The Dikin-Karmarkar Principle
for Steepest Descent

Catherine M. Samuelsen

September, 1992

TR92-29

RICE UNIVERSITY

The Dikin-Karmarkar Principle for Steepest Descent

by

Catherine Marie Samuelsen

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Richard A. Tapia, Chairman
Noah Harding Professor of Mathematical Sciences

John E. Dennis, Jr.
Noah Harding Professor of Mathematical Sciences

Richard D. Young
Professor of Economics and Mathematical Sciences

Houston, Texas

April, 1992

The Dikin-Karmarkar Principle for Steepest Descent

Catherine Marie Samuelsen

Abstract

Steepest feasible descent methods for inequality constrained optimization problems have commonly been plagued by short steps. The consequence of taking short steps is slow convergence or even convergence to non-stationary points (zigzagging). In linear programming, both the projective algorithm of Karmarkar (1984) and its affine-variant, originally proposed by Dikin (1967), can be viewed as steepest feasible descent methods. However, both of these algorithms have been demonstrated to be effective and seem to have overcome the problem of short steps. These algorithms share a common norm. It is this choice of norm, in the context of steepest feasible descent, that we refer to as the Dikin-Karmarkar Principle.

This research develops mathematical theory to quantify the short step behavior of Euclidean norm steepest feasible descent methods and the avoidance of short steps for steepest feasible descent with respect to the Dikin-Karmarkar norm. While the theory is developed for linear programming problems with only nonnegativity constraints on the variables, our numerical experimentation demonstrates that this behavior occurs for the more general linear program with equality constraints added. Our numerical results also suggest that taking longer steps is not sufficient to ensure the efficiency of a steepest feasible descent algorithm. The uniform way in which the Dikin-Karmarkar norm treats every boundary is important in obtaining satisfactory convergence.

How precious to me are your
thoughts, O God!
How vast is the sum of
them!
Were I to count them,
they would outnumber the
grains of sand.
Psalm 139:17-18a (NIV)

Acknowledgments

My greatest thanks are given to Jesus Christ, my personal Lord and Savior. Without His grace and mercy, I would not have been able to complete this work.

I lovingly acknowledge the support and sacrifice of my family—my husband, Fred, who has endured having a student wife for many years; and my children, Micah and Amanda, who have shared their mother with Rice University. I appreciate their love and understanding.

I give my appreciation to some very special friends. My thanks go to Cheryl Bosman for her continual encouragement. I thank all my First Place friends for their unflagging encouragement, prayers, and fellowship. For their prayers and support, I thank Roni Archer and Jacque Truitt.

I acknowledge Amr El-bakry for the help he has given me with test problems for my numerical experimentation and Professor Frank Jones for sharing his expertise in Euclidean geometry.

My deepest thanks go to my thesis advisor, Richard Tapia. He has been not only a source of academic advice and inspiration; but, he has been a wonderful source of encouragement—believing in me and teaching me to believe in myself.

This research was supported in part by NSF Coop. Agr. No. CCR-8809615, AFOSR89-0363, DOE DEFG05-86ER25017, ARO 9DAAL03-90-G-0093, and an NSF Graduate Fellowship.

To Fred, Micah, and Amanda

Illustrations

Contents

Tables

Chapter 1

INTRODUCTION

The announcement of a practical, highly-efficient polynomial-time algorithm for linear programming by Karmarkar [23] in 1984 created much excitement in the mathematical community. This single algorithm sparked a huge amount of related research—to verify numerical claims, to modify and extend the algorithm, to develop new interior point methods (both for linear and nonlinear programming). Soon after Karmarkar's projective algorithm was published, its affine-scaling variant was proposed by several researchers (for example, Barnes [5] and Vanderbei, Meketon, and Freedman[40]). It was later learned that the affine-scaling variant had originally been introduced by Dikin [13] in 1967.

The algorithms proposed both by Dikin and Karmarkar both solve subproblems at each iteration which produce steepest feasible descent directions with respect to a common, well-chosen norm. However, steepest feasible descent methods have been known to produce short steps which may result in very slow convergence or convergence to nonstationary points. The norm used by Dikin and Karmarkar could be considered an optimal choice for steepest feasible descent applied to linear programming since global convergence properties have been proven and good practical results have been demonstrated. It is this choice of norm—in the context of solving problems with nonnegativity constraints—that we refer to as the Dikin-Karmarkar Principle. Steepest feasible descent with respect to this norm has the surprising property that the steps are bounded away from short steps as the boundary is neared, unlike Euclidean norm steepest feasible descent which virtually assures that as the solution is approached short steps will be taken.

This thesis is organized in the following manner. Chapter 2 contains a description of steepest descent methods for unconstrained and constrained optimization and an historical perspective. Chapter 3 points out the difficulty that steepest feasible descent methods encounter because of short steps—the so-called "zigzagging" phenomenon. Chapter 4 presents Karmarkar's projective algorithm and its affine-scaling variant. Dikin's algorithm is also presented. The relation of both the projective algorithm and the affine-scaling algorithm to steepest descent is discussed. Chapter 5 will discuss the Dikin-Karmarkar norm along with a geometric interpretation. The observations concerning the behavior of steepest feasible descent with respect to the Euclidean norm versus the behavior of steepest feasible descent with respect to the Dikin-Karmarkar norm which motivated this research are discussed. Chapter 6 contains the theoretical results of this research. We concern ourselves with the special linear program in which the only constraints are nonnegativity constraints on the variables. We show that as the boundary is approached, the steepest feasible descent step with respect to the Euclidean norm must become progressively shorter and in fact, asymptotically approaches the shortest possible step. Whereas, for steepest feasible descent with respect to the Dikin-Karmarkar norm, asymptotically, as the boundary is approached, the step is bounded away from that shortest step. This behavior is demonstrated numerically, in the more general linear programming problem with linear equality constraints, in Chapter 7. Chapter 8 gives some final remarks and observations.

Chapter 2

STEEPEST DESCENT

2.1 Unconstrained Minimization

We begin by considering the unconstrained minimization problem

$$\text{minimize} \quad f(x), \tag{2.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable.

A natural requirement for an iterative method to solve problem (2.1) is that the objective function value decrease at each iteration, i.e. $f(x + \alpha d) < f(x)$, for some $\alpha > 0$. To obtain decrease, an obvious choice for d is a vector that gives the greatest local decrease in the objective function f. In other words, we ask for a vector that minimizes $\nabla f(x)^T d$ with respect to d. Clearly, to make this minimization well-defined, we must impose some type of normalization on the direction vector d. This notion is formalized in the following definition.

> **Definition 2.1** *(Steepest Descent Direction)* By a steepest descent direction for f at x, *with respect to a given norm,* $\|\cdot\|$, we mean any $\hat{d}$ that solves
>
> $$\begin{array}{ll} \text{minimize} & \nabla f(x)^T d \\ \text{subject to} & \| d \| \leq \delta, \end{array} \tag{2.2}$$
>
> for some $\delta > 0$.

Since $\{ d \mid \| d \| \leq \delta \}$ is a compact set, a solution to Problem (2.2) exists, though it may not be unique.

Clearly, the solutions to (2.2) depend on the choice of norm. When the norm is a weighted Euclidean norm, i.e.

$$\|\cdot\|_W \equiv \|W^{-1} \cdot \|_2, \tag{2.3}$$

where $W \in \mathbb{R}^{n \times n}$ is a symmetric, positive definite matrix, then the steepest descent direction, with respect to the W-norm (2.3), is unique (for a given $\delta > 0$) and it is a positive scalar multiple of

$$- W^2 \nabla f(x). \tag{2.4}$$

For $W = I$, the norm is the Euclidean norm, and the negative gradient is a direction of steepest descent.

We formally define a method of steepest descent as follows:

Definition 2.2 *(Method of Steepest Descent)* By a method of steepest descent for problem (2.1), we mean any iterative method of the form,

$$x^{k+1} = x^k + \alpha_k d^k, \quad \alpha_k > 0,$$

in which d^k is steepest descent direction for f at x^k as described in Definition 2.1.

We refer to a problem of the form (2.2) as a *steepest descent subproblem* for problem (2.1).

2.2 Linearly Constrained Minimization

Consider the optimization problem with linear equality constraints and nonnegativity constraints on the variables:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \\ & x \geq 0, \end{array} \tag{2.5}$$

where $f : I\!R^n \to I\!R$, $x \in I\!R^n$, $b \in I\!R^m$, and $A \in I\!R^{m \times n}$. We say that x is *strictly feasible* for problem (2.5) if $Ax = b$ and $x > 0$.

The concept of steepest descent is generalized to the linearly constrained problem (2.5) as follows:

Definition 2.3 *(Steepest Feasible Descent Direction)* By a steepest feasible descent direction for f at x, *with respect to a given norm,* $\|\cdot\|$, we mean any $\hat{d}$ that solves

$$\begin{array}{ll} \text{minimize} & \nabla f(x)^T d \\ \text{subject to} & Ad = 0 \\ & \| d \| \leq \delta, \end{array} \tag{2.6}$$

for x, a strictly feasible point for problem (2.5), and some $\delta > 0$.

