Linear equations, Inequalities, Linear Programs (LP), and a New Efficient Algorithm

Katta G. Murty
Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48109-2117, USA
Phone: 734-763-3513, fax: 734-764-3451
e-mail:murty@umich.edu
Webpage: http://www-personal.engin.umich.edu/~ murty/

March 2006

Abstract

The dawn of mathematical modeling and algebra occurred well over 3000 years ago in several countries (Babylonia, China, India,...). The earliest algebraic systems constructed are systems of linear equations, and soon after, the famous elimination method for solving them was discovered in China and India. This effort culminated in the writing of two books that attracted international attention by the Arabic mathematician Muhammad ibn-Musa Alkhawarizmi in the first half of 9th century. The first, *Al-Maqala fi Hisab al-jabr w'almuqabilah* (An essay on Algebra and equations), was translated into Latin under the title *Ludus Algebrae*, the name "algebra" for the subject came from this Latin title, and Alkhawarizmi is regarded as the father of algebra. Linear algebra is the branch of algebra dealing with systems of linear equations. The second book *Kitab al-Jam'a wal-Tafreeq bil Hisab al-Hindi* appeared in Latin translation under the title *Algoritmi de Numero Indorum* (meaning *Alkhawarizmi Concerning the Hindu Art of Reckoning*), and the word "algorithm" for procedures for solving algebraic systems originated from this Latin title.

The elimination method for solving linear equations remained unknown in Europe until Gauss rediscovered it in 19th century while approximating by a quadratic formula the orbit of the asteroid Ceres based on recorded observations in tracking it earlier by the Italian astronomer Piazzi. Europeans gave the names "Gaussian elimination method", "GJ (Gauss-Jordan) elimination method" for this method.

However, there was no computationally viable method until recently to solve systems of linear constraints including inequalities. Examples of linear constraints with inequalities started appearing in published literature in mid-18th century. In the 19th and early 20th century Fourier,

De la Vallée Poussin, Farkas, Kantarovich, and others did initial work for solving such systems. This work culminated in the 1947 paper on the Simplex method for linear programming (LP) by George Dantzig. The simplex method is a 1-dimensional boundary method, it quickly became the leading algorithm to solve LPs and related problems. Its computational success made LP a highly popular modeling tool for decision making problems, with numerous applications in all areas of science, engineering, and business management. From the nature of the simplex method, LP can be viewed as the 20th century extension of linear algebra to handle systems of linear constraints including inequalities.

Competing with the simplex method now are a variety of interior point methods for LP developed in the last 20 years stimulated by the pioneering work of Karmarkar, these follow a central path using a logarithmically defined centering strategy. All these methods and also the simplex method need matrix inversions; their success for large scale problem solving requires taking careful advantage of sparcity in the data.

I will discuss a new interior point method based on a much simpler centering strategy that I developed recently. It is a fast efficient descent method that can solve LPs without matrix inversions, hence it can handle dense problems, and is also not affected by redundant constraints in the model.

Key words: Linear programming (LP), Dantzig's simplex method, boundary methods, gravitational methods, interior point methods, solving LPs without matrix inversions.

1 Historical Overview

1.1 Mathematical Modeling, Algebra, Systems of Linear Equations and Linear Algebra

One of the most fundamental ideas of the human mind, discovered more than 5000 years ago by the Chinese, Indians, Iranians, and Babylonians, is to represent the quantities that we like to determine by symbols; usually letters of the alphabet like x, y, z; and then express the relationships between the quantities represented by these symbols in the form of equations, and finally use these equations as tools to find out the true values represented by the symbols. The symbols representing the unknown quantities to be determined are nowadays called **unknowns** or **variables** or **decision variables**.

