\right] Y^{(t)}\right\| \leq \mu' \left\|\nabla f\left(x^{(t)}\right) Y^{(t)}\right\| \qquad (3.17)$$

for a constant $\mu' \in (0..\infty]$ independent of t

$$u_+^{(t)}\left(x^{(t+1)}-\underline{b}\right)=0 \quad \text{and} \quad u_-^{(t)}\left(x^{(t+1)}-\bar{b}\right)=0 \tag{3.18}$$

The $u^{(t)}$ above may be interpreted as Lagrange multiplier estimates in the following manner. If $x^{(t+1)}$ is optimal for (3.15) and the Lagrange multipliers are $u^{(t)}$, the KKT conditions for (3.15) may be written

$$\begin{gathered}\nabla f\left(x^{(t+1)}\right)Y^{(t)}-u^{(t)}Y^{(t)}=\mathbf{0}\ , \\ u_+^{(t)}\left(x^{(t+1)}-\underline{b}\right)=0 \quad \text{and} \quad u_-^{(t)}\left(x^{(t+1)}-\bar{b}\right)=0.\end{gathered} \tag{3.19}$$

Therefore, the quantity inside the norm on the left-hand-side of (3.17) is a "reduced gradient" for (3.15). Thus, the conditions (3.17) and (3.18) may be considered to be approximate optimality conditions for (3.15). Clearly, condition (3.16) is also true at the optimum of (3.15), so that the coordinator conditions (3.16) and (3.17) are approximate optimality condition for the master problem (3.15).

Using an active set method [Fletcher 1987; Gill *et al* 1981], we may generate (for fixed t) sequences $x^{(t,i)}=x^{(t)}+Y^{(t)}w^{(t,i)}$ and $u^{(t,i)}$ that maintain

$$u_+^{(t,i)}\left(x^{(t,i)}-\underline{b}\right)=0 \quad \text{and} \quad u_-^{(t,i)}\left(x^{(t,i)}-\bar{b}\right)=0$$

and such that any limit point of $\left\{w^{(t,i)}\right\}_i$ is an optimal solution for (3.15). Therefore, we may computationally check satisfaction of (3.17), and (3.17) will be true for *all* i sufficiently large (by the optimality conditions (3.19) and because $\mu'>0$). To check satisfaction of (3.16), suppose $\mathbf{0}$ is not optimal for (3.15). For the moment, let $f^{(t)}(w):=f\left(x^{(t)}+Y^{(t)}w\right)$. Using a linear underestimator for the convex function $f^{(t)}$ and the fact that a linear functional over a line segment attains its minimum at one end, we have

$$\begin{aligned}
&\mu\left[\min_{x\in\Xi}f(x)-f\left(x^{(t)}\right)\right]=\mu\left[\min_{x\in\Xi}f(x)-f^{(t)}(\mathbf{0})\right]\\
\geq\ &\mu\Bigg[f^{(t)}\left(w^{(t,i)}\right)-f^{(t)}(\mathbf{0})\\
&\quad+\min\left\{\ \nabla f^{(t)}\left(w^{(t,i)}\right)\left(\mathbf{0}-w^{(t,i)}\right)\ ,\ \nabla f^{(t)}\left(w^{(t,i)}\right)\left(\mathbf{1}-w^{(t,i)}\right)\ \right\}\Bigg]\\
\to\ &\mu\left[\phi^{(t)}-f^{(t)}(\mathbf{0})\right]\\
>\ &\phi^{(t)}-f^{(t)}(\mathbf{0})
\end{aligned}$$

where $\phi^{(t)}$ is the minimum value of (3.15). Therefore, we have a computable sufficient condition for (3.16) which is always true for all i sufficiently large. Once we have verified (3.16) and (3.17), we may truncate the sequence $w^{(t,i)}$ (at iteration $\hat{\imath}$, say) to produce our choice of $x^{(t+1)} = x^{(t)} + Y^{(t)}w^{(t,\hat{\imath})}$. The reader should note that this is the step where convexity is crucial; without convexity, there is no way of bounding the global minimum of f along the segment $\Xi^{(t)}$ without some global information.

The convergence proof of our coordinator is modeled after the standard proof of the Frank–Wolfe method [Frank and Wolfe 1956; Craven 1978].

Theorem 3.5 (Convergence of Decomposition Method) *Suppose*

1. *f is convex and essentially smooth,*
2. *$x^{(0)} \in \mathcal{B} \cap \operatorname{dom} f$,*
3. *$y^{(t)}$ are solutions to (3.9) (where (3.10) and (3.11) hold) for some decoupled resource allocation R,*
4. *$Y^{(t)}$ is given by (3.14),*
5. *$x^{(t+1)} = x^{(t)} + Y^{(t)}w^{(t)} \in \mathcal{B}$ satisfies the coordinator conditions (3.16), (3.17) and (3.18) for all t.*

Then either $f\left(x^{(t)}\right) \xrightarrow{t\to\infty} -\infty$ or $\lim_{t\to\infty} f\left(x^{(t)}\right) = \tilde{f} > -\infty$, and in the latter case the limit points of $x^{(t)}$ are optimal solutions of (3.1).

Proof: From (3.16) we may deduce that the algorithm is monotonic, i.e., $f\left(x^{(i+1)}\right) \leq f\left(x^{(i)}\right)$. Therefore, either the function values diverge to $-\infty$ or else they have a finite limiting value $\tilde{f}$. Suppose the latter is the case and that there is an limit point $\tilde{x}$. We see that $\lim f\left(x^{(t+1)}\right) - f\left(x^{(t)}\right) = 0$. Consider a subsequence $\left\{x^{(\sigma(t))}\right\} \subset \left\{x^{(t)}\right\}$ that has $x^{(\sigma(t))} \to \tilde{x}$. We may also assume, by condition 2 of definition 3.1, that $y^{(\sigma(t))} \to \tilde{y}$ (further thin the sequence). Now, for all $\lambda \in (0..1]$ we have

$$\begin{aligned} f\left((1-\lambda)x^{(\sigma(t))} + \lambda y^{(\sigma(t))}\right) - f\left(x^{(\sigma(t))}\right) &\geq \left[\min_{x \in \Xi(\sigma(t))} f(x) - f\left(x^{(\sigma(t))}\right)\right] \\ &\geq \frac{1}{\mu}\left[f\left(x^{(\sigma(t)+1)}\right) - f\left(x^{(\sigma(t))}\right)\right]. \end{aligned}$$

Since the right-hand-side $\to 0$ as $t \to \infty$, we have $f((1-\lambda)\tilde{x} + \lambda\tilde{y}) - f(\tilde{x}) \geq 0$. Now take the limit of this as $\lambda \downarrow 0$ to get $\nabla f\left(\tilde{x}\right)\left(\tilde{y} - \tilde{x}\right) \geq 0$.

Now, for any subsequence $x^{(\sigma(t))} \to \tilde{x}$, the hypotheses of theorem (3.3) are met, so that $\tilde{x}$ is a KKT point, and hence optimal for the convex problem (3.1).

♠

We may note that the condition (3.17) is not needed in the convergence proof. However, we have noticed in practice that the use of (3.17) tends to force convergence more robustly than using (3.16) alone. Convergence is still assured if $\mu' = +\infty$ and (3.17) is a null condition.

3.2.2 Coordination using a Stabilization Method

We will let the vector $\hat{w}^{(t)}$ be the unique minimizer of the positive definite quadratic form

$$Q^{(t)}(w) := \nabla f\left(x^{(t)}\right) Y^{(t)} w + w^{\mathsf{T}} H^{(t)} w \tag{3.20}$$

subject to the constraints that $\hat{w}^{(t)}$ be feasible for the master problem (3.15). Here the symmetric matrix $H^{(t)}$ satisfies

$$\exists\{\beta\}\forall\{t,w\} \;\; \left|w^{\mathsf{T}} H^{(t)} w\right| \leq \beta \left\|w\right\|^2 \tag{3.21}$$

and

$$H^{(t)} \text{ is positive definite.} \tag{3.22}$$

The condition (3.21) merely says that all eigenvalues of the sequence of matrices $\left\{H^{(t)}\right\}$ are bounded in absolute value by some β. This leaves much latitude as to how one would choose the $H^{(t)}$, but as long as

$$\exists\{\beta'\}\forall\{t,w\} \;\; \left|w^{\mathsf{T}} {Y^{(t)}}^{\mathsf{T}} \nabla^2 f\left(x^{(t)}\right) Y^{(t)} w\right| \leq \beta' \left\|w\right\|^2,$$

```
Given the constants $\gamma_1, \gamma_2$ and $\gamma_3$ (independent of $t$)
        with $0 < \gamma_1 < \gamma_2 \leq \gamma_3 < 1$
If $\nabla f\left(x^{(t)}\right) Y^{(t)} \widehat{w}^{(t)} \geq 0$
        Then set $\alpha^{(t)} = 0$
        Else
              Set $\alpha = 1$
              Repeat until explicitly stopped
                      if $f\left(x^{(t)} + \alpha Y^{(t)} \widehat{w}^{(t)}\right) \leq f\left(x^{(t)}\right) + \gamma_1 \alpha \nabla f\left(x^{(t)}\right) Y^{(t)} \widehat{w}^{(t)}$
                            Then choose $\alpha^{(t)} = \alpha$ and stop
                            Else reset $\alpha \in [\gamma_2 \alpha .. \gamma_3 \alpha]$ and continue
```

Figure 4: The stabilization algorithm.

one may use the reduced Hessian matrix ${Y^{(t)}}^{\mathrm{T}} \nabla^2 f\left(x^{(t)}\right) Y^{(t)}$ or a quasi-Newton approximation. To satisfy (3.22), non positive definite matrices may be perturbed via a modified Cholesky factorization [Gill and Murray 1974; Gill *et al* 1981, pp. 109–111]. Such a modification will not destroy the property (3.21), although the minimal value of β may increase.

Now suppose that we choose the step length $\alpha^{(t)}$ via the stabilization algorithm in figure 4 and choose

$$w^{(t)} = \alpha^{(t)} \widehat{w}^{(t)}.$$

Note that, by setting $\gamma_2 = \gamma_3 = 1/2$, this stabilization algorithm reduces to the familiar Armijo rule. The new primal point is produced by the rule

$$x^{(t+1)} = x^{(t)} + Y^{(t)} w^{(t)} = x^{(t)} + \alpha^{(t)} Y^{(t)} \widehat{w}^{(t)}.$$

We shall prove that limit points of the resulting sequence $\left\{x^{(t)}\right\}$ are KKT points.

Theorem 3.6 (Convergence of Decomposition Method) *Suppose that*

1. *f is essentially smooth and $\mathcal{B}$ is closed and convex in (3.1),*

2. *$x^{(0)} \in \mathcal{B}$ is given,*

3. *$y^{(t)}$ are solutions of the subproblems of the form (3.9), for the decoupled resource allocation R, where (3.10) and (3.11) hold,*

4. *$Y^{(t)}$ is defined by (3.14),*

5. *the sequence $H^{(t)}$ satisfies both (3.21) and (3.22),*

6. *$\widehat{w}^{(t)}$ is defined as the unique minimizer of $Q^{(t)}(w)$ (cf. (3.20)) subject to*

$$\underline{b} \leq x^{(t)} + Y^{(t)}w \leq \overline{b},$$

7. *$\alpha^{(t)}$ is computed via the stabilization algorithm of figure 4, and*

8. *$x^{(t+1)} = x^{(t)} + Y^{(t)}w^{(t)} = x^{(t)} + \alpha^{(t)}Y^{(t)}\widehat{w}^{(t)}$.*

Then

1. *the stabilization algorithm always terminates finitely, producing $\alpha^{(t)} \in [0..1]$,*

2. *the procedure produces an infinite sequence $\left\{x^{(t)}\right\} \subset \mathcal{B}$, and*

3. *either $f\left(x^{(t)}\right) \to -\infty$ or each $\tilde{x} \in \operatorname{lim\,pt} x^{(t)}$ is a KKT point for (3.1).*

Before proving the theorem, we note that it is possible that $\operatorname{lim\,pt} x^{(t)} = \emptyset$, in which case the sequence is unbounded. In practice one usually includes very large box constraints in the definition of the feasible region, so that limit points must exist. Note also that an essentially smooth function is bounded from below on any bounded set, so that if the sequence is bounded, not only must a limit point exist, but all limit points are KKT points.

Proof: 1. If the stabilization algorithm terminates, it does so with $\alpha^{(t)} \in [0..1]$, so we must show that the algorithm must terminate. Suppose the algorithm is in an endless loop, which means that $\nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)} < 0$, and the value of $\alpha > 0$

becomes arbitrarily small (because $\gamma_3 < 1$). In a neighborhood of 0, the function of α given by $f\left(x^{(t)} + \alpha Y^{(t)}\widehat{w}^{(t)}\right)$ is continuously differentiable, so that

$$f\left(x^{(t)} + \alpha Y^{(t)}\widehat{w}^{(t)}\right) - f\left(x^{(t)}\right) - \alpha \nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)} \in O\left(\alpha^2\right).$$

But $\gamma_1 < 1$, so that the algorithm is not in an endless loop and will eventually terminate with positive $\alpha^{(t)}$.

2. By induction, if $x^{(t)} \in \mathcal{B}$, then $\widehat{w}^{(t)}$ is chosen so that $x^{(t)} + Y^{(t)}\widehat{w}^{(t)} \in \mathcal{B}$. Therefore, since $\alpha^{(t)} \in (0..1)$, $x^{(t+1)} \in \mathcal{B}$.