REMARK: Any direction that satisfies Definition 2.3 is a *feasible direction* as described by Zoutendijk [42]. He described a class of solution methods for constrained minimization, so-called *methods of feasible directions.* In this class of iterative methods, the starting point is feasible and all the iterates remain feasible.

Again, the solutions to problem (2.6) depend on the norm. The steepest feasible descent direction with respect to the W-norm,(2.3), is simply the projection onto the null space of A, in the W-norm, of the steepest descent direction for the unconstrained problem. In particular, the steepest feasible descent direction is given by

$$-W^2[I - A^T(AW^2A^T)^{-1}AW^2]\nabla f(x). \tag{2.7}$$

Note that in the case where we are minimizing f with only nonnegativity constraints on the variables, the steepest feasible descent direction, with respect to the W-norm, reduces to

$$-W^2\nabla f(x),$$

for $x > 0$.

Analogous to Definition (2.2), we give the following definition.

Definition 2.4 *(Method of Steepest Feasible Descent)* By a method of steepest feasible descent for problem (2.5), we mean any iterative method of the form,

$$x^{k+1} = x^k + \alpha_k d^k, \quad \alpha_k > 0,$$

in which d^k is steepest feasible descent direction for f at x^k as described in Definition 2.1.

This definition ensures that the iterates remain strictly feasible.* We refer to a problem of the form (2.2) as a *steepest descent subproblem* for problem (2.1).

2.3 Historical Perspective

2.3.1 Cauchy

The gradient method was originally proposed by Cauchy [8] in 1847 and is a method of steepest decent with respect to the Euclidean norm. Cauchy considered the problem of minimizing a function of several variables. Using a first-order Taylor series

*In accordance with contemporary terminology, such a method could be called and ***interior point method.***

approximation he noted that taking a sufficiently small step in the direction of the negative gradient would guarantee decrease in the value of the objective function. Cauchy chose the steplength to give the global minimizer in the negative gradient direction, i.e. α_k solved

$$\underset{\alpha>0}{\text{minimize}}\, f(x^k - \alpha \nabla f(x^k)). \tag{2.8}$$

No convergence analysis was given in this classical paper. Cauchy simply suggested that since the function value would decrease at each step, eventually the minimum would be achieved. (Cauchy made the remark that in order to obtain the new iterates quickly, one could use Newton's method or the secant method on the one dimensional minimization subproblem to obtain a steplength.)

2.3.2 Curry

In 1944, Curry [11] published perhaps the first convergence result for the gradient method for unconstrained optimization. For continuously differentiable functions in $\mathbb{R}^n$, he proved that with the proper choice of steplength, every limit point of the sequence generated by the gradient method is a stationary point of f, i.e. the gradient method cannot converge to a point that is not a stationary point. Curry's choice of steplength α_k was the first stationary point of problem (2.8). Curry's result also holds where the steplength α_k is the first local minimizer in the negative gradient direction. Byrd and Tapia [7] extended Curry's theorem to arbitrary choices of norm and to spaces of arbitrary dimension.

2.3.3 Rosen

In 1957, Rosen extended the gradient method to constrained optimization. His gradient projection method was proposed first for linearly constrained problems [34, 35] and then extended to nonlinearly constrained problems [36] in 1961. Rosen's gradient projection method is based on a projection of the gradient of the objective function onto a subspace of the domain, where the subspace is defined by the intersection of hyperplanes that are determined by the active constraints.

Given a feasible point x^0, Rosen's method generates a sequence of the form

$$x^{k+1} := \quad x^k + \alpha_k P_k(-\nabla f(x^k)) \tag{2.9}$$

where P_k is a linear Euclidean norm projection operator. The steplength is taken to be the minimum between the value for which a new inequality constraint becomes active and the value which minimizes the objective function in the current direction. For the linearly constrained problem (2.5), When the constraint matrix, A, has full rank, the steepest feasible descent with respect to the Euclidean norm is given by the Euclidean norm projection of $\nabla f(x)$ onto the null space of the constraint matrix A. This follows from a straightforward application of the second order necessary conditions to problem (2.6).

2.3.4 Goldstein and Levitin & Poljak

A gradient projection method for convex programming in a Hilbert space setting was proposed by Goldstein [19, 20] in 1964 and independently by Levitin and Poljak [27] in 1965. The method computes the iterative sequence as follows:

$$x^{k+1} := \quad P_{\mathcal{S}}(x^k - \alpha_k \nabla f(x^k)), \tag{2.10}$$

where $P_{\mathcal{S}}$ is the closest point projection operator for the Hilbert space and $\mathcal{S}$ is the convex feasible region.

Goldstein proved that Curry's theorem holds under the assumptions: (1) the objective function, f, is twice continuously differentiable, (2) f is bounded below on the convex feasible set $\mathcal{S}$, and (3) the Hessian of f is uniformly bounded on $\mathcal{S}$. Levitin and Poljak proved Curry's theorem holds under the assumptions: (1) the Jacobian of f is uniformly Lipschitz continuous on the feasible region $\mathcal{S}$, and (2) the convex feasible region is a bounded.

2.3.5 McCormick and Tapia

In 1972, McCormick and Tapia [29] studied Goldstein's gradient projection method for a general objective function. They proved Curry's theorem under less stringent assumptions than needed by Goldstein and Levitin and Poljak. They assumed that (1) the objective function f is continuously Fréchet differentiable on the feasible region $\mathcal{S}$ and (2) the feasible region $\mathcal{S}$ is closed and convex,

Chapter 3

THE CURSE OF SHORT STEPS

3.1 The Phenomenon of Zigzagging

It is natural to ask whether Curry's theorem holds for Rosen's projected gradient method applied to as simple a problem as (2.5), i.e. for steepest feasible descent with respect to the Euclidean norm.

3.1.1 Zoutendijk

Zoutendijk [42] recognized that most feasible direction methods, without careful steplength control, may converge to a point that is not a stationary point. He pointed out that these methods have the potential to generate iterates that bounce between constraints without making adequate progress on the minimization problem. In requiring the iterates to be feasible, the steplength choice often emphasize feasibility at the expense of function decrease. He coined the term *zigzagging* to describe this phenomenon. Zigzagging occurs when the steplength is determined by the constraints rather than the minimization of the objective function. As a result, zigzagging can result in convergence of the iterates to a point which is not a solution to the minimization problem.

3.1.2 Wolfe's Example

Wolfe studied the behavior of Rosen's Gradient Projection method for a special case of problem (2.5) where the only constraints were nonnegativity constraints on the variables:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \geq 0. \end{array} \tag{3.1}$$

He set out to prove that, under mild conditions on the objective function, the gradient projection method would converge. In fact, he was able to construct an example for which Rosen's method produced a sequence of points that converged to a point, $\hat{x}$, that was not a stationary point. (The results are seen graphically in Figure 3.1.†) Hence, Rosen's gradient projection method does not satisfy Curry's theorem.

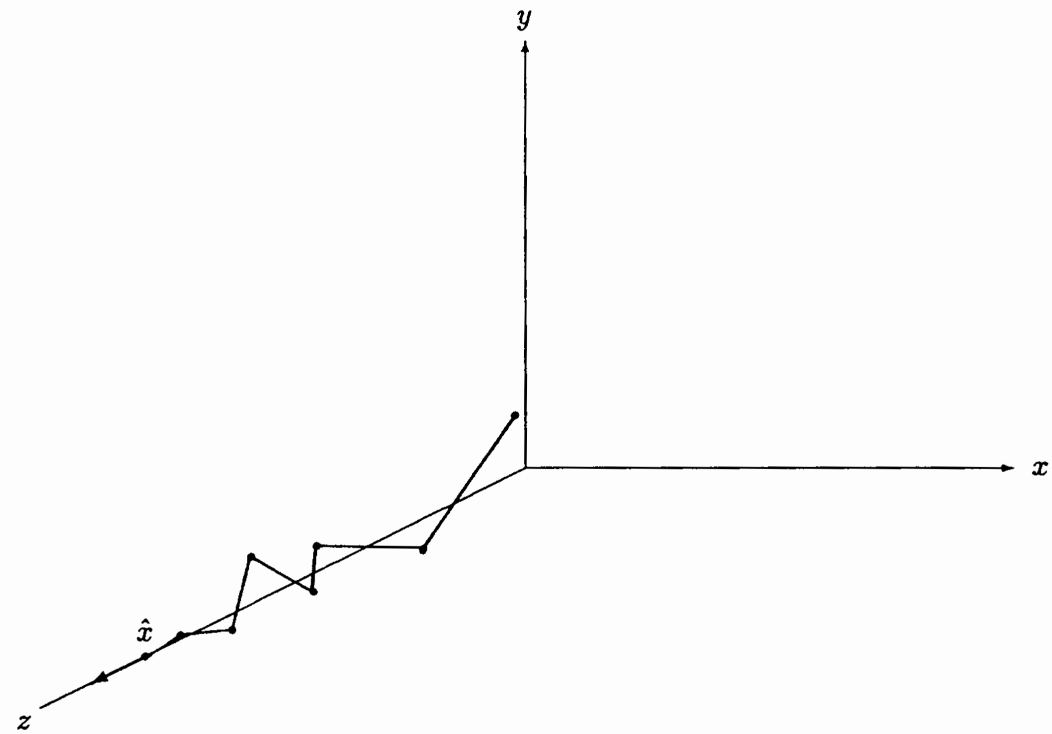

Figure 3.1 Wolfe's Zigzagging Example

3.2 McCormick's Anti-zigzagging Strategy

McCormick recognized that in Wolfe's example, the zigzagging phenomenon occurred because, after a finite number of iterations, the local minimization along the computed step direction did not occur. Instead, the steplength was based entirely on feasibility considerations. In his paper [30] , descriptively entitled, "Anti-zigzagging by Bending," McCormick sought to modify Rosen's method so that longer steps would be taken at each iteration and thus avoid the short steps associated with zigzagging. He proposed, for problem (3.1), to take the steepest feasible descent direction initially.