The process of representing the relationships between the variables through equations or other functional relationships is called **modeling** or **mathematical modeling**. The earliest mathematical models constructed are systems of linear equations, and soon after, the famous **elimination method** for solving them was discovered in China and India. The Chinese text *Chiu-Chang Suanshu* (9 Chapters on the Mathematical Art) composed over 2000 years ago describes the method using a problem of determining the yield (measured in units called "tou") from three types of grain: inferior, medium, superior; given the yield data

from three experiments each using a separate combination of the three types of grain (see [Kangshen, Crossley and Lun, 1999] for information on this ancient work, also a summary of this ancient Chinese text can be seen at the website: http://www-groups.dcs.st-and.ac.uk/˜ history/HistTopics/Nine_chapters.html). Ancient Indian texts *Sulabha suutrah* (Easy Solution Procedures) with origins to the same period describe the method in terms of solving systems of two linear equations in two variables (see [Lakshmikantham and Leela, 2000] for information on these texts, and for a summary and review on this book see: http://www.tlca.com/adults/origin-math.html).

This effort culminated around 825 AD in the writing of two books that attracted international attention, by the Arabic mathematician Muhammad ibn-Musa Alkhawarizmi . The first was *Al-Maqala fi Hisab al-jabr w'almuqabilah* (An essay on Algebra and equations). The term "al-jabr" in Arabic means "restoring" in the sense of solving an equation. In Latin translation the title of this book became *Ludus Algebrae*, the second word in this title surviving as the modern word **algebra** for the subject, and Alkhawarizmi is regarded as the father of algebra. **Linear algebra** is the name given subsequently to the branch of algebra dealing with systems of linear equations. The word *linear* in "linear algebra" refers to the "linear combinations" in the spaces studied, and the *linearity* of "linear functions" and "linear equations" studied in the subject.

The second book *Kitab al-Jam'a wal-Tafreeq bil Hisab al-Hindi* appeared in a Latin translation under the title *Algoritmi de Numero Indorum* meaning *Al-Khwarizmi Concerning the Hindu Art of Reckoning*, it was based on earlier Indian and Arabic treatises. This book survives only in its Latin translation, as all the copies of the original Arabic version have been lost or destroyed. The word **algorithm** for procedures for solving algebraic systems originated from the title of this Latin translation. Algorithms seem to have originated in the work of ancient Indian mathematicians on rules for solving linear and quadratic equations.

1.2 Elimination Method for Solving Linear Equations

We begin with an example application that leads to a model involving simultaneous linear equations. A steel company has four different types of scrap metal (called SM-1 to SM-4) with compositions given in the table below. They need to blend these four scrap metals into a mixture for which the composition by weight is: Al - 4.43%, Si - 3.22%, C - 3.89%, Fe - 88.46%. How should they prepare this mixture ?

Type	% in type, by weight, of element			
	Al	Si	C	Fe
SM-1	5	3	4	88
SM-2	7	6	5	82
SM-3	2	1	3	94
SM-4	1	2	1	96

To answer this question, we first define the decision variables, denoted by x_1, x_2, x_3, x_4, where for $j = 1$ to 4, x_j = proportion of SM-j by weight in the mixture to be prepared. Then the percentage by weight, of the element Al in the mixture will be, $5x_1 + 7x_2 + 2x_3 + x_4$, which is required to be 4.43. Arguing the same way for the elements Si, C, and Fe, we find that the decision variables x_1 to x_4 must satisfy each equation in the following system of linear equations in order to lead to the desired mixture:

$$\begin{aligned} 5x_1 + 7x_2 + 2x_3 + x_4 &= 4.43 \\ 3x_1 + 6x_2 + x_3 + 2x_4 &= 3.22 \\ 4x_1 + 5x_2 + 3x_3 + x_4 &= 3.89 \\ 88x_1 + 82x_2 + 94x_3 + 96x_4 &= 88.46 \\ x_1 + x_2 + x_3 + x_4 &= 1 \end{aligned}$$

The last equation in the system stems from the fact that the sum of the proportions of various ingradients in a blend must always be equal to 1. From the definition of the variables given above, it is clear that a solution to this system of equations makes sense for the blending application under consideration, only if all the variables in the system have nonnegative values in it. The nonnegativity restrictions on the variables are **linear inequality constraints**. They cannot be expressed in the form of linear equations, and since nobody knew how to handle linear inequalities at that time, they ignored them and considered this system of equations as the mathematical model for the problem.