3. The stabilization algorithm ensures that $f\left(x^{(t+1)}\right) \leq f\left(x^{(t)}\right)$. If $f\left(x^{(t)}\right) \to -\infty$, the minimizer is ill defined, so that we may assume that this is not the case. Then $\lim f\left(x^{(t+1)}\right) - f\left(x^{(t)}\right) = 0$ and the stabilization algorithm shows that

$$\liminf_{t\to\infty} \alpha^{(t)} \nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)} \geq 0.$$

Clearly, if $\alpha^{(t)}$ are bounded away from 0, then

$$\liminf_{t\to\infty} \nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)} \geq 0. \tag{3.23}$$

We shall show that (3.23) holds in any event by considering only subsequences for which $\alpha^{(t)} \to 0$. The stabilization algorithm shows that

$$f\left(x^{(t+1)}\right) > f\left(x^{(t)}\right) + \frac{\gamma_1}{\gamma_2}\alpha^{(t)} \nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)},$$

from which we may derive

$$\frac{f\left(x^{(t+1)}\right) - f\left(x^{(t)}\right)}{\alpha^{(t)}} > \frac{\gamma_1}{\gamma_2} \nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)}. \tag{3.24}$$

We may further thin this subsequence so that $x^{(t)} \to \tilde{x}$, $Y^{(t)} \to \tilde{Y}$ and $\widehat{w}^{(t)} \to \tilde{w}$ (the decoupled resource allocation bounds the $Y^{(t)}$ sequence and the boundedness of $H^{(t)}$ and $x^{(t)}$ sequences bounds the sequence of $\widehat{w}^{(t)}$). Because $\alpha^{(t)} \to 0$, (3.24) shows that

$$\nabla f\left(\tilde{x}\right) \tilde{Y}\tilde{w} \geq \underbrace{\frac{\gamma_1}{\gamma_2}}_{<1} \nabla f\left(\tilde{x}\right) \tilde{Y}\tilde{w},$$

which implies that $\nabla f(\tilde{x})\tilde{Y}\tilde{w} = 0$ and proves that (3.23) holds in any case. We claim that

$$\liminf_{t\to\infty} \nabla f\left(x^{(t)}\right) Y^{(t)}\mathbf{1} \geq 0 \tag{3.25}$$

too. To prove the claim, suppose that it is not true and we shall derive a contradiction. Assume that $\lim\operatorname{pt} x^{(t)} \neq \emptyset$, for otherwise the proof is complete. We thin the sequence so that

$$\begin{aligned} x^{(t)} &\longrightarrow \tilde{x} && \text{by assumption} \\ y^{(t)} &\longrightarrow \tilde{y} && \text{by boundedness imposed by resource allocation} \\ Y^{(t)} &\longrightarrow \tilde{Y} && \text{by definition in terms of } x^{(t)} \text{ and } y^{(t)} \\ H^{(t)} &\longrightarrow \tilde{H} && \text{by (3.21)} \end{aligned}$$

and

$$\lim_{t\to\infty} \nabla f\left(x^{(t)}\right) Y^{(t)}\mathbf{1} = \nabla f(\tilde{x})\tilde{Y}\mathbf{1} = -\zeta < 0.$$

Now choose $\lambda \in (0..1]$ (independent of t) so that

$$\lambda\nabla f(\tilde{x})\tilde{Y}\mathbf{1} + \lambda^2\mathbf{1}\tilde{H}\mathbf{1} < -\frac{\lambda\zeta}{2}.$$

Then $\lambda\mathbf{1}$ is feasible for the master problem (3.15), so that the definition of $\widehat{w}^{(t)}$ gives, for t sufficiently large and in the subsequence under consideration,

$$\nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)} + \widehat{w}^{(t)\top}H^{(t)}\widehat{w}^{(t)} \leq \nabla f\left(x^{(t)}\right) Y^{(t)}(\lambda\mathbf{1}) + (\lambda\mathbf{1})H^{(t)}(\lambda\mathbf{1}) < -\frac{\lambda\zeta}{4} < 0.$$

This contradicts (3.23) and proves the claim (3.25). Since $Y^{(t)}\mathbf{1} = y^{(t)} - x^{(t)}$, we may apply theorem 3.3 for every $\tilde{x} \in \lim\operatorname{pt} x^{(t)}$ to see that $\tilde{x}$ is a KKT point.

♠

Chapter 4

Computational Results

This chapter describes our test problems and documents the results produced by our decomposition algorithm. In §4.1 we describe the problems that motivated this decomposition algorithm. These problems are linear, multicommodity network flow problems, so that c and D are linear functions. The results of the previous chapters are, therefore, somewhat more general than we needed for solving these problems, but may be applied to other important problems. §4.2 describes our code and the key parameters used. Finally, §4.3 shows the timing results on the large-scale problems.

4.1 The Patient Distribution System (PDS)

The test problems we used were obtained from the CINCMAC analysis group of the Military Airlift Command (MAC) at Scott Air Force Base [Chmielewski 1989]. The model is called the Patient Distribution System (PDS) and is a logistics model designed to help make decisions about how well MAC can evacuate patients from Europe. The PDS problems are a class of problems; PDS-$\mathcal{D}$ denoting the problem that models a scenario lasting $\mathcal{D}$ days, for integers $\mathcal{D} \in [1..85]$. As $\mathcal{D}$ becomes larger, the size of PDS-$\mathcal{D}$ grows quite large, as may be seen in table 1. The PDS problems are linear multicommodity network flow problems, which are block angular linear programs where each A_k is a node-arc-incidence matrix. (See figure 5.)

These PDS problems have received considerable attention lately, partly because they are a real-world application, and partly because they seem to be quite challenging. Carolan *et al* [1990] used the KORBX system at Scott AFB to solve numerous problems, including some of the smaller PDS problems. It took the KORBX system (using default parameters) between 3.3 hours and 4.5 hours to solve PDS-10. Only one out of the four KORBX codes finished within 24 hours on PDS-20. Setiono [1990] has solved small and medium-sized PDS problems using a dual proximal point linear programming algorithm, solving the resulting linear systems with the preconditioned conjugate gradient method. Setiono solved PDS-20 in 25.5 hours on an Astronautics ZS-1 computer. (For reference we note that Smith and Klinger [1988] claim that the ZS-1 achieves 3.0 Mflops on the 24 Livermore loops and 6.3 Mflops on the 100×100 Linpack benchmark.) Using a decomposition technique, Pinar and Zenios [1990] solved many of the PDS problems. The largest problem they report on is PDS-30, taking slightly more than 2 hours to solve on a CRAY Y-MP. Cheng *et al* [1989] report the solution of problems as large as PDS-50 in 10.2 hours using the KORBX system (apparently not the same version as Carolan *et al* [1990] use). In unpublished work, De Leone [1990] has solved PDS-40 using an SOR-based technique in approximately 27 hours on a DECstation 3100.

4.2 Description of the Code and Parameters

We ran our code on two machines: a DECstation 3100 running the ULTRIX operating system, and a 20 processor Sequent Symmetry S81 running the DYNIX operating system. Most of the code was written in C, with the portion used to solve the network subproblems being written in FORTRAN (see further discussion below). The C programming language was chosen primarily because of its ability to work properly with modern data structures. The code was compiled with the default code optimization (`-O1`). Double precision was used for all calculations.

To construct the barrier problem approximations (§2) we used the logarithmic barrier function ($\rho_j(\xi) = -\ln \xi$ for all j) and the following parameter values (see figure 2):

- $\Theta = 1$ cf. (2.13)
  Assumes right-hand-side of moderate size ($d_j \in [-10^6, 10^6]$, for example).

- $\lambda_\theta = 0.9$ cf. (2.14)
  We found that if this parameter is smaller, the method does not find feasible solutions as quickly. Such seemingly large values of λ_θ do not seem to cause numerical problems for our test problems.

- $\tau^1 = 10$ cf. (2.23)
  We normalize the cost coefficients so that $\|c\|_\infty = 1$, making τ^1 ten times the maximum absolute value of the cost coefficients.

- $\tau_{\text{inf}}J = 10^{-8}$ cf. (2.23) and theorems 2.18 and 2.20
  Eight places of accuracy in the objective function is a fairly ambitious goal for problems as large as our test problems. This was not always achieved in 50 iterations.

- $\lambda_\tau = 0.5$ cf. (2.23)
  Smaller values tend to introduce the problems of ill-conditioning earlier, so that convergence is hampered. Larger values require too many iterations until τ^i is sufficiently small.

The code takes full advantage of the network structure of the A_k, since solving multicommodity networks was the initial goal of this work. The code also takes advantage of the special structure of the matrix D. Typically for multicommodity networks, a constraint $D_{j\bullet}x \leq d_j$ represents a physical situation where the flow of a given "topological arc" (an arc appearing at most once in each commodity) can only handle a certain capacity of flow. In this case ${D_{[k]}}_{j,n} \in \{0, 1\}$, and for each j

and k, there is at most one $\hat{n}$ such that $D_{[k]_{j,\hat{n}}} = 1$. Our code has a $J \times K$ array of pointers to this $\hat{n}$ (it stores 0 if no such $\hat{n}$ exists). This saves space because J and K are usually much smaller than N.

For our decoupled resource allocation $R(x)$ defined in §3.1, we used (3.7) with the following parameter values:

- $\Omega = 10^{-8}$ cf. (3.7)
  Somewhat to our surprise, we found that using smaller values for Ω seemed to make the algorithm perform better. Using a smaller Ω means that we let the current point get closer to the boundary of the trust region, possibly at the expense of being able to move less within the null space of D. We found that $\Omega = 0$ (in which case our convergence proof fails) does work in practice (on these problems).

- $\omega = 0.7$ cf. (3.7)
  We found that $\omega \in [0.5..0.9]$ worked reasonably well.

- $\kappa = 10^{10}$ cf. (3.7)
  This bound on $\left\| y^{(i,t)} - x^{(i,t)} \right\|_\infty$ turned out to be redundant on these problems.

To solve the network subproblems we use a modified version of RNET, written in FORTRAN by Grigoriadis and Hsu [1979]. RNET is an implementation of the network simplex method. (Lustig [1990] shows that the C programming language is typically much better for network codes because of the lack of pointer arithmetic in FORTRAN.) This code was modified by the authors to work in double precision rather than integer arithmetic, and to use parameters to specify input data. Because of the addition of the decoupled resource allocation $R(x)$, each subproblem is a network with the cost *and* the bounds modified. Since no primal or dual feasible solution is immediately available from the previous iteration, we begin RNET with an all artificial basis at each iteration. The parameters given to RNET are mostly determined by the suggestions in Grigoriadis and Hsu [1979]. We allow for a large number of pivots.

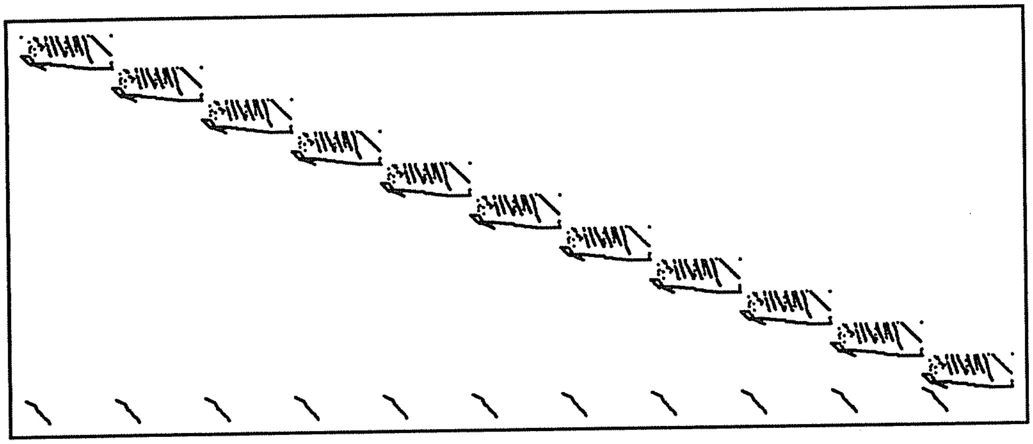


Figure 5: Sparsity structure of the constraint matrix for PDS-01.

The coordinator algorithm uses an active set method in conjunction with the steepest descent direction. We stop when both the function and the norm of the projected gradient have been sufficiently decreased, as is explained at the end of §3.2. The code uses $\mu = 0.4$ and $\mu' = 0.03$. If we stop when the function values have been sufficiently decreased, but ignore the projected gradient, then the method converges in theory, but in practice it seems somewhat problematic. Using a larger μ' would allow the algorithm to terminate, when in fact the coordinator could probably find a significantly better point at low cost. We run the coordinator algorithm for at most 15 iterations within each major iteration.

As part of this coordinator method we need to use a one-dimensional line search method. The one-dimensional line search algorithm we used is (2.6.4) in Fletcher [1987]. Special structure of the objective function allows us to use Newton's method in place of the usual minimization of a quadratic or cubic interpolant. Parameters were set to attain a line search of medium accuracy.

4.3 Results of Numerical Experiments

We shall now present performance results of our codes on a subset of the PDS problems. We were interested in two things when beginning these tests. First we

Problem	max size of block		coupling	dimension of A	
Name	$\max_k M(k)$	$\max_k N(k)$	J	M	N
PDS-01	126	339	87	1,386	3,729
PDS-02	252	685	181	2,772	7,535
PDS-03	390	1117	303	4,290	12,287
PDS-05	686	2,149	553	7,546	23,639
PDS-06	835	2,605	696	9,185	28,655
PDS-10	1,399	4,433	1,169	15,389	48,763
PDS-20	2,857	10,116	2,447	31,427	105,728
PDS-30	4,223	15,126	3,491	46,453	154,998
PDS-40	5,652	20,698	4,672	62,172	212,859
PDS-50	7,031	26,034	5,719	77,341	270,095
PDS-60	8,423	31,474	6,778	92,653	329,643
PDS-70	9,750	36,262	7,694	107,250	382,311

Each PDS problem has eleven blocks (i.e. $K = 11$).

Table 1: Sizes of some of the PDS problems.

wanted to develop algorithms that compute approximate solutions to multicommodity network flow problems quickly. Second we want our method to compute accurate solutions. Although our solution is primal feasible (always $\underline{b} \leq x \leq \overline{b}$, $Dx < d$ and $\|Ax - a\|_\infty / \|a\|_\infty \approx$ machine epsilon), our method does not provide good bounds on the gap between the objective value obtained and the optimal value. We compute duals (see (1.12)) on the network constraints and the dual estimates $p^{(i)}$ (defined in (2.16)), and these will give convergent lower bounds as in theorems 2.11 and 2.20. When we have computed these lower bounds, however, they are not even as good as the lower bound given by solving the relaxed problem (2.1). We believe that this occurs for two reasons. First, we are using linearizations for the subproblems and the $p^{(i)}$ may converge very slowly; a suitable quadratic approximation may improve this situation, but in the interest of speed we prefer solving the linear problems. Second, an inaccurate $p^{(i)}$ will lead to inaccurate estimate $r^{(i)}$ for the optimal r in (1.18). Since the bounds b are quite large, the value of (1.18) may differ significantly from its optimum. However, when we compare our final objective values with those obtained by others, they typically match to between 5 and 7 significant figures. We further point out that highly accurate solutions are not needed for planning models like the PDS problems, where the models are already very rough approximations of reality. This may be contrasted to problems in mathematics [Varga 1990] where very accurate models require highly accurate floating point approximations of the solution.