† minimize $\frac{4}{3}(x^2 - xy + y^2)^{\frac{3}{4}}$, subject to $x, y, z \geq 0$.

However, when a boundary was encountered, instead of stopping, the step direction vector was "bent" to follow the newly encountered constraint. The next iterate was chosen to minimize the objective function along this bent vector. Of course, several "bendings" might be required. This approach allowed longer steps to be taken and prevented the problem of the steplength being dictated by the constraints and not the minimization. McCormick demonstrated that this strategy prevented zigzagging, i.e., he proved a version of Curry's theorem. McCormick and Tapia [28] noted that the "bending" method was equivalent to Goldstein's gradient projection method for the special case where the feasible region is the nonnegative orthant. Then they extended Curry's theorem to the general gradient projection method [29].

3.3 Observations

Initially, both Goldstein's and Rosen's methods take a step in the steepest feasible descent direction. (See the corollary to Proposition 2 in McCormick and Tapia [29]). It is when a new boundary is encountered that the difference occurs. The gradient projection direction adaptively changes as it meets a boundary, while the projected gradient method stops at the boundary. It is this seemingly small distinction which allows one to zigzag while the other cannot.

These observations indicate the importance of considering not only the active constraints, but also the inactive constraints when making a choice of direction at any iteration. The Euclidean norm steepest feasible descent does not use information about the inactive constraints in determining the direction—it only considers which direction gives the greatest amount of *local* decrease. In choosing the norm in which decrease is measured, we believe that it is correct to include information about distance from the inactive constraints. It is this property that the Dikin-Karmarkar norm possesses which contributes to its good convergence properties.

Chapter 4

THE DIKIN & KARMARKAR ALGORITHMS

4.1 Karmarkar's Algorithm

In 1984, Karmarkar [23] proposed a polynomial-time method for the solution of linear programming problems of the form

$$\begin{array}{ll} \text{minimize} & c^Tx \\ \text{subject to} & Ax = 0 \\ & e^Tx = 1 \\ & x \geq 0, \end{array} \tag{4.1}$$

where c, x, $e \in \mathbb{R}^n$, $e = (1, 1, \ldots, 1, 1)^T$, $A \in \mathbb{R}^{m \times n}$ is of full row rank, and the optimal objective function value is zero.

KARMARKAR'S ALGORITHM: Given an initial, strictly feasible point, x^0, for problem (4.1) and a tolerance for the objective function, $\epsilon > 0$, let $k = 0$.

WHILE $c^Tx^k > \epsilon$ **DO**

- $D_k \leftarrow \text{diag}(x^k)$
- Compute $\hat{x} \in \mathbb{R}^n$ as the solution to

$$\begin{array}{ll} \text{minimize} & c^TDx' \\ \text{subject to} & ADx' = 0 \\ & e^Tx' = 1 \\ & \| e - x' \|_2 \leq \delta \end{array} \tag{4.2}$$

- $x^k \leftarrow D\hat{x}/e^TD\hat{x}$
- $k \leftarrow k + 1$

END DO

Theoretically the algorithm was appealing because it was a polynomial-time algorithm. Karmarkar's algorithm was not the first algorithm for linear programming to have a theoretical polynomial-time bound. In 1979, Khachiyan [24] proposed a modification to the ellipsoid method which led to the first polynomial-time algorithm for linear programming. Unfortunately, the practical performance of the ellipsoid method was disappointing—it was not competitive with the simplex method. However, Karmarkar's method was practically appealing because, in some cases, its performance did rival that of the simplex method. The approach of the algorithm was much different than that of the simplex algorithm, the iterates moving through the interior of the feasible region rather than along the boundaries.

While Karmarkar's algorithm is not a straightforward steepest feasible descent method for problem (4.1), the subproblem solved at each iteration has the form of a steepest feasible descent subproblem with respect to a weighted Euclidean norm. That norm is

$$\| \cdot \|_D \equiv \|D^{-1} \cdot \|_2, \qquad D = \mathrm{diag}\,(x^k), \tag{4.3}$$

where x^k is the current, strictly feasible iterate. It has been shown by Morshedi and Tapia [31] and by Tapia and Zhang [38] that Karmarkar's algorithm is actually a steepest feasible descent method applied to the nonlinear program which results from a simple transformation of the linear program.

4.2 The Affine-Variant

Subsequent to the announcement of Karmarkar's algorithm, researchers considered modifications to the algorithm. Motivated to simplify Karmarkar's algorithm and and to develop an algorithm that gave monotone decrease in the objective function, the affine-scaling variant was introduced. It provided a simpler scaling of the problem, decrease in the objective function at each iteration, and no longer required that the right hand side of the linear equality constraints be zero or that objective function be zero at the solution.

The subproblem that is solved at each iteration is

$$\begin{aligned} \text{minimize} \quad & c^T D x' && (4.4) \\ \text{subject to} \quad & A D x' = 0 && (4.5) \\ & e^T D x' = 1 && (4.6) \\ & \|e - x'\|_2 \le \delta. && (4.7) \end{aligned}$$

We can, via a change of variables, produce an equivalent subproblem where the the matrix D does not appear in (4.4), (4.5), and (4.6) and (4.7) becomes

$$\|D^{-1}(x_c - x)\|_2 \leq \delta.$$

In this manner, we observe that the affine variant can be viewed as a method of steepest feasible descent with respect to the norm (4.3).

4.3 Dikin's Algorithm

In 1967, Dikin [13] considered an extension of the method of steepest descent to linear and quadratic programming problems with inequality constraints—specifically nonnegativity of the variables. He proposed an iterative method to solve linear programming problems of the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0, \end{array} \tag{4.8}$$

where $c, x \in I\!R^n$, $b \in I\!R^m$, and $A \in I\!R^{m \times n}$ is of full row rank.

DIKIN'S ALGORITHM: Given an initial, strictly feasible point, x^0, for problem (4.8), let $k = 0$.

1. $D_k \leftarrow \text{diag}(x^k)$
2. Compute $\mu^k \in I\!R^m$ as the solution to

$$\text{minimize} \sum_{j=1}^{n} [\, x_j^k \, (\sum_{i=1}^{m} (a_{ij}\mu_j - c_j)]^2 \tag{4.9}$$

3. $\delta^k \leftarrow A^T \mu^k - c$
4. $\phi_k \leftarrow (x^{k\,T} \delta^k)^2$
5. **WHILE** $\phi_k \neq 0$ **DO**
   - $\lambda_k \leftarrow 1/\sqrt{\phi_k}$
   - $x^{k+1} \leftarrow x^k + \lambda_k D_k^2 \delta$
   - $k \leftarrow k + 1$
   - GO TO 1

   END DO

Note that problem (4.9) is exactly the least squares problem,

$$\underset{\mu \in R^m}{\text{minimize}} \ \| \hat{A}_k{}^T \mu - \hat{c} \|_2^2, \tag{4.10}$$

for $\hat{A} = AD_k$ and $\hat{c} = D_k c$. When A has full row rank, the unique solution to problem (4.10) is given by

$$\mu^k = (AD_k^2 A^T)^{-1} AD_k^2 c. \tag{4.11}$$

Therefore, the step taken at each iteration is

$$-\,D^2\,[I - A^T(A\,D^2 A^T)^{-1} A\,D^2]\,c, \tag{4.12}$$

which is exactly the steepest feasible descent direction given in (2.7).

Dikin [14] proved a version of Curry's theorem for his algorithm, namely that any limit point of the iterative sequence is a solution of the linear programming problem with the only requirement being primal nondegeneracy. So we find that the weighted Euclidean norm chosen by Dikin and Karmarkar overcomes the zigzagging problem associated with Euclidean norm steepest feasible descent. We will refer to the common norm (4.3) as the Dikin-Karmarkar or DK-norm.

We know that our iterates x^k may have some components that are converging to zero. So any measurement of distance should be a relative one [38]. It is this relative weighting of the steps that allows us to look equally at components that are converging to zero that we believe contributes to making the Dikin-Karmarkar norm an ideal choice. We refer to the choice of the Dikin-Karmarkar norm in the context of steepest feasible descent for problems with nonnegativity constraints as the Dikin-Karmarkar principle.