To solve a system of linear equations, each step in the elimination method uses one equation to express one variable in terms of the others, then uses that expression to eliminate that variable and that equation from the system leading to a smaller system. The same process is repeated on the remaining system. The work in each step is organized conveniently through what is now-a-days called the **Gauss-Jordan (GJ) pivot step**. We will illustrate this step on the following system of 3 linear equations in 3 decision variables given in the following detached coefficient tableau at the top. In this representation, each row in the tableau corresponds to an equation in the system, and RHS is the column vector of right hand side constants in the various equations. Normally the equality symbol for the equations is omitted.

Basic Variable	x_1	x_2	x_3	RHS
	$\boxed{1}$	-1	-1	10
	-1	2	-2	20
	1	-2	-4	30
x_1	1	-1	-1	10
	0	1	-3	30
	0	-1	-3	20

In this step on the system given in the top tableau, we are eliminating the

variable x_1 from the system using the equation corresponding to the first row. The column vector of the variable eliminated, x_1, is called the *pivot column*, and the row of the equation used to eliminate the variable is called the *pivot row* for the pivot step, the element in the pivot row and pivot clolumn, known as the *pivot element*, is boxed. The pivot step converts the pivot column into the unit column with "1" entry in the pivot row and "0" entries in all the other rows. In the resulting tableau after this pivot step is carried out, the variable eliminated, x_1, is recorded as the *basic variable* in the pivot row. This row now contains an expression for x_1 as a function of the remaining variables. The other rows contain the remaining system after x_1 is eliminated, the same process is now repeated on this system.

When the method is continued on the remaining system two things may occur: (a): all the entries in a row may become 0, this is an indication that the constraint in the corresponding row in the original system is a redundant constraint, such rows are eliminated from the tableau; and (b): the coefficients of all the variables in a row may become 0, while the RHS constant remains nonzero, this indicates that the original system of equations is inconsistent, i.e., it has no solution, if this occurs the method terminates.

If the inconsistency termination does not occur, the method terminates after performing pivot steps in all the rows. If there are no nonbasic variables at that stage, equating each basic variable to the RHS in the final tableau gives the unique solution of the system. If there are nonbasic variables, from the rows of the final tableau we get the general solution of the system in parametric form in terms of the nonbasic variables as parameters.

The elimination method remained unknown in Europe until Gauss rediscovered it at the beginning of the 19th century while calculating the orbit of the asteroid Ceres based on recorded observations in tracking it earlier. It was lost from view when the astronomer tracking it, Piazzi, fell ill. Gauss got the data from Piazzi, and tried to approximate the orbit of Ceres by a quadratic formula using that data. He designed the method of least squares for estimating the best values for the parameters to give the closest fit to the observed data, this gives rise to a system of linear equations to be solved. He rediscovered the elimination method to solve that system. Even though the system was quite large for hand computation, Gauss's accurate computations helped in relocating the asteroid in the skies in a few months time, and his reputation as a mathematician soared.

Europeans gave the names **Gaussian elimination method**, **Gauss-Jordan elimination method** to two variants of the method at that time. These methods are still the leading methods in use today for solving systems of linear equations.

1.3 Lack of a Method To Solve Linear Inequalities Until Modern Times

Even though linear equations had been conquered thousands of years ago, systems of linear inequalities remained inaccessible until modern times. The set of feasible solutions to a system of linear inequalities is called a **polyhedron** or

convex polyhedron, and geometric properties of polyhedra were studied by the Egyptians earlier than 2000 BC while building the pyramids, and later by the Greeks, Chinese, Indians and others.

The following theorem (for a proof see [Murty, 2006]) relates systems of linear inequalities to systems of linear equations.