We ran the code for 50 iterations in all cases. Figures 6, 8 and 10 show the original objective functions $cx^{(i)}$ as a function of iteration i for three of the larger PDS problems. Note that the objective function increases during the FEASIBILITY PHASE and then decreases steadily during the REFINE PHASE. Figures 7, 9 and 11 graph the improvement

$$\kappa^{(i)} := \begin{cases} -\log_{10}\left(\dfrac{cx^{(i-1)} - cx^{(i)}}{|cx^{(i)}|}\right) & \text{if } cx^{(i-1)} > cx^{(i)} \\ \\ \text{undefined} & \text{otherwise .} \end{cases}$$

as a function of the iteration i for the same three problems. (One might say that "the $\kappa^{(i)}$th digit of the objective function changed from iteration $i-1$ to iteration i.") These figures show that the improvement in the objective function $cx^{(i)}$ tends to become small after 50 iterations. The large improvement for PDS-60 in iteration 49 illustrates that convergence is not always predictable from objective functions alone. The primal objective function values we obtained matched the optimal values of De Leone [1990] and Setiono [1990] to between 5 and 7 digits.

Tables 2, 3 and 4 contain timings and optimal objective function values for the PDS problems we solved. The column of the table labeled "Relaxed" contains statistics for computing the solution to (2.1). The column labeled "Feasible" contains statistics for computing a feasible point. The number of iterations required to obtain a feasible solution via the shifted barrier approach varied between 11 and 16. The column labeled "Final" contains statistics for computing the final approximation of the optimal solution. The row of the table labeled "Total iterations" is the total number of iterations the method has taken to attain a given phase. The row labeled "Objective $\times 10^{-10}$" is the value of $10^{-10}cx^{(i)}$ where $x^{(i)}$ is the current point and c is the *original* cost vector ($\|c\|_\infty$ is *not* necessarily 1). A row labeled "DECstation" shows the performance on the DECstation 3100 . A row labeled "Sequent($\wp$)" shows the performance on the Sequent Symmetry using $\wp$ processors. All times reported are wall clock time.

The version of the code on the Sequent Symmetry uses the most obvious parallel strategy; at each iteration a separate subproblem for each commodity is solved on a separate processor. While there are other possibilities for parallelism in the program (e.g. parallel function evaluation in the coordinator problem or overlapping coordinator and subproblems in a chaotic fashion as in §5), most of the work is done in solving the (large scale) subproblems. Speedups of 4 or 5 are typical with 11 processors (and 11 commodities). This corresponds to perfect (linear) speedup of the subproblem solutions if between 80% and 90% of the work is done in solving the subproblems.

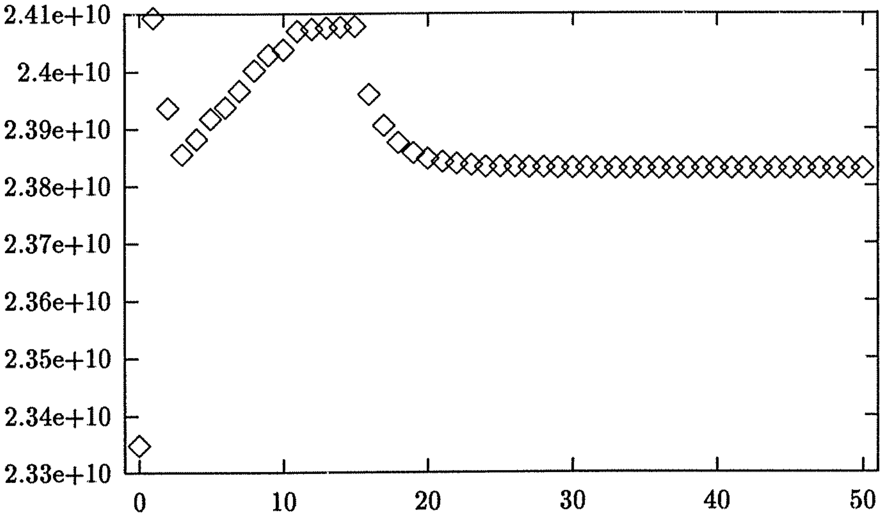


Figure 6: PDS-20: objective function vs. iterations.

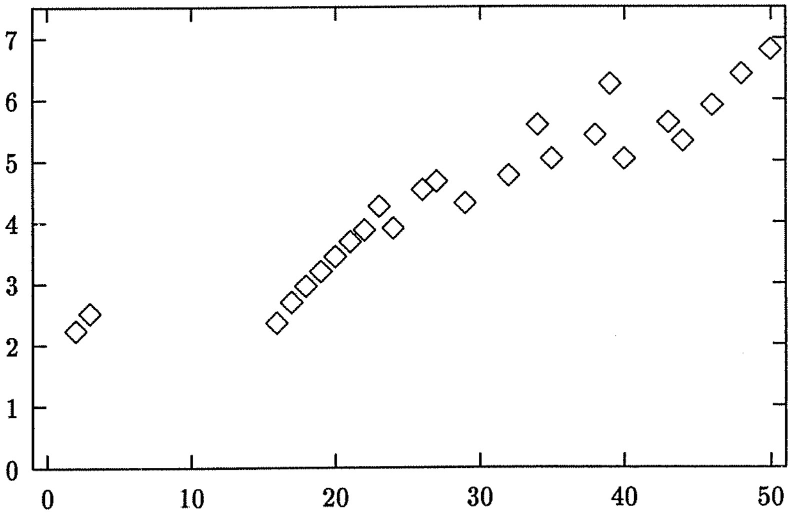


Figure 7: PDS-20: −log of improvement (κ) vs. iterations.

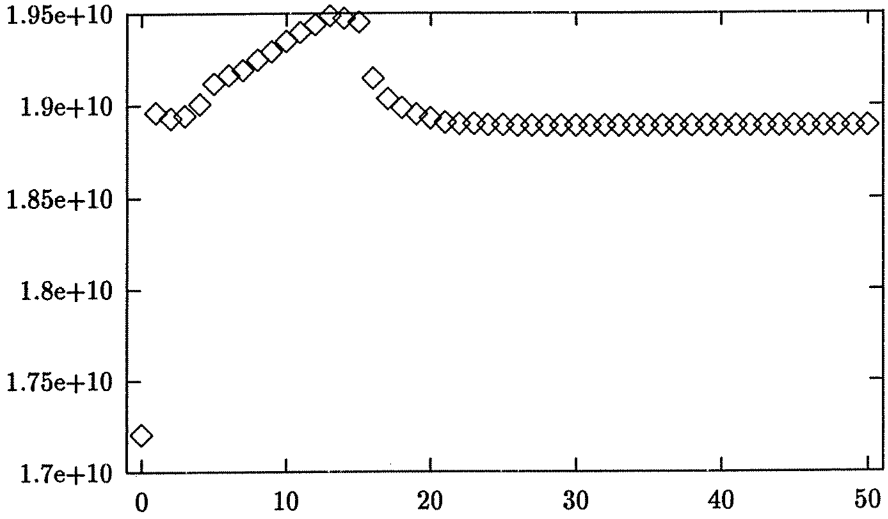


Figure 8: PDS-40: objective function vs. iterations.

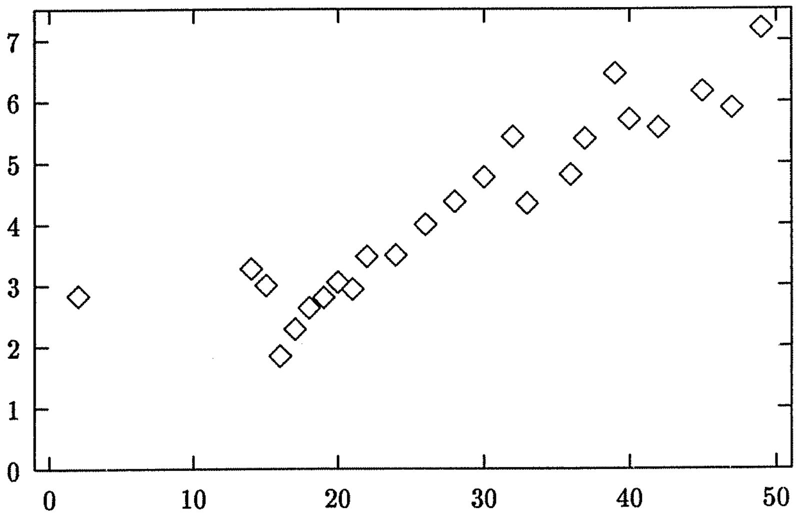


Figure 9: PDS-40: $-\log$ of improvement (κ) vs. iterations.

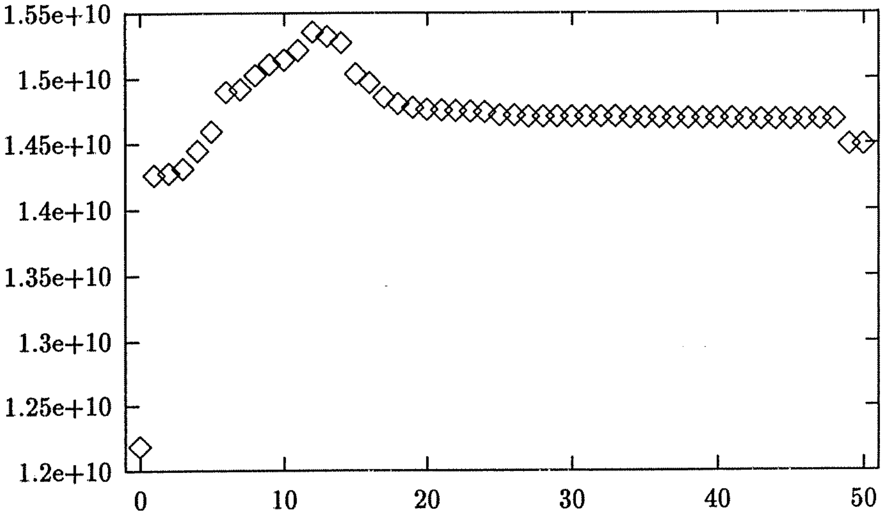


Figure 10: PDS-60: objective function vs. iterations.

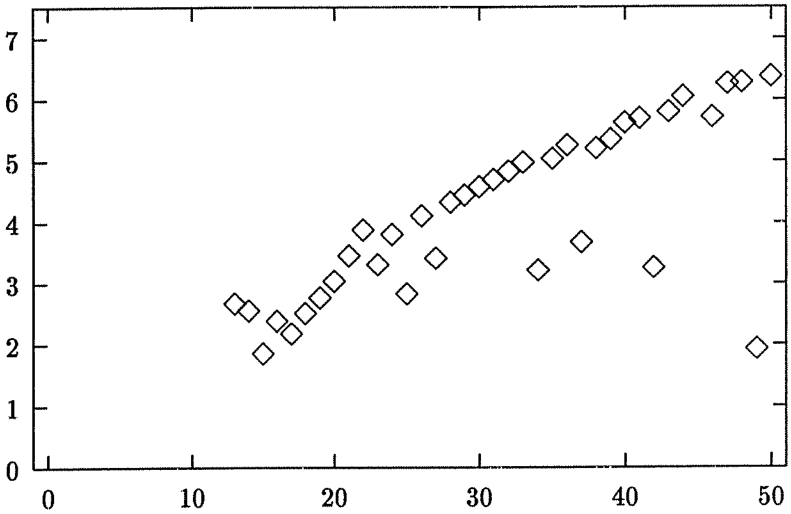


Figure 11: PDS-60: −log of improvement (κ) vs. iterations.

PDS-01			
Phase	Relaxed	Feasible	Final
Total iterations	0	11	50
Objective $\times 10^{-10}$	2.9033	2.9096	2.9084
DECstation	1.4sec	18sec	1min 30sec
Sequent(11)	1.1sec	12sec	1min 4sec
PDS-02			
Phase	Relaxed	Feasible	Final
Total iterations	0	12	50
Objective $\times 10^{-10}$	2.8758	2.8876	2.8858
DECstation	2.9sec	42sec	3min 17sec
Sequent(11)	1.5sec	22sec	2min 9sec
PDS-03			
Phase	Relaxed	Feasible	Final
Total iterations	0	13	50
Objective $\times 10^{-10}$	2.8442	2.8622	2.8597
DECstation	4.7sec	1min 20sec	5min 47sec
Sequent(11)	2sec	39sec	3min 38sec
PDS-05			
Phase	Relaxed	Feasible	Final
Total iterations	0	16	50
Objective $\times 10^{-10}$	2.7824	2.8125	2.8054
DECstation	15sec	3min 14sec	11min 28sec
Sequent(11)	4.2sec	1min 33sec	6min 43sec
PDS-06			
Phase	Relaxed	Feasible	Final
Total iterations	0	16	50
Objective $\times 10^{-10}$	2.7526	2.7846	2.7761
DECstation	17sec	4min 17sec	15min 4sec
Sequent(11)	4.6sec	1min 56sec	8min 44sec

Table 2: Timing and objective value results for small PDS problems.

PDS-10			
Phase	Relaxed	Feasible	Final
Total iterations	0	16	50
Objective $\times 10^{-10}$	2.6333	2.6857	2.6727
DECstation	30sec	8min 15sec	28min 31sec
Sequent(11)	9sec	3min 57sec	16min 39sec
PDS-20			
Phase	Relaxed	Feasible	Final
Total iterations	0	14	50
Objective $\times 10^{-10}$	2.3342	2.4069	2.3822
DECstation	2min 22sec	33min 19sec	2hr 12min
Sequent(11)	40sec	9min 44sec	50min 43sec
PDS-30			
Phase	Relaxed	Feasible	Final
Total iterations	0	12	50
Objective $\times 10^{-10}$	2.0284	2.1818	2.1390
DECstation	4min 38sec	1hr 9min	5hr 23min
Sequent(11)	1min 40sec	16min 43sec	1hr 48min
PDS-40			
Phase	Relaxed	Feasible	Final
Total iterations	0	14	50
Objective $\times 10^{-10}$	1.7188	1.9452	1.8866
DECstation	8min 53sec	2hr 39min	10hr 27min
Sequent(11)	3min 45sec	32min 32sec	2hr 54min

Table 3: Timing and objective value results for large PDS problems.

PDS-50			
Phase	Relaxed	Feasible	Final
Total iterations	0	13	50
Objective $\times 10^{-10}$	1.5002	1.7336	1.6625
DECstation	13min 28sec	3hr 50min	16hr 46min
Sequent(11)	4min 35sec	54min 2sec	5hr 30min
PDS-60			
Phase	Relaxed	Feasible	Final
Total iterations	0	13	50
Objective $\times 10^{-10}$	1.2159	1.5288	1.4462
DECstation	18min 40sec	5hr 27min	24hr 6min
Sequent(11)	5min 38sec	1hr 19min	6hr 55min
PDS-70			
Phase	Relaxed	Feasible	Final
Total iterations	0	16	50
Objective $\times 10^{-10}$	0.9309	1.3191	1.2311
Sequent(11)	8min 12sec	2hr 9min	9hr 24min

Table 4: Timing and objective value results for very large PDS problems.