Chapter 5

THE DIKIN-KARMARKAR PRINCIPLE

The choice of norm by Dikin and Karmarkar is specifically suited to problems with nonnegativity constraints. For this reason, as we look at the role of the norm in this context, we will restrict our attention to linear programming problems with nonnegativity constraints on all the variables:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x \geq 0. \end{array} \tag{5.1}$$

We are specifically interested in steepest descent directions for this problem—with respect to the Euclidean norm and with respect to the Dikin-Karmarkar norm.

5.1 The Choice of Norm

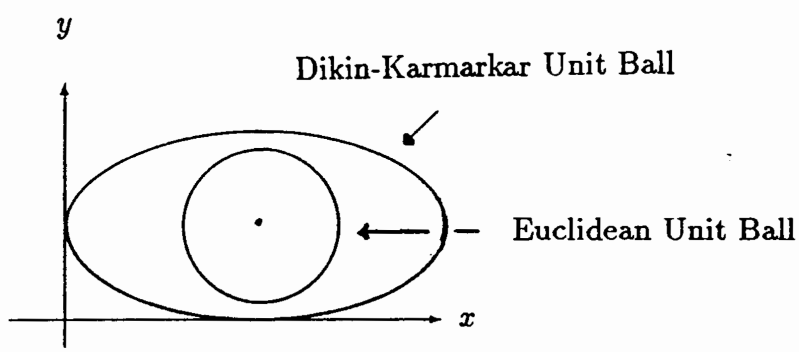

Figure 5.1 The Dikin-Karmarkar and Euclidean Unit Balls in $\mathbb{R}^2$

In Figure 5.1, we illustrate the unit balls in both the Euclidean norm and the Dikin-Karmarkar norms in $\mathbb{R}^2$. The geometry of the Dikin-Karmarkar unit ball changes based on the distance the current iterate is from the boundaries, while the Euclidean ball is fixed, regardless of the boundaries.

5.2 The Dikin-Karmarkar Principle

When using an iterative method to solve problems with inequality constraints, it is important that the direction chosen at each iteration take into account all the boundaries of the feasible region. When using steepest feasible descent, the choice of the Euclidean norm ignores the boundaries in the choice of direction—the direction is always the negative gradient. However, the Dikin-Karmarkar norm is such that the distance of the current strictly feasible iterate from each of the boundaries is taken into account in the norm itself. It is this choice of norm, in the context of solving problems with inequality constraints, that we call the Dikin-Karmarkar Principle. By taking all the boundaries into account, the norm allows the direction taken to not only focus on the amount of local decrease, but also how far we can move in the direction chosen before a boundary is encountered. We believe and demonstrate in the theory and numerical results that follow, that it is this consideration of the boundary that results in steepest feasible descent with respect to the Dikin-Karmarkar norm being a more effective algorithm than steepest feasible descent with respect to the Euclidean norm.

5.3 Behavior of Steepest Feasible Descent Near the Boundary

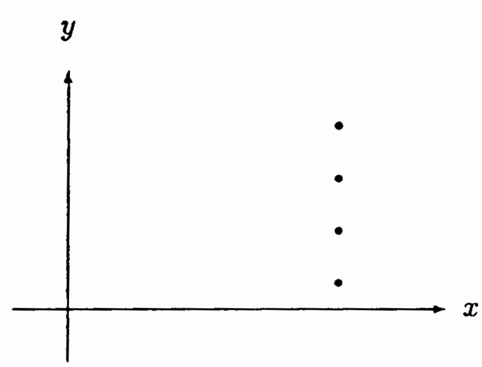

Figure 5.2 Sequence Converging to the Boundary

We have seen in Chapter 3 that steepest feasible descent methods may encounter problems with convergence as a result of taking short steps. In particular, zigzagging can occur as steps are taken toward the boundary. However, in linear programming problems, we know that the solution lies on the boundary; so, taking such steps is necessary.

We gained intuition about the geometry of steepest feasible descent with respect to both norms by looking at what directions would be generated by the algorithms at each point of a sequence converging to the boundary. In Figure 5.2, we see a particular sequence of points converging to the boundary.

We choose a particular linear functional $c^T x$. Figure 5.3 illustrates the directions generated when we use steepest descent with respect to the Euclidean norm at each point of this particular sequence. Figure 5.4 illustrates the directions generated when we use steepest descent with respect to the Dikin-Karmarkar norm at each point in this same sequence.

Note that the directions generated using the Euclidean norm produce relatively shorter and shorter steps to the closest boundary; while, the directions generated using the Dikin-Karmarkar norm produce relatively longer steps to the boundary. These observations lead us to examine the phenomenon of short steps in steepest descent.

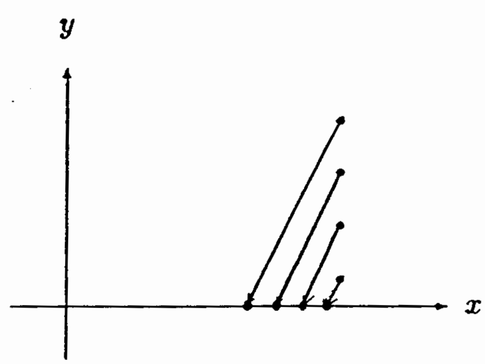

Figure 5.3 Steepest Descent Directions with respect to the Euclidean Norm

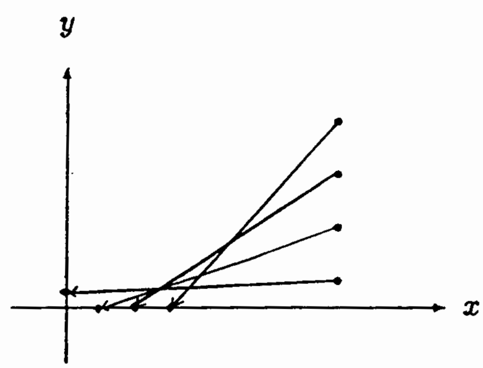

Figure 5.4 Steepest Descent Directions with respect to the Dikin-Karmarkar Norm

Chapter 6

THEORETICAL RESULTS

Before a formal statement of theorems, we first set the stage. We begin by restricting our attention to linear programming problems with nonnegativity constraints:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x \geq 0, \end{array} \tag{6.1}$$

where $c > 0$. We are interested in well-posed problems. It is for this reason that we are restricting ourselves to problems in which the vector c is strictly positive, otherwise problem (6.1) would not have a solution. We refer to such linear functionals, where $c > 0$, as *valid* linear functionals. With the problem we are addressing now clearly stated, we examine short steps in steepest descent methods applied to this problem.

From an interior point $x > 0$, we consider the direction of the *shortest* step to the boundary of $\{ x \,|\, x \geq 0 \}$. This short step is illustrated for $\mathbb{R}^2$ in Figure 6.1.

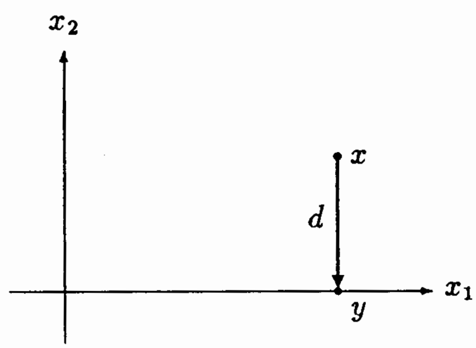

Figure 6.1 A Short Step in $\mathbb{R}^2$

Definition 6.1 *(A Shortest Step Direction)* Consider a point $x > 0$. We say that d is *a shortest step direction* from x if

$$d = \alpha \vec{e_j},$$

where $\alpha > 0$, $\vec{e_j}$ is the j^{th} standard basis vector, and j is the index of the smallest component of x.

We would like to stay away from moving in a shortest step direction at any particular iteration; in fact, we wish to stay away from a neighborhood of such undesirable directions. These directions are illustrated for $\mathbb{R}^3$ in Figure 6.2 and are defined as

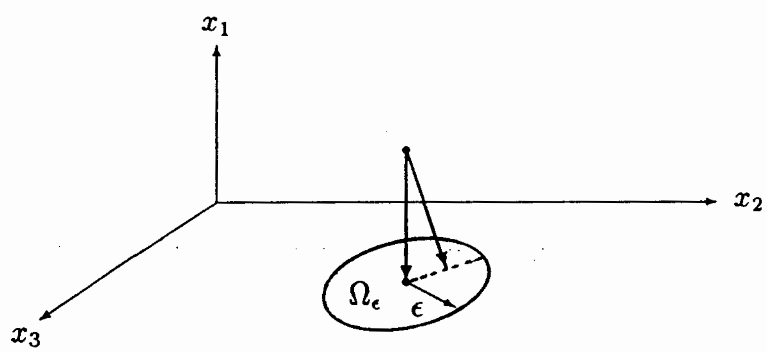

Figure 6.2 ϵ–Short Step Direction in $\mathbb{R}^3$

follows:

Definition 6.2 *(ϵ–Short Step Direction)* Given a point $x > 0$, let $d = \vec{e_j}$ be a shortest step direction from x. Choose $\beta > 0$, so that $y = x + \beta d$ is on the boundary $\mathcal{F}_j = \{ z : z_j = 0 \}$. For $\epsilon > 0$, let

$$\Omega_\epsilon = \{ z \in \mathcal{F}_j \mid \| z - y \|_2 \leq \epsilon \}. \tag{6.2}$$

We say that any $s \in \mathbb{R}^n$ is an *ϵ–short step direction* if

$$x + \alpha s \in \Omega_\epsilon \tag{6.3}$$

for some $\alpha > 0$.