Theorem 1: If the system of linear inequalities: $A_{i.}x \geq b_i, \quad i = 1$ to m in variables $x = (x_1, \ldots, x_n)^T$ has a feasible solution, then there exists a subset $\mathbf{P} = \{p_1, \ldots, p_s\} \subset \{1, \ldots, m\}$ such that every solution of the system of linear equations: $A_{i.}x = b_i, \quad i \in \mathbf{P}$ is also feasible to the orignal system of linear inequalities.

A paradox: Theorem 1 presents an interesting paradox.

As you know, linear equations can be transformed into linear inequalities by replacing each equation with the opposing pair of inequalities. But there is no way a linear inequality can be transformed into linear equations. This indicates that linear inequalities are more fundamental than linear equations.

But this theorem shows that linear equations are the key to solving linear inequalities, and hence are more fundamental. This is the paradox.

Theorem 1 provides an enumerative approach for solving a system of linear inequalities, involving enumeration over subsets of the inequalities treated as equations. But the effort required by the method grows exponentially with the number of inequalities in the system in the worst case.

1.4 The Importance of Linear Inequality Constraints, and Their Relation to Linear Programs

The first interest in inequalities arose from studies in mechanics, beginning with the 18th century.

Linear programming (LP) involves optimization of a linear objective function subject to linear inequality constraints. Crude examples of LP models started appearing in published literature from about the mid-18th century. We will now present an example of a simple application of LP from the class of **product mix models** from [Murty, 2005-2, or 1983].

A fertilizer company makes two kinds of fertilizers called Hi-phosphate (Hi-ph) and Lo-phosphate (Lo-ph). The manufacture of these fertilizers requires three raw materials called RM 1, 2, 3. At present their supply of these raw materials comes from the company's own quarry which is only able to supply maximum amounts of 1500, 1200, 500 tons/day respectively of RM 1, RM 2, RM 3. Even though there are other vendors who can supply these raw materials if necessary, at the moment they are not using these outside suppliers.

They sell their output of Hi-ph and Lo-ph fertilizers to a wholesaler who is willing to buy any amount that they can produce, so there are no upper bounds on the amounts of Hi-ph and Lo-ph manufactured daily.

At the present rates of operation their Cost Accounting Department estimates that it is costing the quarry \$50, 40, 60/ton respectively to produce and deliver RM 1, RM 2, RM 3 at the fertilizer plant. Also, at the present rates of operation, all other production costs (for labor, power, water, maintenance, depreciation of plant and equipment, floorspace, insurance, shipping to the wholesaler, etc.) come to \$7/ton to manufacture Hi-ph or Lo-ph and to deliver them to the wholesaler.

The sale price of the manufactured fertilizers to the wholesaler fluctuates daily, but their averages over the last one month have been \$222, 107/ton respectively for Hi-Ph, Lo-ph fertilizers.

The Hi-ph manufacturing process needs as inputs 2 tons of RM 1, and 1 ton each of RM 2, RM 3 for each ton of Hi-ph manufactured. Similarly the Lo-ph manufacturing process needs as inputs 1 ton of RM 1, and 1 ton of RM 2 for each ton of Lo-ph manufactured. So, the net profit/ton of fertilizer manufactured is $\$(222 - 2 \times 50 - 1 \times 40 - 1 \times 60 - 7) = 15$, $(107 - 1 \times 50 - 1 \times 40 - 7) = 10$/respectively for Hi-ph, Lo-ph.

We will model the problem with the aim of determining how much of Hi-ph and Lo-ph to make daily to maximize the total daily net profit from these fertilizer operations. There are clearly two decision variables; these are:

$$\begin{aligned} x_1 &= \text{the tons of Hi-ph made per day} \\ x_2 &= \text{the tons of Lo-ph made per day} \end{aligned}$$

Since all the data is given on a per ton basis, it provides an indication that the linearity assumptions (proportionality, additivity) are quite reasonable in this problem to express each of the constraint and the objective functions. Also, the amount of each fertilizer manufactured can vary continuously within its present range. So, LP is an appropriate model for this problem. The LP formulation of this fertilizer product mix problem is given below. Each costraint in the model is the material balance inequality of the *item* shown against it.