Chapter 5

Multiple Updates and Asynchronous Parallel Methods

This chapter will extend the results of §3 to a more general setting in which many updates—computed with regard to different basepoints—are considered asynchronously. For example, if one has a computational environment that has $K + 1$ independent CPUs, one could assign the coordinator problem to CPU 0 and the kth subproblem to CPU k. The method that we develop in this chapter allows the programmer to let each CPU execute without waiting for any other CPU to finish execution, thus enhancing the load balance.

5.1 Subproblems and Current Information

The subproblems, as discussed in §3.1, remain of the form (3.9) where (3.10) and (3.11) hold and R is a decoupled resource allocation (cf. definition 3.1). Again, we let $y^{(t)}$ be the solution of (3.9), so that $y^{(t)}$ depends upon the *basepoint* $x^{(t)}$ used in the subproblem (3.6).

We say that a set

$$\Upsilon_{[k]}^{(t)} \subset \{0, \ldots, t\} \subset \mathbb{Z}$$

determines the **current information for the kth block at iteration** t if the information that the algorithm has stored for the kth block at iteration t is

$$\left\{ \ y_{[k]}^{(\ell)} \ \middle| \ \ell \in \Upsilon_{[k]}^{(t)} \ \right\}.$$

The assumption here is that the vectors $\left\{ \ y_{[k]}^{(\ell)} \ \middle| \ \ell \notin \Upsilon_{[k]}^{(t)} \ \right\}$ have been discarded to limit the memory usage of the algorithm. We shall let $v_{[k]}^{(t)}$ represent the number of elements of the current information set $\Upsilon_{[k]}^{(t)}$ and let $v^{(t)} := \sum_{k=1}^{K} v_{[k]}^{(t)}$. The results of this chapter subsume those of §3, which may be seen by letting $\Upsilon_{[k]}^{(t)} = \{t\}$ for all blocks k and iterations t.

We will need some assumptions on $\Upsilon_{[k]}^{(t)}$ to rule out cases where $\Upsilon_{[k]}^{(t)}$ consists entirely of information that is too old for all large t. This will be done by making another assumption in §5.2.2.

5.2 The Master Problem and Coordination

Define the $N_{[k]} \times v_{[k]}^{(t)}$ matrix $Y_{[k]}^{(t)}$ as that matrix having columns

$$\left\{ \ y_{[k]}^{(\ell)} - x_{[k]}^{(t)} \ \middle| \ \ell \in \Upsilon_{[k]}^{(t)} \ \right\},$$

the $N \times v^{(t)}$ matrix

$$Y^{(t)} := \begin{pmatrix} Y_{[1]}^{(t)} & & \\ & \ddots & \\ & & Y_{[K]}^{(t)} \end{pmatrix}. \tag{5.1}$$

Each column of this $N \times v^{(t)}$ matrix corresponds to a feasible search direction, non-zero in precisely one block. As in §3 we have $AY^{(t)} = \mathbf{0}$ for all t so that the constraints $Ax = a$ may be ignored in defining the master problem:

$$\underset{w}{\text{minimize}} \ \ f\left(x^{(t)} + Y^{(t)}w\right) \quad \text{subject to} \quad w \in F^{(t)}, \tag{5.2}$$

with the feasible set defined by

$$F^{(t)} := \left\{ \ w \ \middle| \ \underline{b} \leq x^{(t)} + Y^{(t)}w \leq \overline{b} \ \right\}. \tag{5.3}$$

Note that this is exactly the same master problem as (3.15) in §3.2 except that $Y^{(t)}$ is more general and the corresponding set of step-size variables w is allowed to have cardinality $v^{(t)} \geq K$. In the case where $v_{[k]}^{(t)} > 1$ for some k, the columns of $Y^{(t)}$ need not be complementary, implying that $F^{(t)}$ is not a box. This causes us to look at **inner approximations** of $F^{(t)}$, i.e., approximations $W^{(t)}$ of $F^{(t)}$ with $W^{(t)} \subset F^{(t)}$.

5.2.1 Inner Approximations for the Master

Clearly, $F^{(t)}$ is block separable, so that we may consider these constraints as K independent sets of constraints. However, (5.2) has the property that the constraint matrix $Y^{(t)}$ has unstructured blocks $Y_{[k]}^{(t)}$. We shall consider the structure of the constraints of (5.2) in order to approximate them with simple constraints useful in developing a suitable coordinator.

Proposition 5.1 *We may use a simple ratio test to compute (possibly infinite) bounds* $\underline{w}^{(t)} \leq \mathbf{0}$ *and* $\overline{w}^{(t)} \geq \mathbf{1}$ *such that*

$$\forall \left\{\ell = 1, \ldots, v^{(t)}\right\} \quad \left[\ \underline{w}_\ell^{(t)} \leq w_\ell \leq \overline{w}_\ell^{(t)} \quad \Longleftrightarrow \quad \underline{b} \leq x^{(t)} + Y_{\bullet\ell}^{(t)} w_\ell \leq \overline{b}\ \right]. \qquad (5.4)$$

Proof: The bounds are given by

$$\underline{w}_\ell^{(t)} = \max \left\{ \max_{\{n | Y_{n,\ell}^{(t)} > 0\}} \frac{\left(\underline{b} - x^{(t)}\right)_n}{Y_{n,\ell}^{(t)}} \ , \ \max_{\{n | Y_{n,\ell}^{(t)} < 0\}} \frac{\left(\overline{b} - x^{(t)}\right)_n}{Y_{n,\ell}^{(t)}} \right\}$$

and

$$\overline{w}_\ell^{(t)} = \min \left\{ \min_{\{n | Y_{n,\ell}^{(t)} > 0\}} \frac{\left(\overline{b} - x^{(t)}\right)_n}{Y_{n,\ell}^{(t)}} \ , \ \min_{\{n | Y_{n,\ell}^{(t)} < 0\}} \frac{\left(\underline{b} - x^{(t)}\right)_n}{Y_{n,\ell}^{(t)}} \right\}.$$

Note that, to show that $\underline{w}_\ell^{(t)} \leq \zeta \leq \overline{w}_\ell^{(t)}$ it suffices to show that $\zeta e_\ell \in F^{(t)}$, where e_ℓ is the ℓth unit vector. We then see that $\underline{w}^{(t)} \leq \mathbf{0}$ because $\mathbf{0} \in F^{(t)}$. Also, $\overline{w}^{(t)} \geq \mathbf{1}$ because the kth block of the right-hand set of inequalities in (5.4) with $w = e_\ell$ either reduces to

$$\underline{b}_{[k]} \leq x_{[k]}^{(t)} + y_{[k]}^{(i)} - x_{[k]}^{(t)} \leq \overline{b}_{[k]} \qquad \text{or} \qquad \underline{b}_{[k]} \leq x_{[k]}^{(t)} + \mathbf{0} \leq \overline{b}_{[k]},$$

(for some $i \in \Upsilon^{(t)}_{[k]}$) depending on whether column ℓ of $Y^{(t)}$ is or is not associated with the kth block. Either option is true because all $x^{(t)}$ and $y^{(t)}$ are contained in $\mathcal{B}$.

♠

The following theorem gives a class of useful inner approximations for the feasible region of (5.2).

Theorem 5.2 (Inner Approximation) *Suppose* $\underline{w}^{(t)}$ *and* $\overline{w}^{(t)}$ *are given by (5.4) in proposition 5.1. Then* w *is feasible for (5.2) (i.e.,* $w \in F^{(t)}$*) if*

$$\forall \left\{ \ell = 1, \ldots, v^{(t)} \right\} \qquad w_\ell = \lambda_\ell \underline{w}^{(t)}_\ell + \mu_\ell \overline{w}^{(t)}_\ell \qquad \textit{where} \tag{5.5}$$

$$\lambda \geq \mathbf{0} \ , \ \mu \geq \mathbf{0} \qquad \textit{and} \tag{5.6}$$

$$\forall \{k = 1, \ldots, K\} \qquad \mathbf{1} \left(\max \left\{ \lambda_{[k]}, \mu_{[k]} \right\} \right) \leq 1. \tag{5.7}$$

(The maximum in (5.7) is taken componentwise.)

Proof: We begin by defining

$$W^{(t)}_0 := \left\{ \ w \ \middle| \ \exists\{\lambda, \mu\} \quad \text{(5.5), (5.6) and (5.7) are satisfied} \quad \right\}.$$

Suppose that $w \in W^{(t)}_0$ with the corresponding vectors $\lambda, \mu \in \mathbb{R}^{v^{(t)}}$. If $\lambda_\ell = \mu_\ell = 0$, we have $w_\ell = 0$ and we define $w'_\ell = 0$. If, on the other hand, $\max\{\lambda_\ell, \mu_\ell\} > 0$, define $w'_\ell = w_\ell / \max\{\lambda_\ell, \mu_\ell\}$. Then, for $w'_\ell \neq 0$,

$$w'_\ell = \frac{\lambda_\ell \underline{w}^{(t)}_\ell + \mu_\ell \overline{w}^{(t)}_\ell}{\max\{\lambda_\ell, \mu_\ell\}} \left\{ \begin{array}{lcll} \geq & \dfrac{\lambda_\ell \underline{w}^{(t)}_\ell}{\max\{\lambda_\ell, \mu_\ell\}} & \geq & \underline{w}^{(t)}_\ell \\ \leq & \dfrac{\mu_\ell \overline{w}^{(t)}_\ell}{\max\{\lambda_\ell, \mu_\ell\}} & \leq & \overline{w}^{(t)}_\ell \end{array} \right.$$

which is to say that w' satisfies the simple bound constraints in (5.4). Therefore, by (5.4),

$$\forall \left\{ \ell = 1, \ldots, v^{(t)} \right\} \ \underline{b} - x^{(t)} \leq Y^{(t)}_{\bullet \ell} w'_\ell \leq \overline{b} - x^{(t)}.$$

After multiplying each of these by $\max\{\lambda_\ell, \mu_\ell\}$ and summing over ℓ, we get, for each block k

$$\underbrace{\left(\underline{b}_{[k]} - x^{(t)}_{[k]}\right)}_{\leq \mathbf{0}} \underbrace{\sum_{\ell=1}^{v^{(t)}_{[k]}} \max\left\{\lambda_{[k]_\ell}, \mu_{[k]_\ell}\right\}}_{\in[0..1]} \leq Y^{(t)}_{[k]} w_{[k]} \leq \underbrace{\left(\bar{b}_{[k]} - x^{(t)}_{[k]}\right)}_{\geq \mathbf{0}} \underbrace{\sum_{\ell=1}^{v^{(t)}_{[k]}} \max\left\{\lambda_{[k]_\ell}, \mu_{[k]_\ell}\right\}}_{\in[0..1]},$$

which implies that $w \in F^{(t)}$.

♠

Note that the converse of the theorem is false, for consider the case where $v^{(t)}_{[1]} = 2$ and the two columns $Y^{(t)}_{[1]\,\bullet 1}$ and $Y^{(t)}_{[1]\,\bullet 2}$ happen to be complementary. Then the constraint

$$\sum_{\ell=1}^{2} \max\left\{\lambda_{[1]_\ell}, \mu_{[1]_\ell}\right\} \leq 1$$

could be replaced by the less restrictive

$$\max\left\{\lambda_{[1]_1}, \mu_{[1]_1}\right\} \leq 1 \qquad \text{and} \qquad \max\left\{\lambda_{[1]_2}, \mu_{[1]_2}\right\} \leq 1.$$

The constraints (5.7) may be written as

$$\lambda \leq \gamma \ , \ \ \mu \leq \gamma \tag{5.8}$$

and

$$\forall\{k\} \ \ \mathbf{1}\gamma_{[k]} = 1. \tag{5.9}$$

Then the representation of theorem 5.2 is a system of $v^{(t)} + K$ simple equality constraints (one could eliminate the first $v^{(t)}$ by an easy change of variables), $2v^{(t)}$ nonnegativity constraints and $2v^{(t)}$ other inequality constraints in $4v^{(t)}$ variables.

If we add the slack variables $\tilde{\lambda}$ and $\tilde{\mu}$, then the constraint stytem for this is

$$\left.\begin{array}{r}
\begin{pmatrix} \mathbf{I} & -\operatorname{diag} \underline{w}^{(t)} & -\operatorname{diag} \overline{w}^{(t)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{I} & \mathbf{0} & -\mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{I} & -\mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & T^{(t)} \end{pmatrix} \begin{pmatrix} w \\ \lambda \\ \mu \\ \tilde{\lambda} \\ \tilde{\mu} \\ \gamma \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{pmatrix}, \\
\\
w \text{ unrestricted}, \quad \lambda \geq \mathbf{0}, \quad \mu \geq \mathbf{0}, \\
\gamma \geq \mathbf{0} \text{ (implicitly)}, \quad \tilde{\lambda} \geq \mathbf{0} \quad \text{and} \quad \tilde{\mu} \geq \mathbf{0},
\end{array}\right\} \tag{5.10}$$

where $T^{(t)} \in \{0,1\}^{K \times v^{(t)}}$ is derived from (5.9). E.g., if $K = 2$, $v_{[1]}^{(t)} = 2$ and $v_{[2]}^{(t)} = 3$, then

$$T^{(t)} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Special basis partitioning techniques or general sparse matrix techniques may be used to handle such constraints. In summary,

$$W_0^{(t)} = \left\{ \; w \;\middle|\; \exists\{\lambda, \mu, \tilde{\lambda}, \tilde{\mu}, \gamma\} \; \text{(5.10) is satisfied} \; \right\} \subset F^{(t)},$$

and the containment may be strict.

We may approximate the feasible region of (5.2) with *simple bounds* in the following way. Choose a $\gamma \in \mathbb{R}_+^{v^{(t)}}$ that satisfies (5.9). Then it is clear that w satisfies (5.5) for some λ and μ as in (5.6) and (5.8) iff

$$w \in W_0^{(t)}(\gamma) := \left\{ \; w \;\middle|\; \forall\left\{\ell = 1, \ldots, v^{(t)}\right\} \; \gamma_\ell \underline{w}_\ell^{(t)} \leq w_\ell \leq \gamma_\ell \overline{w}_\ell^{(t)} \; \right\}. \tag{5.11}$$

Note that $W_0^{(t)}(\gamma) \subset W_0^{(t)} \subset F^{(t)}$ if γ satisfies (5.9).