6.1 Tools Necessary for Proof of Theorems

We wish to compare steepest feasible descent for problem (6.1), with respect to both the Euclidean norm and the Dikin-Karmarkar norm. For linear programming problems of this form, it is impossible to say that at a particular iteration one norm choice will always give better performance than the other. However, if we look at all valid linear functionals, what can we say about how these two choices of norm will effect our performance overall?

At a given point $x > 0$, we consider the proportion of all valid linear functionals that will give us an ϵ–short step direction when we use a method of steepest descent, i.e. we want to find

$$\frac{\text{measure of valid linear functionals giving an } \epsilon\text{–short step direction}}{\text{measure of valid linear functionals}}.$$

So we must define a measure for the set of linear functionals.

6.1.1 Parametrization of Linear Functionals

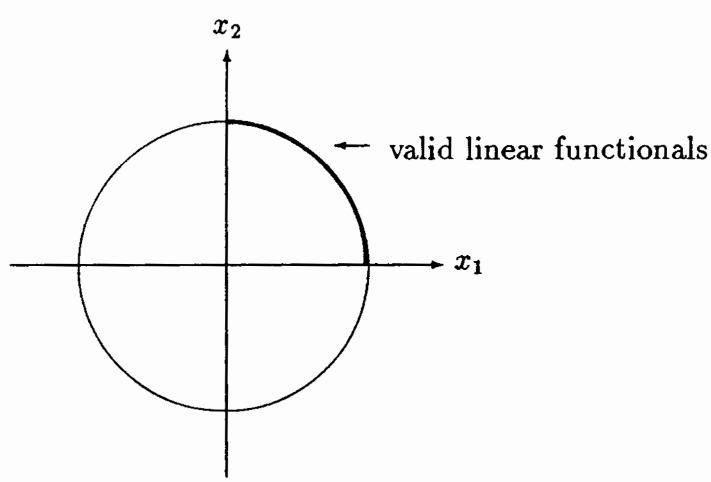

Figure 6.3 Parametrization of Linear Functionals in $\mathbb{R}^2$

We begin by parametrizing linear functionals in terms of their unit normals. Thus we represent a particular linear functional $c^T x$ by $\tilde{c}$, where

$$\tilde{c} = \frac{c}{\| c \|_2}. \tag{6.4}$$

This leads to a parameter space which is the unit sphere.

6.1.2 Measure of Linear Functionals

Thus we will define the measure of a particular set of linear functionals to be the surface area of the portion of the unit sphere which represents that set of linear functionals. For the set of linear functionals valid for problem (6.1), the measure is the surface area of the unit sphere *in the positive orthant.*

The surface area can be easily be computed using spherical coordinates and integrating over the representative area of the unit sphere in $I\!R^n$. The angles that will be integrated over, θ_i, will be taken from

$$0 < \theta_i < \pi/2$$

rather than $0 < \theta_i < \pi$ since that valid linear functionals lie only in the positive orthant. Let $S(n)$ denote the surface area of the unit sphere that represents the valid linear functionals, then

$$S(n) = \int_0^{\pi/2} \{\int dV_{n-2}\}\, d\theta_1, \tag{6.5}$$

where dV_{n-2} is the $(n-2)$-dimensional volume differential.

We will denote the surface area of the valid linear functionals for which the steepest feasible descent direction at x is an ϵ–short step direction by $S_\epsilon(x, n)$.

Thus, the proportion of valid linear functionals in $I\!R^n$ for which the steepest descent direction at $x > 0$, is an ϵ–short step direction is given by

$$m_\epsilon(x) = \frac{S_\epsilon(x,n)}{S(n)}. \tag{6.6}$$

For the Euclidean Norm

The surface area of the linear functionals that will produce ϵ–short step directions when the Euclidean norm is chosen can be seen in Figure 6.4. Consider a point $\bar{x} > 0$. Without loss of generality, let

$$\bar{x}_n = \min\{\, \bar{x}_i,\, i = 1, \ldots, n\,\}.$$

For ease of notation, we will let $r = \bar{x}_n$. View the x′ axis as representing the $(n-1)$–dimensional surface in $I\!R^n$ where $x_n = 0$.

The surface area of the linear functionals for which the steepest descent direction with respect to the Euclidean norm is an ϵ–short step direction can be computed by

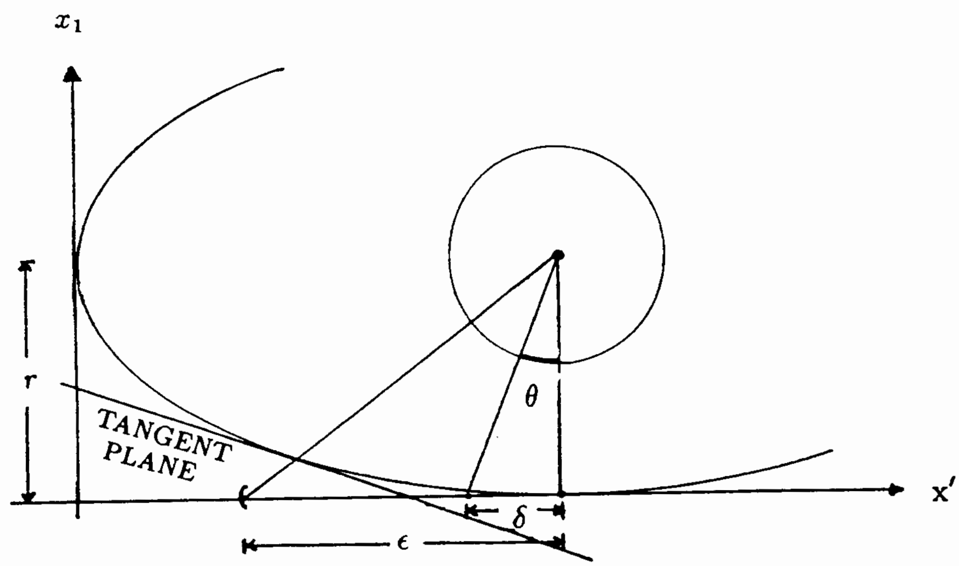

Figure 6.4 Computing the Measure for the Euclidean Norm

integrating θ_n from 0 to θ, i.e.

$$S_\epsilon(x,n) = \int_0^\theta \{\int dV_{n-2}\}\, d\theta_1 \tag{6.7}$$

$$= \theta\{\int dV_{n-2}\}. \tag{6.8}$$

The angle θ is determined by ϵ and r:

$$\theta = \arctan(\epsilon/\bar{x}_n). \tag{6.9}$$

Thus, for $\bar{x} > 0$ for which $\bar{x}_i = \min\{\bar{x}_j,\, j = 1,\ldots,n\}$., the proportion we are interested in is given by

$$\frac{S_\epsilon(\bar{x},n)}{S(n)} = \frac{\arctan(\epsilon/\bar{x}_i)}{\pi/2}. \tag{6.10}$$

For the Dikin-Karmarkar Norm

The problem that we need to solve is as follows—given a strictly positive point $\bar{x} \in I\!R^n$ and $\epsilon > 0$, find the set of all linear functionals for which steepest descent with respect to the Dikin-Karmarkar norm will produce an ϵ–short step at $\bar{x}$.

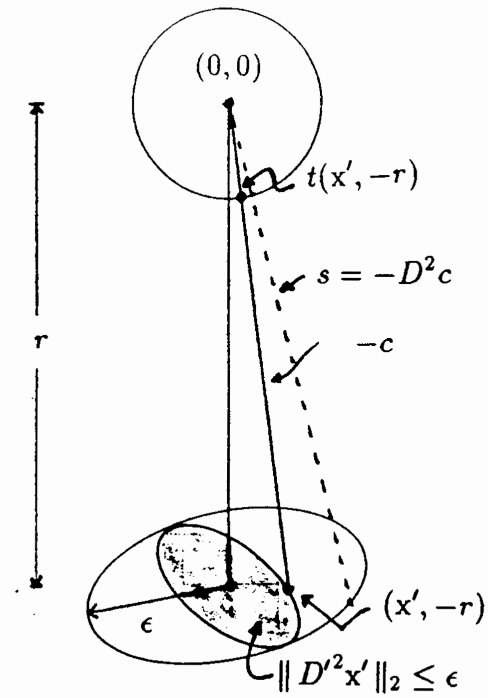

Figure 6.5 Computing the Measure for the Dikin-Karmarkar Norm

Without loss of generality, suppose that

$$\bar{x}_n = \min\{\, \bar{x}_j,\ j = 1, \ldots, n\,\}.$$

Let $r = \bar{x}_n$. Consider a unit sphere centered at $\bar{x}$. We make a change of coordinate systems by translating the entire space by $\bar{x}$ so that our sphere is now centered at the origin. We will denote all points $\mathrm{x} \in I\!R^n$ as

$$\mathrm{x} = (\mathrm{x}', x_n),$$

where $\mathrm{x}' = (x_1, x_2, \ldots, x_{n-1}) \in I\!R^{n-1}$. Our closest face $\hat{\mathcal{F}}_n$ is now the surface at which $x_n = -r$. The center of our Ω_ϵ region is $(0', -r)$. See Figure 6.5.