$$\begin{array}{lrcrcrll} \text{Maximize} & z(x) = 15x_1 & + & 10x_2 & & & \text{Item} & \\ \text{Subject to} & 2x_1 & + & x_2 & \leq & 1500 & \text{RM 1} & \\ & x_1 & + & x_2 & \leq & 1200 & \text{RM 2} & (1) \\ & x_1 & & & \leq & 500 & \text{RM 3} & \\ & x_1 & \geq 0, & x_2 & \geq & 0 & & \end{array}$$

In this example all the constraints on the variables are inequality constraints. In the same way, inequality constraints appear much more frequently and prominently than equality constraints in most real world applications. In fact we can

go as far as to assert that in most applications in which a linear model is the appropriate one to use, most of the constraints are actually linear inequalities, and linear equations play only the role of a computational tool through approximations, or through results similar to Theorem 1. Linear equations were used to model problems mostly because an efficient method to solve them is known.

Fourier was one of the first to recognize the importance of inequalities as opposed to equations for applying mathematics. Also, he is a pioneer who observed the link between linear inequalities and linear programs, in early the 19th century.

For example, the problem of finding a feasible solution to the following system of linear inequalities (2) in x_1, x_2, can itself be posed as another LP for which an initial feasible solution is readily available. Formulating this problem known as a **Phase I problem** introduces one or more nonnegative variables known as **artificial variables** into the model. All successful LP algorithms require an initial feasible solution at the start, so the Phase I problem can be solved using any of those algorithms, and at termination it either outputs a feasible solution of the original problem, or an evidence for its infeasibility. The Phase I model for finding a feasible solution for (2) is (3), it uses one artificial variable x_3.

$$\begin{array}{rrcr} x_1 & +2x_2 & \geq & 10 \\ 2x_1 & -4x_2 & \geq & 15 \\ -x_1 & +10x_2 & \geq & 25 \end{array} \qquad (2)$$

$$\begin{array}{lrrrcr} \text{Minimize} & & & x_3 & & \\ \text{Subject to} & x_1 & +2x_2 & +x_3 & \geq & 10 \\ & 2x_1 & -4x_2 & +x_3 & \geq & 15 \\ & -x_1 & +10x_2 & +x_3 & \geq & 25 \\ & & & x_3 & \geq & 0 \end{array} \qquad (3)$$

For the Phase I problem (3), $(x_1, x_2, x_3)^T = (0, 0, 26)^T$ is a feasible solution. In fact solving such a Phase I problem provides the most efficient approach for solving systems of linear inequalities.

Also, the duality theory of linear programming shows that any linear program can be posed as a problem of solving a system of linear inequalities without any optimization. Thus solving linear inequalities, and LPs, are mathematically equivalent problems. Both problems of comparable sizes can be solved with comparable efficiencies by available algorithms. So, the additional aspect of "optimization" in linear programs does not make LPs any harder either theoretically or computationally.

1.5 Elimination Method of Fourier for Linear Inequalities

By 1827 Fourier generalized the elimination method to solve a system of linear inequalities. The method now known as the **Fourier** or **Fourier-Motzkin elimination method** is one of the earliest methods proposed for solving systems of linear inequalities. It consists of successive elimination of variables from the system. We will illustrate one step in this method using an example in which we will eliminate the variable x_1 from the following system.

$$\begin{array}{rcr} x_1 - 2x_2 + x_3 & \leq & 6 \\ 2x_1 + 6x_2 - 8x_3 & \leq & -6 \\ -x_1 - x_2 - 2x_3 & \leq & 2 \\ -2x_1 - 6x_2 + 2x_3 & \leq & 2 \end{array}$$

x_1 appears with a positive coefficient in the 1st and 2nd constraints; and a negative coefficient in the 3rd and 4th constraints. By making the coefficient of x_1 in each constraint into 1, these constraints can be expressed as:

$$\begin{array}{rcl} & x_1 & \leq 6 + 2x_2 - x_3 \\ & x_1 & \leq -3 - 3x_2 + 4x_3 \\ -2 - x_2 - 2x_3 \leq & x_1 & \\ -1 - 3x_2 + x_3 \leq & x_1 & \end{array}$$