Any of these results will hold if $\underline{w}^{(t)}$ is replaced by $\mathbf{0}$ and $\overline{w}^{(t)}$ is replaced by $\mathbf{1}$. In particular, the product of simplices

$$\left\{ \; w \geq \mathbf{0} \;\middle|\; \forall\{k = 1, \ldots, K\} \; \mathbf{1} w_{[k]} \leq 1 \; \right\} \subset F^{(t)}. \tag{5.12}$$

These are a few of the more interesting inner approximations to the feasible region of (5.2).

5.2.2 The Culling Function

At a given iteration t and for each block k the columns of $Y_{[k]}^{(t)}$ contain information about updates $y_{[k]}^{(\ell)}$ for $\ell \in \Upsilon_{[k]}^{(t)}$. In order to prove convergence, we must guarantee that this $Y^{(t)}$ contains information about acceptable descent directions about the current point. This will be guaranteed by selecting or *culling* one special column in each block. This special column will give us the up-to-date information needed to show convergence. Now the technical details:

Definition 5.3 *We say that*

$$\eta : \{1, \ldots, K\} \times \mathbb{Z}_+ \to \mathbb{Z}_+ \tag{5.13}$$

is a **culling function** *if*

$$\forall\{k,t\} \;\; \eta(k,t) \in \Upsilon_{[k]}^{(t)} \tag{5.14}$$

and

$$\forall\{k\} \;\; \lim_{t\to\infty} \left\| x_{[k]}^{(t)} - x_{[k]}^{(\eta(k,t))} \right\| = 0. \tag{5.15}$$

Therefore, in each block we are choosing the information $y_{[k]}^{(\eta(k,t))}$, computed with respect to $x_{[k]}^{(\eta(k,t))}$ which becomes arbitrarily close to the current point $x_{[k]}^{(t)}$. Condition (5.15) is holds in certain reasonable situations.

Proposition 5.4 *Each of the following implies that (5.15) holds:*

1. $\left\{x^{(t)}\right\}$ *converges* *and* $\forall\{k\} \;\; \lim_{t\to\infty} \eta(k,t) = +\infty$

2. $\lim_{t\to\infty} \left\| x^{(t+1)} - x^{(t)} \right\| \to 0$ *and* $\exists\{T\}\forall\{k,t\} \;\; t - \eta(k,t) \leq T$

Proof: 1. If the sequence converges, it is Cauchy and so

$$\lim_{t,\eta(k,t)\to\infty} \left\| x_{[k]}^{(t)} - x_{[k]}^{(\eta(k,t))} \right\| = 0.$$

2.

$$\left\| x_{[k]}^{(t)} - x_{[k]}^{(\eta(k,t))} \right\| \leq \sum_{\ell=1}^{t-\eta(k,t)} \left\| x_{[k]}^{(t-\ell+1)} - x_{[k]}^{(t-\ell)} \right\| \leq \sum_{\ell=1}^{T} \left\| x_{[k]}^{(t-\ell+1)} - x_{[k]}^{(t-\ell)} \right\|,$$

which converges to 0 because T is a constant.

♠

The above proposition showed, in the first instance, that if the sequence converges and the culling function satisfies a "totally asynchronous" assumption [Bertsekas and Tsitsiklis 1989, §6.1], then (5.15) holds. In the second instance, if successive iterates become arbitrarily close, then a "partially asynchronous" assumption [Bertsekas and Tsitsiklis 1989, §7.1] suffices.

The following lemma will be used in the convergence theorem below.

Lemma 5.5 *Suppose that (5.15) holds and that $y^{(t)}$ solve subproblems of the form (3.9) with basepoint $x^{(t)}$ where (3.10) and (3.11) hold, and where the $G^{(t)}$ are block diagonal with blocks corresponding to the original blocks, i.e., no coupling whatsoever between blocks in the subproblems. Then, for t sufficiently large, the subproblem solutions with kth block given by $y_{[k]}^{(\eta(k,t))}$ solve subproblems of the form (3.9) with basepoint $x^{(t)}$ where (3.10) and (3.11) hold.*

The lemma says that, so long as $\left\| x_{[k]}^{(t)} - x_{[k]}^{(\eta(k,t))} \right\| \to 0$, the *old information* $y_{[k]}^{(\eta(k,t))}$ may play the role of *new information* computed relative to the current point $x_{[k]}^{(t)}$.

Proof: First, we see that the same decoupled resource allocation R may be used relative to the current point. Conditions 2 and 3 of definition 3.1 follow clearly from (5.15). The condition 3 and (5.15) then show that, for t large enough, $x^{(t)} \in R\left(\hat{x}^{(t)}\right)$, where $\hat{x}^{(t)}$ denotes the vector with kth block equal to $x_{[k]}^{(\eta(k,t))}$.

Now, we consider the objective function. For simplicity of notation, we let $\hat{g}^{(t)}$ denote that vector having kth block equal to $g_{[k]}^{(\eta(k,t))}$ and $\hat{G}^{(t)}$ denote that block-diagonal matrix having kth block equal to $G_{[k]}^{(\eta(k,t))}$. Using the word "constant" to denote any constant that does not depend on the argument y, we have

$$\begin{aligned} & g_{[k]}^{(\eta(k,t))} \left(y_{[k]} - x_{[k]}^{(\eta(k,t))}\right) + \left(y_{[k]} - x_{[k]}^{(\eta(k,t))}\right)^{\top} G_{[k]}^{(\eta(k,t))} \left(y_{[k]} - x_{[k]}^{(\eta(k;t))}\right) \\ = \; & \text{constant} + \left[g_{[k]}^{(\eta(k,t))} - 2 x_{[k]}^{(\eta(k,t))^{\top}} G_{[k]}^{(\eta(k,t))} \right] y_{[k]} + y_{[k]}{}^{\top} G_{[k]}^{(\eta(k,t))} y_{[k]} \\ = \; & \text{constant} + \left[g_{[k]}^{(\eta(k,t))} - 2 \left(x_{[k]}^{(\eta(k,t))} - x_{[k]}^{(t)} \right)^{\top} G_{[k]}^{(\eta(k,t))} \right] \left(y_{[k]} - x_{[k]}^{(t)} \right) \end{aligned}$$

$$+\left(y_{[k]} - x^{(t)}_{[k]}\right)^{\top} G^{(\eta(k,t))}_{[k]} \left(y_{[k]} - x^{(t)}_{[k]}\right).$$

Since these functions differ by a constant, their minimizers over the same constraint set are the same, so that $y^{(\eta(k,t))}_{[k]}$ minimizes the latter. Then the choice of $G^{(t)}_{[k]} = G^{(\eta(k,t))}_{[k]}$ satisfies (3.11) and the choice of

$$g^{(t)}_{[k]} = \left[g^{(\eta(k,t))}_{[k]} - 2\left(x^{(\eta(k,t))}_{[k]} - x^{(t)}_{[k]}\right)^{\top} G^{(\eta(k,t))}_{[k]}\right]$$

satisfies (3.10).

♠

We now discuss the culling function's relationship with the inner approximations $W^{(t)}$. Suppose that we have a culling function η and a constant $\xi > 0$ (independent of t). We define the line segment $\Xi^{(t)}$ blockwise by

$$\Xi^{(t)}_{[k]} := \left\{ \ x_{[k]} = x^{(t)}_{[k]} + \lambda\left(y^{(\eta(k,t))}_{[k]} - x^{(t)}_{[k]}\right) \ \middle| \ 0 \le \lambda \le \xi \ \right\}. \tag{5.16}$$

We are interested in inner approximations $W^{(t)}$ that satisfy both

$$\Xi^{(t)} \subset W^{(t)} \subset F^{(t)} \tag{5.17}$$

and

$$\exists\{\,\beta\,\}\forall\left\{\,t\,,\ w \in W^{(t)}\,\right\}\ \left\|Y^{(t)}w\right\| \le \beta. \tag{5.18}$$

We shall consider modifications of the inner approximations of §5.2.1 for this purpose.

First, consider the inner approximation $W^{(t)}_0$ given by theorem 5.2. It may be shown that, by taking w equal to any unit vector e_ℓ, we may produce a point $(w, \gamma, \lambda, \mu)$ feasible for (5.10) (cf. the proof of proposition 5.1). Therefore, by choosing $\xi = 1$, for any $w \in \Xi^{(t)}$, there exists (γ, λ, μ) such that $(w, \gamma, \lambda, \mu)$ is feasible for (5.10). Therefore, the choice $W^{(t)} = W^{(t)}_0$ satisfies (5.17). To show that (5.18) holds, We may need to further restrict $W^{(t)}$ by modifying $\underline{w}^{(t)}$ and $\overline{w}^{(t)}$. This could be done by assuring that

$$w \in W^{(t)} \implies -\frac{\beta'}{\left\|Y^{(t)}_{\bullet\ell}\right\|} \le w_\ell \le \frac{\beta'}{\left\|Y^{(t)}_{\bullet\ell}\right\|} \tag{5.19}$$

for columns $Y_{\bullet\ell}^{(t)}$ not chosen by the culling function. Note that (5.15) and condition 2 of definition 3.1 imply that the vectors $y_{[k]}^{(\eta(k,t))} - x_{[k]}^{(t)}$ are bounded independently of t. Therefore, the bound

$$w \in W^{(t)} \implies -\beta' \le w_\ell \le \beta'$$

will suffice for columns $Y_{\bullet\ell}^{(t)}$ that are chosen by the culling function. This final point is important because, if these artificial bounds were allowed to restrict w_ℓ too much, the first containment of (5.17) might no longer hold.

Second, consider the simple bound inner approximation $W_0^{(t)}(\gamma)$ of (5.11). Say that $\eta(k,t)$ corresponds to column $\ell(k,t)$ of $Y^{(t)}$. Then it is easy to see that (5.17) holds for the choice $W^{(t)} = W_0^{(t)}(\gamma)$ (where γ satisfies (5.9)) so long as

$$\forall\{t,k\} \;\; \xi \le \gamma_{\ell(k,t)}^{(t)} \overline{w}_{\ell(k,t)}^{(t)}.$$

A similar modification made to $\underline{w}^{(t)}$ and $\overline{w}^{(t)}$ may again be done to satisfy (5.18).

As the third possibility, consider the inner approximation given by the product of simplices (5.12). If the maximal cardinality of the information sets is bounded uniformly, i.e., if

$$\exists\{\delta\}\forall\{k,t\} \;\; \left|\Upsilon_{[k]}^{(t)}\right| \le \delta,$$

then we may let $\xi = 1/\delta$ to force $\Xi^{(t)}$ to be contained in the inner approximation (5.12). Artificial bounds of the form (5.19) will need to be imposed for columns $Y_{\bullet\ell}^{(t)}$ that grow arbitrarily large.

5.2.3 Asynchronous Coordination using a Stabilization Method

This section generalizes the results of §3.2.2. In particular, we use the new definitions of $Y^{(t)}$ and $\Xi^{(t)}$ and make use of the convex inner approximation $W^{(t)}$ that satisfies (5.17) and (5.18). Suppose that $Q^{(t)}$ is defined by (3.20) where (3.21) and (3.22) are satisfied. Then we define

$$\widehat{w}^{(t)} = \arg\min\left\{ \; Q^{(t)}(w) \;\middle|\; w \in W^{(t)} \; \right\}. \tag{5.20}$$

```
Given the constants $\gamma_1, \gamma_2$ and $\gamma_3$ (independent of $t$)
    with $0 < \gamma_1 < 1$ and $0 < \gamma_2 \leq \gamma_3 < 1$
If $\nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)} \geq 0$
    Then set $\alpha^{(t)} = 0$
    Else
        Set $\alpha = 1$
        Repeat until explicitly stopped
            if $f\left(x^{(t)} + \alpha Y^{(t)}\widehat{w}^{(t)}\right) - f\left(x^{(t)}\right)$
                $\leq \gamma_1 \left[\alpha \nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)} - \left\|\alpha Y^{(t)}\widehat{w}^{(t)}\right\|^2\right]$
                Then choose $\alpha^{(t)} = \alpha$ and stop
                Else reset $\alpha \in [\gamma_2\alpha..\gamma_3\alpha]$ and continue
```

Figure 12: The stabilization algorithm for the asynchronous case.

(The argmin is unique by (3.22).) The stepsize $\alpha^{(t)}$ is computed by the stabilization algorithm in figure 12. We will show that this converges subsequentially under reasonable assumptions.

Theorem 5.6 (Partially Asynchronous Convergence) *Suppose that*

1. *f is essentially smooth and $\mathcal{B}$ is closed and convex in (3.1),*

2. *$x^{(0)} \in \mathcal{B} \cap \operatorname{dom} f$ is given,*

3. *$y^{(t)}$ are solutions of the subproblems of the form (3.9) relative to the base-points $x^{(t)}$, for the decoupled resource allocation R, where (3.10) and (3.11) hold,*

4. *$Y^{(t)}$ is defined by (5.1),*

5. *η satisfies (5.13) and (5.14)*

6. *there is some T for which $\forall\{k,t\}$ $t - \eta(k,t) \leq T$,*

7. $\Xi^{(t)}$ is defined as in (5.16),

8. $W^{(t)}$ satisfies (5.17) and (5.18),

9. the sequence $H^{(t)}$ satisfies, both (3.21) and (3.22),

10. $\widehat{w}^{(t)}$ is defined by (5.20)

11. $\alpha^{(t)}$ is computed via the stabilization algorithm of figure 12, and

12. $x^{(t+1)} = x^{(t)} + Y^{(t)}w^{(t)} = x^{(t)} + \alpha^{(t)}Y^{(t)}\widehat{w}^{(t)}$.

Then

1. the stabilization algorithm always terminates finitely, producing $\alpha^{(t)} \in [0..1]$,

2. the procedure produces an infinite sequence $\left\{x^{(t)}\right\} \subset \mathcal{B}$, and

3. the limit $\tilde{f} := \lim_{t\to\infty} f\left(x^{(t)}\right)$ exists.

Moreover, if $\tilde{f}$ is finite, then

4. $\lim_{t\to\infty} \left\|x^{(t+1)} - x^{(t)}\right\| = 0$,

5. η is a culling function, and

6. each $\tilde{x} \in \lim\text{pt}\, x^{(t)}$ is a KKT point for (3.1).

The theorem is called partially asynchronous because of assumption (6) [Bertsekas and Tsitsiklis 1989, §7.1].

Proof: 1. Assume that the stabilization algorithm is in the loop. We know that

$$\left\|\alpha Y^{(t)}\widehat{w}^{(t)}\right\|^2 + f\left(x^{(t)} + \alpha Y^{(t)}\widehat{w}^{(t)}\right) - f\left(x^{(t)}\right) - \alpha\nabla f\left(x^{(t)}\right) Y^{(t)}\widehat{w}^{(t)} \in O\left(\alpha^2\right)$$

by Taylor's theorem. Therefore, with $\gamma_1 < 1$, the stabilization algorithm will terminate for α sufficiently small.