Every ϵ–short step direction from $\hat{x}$ produced by steepest descent with respect to the norm can be written

$$s = -D^2c,$$

where $\mathrm{D} = \mathrm{diag}\,(\bar{x})$. Since $\hat{x} + s = -D^2c \in \Omega_\epsilon$, then $\|(0', -r) - D^2c\,\|_2 \leq \epsilon$ and $(D^2c)_n = r$. Thus

$$\Omega_\epsilon \;=\; \{(\mathrm{x}', x_n) \mid \|\mathrm{D}'^2\mathrm{x}'\| \leq \epsilon \text{ and } x_n = r\,\}. \qquad (6.11)$$

where $D' = \mathrm{diag}(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{n-1})$. The linear functionals that give rise to Ω_ϵ are described by

$$F = \{(x', x_n) \mid \|D'^2 x'\| \leq r^2\epsilon \text{ and } x_n = r\}. \tag{6.12}$$

Let $(x', -r) \in F_x$. With the unit sphere expressed as

$$S = \{t(x', r) \mid \|(x', r)\|_2 = 1\}, \qquad t = \frac{1}{\sqrt{\|x'\|_2^2 + r^2}}, \tag{6.13}$$

the surface of the unit sphere that describes the set of all linear functionals that will give rise to ϵ–short step is

$$X_\epsilon = \left\{ \frac{(x', r)}{\sqrt{\|x'\|_2^2 + r^2}} \mid \|D'x'\|_2 \leq r^2\epsilon \right\}. \tag{6.14}$$

We make the change of variables $y = x/\|x\|_2$:

$$y' = \frac{x'}{\sqrt{\|x'\|^2 + r^2}} \tag{6.15}$$

$$y_n = \frac{r}{\sqrt{\|x'\|^2 + r^2}} \tag{6.16}$$

From (6.15), (6.16), and since $\|y'\| + y_n^2 = 1$, we have

$$x' = \frac{ry'}{\sqrt{1 - \|y'\|^2}}. \tag{6.17}$$

Thus our surface of interest, (6.14), can be described by

$$Y = \left\{ y' \mid \frac{r}{\sqrt{1 - \|y'\|^2}} \| D'^2 y' \|_2 \leq \epsilon r^2 \right\} \tag{6.18}$$

$$Y_n = \left\{ y_n = \sqrt{1 - \|y'\|^2} \right\}. \tag{6.19}$$

The surface increment we wish to integrate over is

$$dS = dy'/y_n. \tag{6.20}$$

So our surface area is given by

$$S_\epsilon(x, n) = \int_Y \frac{dy'}{\sqrt{1 - \|y'\|^2}}. \tag{6.21}$$

Now, we consider Y that we are integrating over. From (6.18) we have

$$\sum_{i=1}^{n-1} \left(\frac{d_i^4 + r^2\epsilon^2}{r^2\epsilon^2} \right) y_i'^2 \;\leq\; 1. \tag{6.22}$$

Let

$$\lambda_i^2 \;=\; \left(\frac{d_i^4 + r^2\epsilon^2}{r^2\epsilon^2} \right). \tag{6.23}$$

Note that $\lambda_i > 1$.

We make a final change of variables, $z = \Lambda y'$, where $\Lambda = \mathrm{diag}(\lambda_i)$. From (6.21), our surface integral is now

$$\frac{1}{\lambda_1\lambda_2\cdots\lambda_{n-1}} \int_Z \frac{dz}{\sqrt{1-(z_1/\lambda_1)^2-(z_2/\lambda_2)^2-\cdots-(z_{n-1}/\lambda_{n-1})^2}}, \tag{6.24}$$

where $Z = \{ \| z \|_2 \leq 1 \}$. Note that the integrand is bounded on Z.

$$1 \;\leq\; \frac{1}{\sqrt{1-\|\Lambda^{-1}z\|^2}} \;\leq\; \frac{1}{\sqrt{1-\|z'\|^2}}. \tag{6.25}$$

So (6.24) can be bounded.

6.1.3 Converging Sequence Described

With this concept of how to measure the effect of a particular norm choice, we again consider a sequence of points $\{x^k\}$ which converge to the boundary; in particular, we look at $\{x^k\}$ for which

$$x^k \to x^*$$

where

$$\begin{aligned} x_i^* &= 0, \quad i = j \\ x_i^* &> 0, \quad i \neq j. \end{aligned} \tag{6.26}$$

for some $1 \leq j \leq n$.

6.2 Euclidean Steepest Feasible Descent Gives Short Steps

We consider the performance of steepest descent with respect to the Euclidean norm and the Dikin-Karmarkar norm for this sequence converging to the boundary.

Finally, we give a formal statement of our theoretical results for steepest feasible descent, with respect to the Euclidean norm, for problem (6.1).

Theorem 6.1 Given a sequence $\{x^k\}$, which converges to a point x^* satisfying (6.26) and an $\epsilon > 0$, for steepest descent with respect to the Euclidean norm,

$$\lim_{k\to\infty} m_\epsilon(x^k, n) \;=\; 1. \tag{6.27}$$

Therefore, for the Euclidean norm, the proportion of linear functionals for which the steepest descent direction is an ϵ–short step direction is one in the limit.

Proof Without loss of generality, assume that $x^k \to x^*$ for which $x_1^* = 0$. From (6.7) we see that

$$\theta = \arctan\left(\frac{\epsilon}{x_1^k}\right). \tag{6.28}$$

Letting $k \to \infty$,

$$\theta = \arctan\left(\frac{\epsilon}{x_1^k}\right) \to \pi/2. \tag{6.29}$$

So from (6.6) and (6.7), we see that

$$\lim_{k\to\infty} m_\epsilon(x^k, n) = 1. \tag{6.30}$$

□

6.3 Dikin-Karmarkar Steepest Feasible Descent Avoids Short Steps

We consider the same sequence $\{x^k\}$ converging to a point on the boundary and look at m_ϵ for the Dikin-Karmarkar norm.

Theorem 6.2 Given a sequence $\{x^k\}$, which converges to a point x^* satisfying (6.26) and an $\epsilon > 0$, for steepest descent with respect to the Dikin-Karmarkar norm,

$$\lim_{k\to\infty} m_\epsilon(x^k, n) \;=\; 0. \tag{6.31}$$

Therefore, for the Dikin-Karmarkar norm, the proportion of linear functionals for which the steepest descent direction is an ϵ–short step direction is zero in the limit.

Proof Consider (6.24). Note that the integral is bounded so that

$$0 < m \leq \int_Z \frac{dz}{\sqrt{1 - (z_1/\lambda_1)^2 - (z_2/\lambda_2)^2 - \cdots - (z_{n-1}/\lambda_{n-1})^2}} \leq M < \infty. \tag{6.32}$$

However, we consider the quantity multiplying the integral:

$$\frac{1}{\lambda_1 \lambda_2 \cdots \lambda_{n-1}} = \sqrt{\prod_{i=1}^{n-1} \left(\frac{r^2 \epsilon^2}{d_i^4 + r^2 \epsilon^2} \right)}. \tag{6.33}$$

So for x^k such that $x_n^k \longrightarrow 0$, letting $r \longrightarrow 0$, we see that (6.33) converges to zero. □

Thus we find that as our iterates approach the boundary, we are assured that our iterates will be bounded away from a region of short steps.

Chapter 7

NUMERICAL RESULTS

Our theory gives an explanation of the behavior of steepest feasible descent with respect to to the Euclidean and the Dikin-Karmarkar norms for the simplified case with only nonnegativity constraints. We wanted to discover whether this behavior extended to linear programming problems with linear constraints added. We found in our numerical experimentation that, indeed, the behavior described in our theory occurred in this more general case.

In our numerical testing, steepest feasible descent with respect to the Dikin-Karmarkar norm was compared to steepest feasible descent with respect to the Euclidean norm. With the goal of discovering how this choice of norm in a steepest feasible descent method affected length of the step to the boundary, we made the following comparisons. For each linear programming problem tested, we applied the steepest feasible descent method as described in Section 2.2, for both the Dikin-Karmarkar norm and the Euclidean norm. The steps were taken a fixed fraction of the distance to the boundary. At each iteration, a comparison was made of the length of the steepest feasible descent step to the boundary for the solution method being applied, and length of the steepest feasible descent step to the boundary for the other norm; a comparison was also made between the amount of decrease in the objective function given by each steepest feasible descent step to the boundary.