2. By induction, if $x^{(t)} \in \mathcal{B}$, then $\widehat{w}^{(t)} \in W^{(t)} \subset F^{(t)}$ so that $x^{(t)} + Y^{(t)}\widehat{w}^{(t)} \in \mathcal{B}$. Therefore, since $\alpha^{(t)} \in [0..1]$, $x^{(t+1)} \in \mathcal{B}$.

3. The stabilization algorithm shows that $f\left(x^{(t+1)}\right) \leq f\left(x^{(t)}\right)$, so that either $f\left(x^{(t)}\right) \to -\infty$ or $\left\{f\left(x^{(t)}\right)\right\}$ is bounded below. In either case $\tilde{f}$ exists.

4. The rest of the proof assumes that $\tilde{f}$ is finite. For the subsequence characterized by $\alpha^{(t)} = 0$ the result follows from hypothesis (8). If $\alpha^{(t)} > 0$, then the stabilization algorithm guarantees that

$$0 < \gamma_1 \leq \frac{f\left(x^{(t)}\right) - f\left(x^{(t)} + \alpha^{(t)}Y^{(t)}\widehat{w}^{(t)}\right)}{\left\|\alpha^{(t)}Y^{(t)}\widehat{w}^{(t)}\right\|^2 - \alpha^{(t)}\nabla f\left(x^{(t)}\right)Y^{(t)}\widehat{w}^{(t)}}. \tag{5.21}$$

Since $\tilde{f}$ is finite, the numerator of the fraction in (5.21) converges to 0, and so the denominator also converges to 0. Since the denominator is the sum of two positive terms, both terms converge to zero. That is to say that

$$\alpha^{(t)}Y^{(t)}\widehat{w}^{(t)} = x^{(t+1)} - x^{(t)} \to \mathbf{0} \tag{5.22}$$

and

$$-\alpha^{(t)}\nabla f\left(x^{(t)}\right)Y^{(t)}\widehat{w}^{(t)} \to 0.$$

5. We see that η is a culling function from part 2 of proposition 5.4.

6. The rest of the proof assumes that there is a limit point of $\{x^{(t)}\}$, for otherwise the theorem is proven. We will first show that

$$\liminf_{t\to\infty} \nabla f\left(x^{(t)}\right)Y^{(t)}\widehat{w}^{(t)} \geq 0. \tag{5.23}$$

The stabilization algorithm shows that (5.23) is true for the subsequence characterized by $\alpha^{(t)} = 0$. For all subsequences σ for which $\lim \alpha^{(\sigma(t))} > 0$, (5.23) is implied by (5.22). Consider, therefore a subsequence σ for which $\alpha^{(\sigma(t))} > 0$ and $\alpha^{(\sigma(t))} \to 0$. Then the stabilization algorithm shows that

$$\frac{f\left(x^{(\sigma(t))}\right) - f\left(x^{(\sigma(t))} + \dfrac{\alpha^{(\sigma(t))}}{\gamma^{(\sigma(t))}}Y^{(\sigma(t))}\widehat{w}^{(\sigma(t))}\right)}{\left\|\dfrac{\alpha^{(\sigma(t))}}{\gamma^{(\sigma(t))}}Y^{(\sigma(t))}\widehat{w}^{(\sigma(t))}\right\|^2 - \dfrac{\alpha^{(\sigma(t))}}{\gamma^{(\sigma(t))}}\nabla f\left(x^{(\sigma(t))}\right)Y^{(\sigma(t))}\widehat{w}^{(\sigma(t))}} \leq \gamma_1, \tag{5.24}$$

for some $\gamma^{(t)} \in [\gamma_2..\gamma_3]$. If $\lim \alpha^{(\sigma(t))} = 0$, then (5.24) shows that

$$\begin{aligned} & \frac{f\left(x^{(\sigma(t))}\right) - f\left(x^{(\sigma(t))} + \frac{\alpha^{(\sigma(t))}}{\gamma^{(\sigma(t))}} Y^{(\sigma(t))} \widehat{w}^{(\sigma(t))}\right)}{\alpha^{(\sigma(t))}} \\ \leq \; & \frac{\gamma_1}{\gamma^{(\sigma(t))}} \left[\frac{\alpha^{(\sigma(t))}}{\gamma^{(\sigma(t))}} \left\| Y^{(\sigma(t))} \widehat{w}^{(\sigma(t))} \right\|^2 - \nabla f\left(x^{(\sigma(t))}\right) Y^{(\sigma(t))} \widehat{w}^{(\sigma(t))} \right]. \end{aligned} \tag{5.25}$$

We may further thin the subsequence σ to have

$$\begin{array}{rcll} x^{(\sigma(t))} & \rightarrow & \tilde{x} \in \text{lim pt}\, x^{(t)} & \text{by assumption} \\ Y^{(\sigma(t))} \widehat{w}^{(\sigma(t))} & \rightarrow & \tilde{z} & \text{by hypothesis 8} \\ \gamma^{(\sigma(t))} & \rightarrow & \tilde{\gamma} \in [\gamma_2..\gamma_3] & \text{by compactness.} \end{array}$$

Then taking the limit of (5.25) shows that

$$\frac{1}{\tilde{\gamma}} \nabla f(\tilde{x}) \tilde{z} \leq \frac{\gamma_1}{\tilde{\gamma}} \nabla f(\tilde{x}) \tilde{z},$$

but because $\gamma_1 < 1$, this implies that $\nabla f(\tilde{x}) \tilde{z} = 0$. Therefore, in any case, (5.23) holds.

Define $\hat{y}^{(t)}$ as that vector having the kth block equal to $y_{[k]}^{(\eta(k,t))}$. We claim that

$$\liminf_{t \to \infty} \nabla f\left(x^{(t)}\right) \left(\hat{y}^{(t)} - x^{(t)}\right) \geq 0 \tag{5.26}$$

in the case when $\left\{x^{(t)}\right\}$ has a limit point. To prove the claim, suppose that it is not true and we shall derive a contradiction. We thin the sequence so that

$$\begin{array}{ll} x^{(t)} \rightarrow \tilde{x} & \text{by assumption} \\ \hat{y}^{(t)} \rightarrow \tilde{y} & \text{by lemma 5.5 and condition 2 of definition 3.1} \end{array}$$

and

$$\lim_{t \to \infty} \nabla f\left(x^{(t)}\right) \left(\hat{y}^{(t)} - x^{(t)}\right) = \nabla f\left(\tilde{x}\right) \left(\tilde{y} - \tilde{x}\right) = -\zeta < 0.$$

Define $u^{(t)}$ to be that vector in $\{0,1\}^{v^{(t)}}$ for which $Y^{(t)} u^{(t)} = \hat{y}^{(t)} - x^{(t)}$. Now, because of (3.21) and (3.22), we may choose $\lambda \in (0..\xi]$ independent of t so that

$$\exists\{t_0 > 0\} \forall\{t > t_0\} \;\; \lambda \nabla f\left(x^{(t)}\right) Y^{(t)} u^{(t)} + \lambda^2 u^{(t)\mathsf{T}} \tilde{H} u^{(t)} < -\frac{\lambda\zeta}{2}.$$

Then $\lambda u^{(t)} \in \Xi^{(t)} \subset W^{(t)}$, so that the definition of $\widehat{w}^{(t)}$ gives, for t sufficiently large and in the subsequence under consideration,

$$\begin{aligned} & \nabla f\left(x^{(t)}\right) Y^{(t)} \widehat{w}^{(t)} + \widehat{w}^{(t)\top} H^{(t)} \widehat{w}^{(t)} \\ \leq \; & \nabla f\left(x^{(t)}\right) Y^{(t)} \left(\lambda u^{(t)}\right) + \left(\lambda u^{(t)}\right)^{\top} H^{(t)} \left(\lambda u^{(t)}\right) \\ < \; & -\frac{\lambda \zeta}{2} < 0. \end{aligned}$$

This contradicts (5.23) and proves the claim (5.26).

Lemma 5.5 now allows us to apply theorem 3.3 for every $\tilde{x} \in \lim \operatorname{pt} x^{(t)}$ to see that all limit points are KKT points.

♠

5.3 Specific Instances of the Algorithm

The methods of §5 are general and include many of the known decomposition methods. It is easy to see that the methods of §3 are contained in the more general methods of this chapter: let $\Upsilon^{(t)}_{[k]} = \{t\}$ and $\eta(k,t) = t$ for all k and t and use the inner approximation given by the entire feasible set for the master

$$W^{(t)} = \left\{ \; w \;\middle|\; \underline{w}^{(t)} \leq w \leq \overline{w}^{(t)} \; \right\} = F^{(t)}.$$

We now consider a few other special cases of the methods developed so far in §5.

5.3.1 Restricted Simplicial Decomposition

Recall the simplicial decomposition method, discussed in §1.5. One may further restrict the number of columns to be stored by insisting *a priori* that the method store no more than $L+1$ columns—L of the form $y^{(\ell)}$ in addition to $x^{(t)}$. This method is called **restricted simplicial decomposition** and was developed by Hearn *et al* [1987]. As one would expect, a trade-off occurs as the parameter L varies. When $L = N+1$ the simplicial decomposition algorithm results. When

$L \approx N$ the excellent convergence properties of simplicial decomposition are retained at the expense of large storage requirements. On the other hand, for $L \approx 1$, the storage requirements are reduced, but the speed of convergence is adversely effected. This is consistent with the slow convergence of the Frank–Wolfe method [Frank and Wolfe 1956], which corresponds to the special case $L = 1$. Hearn *et al* [1987] show that the convergence rate is still good so long as L is at least the dimension of the facet containing the optimal solution (assuming $\mathcal{B}$ is polyhedral and the optimal solution is unique).

We shall now see that a very close relative of the restricted simplicial decomposition algorithm of Hearn *et al* [1987] is a particular instance of the method of this chapter. In order to eliminate some of the complexities of the notation, we shall consider a problem with only one block (i.e., $K = 1$) and eliminate all reference to block indices. The version where $K > 1$ merely does the same bookkeeping for each block in the obvious way. We will also choose the culling function $\eta(t) = t$ implicitly in this section.

Suppose $\mathcal{B}$ is bounded and the decoupled resource allocation R is given by the simple bounds

$$\underline{R} = -\kappa\mathbf{1} \quad \text{and} \quad \overline{R} = \kappa\mathbf{1} \quad \text{where} \quad \kappa \geq \sup_{x,y\in\mathcal{B}} \|x - y\| \, .$$

(This removes the effect of R on the updates $y^{(t)}$ while still retaining the boundedness properties needed in the convergence proofs above.) Therefore, if the subproblems have linear objective functions, the $y^{(t)}$ may be computed as extreme points of $\mathcal{B}$. Moreover, we shall choose the coordinator constraints $W^{(t)}$ to be those given by the simplex (5.12). Since we are assuming that $\mathcal{B}$ is bounded, we need not modify $W^{(t)}$ in order to guarantee that $\left\{Y^{(t)}\widehat{w}^{(t)}\right\}$ is bounded.

Consider the restricted simplicial decomposition algorithm of figure 13 where at most L updates are kept. The element of "least weight" is that for which the component of $w^{(t-1)}$ is nearest to zero. The algorithm of figure 13 differs from the method of Hearn *et al* [1987] in that:

- Hearn *et al* [1987] allow the algorithm to keep old points $x^{(\ell)}$ (for $\ell < t$) as

```
Given $x^{(0)} \in \mathcal{B}$ and $\Upsilon^{(-1)} = \emptyset$
For $t = 0, 1, \ldots$
    Solve the linear subproblems to get $y^{(t)}$ and construct $Y^{(t)}$
    If $\left|\Upsilon^{(t-1)}\right| < L$
        Then set $\Upsilon^{(t)} = \Upsilon^{(t-1)} \cup \{t\}$
        Else construct $\Upsilon^{(t)}$ by replacing the element
            of "least weight" in $\Upsilon^{(t-1)}$ with $t$
    Solve the coordinator subject to the constraints (5.12)
        to get $w^{(t)}$ and set $x^{(t+1)} = x^{(t)} + Y^{(t)} w^{(t)}$
```

Figure 13: Restricted Simplicial Decomposition.

well as points $y^{(\ell)}$. However, when the number of updates nears L, only $x^{(t)}$ is retained from this set at iteration t.

- The coordinator of Hearn *et al* [1987] must find an exact minimizer of $f\left(x^{(t)} + Y^{(t)}w\right)$ subject to the constraints (5.12). Our method solves a computable coordinator problem.

Hearn *et al* [1987, §6.2] give an extension using a convex quadratic program as a coordinator. The basic technique is to *majorize* [Ortega and Rheinboldt 1970, §8.3(d)] the function f by a quadratic function $gx + x^{\mathsf{T}}Gx$ where the smallest eigenvalue of the real symmetric matrix G is larger than the maximal curvature of f over some level set. One problem with this approach is that the maximum curvature is a non-local quantity that is in general impossible to compute. Even if we were able to compute the maximal curvature of f, this property of G makes convergence very slow because the steps taken are very small.

Mulvey *et al* [1990] introduce **truncated simplicial decomposition**, where the coordinator problem (the same coordinator as Hearn *et al* [1987] use) is solved more and more accurately, as determined by a forcing sequence $\phi^{(t)} \downarrow 0$. Contrary to this situation, *the coordinator methods in this thesis do not need to be solved* *more and more accurately.*

Hearn *et al* [1987] show that the sequence $\left\{x^{(t)}\right\}$ converges in a finite number of iterations. However, each iteration of their coordinator requires the exact solution of the master problem. For general functions, this exact solution requires an infinite number of steps, so this result is primarily of interest in the case where the original problem is quadratic.

Hearn *et al* [1987] also give an extension of their basic method to handle cases where $\mathcal{B}$ is unbounded. This extension involves computing extreme rays of the feasible region when necessary and modifying the coordinator constraints appropriately. An alternate method of dealing with unboundedness is given by our decoupled resource allocation R, which forces the sequence $\left\{y^{(t)} - x^{(t)}\right\}$ to be bounded. Therefore, we may handle unbounded problems by solving a sequence of bounded subproblems.

5.3.2 A Simple Asynchronous Model

We will consider an instance of the asynchronous decomposition algorithm and how it might be implemented in a distributed computing environment. Specifically, if we have an environment with $\wp$ processors, labeled $1, \ldots, \wp$, we wish to consider mapping the asynchronous algorithm onto this system. First, we will designate processor 1 as the processor that does the coordinating phase of the work and we will use processors $2, \ldots, \wp$ to solve the subproblems. Processors $2, \ldots, \wp$ will each execute the pseudo-code given in figure 14. This does nothing more than receive subproblems (waiting for a message if necessary), solve subproblems and report the answer back to processor 1.

Processor 1 will execute the pseudo-code in figure 15. (In this figure, the dependence on the iteration counter t is implicit for the current data x, Υ and η.) This algorithm begins with a single update $y^{(0)}$ and the information sets $\Upsilon_{[k]}$ corresponding to this update. At each iteration all of the other processors $2, \ldots, \wp$ are checked for outstanding messages. If there is a message from a certain

```
Repeat forever
    Receive a subproblem "$k, t, G, g, \underline{R}, \overline{R}, A_{[k]}, a_{[k]}$" from processor 1
    Compute $y^{(t)}_{[k]} \in \arg\min \left\{ \tfrac{1}{2} y^\top G y + gy \;\middle|\; \begin{array}{c} A_{[k]} y = a_{[k]} \\ \underline{R} \le y \le \overline{R} \end{array} \right\}$
    Send the message "$k, t, y^{(t)}_{[k]}$" to processor 1
```

Figure 14: The subproblem pseudo-code.

```
Given $x = x^{(0)} \in \mathcal{B} \cap \operatorname{dom} f$ and $y^{(0)} \in \mathcal{B}$

For each $k = 1, \ldots, K$ set $\Upsilon_{[k]} = \{0\}$ and $\eta(k) = 0$

For $t = 0, 1, \ldots$ ad infinitum

    For $p = 2, \ldots, \wp$
        If there is a message "$k', t', y^{(t')}_{[k']}$" from processor $p$
            If $\left|\Upsilon_{[k']}\right| = L$ then remove some element $t''$ from $\Upsilon_{[k']}$
              and delete $y^{(t'')}_{[k']}$ from memory
            Set $\Upsilon_{[k']} = \Upsilon_{[k']} \cup \{t'\}$ and $\eta(k') = t'$
            Find $k''$ with $\eta(k'') \le \eta(k)$ for all $k$
            Compute an appropriate subproblem and
              send it to processor $p$ in the form
              "$k'', t, G^{(t)}_{[k'']}, g^{(t)}_{[k'']}, \underline{R}^{(t)}_{[k'']}, \overline{R}^{(t)}_{[k'']}, A_{[k'']}, a_{[k'']}$"

    Solve the coordinator and update the primal point $x$
```

Figure 15: The coordinator pseudo-code.

subproblem processor, the message is read and the update vector incorporated into the current information. Note that the current information for a given block will have at most L members. The culling function $\eta(k')$ is then reset to t' so that the last update $y_{[k']}^{(t')}$ to arrive is chosen. The algorithm then sets up a subproblem and sends it to the idle processor for solution. We are assuming, of course, that the subproblem is set up to satisfy all of the hypotheses of the convergence criterion and we are using the notation $\underline{R} = \underline{R}(x)$ and $\overline{R} = \overline{R}(x)$ for a decoupled resource allocation R.

Let us suppose that the amount of time (measured in terms of iterations by processor 1) for any subproblem to be solved by a subproblem processor (including the time for data transfer) is bounded by T iterations. Then we have $T \geq t - \eta(k)$ for each block k, showing that assumption (6) holds. (This is a reasonable assumption to make in practice and may always be forced to occur if the programmer is willing to make processor 1 do some extra checking.)

It is beneficial in this context to use a coordinator that is as fast as possible, so that the subproblem processors are not waiting long periods of time for useful work. In particular, the coordinator in §5.2.3 seems suitable for this. This coordinator involves solving a quadratic program with simple constraints in at most KL variables, and then doing a line search. If the quadratic programming problems are suitably chosen, the number of iterations in the line search will be reasonable, leading to a very fast coordinator method. (For example, the quadratic term could approximate the Hessian matrix, perhaps including a proximal point term.)

Chapter 6

Conclusions

6.1 The Motivation for This Work

This work began when the author's principal advisor obtained the PDS problem suite of §4.1. Based on experience solving pure network flow problems, we realized that a decomposition method requiring the solution of each pure network block (commodity) approximately one hundred times would be a reasonable solution procedure for the entire problem. From this empirical calculation, we developed a rough decomposition method and experimented with it computationally. This thesis contains the results of the final product of this endeavor, placed on a theoretically sound foundation.

6.2 A Summary of the Earlier Chapters

The barrier function theory of Fiacco and McCormick [1968] was enlarged to facilitate the removal of some, but not all, constraints. In as much as the methods of Fiacco and McCormick [1968] are "sequential unconstrained minimization techniques," the methods of §2 might be called "sequential less-constrained minimization techniques."

A theory of shifted barriers was needed to produce points feasible for the original problem, since barrier functions are ill-defined off of the interior of the feasible region. We have shown that, if a point interior to the barrier exists and if the barrier problems are solved in an approximate sense, it is possible to generate a point feasible to the original problem in a *finite number of steps.*

We also showed that some barriers which use the logarithm have excellent approximate optimality properties. In particular, the KKT conditions of the original problem are approximately satisfied in a quantifiable way by barrier problem solutions. For convex problems, this yields feasible solutions at which the objective function value differs from the optimal by some arbitrarily small parameter.

We chose barrier functions, instead of penalty functions, because of the excellent feasibility properties of the approximate solutions generated. Indeed, in our computational experience with the PDS problems, *all* of the inequality constraints are satisfied *perfectly*, while the affine (network) constraints are satisfied to the floating point tolerance. This is an important consideration when the user of the optimization software needs feasible solutions.

Considering the excellent properties of the barrier problems, we then developed a class of decomposition schemes to handle this problem structure. These decomposition methods use the block structure of the barrier problems to bring to bear the technology of large-grained concurrent computing. These schemes are applicable for the computation of KKT points because they do not require convexity of the objective function or the coupling constraints. However, special structure within each block—such as network structure—may be exploited by the method.

Computational results are given for a subset of the PDS problem suite, including the largest problems of this suite that the author is aware of having been solved. Our source code was compiled on both a DECstation 3100 (a workstation running ULTRIX) and a Sequent Symmetry S81 with 20 processors. The results show not only the benefit of using parallel processing, but that decomposition methods themselves can be very efficient on large-scale problems.

The decomposition method was then generalized to include multiple updates in an asynchronous manner. This very flexible scheme contains the restricted simplicial decomposition algorithm of Hearn *et al* [1987] as a concrete case. In particular, we have generalized restricted simplicial decomposition to the non-pseudoconvex case. This asynchronous convergence theory holds great promise for asynchronous parallel implementations of decomposition methods.

6.3 Future Directions

An asynchronous implementation of the decomposition method is yet to be constructed. This is an important open problem because of the enhanced opportunities for more even load balancing. Moreover, the theory developed here is quite flexible, so that it could map effectively to many different concurrent computational environments. We hope that these avenues will be explored in the future.

Another open problem is suggested in the appendix to this thesis. The barrier-zero functions introduced earlier in the thesis are shown to possess some interesting properties when applied to the standard interior point methods for linear programming in standard form. In particular, these new barrier functions allow one to use a partial basis factorization to compute the Newton direction. This suggests that the Cholesky factorizations needed for the final iterations of the interior point method need only be done on smaller matrices. This technique should also be considered in conjunction with methods for guessing an optimal basis.

Appendix A

The Logarithmic Barrier-Zero Function and Interior Point Methods for Linear Programming

In this section, we consider the logarithmic barrier-zero function of §2.2.3 applied to the linear programming problem in standard form:

$$\underset{x}{\text{minimize}} \; cx \quad \text{subject to} \quad Ax = a \quad \text{and} \quad x \geq \mathbf{0}, \tag{A.1}$$

where A is $M \times N$ with full row rank. Recall that the dual problem is

$$\underset{p}{\text{maximize}} \; pa \quad \text{subject to} \quad pA \leq c.$$

Suppose that the complement of the index set $\mathbf{B} \subset \{1, \ldots, N\}$ is represented as $\mathbf{N}$. Then we use the notation $A_{\mathbf{B}}$ and $A_{\mathbf{N}}$ to denote the corresponding columns of A and $x_{\mathbf{B}}$, $c_{\mathbf{B}}$, etc. to denote the corresponding components of the vectors. Let $\#\mathbf{B}$ and $\#\mathbf{N}$ denote the sizes of the respective index sets.

We say that the set $\mathbf{B}$ is a **basic (optimal) solution of (A.1)** if

$$A_{\mathbf{B}} \text{ is invertible }, \quad A_{\mathbf{B}}^{-1}a \geq \mathbf{0} \quad \text{and} \quad c_{\mathbf{B}}A_{\mathbf{B}}^{-1}A_{\mathbf{N}} \leq c_{\mathbf{N}}.$$

Clearly, if $\mathbf{B}$ is a basic solution of (A.1) then $\#\mathbf{B} = M$. The following lemma provides the motivation for calling the *set* $\mathbf{B}$ the solution to (A.1).

Lemma A.1 (Basic Solutions) *If* $\mathbf{B}$ *is a basic solution to (A.1), then one may construct optimal primal and dual solutions to (A.1) by*

$$x_{\mathbf{B}} = A_{\mathbf{B}}^{-1} a \quad , \quad x_{\mathbf{N}} = \mathbf{0} \quad \textit{and} \quad p = c_{\mathbf{B}} A_{\mathbf{B}}^{-1}.$$

Proof: The reader may easily verify that x and p are primal and dual feasible with the same objective function values.

♠

Recall the logarithmic barrier-zero function discussed in general in §2.2.3 and in particular in §2.2.4. For any fixed $t \in \mathbb{R}^N_{>\mathbf{0}}$ we define the function

$$\rho((\cdot);t) \ : \ \mathbb{R}^N_{>\mathbf{0}} \to \mathbb{R} \ : \ \rho(x;t) = \sum_{n=1}^{N} \rho_n(x_n;t_n)$$

with components defined by

$$\rho_n(x_n;t_n) := \begin{cases} -\ln x_n - \dfrac{x_n^2}{2t_n^2} + \ln t_n + \dfrac{1}{2} & \text{if } x_n < t_n \\ 0 & \text{if } x_n \geq t_n. \end{cases}$$

We have already seen that $\rho((\cdot);t)$ is a twice continuously differentiable, convex barrier function. We also point out that the logarithmic barrier function (with $\rho_n(\cdot) = -\ln(\cdot)$) is the limiting case as $t \to \infty\mathbf{1}$ in the sense that the gradient and Hessian values take on the corresponding values in the limit. Therefore, $\rho((\cdot);t)$ may be considered a perturbation of the logarithmic barrier function.

Define the barrier problem

$$\underset{x}{\text{minimize}} \ \ cx + \tau\rho(x;t) \ \ \text{subject to} \ \ Ax = a \tag{A.2}$$

($x > \mathbf{0}$ implicitly) where τ and t are parameters. The Newton step for (A.2) at the point x is then defined to be the solution of the quadratic program

$$\underset{y}{\text{minimize}} \ \ (c + \tau\nabla\rho(x;t))\, y + \frac{\tau}{2} y^{\top} \nabla^2 \rho(x;t) y \ \ \text{subject to} \ \ Ay = \mathbf{0}.$$

Since $\nabla^2\rho(x;t)$ is positive semi-definite and the quadratic program is affine constrained, y is optimal for the quadratic program iff y and some p satisfy the linear equations

$$\begin{pmatrix} \tau\nabla^2\rho(x;t) & -A^\top \\ -A & \mathbf{0} \end{pmatrix} \begin{pmatrix} y \\ p^\top \end{pmatrix} = \begin{pmatrix} -c^\top - \tau\nabla\rho(x;t)^\top \\ \mathbf{0} \end{pmatrix}. \qquad \text{(A.3)}$$

Therefore, for given values of the parameters τ and t, solving the linear system (A.3) for y and p yields the Newton step y for the barrier problem (A.2) along with the associated Lagrange multiplier estimate p. The theory in §1.7 says that, for fixed $t > \mathbf{0}$, the minimizers of (A.2) approach the minimizers of (A.1) as $\tau \downarrow 0$.

Suppose that we are given a set $\mathbf{B}$ with $A_{\mathbf{B}}$ of full column rank, i.e., rank $A_{\mathbf{B}} = \#\mathbf{B}$. Let t be given by

$$t_n = \begin{cases} x_n & \text{if } n \in \mathbf{B} \\ x_n + T & \text{if } n \notin \mathbf{B}. \end{cases} \qquad \text{(A.4)}$$

Since the corresponding component of the barrier function is zero in such an instance, we see that

$$\nabla\rho_{\mathbf{B}}(x_{\mathbf{B}}) = \mathbf{0}\ , \quad \nabla^2\rho_{\mathbf{B}}(x_{\mathbf{B}}) = \mathbf{0} \quad \text{and} \quad \nabla^2\rho_{\mathrm{N}}(x_{\mathrm{N}}) \text{ is positive definite.}$$

Using this information, and letting H_{N} denote $\nabla^2\rho_{\mathrm{N}}(x_{\mathrm{N}})$, we see that (A.3) may be written as

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & -A_{\mathbf{B}}^\top \\ \mathbf{0} & \tau H_{\mathrm{N}} & -A_{\mathrm{N}}^\top \\ -A_{\mathbf{B}} & -A_{\mathrm{N}} & \mathbf{0} \end{pmatrix} \begin{pmatrix} y_{\mathbf{B}} \\ y_{\mathrm{N}} \\ p^\top \end{pmatrix} = \begin{pmatrix} -c_{\mathbf{B}}^\top \\ -c_{\mathrm{N}}^\top - \tau\nabla\rho_{\mathrm{N}}(x_{\mathrm{N}}) \\ \mathbf{0} \end{pmatrix}. \qquad \text{(A.5)}$$

The sparsity pattern of this matrix is given in figure 16 for the situation where $M \approx 500$, $N \approx 1000$ and $\#\mathbf{B} \approx 300$.