The tables contain the following notation and information. The step taken to the boundary in the steepest feasible descent direction with respect to the Euclidean norm is denoted by s_E. Likewise, s_{dk} denotes the step taken to the boundary in the steepest feasible descent direction with respect to the Dikin-Karmarkar norm. The new iterate The step was taken a fixed fraction ($0 < \alpha < 1$) of the distance to the boundary. In each table, the first column gives the iteration count ITN. The second column is the ratio of the length of the Dikin-Karmarkar step, s_{dk}, to the length of the Euclidean step, s_E. Thus, a ratio greater than one indicates that the Dikin-Karmarkar step is longer. The second column compares the amount of decrease in the objective function given by taking the Dikin-Karmarkar step to the amount of decrease possible by taking the Euclidean step. Again, a value greater than one indicates that the Dikin-Karmarkar step provided the greater decrease.

Two types of problems were used in our testing. First, small, randomly generated problems were tested. Second, a subset of the Netlib linear programming test problems were tested.

7.1 Small Dense Problems

The random problems generated had from 3 to 10 variables. The linear constraint matrices were full rank and dense. The random problems tested were run with the steplength parameter α varying from 0.8 to 0.99. There was not a significant difference in the results for the different parameter values. As could be expected, with a smaller steplength the number of iterations was slightly greater than with a longer steplength. The stopping criterion utilized was that the relative error, $\| y - x^{*}\|_2/\| x^{*}\|_2 < 10^{-6}$. Representative results are for five problems are given in Tables 7.1 through 7.5, with a summary in Table 7.6. In Table 7.1, we see that for

ITN	$c^T s_{dk}/c^T s_E$	$\|s_{dk}\|/\|s_E\|$
1	1.4677	2.5484
2	5.7968	10.6109
3	2.7673	5.8767
4	1.5890	2.4118
5	2.7354	4.9937
6	1.6597	2.3401
7	3.3348	6.8309
8	1.9356	2.8371
9	1.7056	3.4690
10	1.7358	2.4666
11	3.4036	7.0678
12	1.8937	2.7170
13	1.8691	3.8283

Table 7.1 Comparison DK step and Euclidean step for RAND01

RAND01, at every iteration, the Dikin-Karmarkar step is longer and gives greater decrease. Likewise, for RAND02 and RAND03, (Tables 7.2 and 7.3). Note that in

ITN	$c^T s_{dk}/c^T s_E$	$\|s_{dk}\|/\|s_E\|$
1	1.0217	1.0287
2	6.2182	32.2994
3	1.9711	5.7909
4	1.7690	4.5970
5	2.2801	9.6942
6	1.1108	1.2678

Table 7.2 Comparison DK step and Euclidean step for RAND02

RAND03, at iteration 7, the Dikin-Karmarkar steplength is over 2,000 times greater than the Euclidean steplength and the function decrease at that iteration is more that 150 times greater. In both RAND04 and RAND05, within the first iterations, the Euclidean norm step is longer and gives greater decrease, but as the solution (and thus the boundary) is approached, in both problems, the Dikin-Karmarkar step becomes longer and gives greater decrease.

Table 7.6 gives a summary of these five problems. The first two columns give problem dimensions. The next two columns give the average function decrease ratios and steplength ratios for each problem. Note that on all problems, on the average, the Dikin-Karmarkar step was longer and gave greater decrease.

7.2 Netlib Test Problems

A subset of the smaller Netlib linear programming test set was tested. The problems are large and sparse. The results for AFIRO are shown in Table 7.7 and are representative of that obtained for this test set. (For this particular example, the step taken was 0.9 of the distance to the boundary.) We see that the relative decrease in the objective function is superior for the Dikin-Karmarkar norm and the lengths of the steps that can be taken are significantly longer than those for the Euclidean norm. On the average, the amount of objective function decrease possible from the Dikin-Karmarkar step was more that 22 times that possible from the Euclidean step; and the Dikin-Karmarkar steplength to the boundary was on the average more than 300 times that of the Euclidean steplength to the boundary.

ITN	$c^T s_{dk} / c^T s_E$	$\|s_{dk}\| / \|s_E\|$
1	3.3816	5.1067
2	1.2347	2.0131
3	5.6713	35.6394
4	3.0492	32.2240
5	2.1973	5.0041
6	7.2994	66.6234
7	156.2428	2132.7605
8	16.1168	217.8787
9	3.4182	25.2341
10	2.9379	7.6070
11	3.3591	11.9794
12	4.8396	15.6913
13	3.5877	15.4935
14	3.6922	9.7217
15	3.7458	13.8619
16	4.5560	15.0198
17	3.5331	14.6335

Table 7.3 Comparison DK step and Euclidean step for RAND03

ITN	$c^T s_{dk} / c^T s_E$	$\|s_{dk}\| / \|s_E\|$
1	0.9027	0.9073
2	2.7223	3.4063
3	0.9370	0.9389
4	1.6543	2.0211
5	1.2457	1.3742
6	1.6943	2.0838
7	1.2348	1.3599
8	1.7011	2.0941
9	1.2326	1.3570
10	1.7024	2.0962
11	1.2322	1.3564

Table 7.4 Comparison DK step and Euclidean step for RAND04

ITN	$c^T s_{dk} / c^T s_E$	$\|s_{dk}\| / \|s_E\|$
1	0.9134	0.9523
2	2.6809	6.9676
3	14.4369	52.8578
4	3.6554	60.7037
5	2.6453	96.2155
6	2.1560	13.9415
7	4.3167	114.4900
8	1.7907	11.9714
9	5.0727	133.9721
10	1.9639	12.1541
11	3.0697	67.0537
12	1.9113	22.2856
13	5.2271	140.1575
14	2.1218	12.7265

Table 7.5 Comparison DK step and Euclidean step for RAND05

PROBLEM	NUMBER of VARIABLES	NUMBER of CONSTRAINTS	AVERAGE FUNCTION RATIO	AVERAGE STEP RATIO
RAND01	9	3	2.6528	4.8332
RAND02	5	3	1.8583	5.8985
RAND03	9	1	13.2515	153.1547
RAND04	6	4	1.4781	1.4131
RAND05	10	6	3.7116	52.4497

Table 7.6 RANDOM PROBLEM SUMMARIES ($\alpha = 0.9$)

ITN	$c^T s_{dk}/c^T s_E$	$\|s_{dk}\|/\|s_E\|$
1	11.5739	175.4462
2	19.8760	282.9671
3	24.3338	356.0513
4	20.5237	291.9174
5	37.8911	567.2736
6	22.6763	318.1621
7	19.0612	275.0813
8	22.7428	321.0754
9	24.5948	356.2554
AVG	22.5860	327.1366

Table 7.7 Comparison DK step and Euclidean step for AFIRO; ($n = 51$; $m = 27$)

Chapter 8

CONCLUDING REMARKS

We have developed mathematical theory that describes both the asymptotic short step behavior of steepest feasible descent with respect to the Euclidean norm and the avoidance of short steps in steepest feasible descent with respect to the Dikin-Karmarkar norm as the boundary is approached. This theoretical behavior is borne out in practice on problems with linear equality constraints added.

We conjectured that if information about all the boundaries is incorporated into the norm, then finding such a norm that would also give us the longest step possible, might give even better numerical results for steepest feasible descent than with the Dikin-Karmarkar norm.

As we developed the theory, we restricted our attention to problems with only nonnegativity constraints:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x \geq 0, \end{array} \tag{8.1}$$

where $c > 0$. Observe that $x^* = 0$ solves this simple problem. Hence, the step s that would solve the problem in one iteration from a strictly feasible point x would be $s = x$ and this is also the longest step that can be taken among all steps that maintain feasibility and give descent. Furthermore, this is the steepest descent step for the weighted ℓ_∞ norm:

$$\|\cdot\| \equiv \|D_x^{-1} \cdot \|_\infty, \tag{8.2}$$

where $D_x = \text{diag}(x)$.

We might expect long steps and good convergence behavior if we were to use a steepest feasible descent method with respect to this weighted norm to solve the more general problem (8.3):

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0, \end{array} \tag{8.3}$$

where c, $x \in I\!R^n$, $A \in I\!R^{m \times n}$, and $b \in I\!R^m$.

However, if we use this norm in a steepest feasible descent method for problem (8.3) it is not clear how the steepest descent subproblem should be solved. (The obvious approach would require the solution of a linear programming problems and this would not lead to an efficient algorithm.) We therefore restricted our attention to weighted Euclidean norms so that the steepest feasible descent direction can be computed by evaluating a linear projection, i.e. solving a system of linear equations. We therefore considered a weighted Euclidean norm. Interestingly, we discovered that it was possible to use a weighted Euclidean norm that would give us the same "ideal" direction as our weighted infinity norm (8.2).

We consider the following weighted Euclidean norm:

$$\|\cdot\| \equiv \|W^{-1} \cdot\|_2, \tag{8.4}$$

where

$$W^2 \equiv D_x C^{-1}, \tag{8.5}$$

for $C = \operatorname{diag}(c)$ and $D_x = \operatorname{diag}(x)$. The vector, $-x$, is a steepest feasible descent direction with respect to this norm. In other words, the steepest feasible descent direction will reach the solution in one step, as does the weighted infinity norm (8.2). Utilizing this norm, the computational effort to solve the steepest feasible descent subproblem involves a matrix factorization versus the solution to a complete linear programming problem as in the case of a weighted infinity norm. We will refer to then norm satisfying defined by (8.5) as the *long-step* norm. It is clear that steepest feasible descent with respect to the long-step norm satisfies Theorem 6.2.

Using the same set of test problems discussed in Chapter 7, we ran steepest feasible descent with respect to the long-step norm. Comparisons were made between the behavior of steepest feasible descent with respect to the long-step norm and steepest feasible descent with respect to the Euclidean norm; and between the behavior for the long-step norm and the Dikin-Karmarkar norm. As expected, when comparison was made with Euclidean norm steepest feasible descent the long-step norm gave steps that were significantly longer, and also greater decrease in the objective function than was possible possible by using the Euclidean norm. However, in half of the problems tested, the long-step norm steepest feasible descent was unable to converge to the solution.

When compared to steepest feasible descent with respect to the Dikin-Karmarkar norm, when the long-step norm steepest feasible descent was able to find the solution, the long-step norm took longer steps and had greater function decrease. However, it took on the average 46% more iterations.

The major obstacle that the implementation of steepest feasible descent with respect to the long-step norm encountered was that the weighting matrix (8.5) tended to become numerically singular before a solution could be found.

Our experience with this long-step norm leads us to believe that the good convergence behavior exhibited when using the Dikin-Karmarkar norm is not solely due to the fact that it takes longer steps than the Euclidean norm. Neither can the behavior be attributed to only the fact that boundary information is incorporated into the norm. We believe that an important factor in the success of the Dikin-Karmarkar norm is the fact that all components are scaled uniformly, including those components that are zero at the solution [3].

Bibliography

[1] L. Armijo. Minimization of functions having Lipschitz-continuous first partial derivatives. *Pacific Journal of Mathematics*, 16:1–3, 1966.

[2] Mordecai Avriel. *Nonlinear Programming: Analysis and Methods.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1967.

[3] Miroslav D. Ašić, Vera V. Kovačević-Vujčić, and Mirjana D. Radosavljević-Nikolić. A note on limiting behavior of the projective and the affine rescaling algorithms. In Jeffrey C. Lagarias and Michael J. Todd, editors, *Mathematical Developments Arising from Linear Programming*, volume 114 of *Contemporary Mathematics*, pages 151—157, Providence, Rhode Island, 1990. Joint Summer Research Conference held at Bowdoin College, June 25—July 1, 1988, American Mathematical Society.

[4] A. V. Balakrishnan. An operator theoretic formulation of a class of control problems and a steepest descent method of solution. *Journal of the Society of Industrial and Applied Mathematics on Control, Series A*, 1(2):109–127, 1962.

[5] Earl R. Barnes. A variation on Karmarkar's algorithm for solving linear programming problems. *Mathematical Programming*, 36:174–182, 1986.

[6] Earl R. Barnes. Some results concerning convergence of the affine scaling algorithm. In Jeffrey C. Lagarias and Michael J. Todd, editors, *Mathematical Developments Arising from Linear Programming*, volume 114 of *Contemporary Mathematics*, pages 131—139, Providence, Rhode Island, 1990. Joint Summer Research Conference held at Bowdoin College, June 25—July 1, 1988, American Mathematical Society.

[7] R. H. Byrd and R. A. Tapia. An extension of Curry's theorem to steepest descent in normed linear spaces. *Mathematical Programming*, 9:247—254, 1975.

[8] Augustin Cauchy. Méthode générale pour la résolution des systèmes d'équations simultanée. *Compte rendu des séances de l'académie des sciences*, 25:536–538, 1847.

[9] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49:1–23, January 1943.

[10] Jean Bronfenbrenner Crockett and Herman Chernoff. Gradient methods of maximization. *Pacific Journal of Mathematics*, 5(1):33–50, March 1955.

[11] Haskel B. Curry. The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2(3):258–261, 1944.

[12] J. E. Dennis, Jr. and Robert B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1983.

[13] I. I. Dikin. Iterative solution of problems of linear and quadratic programming. *Soviet Mathematics—Doklady*, 8(3):674-675, May–June 1967.

[14] I. I. Dikin. On the speed of an iterative process. *Upravlyaemye Sistemi*, 12:54–60, 1974.

[15] Wendell Fleming. *Functions of Several Variables.* Springer-Verlag, New York, 2nd edition, 1977.

[16] R. Fletcher. Minimizing general functions subject to linear constraints. In F. A. Lootsma, editor, *Numerical Methods for Non-linear Optimization*, pages 279–296. Academic Press, London, 1972.

[17] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical Optimization.* Academic Press, London, 1981.

[18] A. A. Goldstein. Cauchy's method of minimization. *Numerische Mathematik*, 4:146–150, 1962.

[19] A. A. Goldstein. Convex programming in Hilbert space. *Bulletin of the American Mathematical Society*, 70(5):709–710, September 1964.

[20] A. A. Goldstein. Minimizing functionals on Hilbert space. In A. V. Balakrishnan and Lucien W. Neustadt, editors, *Computing methods in Optimization Problems*, pages 159–165. Academic Press, New York, 1964.

[21] A. A. Goldstein. On steepest descent. *Journal of the Society for Industrial and Applied Mathematics, Series A. Control*, 3(1):147–151, 1965.

[22] John Greenstadt. On the relative efficiencies of gradient methods. *Mathematics of Computation*, 21(99):360–367, July 1967.

[23] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[24] L. G. Khachiyan. A polynomial algorithm in linear programming. *Soviet Mathematics—Doklady*, 20(1):191–194, 1979.

[25] K. O. Kortanek and M. Shi. Convergence results and numerical experiments on a linear programming hybrid algorithm. *European Journal of Operational Research*, 1985.

[26] Vera V. Kovačević-Vujčić. Improving the rate of convergence of interior point methods for linear programming. Belgrade University, Belgrade, Yugoslavia.

[27] E. S. Levitin and B. T. Poljak. Constrained minimization techniques. *U.S.S.R. Computational Mathematics and Mathematical Physics*, 6:1–50, 1966.

[28] G. P. McCormick and R. A. Tapia. The gradient projection method and Curry's theorem. Technical report, Mathematics Research Center, University of Wisconsin, Madison, 1970.

[29] G. P. McCormick and R. A. Tapia. The gradient projection method under mild differentiability conditions. *SIAM Journal on Control*, 10(1):93–98, February 1972.

[30] Garth P. McCormick. Anti-zig-zagging by bending. *Management Science*, 15:315–320, 1969.

[31] A. M. Morshedi and R. A. Tapia. Karmarkar as a classical method. Technical Report 87-7, Departments of Mathematical Sciences, Rice University, Houston, Texas, March 1987.

[32] M. Z. Nashed. On general iterative methods for the solutions of a class of nonlinear operator equations. *Mathematics of Computation*, 19(89):14–24, January 1965.

[33] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, 1970.

[34] J. B. Rosen. Nonlinear programming. The gradient projection method. *Bulletin of the American Mathematical Society*, 63:25–26, 1957.

[35] J. B. Rosen. The gradient projection method for nonlinear programming. Part I. Linear constraints. *Journal of the Society for Industrial and Applied Mathematics*, 8(1):181–217, March 1960.

[36] J. B. Rosen. The gradient projection method for nonlinear programming. Part II. Nonlinear constraints. *Journal of the Society for Industrial and Applied Mathematics*, 9:514—532, 1961.

[37] D. F. Shanno. An accelerated gradient projection method for linearly constrained nonlinear estimation. *SIAM Journal of Applied Mathematics*, 18(2):322–334, 1970.

[38] Richard A. Tapia and Yin Zhang. Interpretation and optimality properties of the Karmarkar algorithm. Technical report, Rice University, Department of Mathematical Sciences, Houston, Texas, 1992. In Preparation.

[39] R. J. Vanderbei and J. C. Lagarias. I. I. Dikin's convergence result for the affine-scaling algorithm. In Jeffrey C. Lagarias and Michael J. Todd, editors, *Mathematical Developments Arising from Linear Programming*, volume 114 of *Contemporary Mathematics*, pages 109-119, Providence, Rhode Island, 1990. Joint Summer Research Conference held at Bowdoin College, June 25—July 1, 1988, American Mathematical Society.

[40] Robert J. Vanderbei, Marc S. Meketon, and Barry A. Freedman. A modification of Karmarkar's linear programming algorithm. Technical report, AT & T Bell Laboratories, Holmedel, New Jersey, 07733, June 1985.

[41] Yinyu Ye and Edison Tse. An extension of Karmarkar's projective algorithm for convex quadratic programming. *Mathematical Programming*, 34:157–179, 1989.

[42] G. Zoutendijk. *Methods of Feasible Directions: A Study in Linear and Non-Linear Programming.* Elsevier Publishing Company, Amsterdam, 1960.