Recall that we are assuming that $A_{\mathbf{B}}$ has full column rank. Then we may use Gaussian elimination to construct a factorization of the form

$$\begin{pmatrix} -A_{\mathbf{B}}^\top \\ -A_{\mathrm{N}}^\top \end{pmatrix} = \begin{pmatrix} L & \mathbf{0} \\ Z & \mathbf{I} \end{pmatrix} \begin{pmatrix} U & V \\ \mathbf{0} & W \end{pmatrix}, \qquad \text{(A.6)}$$

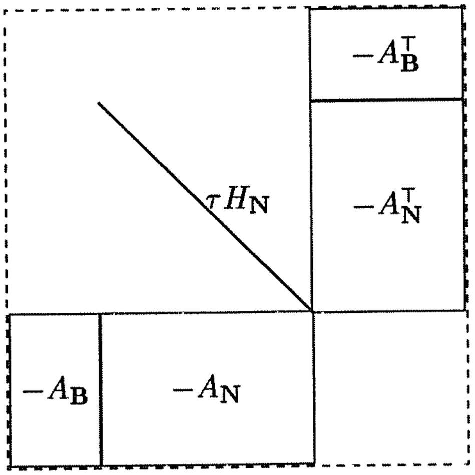

$-A_{\mathbf{B}}^{\mathsf{T}}$
τH_{N}
$-A_{\mathrm{N}}^{\mathsf{T}}$
$-A_{\mathbf{B}}$
$-A_{\mathrm{N}}$

Figure 16: Sparsity pattern of the matrix in (A.5).

where L and U are, respectively, lower and upper triangular nonsingular matrices and

$$\begin{array}{ll} L \text{ is } \#\mathbf{B} \times \#\mathbf{B}, & U \text{ is } \#\mathbf{B} \times \#\mathbf{B}, \\ V \text{ is } \#\mathbf{B} \times (M - \#\mathbf{B}), & W \text{ is } (N - \#\mathbf{B}) \times (M - \#\mathbf{B}), \\ \text{and} & Z \text{ is } (N - \#\mathbf{B}) \times (N - \#\mathbf{B}). \end{array}$$

This factorization may then be used to construct a partial triangular factorization for (A.5) as shown pictorially in figure (17). Note that the partial factorization of figure (17) is trivial to update if $\mathbf{B}$ does not change, but τH_{N} does.

Given the partial triangular factorization shown in figure (17), we see that the problem is reduced to solving a linear system with the *symmetric, positive definite* constraint matrix

$$\begin{pmatrix} \tau H_{\mathrm{N}} & W \\ W^{\mathsf{T}} & \mathbf{0} \end{pmatrix}. \tag{A.7}$$

(Clearly the matrix is symmetric and positive semi-definite, so it is positive definite if it is nonsingular. Since τH_{N} is positive definite factorizing (A.7) reduces to factorizing $W^{\mathsf{T}} \left[\tau H_{\mathrm{N}}\right]^{-1} W$, which is positive definite because (A.6) shows that W has full column rank.)

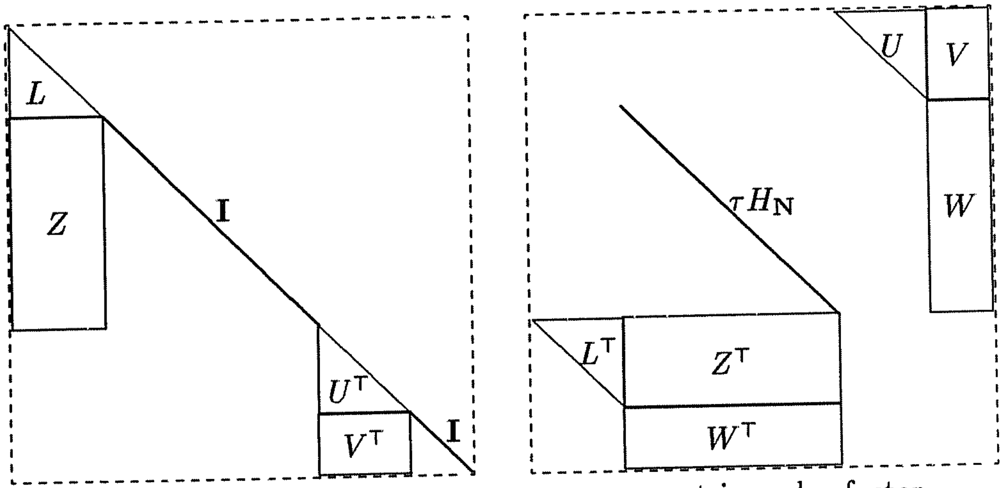


Figure 17: Schematic of partial factorization for (A.5).

It is worthwhile to point out what is going on in some special instances. If $\mathbf{B} = \emptyset$, so that $\mathbf{N}$ is the entire set of variables, then $W = -A^{\mathsf{T}}$ and (A.7) is exactly the form of the matrix of the linear system which the familiar affine scaling methods must solve at each iteration. That is because $\nabla^2 \rho(x;t)$ has no zero diagonals to exploit. If $\#\mathbf{B} = M$ so that an entire basis is chosen, then W has no columns. In this case the matrix (A.7) is diagonal. In cases when $\#\mathbf{B} \approx M$, then the size of $W^{\mathsf{T}} [\tau H_{\mathrm{N}}]^{-1} W$ is small and a dense Cholesky factorization may be efficiently computed.

The linear system with matrix given by (A.7) may be solved by any method; there is some good computational evidence that the method of conjugate gradients is very effective on very large sparse problems [Setiono 1990]. However, we will consider factoring (A.7) by computing the Cholesky factorization

$$W^{\mathsf{T}} [\tau H_{\mathrm{N}}]^{-1} W = CC^{\mathsf{T}}. \tag{A.8}$$

Then a full triangular factorization is represented by figure (18).

Assuming that all of the factorizations are computed in this way, the computation of the solution $(y^{\mathsf{T}}, p)^{\mathsf{T}}$ of (A.5) is accomplished by forward and backward

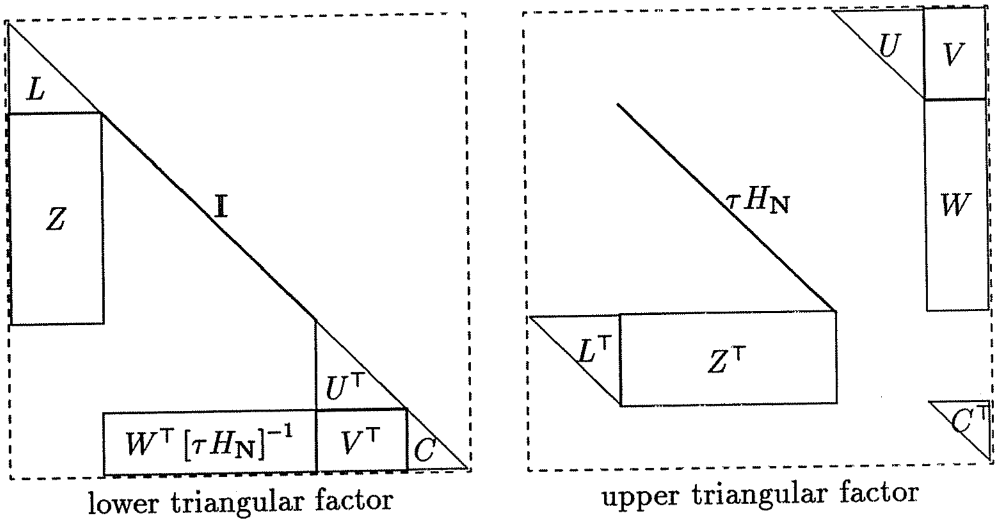


Figure 18: Schematic factorization for (A.5).

substitutions involving the triangular factors in figure 18. Moreover, if the set **B** does not change too drastically from one iteration to another, the factorization (A.6) may be efficiently updated. Once this factorization is updated, the partial factorization depicted in figure 17 is known. Then only the linear system with coefficient matrix (A.7) needs be solved, which may be done by computing the Cholesky factorization (A.8) from scratch during each iteration. For best efficiency, this barrier function should be used with a method for heuristically guessing an optimal basis [Kojima 1986; Megiddo 1988; Tapia and Zhang 1989; Gay 1989; Ye 1990].

Bibliography

D.P. Bertsekas and J.N. Tsitsiklis [1989]. *Parallel and Distributed Computation: Numerical Methods.* Prentice Hall, Englewood Cliffs, N.J.

W.J. Carolan, J.E. Hill, J.L. Kennington, S. Niemi, and S.J. Wichmann [1990]. An empirical evaluation of the KORBX algorithms for military airlift applications. *Operations Research*, 38(2):240–248.

Y.-C. Cheng, D.J. Houck Jr., J.-M. Liu, M.S. Meketon, L. Slutsman, R.J. Vanderbei, and P. Wang [1989]. The AT&T KORBX^{TM} system. *AT&T Technical Journal*, 68(3):7–19.

R. Chmielewski [1989]. Private Communication.

B.D. Craven [1978]. *Mathematical Programming and Control Theory.* Chapman and Hall, London.

J.M. Danskin [1967]. Note on "the sequential maximization technique". *Operations Research*, 15:828–829.

G.B. Dantzig and P. Wolfe [1960]. A decomposition principle for linear programs. *Operations Research*, 8:101–111.

R. De Leone [1990]. Private Communication.

Gy. Farkas (1894). A Fourier-féle mechanikai elv alkalmazásai (Hungarian: On the applications of the mechanical principal of Fourier). *Mathematikai és Természettudományi Értesitö*, 12:457–472.

A.V. Fiacco and G.P. McCormick [1968]. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Research Analysis Corporation, McLean, VA. Re-printed as volume 4 of the SIAM Classics in Applied Mathematics series, SIAM publications, Philadelphia [1990].

R. Fletcher [1987]. *Practical Methods of Optimization.* John Wiley and Sons, New Yrok, second edition.

L.R. Ford and D.R. Fulkerson [1962]. *Flows in Networks.* Princeton University Press, Princeton, N.J.

M. Frank and P. Wolfe [1956]. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(2):95–110.

K.R. Frisch [1954]. Principles of linear programming—with particular reference to the double gradient form of the logarithmic potential method. Technical report, University Institute of Economics, Oslo, October 18.

K.R. Frisch [1955]. The logarithmic potential method of convex programming. Technical report, University Institute of Economics, Oslo, May 13.

J. Gauvin [1977]. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, 12:136–138.

D.M. Gay [1989]. Stopping tests that compute optimal solutions for interior-point linear programming algorithms. Technical report, AT&T Bell Laboratories, Murray Hill, NJ, December.

P.E Gill and W. Murray [1974]. Newton-type methods for unconstrained and linearly constrained optimization. *Mathematical Programming*, 28:311–350.

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

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

M.D. Grigoriadis and Tau Hsu [1979]. *RNET, The Rutgers Minimum Cost Network Flow Subroutines, Users Documentation.* Department of Computer Science, Hill Center for the Mathematical Sciences, Rutgers University—Busch Campus, New Brunswick, New Jersey 08903, version 3.6.

W.W. Hager and D.W. Hearn [1990]. The dual active set algorithm and quadratic networks. Technical report, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL.

D.W. Hearn, S. Lawphongpanich, and J.A. Ventura [1987]. Restricted simplicial decomposition: computation and extensions. *Mathematical Programming Studies*, 31:99–118.

J.K. Ho, T.C. Lee, and R.P. Sundarraj [1988]. Decomposition of linear programs using parallel computation. *Mathematical Programming, Series B*, 42(2):391–405.

B. von Hohenbalken [1977]. Simplicial decomposition in nonlinear programming algorithms. *Mathematical Programming*, 13:49–68.

C.A. Holloway [1974]. An extension of the Frank and Wolfe method of feasible directions. *Mathematical Programming*, 6:14–27.

N. Karmarkar [1984]. A new polynomial-time algorithm for linear programming. *Combinatorica.*

W. Karush [1939]. Minima of functions of several variables with inequalities as side conditions. Master's thesis, University of Chicago.

J.L. Kennington and R.V. Helgason [1980]. *Algorithms for Network Programming.* John Wiley and Sons, New York.

M. Kojima [1986]. Determining basic variables of optimal solutions in Karmarkar's new LP algorithm. *Algorithmica*, 1(4):499–515.

H.W. Kuhn and A.W. Tucker [1951]. Non-linear programming. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–493. University of California Press, Berkeley.

L.S. Lasdon [1970]. *Optimization Theory for Large Systems.* The McMillan Company, New York.

D.N. Lee, K.T. Medhi, J.L. Strand, R.G. Cox, and S. Chen [1989]. Solving large telecommunications network loading problems. *AT&T Technical Journal*, 68(3):48–56.

I.J. Lustig [1990]. The influence of computer language on computational experiments: An example from network optimization. *ORSA Journal on Computing*, 2(2):152–161.

O.L. Mangasarian and S. Fromovitz [1967]. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17:37–47.

D. Medhi [1987]. *Decomposition of Structured Large-Scale Optimization Problems and Parallel Optimization.* PhD thesis, University of Wisconsin—Madison. Also appeared as computer sciences tech report # 718.

N. Megiddo [1988]. On finding primal- and dual-optimal bases. Technical Report RJ 6328 (61997), IBM Almaden Research Center, San Jose, CA, July.

J.M. Mulvey, S.A. Zenios, and D.P. Ahlfeld [1990]. Simplicial decomposition for convex generalized networks. *Journal of Information and Optimization Sciences*, 11(2):359–387.

J.M. Ortega and W.C. Rheinboldt [1970]. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York.

M.Ç. Pinar and S.A. Zenios [1990]. Parallel decomposition of multicommodity network flows using smooth penalty functions. Technical Report 90-12-06, Decision Sciences Department, The Wharton School, University of Pennsylvania.

R.T. Rockafellar [1970]. *Convex Analysis.* Princeton University Press, Princeton, N.J.

A. Schrijver [1986]. *Theory of Linear and Integer Programming.* John Wiley & Sons, New York.

R. Setiono [1990]. Interior dual proximal point algorithm using preconditioned conjugate gradient. Technical Report 951, Computer Sciences Department, University of Wisconsin—Madison.

M. Slater [1950]. Lagrange multipliers revisited: A contribution to nonlinear programming. Cowles Commission Discussion Paper, Mathematics 403, November.

J.E. Smith and S.D. Klinger [1988]. Performance of the Astronautics ZS-1 central processor. Technical report, Astronautics Corporation of America, 5800 Cottage Grove Rd., Madison, WI 53716, March.

R.A. Tapia and Y. Zhang [1989]. A fast optimal basis identification technique for interior point linear programming methods. Technical Report TR89-1, Dept. of Mathematical Sciences, Rice Univ., Houston, TX, October.

P. Tseng [1990]. Dual ascent methods for problems with strictly convex costs and linear constraints: A unified approach. *SIAM Journal on Control and Optimization*, 28(1):214–242.

R.S. Varga [1990]. *Scientific Computation on Mathematical Problems and Conjectures.* SIAM publications, Philadelphia.

Y. Ye [1990]. Recovering optimal basic variables in Karmarkar's polynomial algorithm for linear programming. *Mathematics of Operations Research*, 15(3):564–572.

S. A. Zenios and R. A. Lasken [1988]. Nonlinear network optimization on a massively parallel connection machine. *Annals of Operations Research*, 14.