The remaining system after x_1 is eliminated is therefore:

$$\begin{array}{rcl} -2 - x_2 - 2x_3 & \leq & 6 + 2x_2 - x_3 \\ -2 - x_2 - 2x_3 & \leq & -3 - 3x_2 + 4x_3 \\ -1 - 3x_2 + x_3 & \leq & 6 + 2x_2 - x_3 \\ -1 - 3x_2 + x_3 & \leq & -3 - 3x_2 + 4x_3 \end{array}$$

and then $\max\{-2-x_2-2x_3, -1-3x_2+x_3\} \leq x_1 \leq \min\{6+2x_2-x_3, -3-3x_2+4x_3\}$ is used to get a value for x_1 in a feasible solution when values for other variables are obtained by applying the same steps on the remaining problem successively.

However starting with a system of m inequalities, the number of inequalities can jump to $O(m^2)$ after eliminating only one variable from the system, so this method is not practically viable except for very small problems.

1.6 History of the Simplex Method for LP

In 1827 Fourier published a geometric version of the principle behind the simplex algorithm for a linear program (vertex to vertex descent along the edges to an optimum, a rudimentary version of the simplex method) in the context

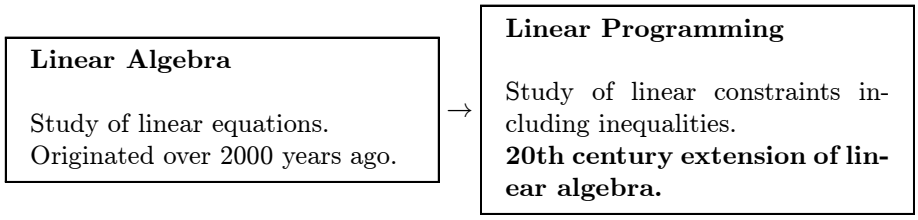
of a specific LP in 3 variables (an LP model for a Chebyshev approximation problem), but did not discuss how this descent can be accomplished computationally on systems stated algebraically. In 1910 De la Vallée Poussin designed a method for the Chebyshev approximation problem that is an algebraic and computational analogue of this Fourier's geometric version, this procedure is essentially the primal simplex method applied to that problem.

In a parallel effort [Gordon, 1873], [Farkas, 1895], and [Minkowski, 1896] studied linear inequalities, and laid the foundations for the algebraic theory of polyhedra and derived necessary and sufficient conditions for a system of linear constraints including linear inequalities to have a feasible solution.

Studying LP models for organizing and planning production, [Kantarovich, 1939] developed ideas of dual variables ('resolving multipliers') and derived a dual-simplex type method for solving a general LP.

Full citations for references before 1939 mentioned so far can be seen from the list of references in [Danizig, 1963] or [Schrijver, 1986].

This work culminated in mid-20th century with the development of the primal simplex method by Dantzig. This was the first complete, practically and computationally viable method for solving systems of linear inequalities. So, LP can be considered as the branch of mathematics which is an extension of linear algebra to solve systems of linear inequalities. The development of LP is a landmark event in the history of mathematics, and its application brought our ability to solve general systems of linear constraints (including linear equations, inequalities) to a state of completion.



2 The Importance of LP

LP has now become a dominant subject in the development of efficient computational algorithms, study of convex polyhedra, and in algorithms for decision making. But for a short time in the beginning, its potential was not well recognized. Dantzig tells the story of how when he gave his first talk on LP and his simplex method for solving it at a professional conference, Hotelling (a burly person who liked to swim in the sea, the popular story about him was that when he does, the level of the ocean raises perceptibly, see Figures 1, 2; my thanks to Katta Sriramamurthy for these figures) dismissed it as unimportant since everything in the world is nonlinear. But Von Neumann came to the

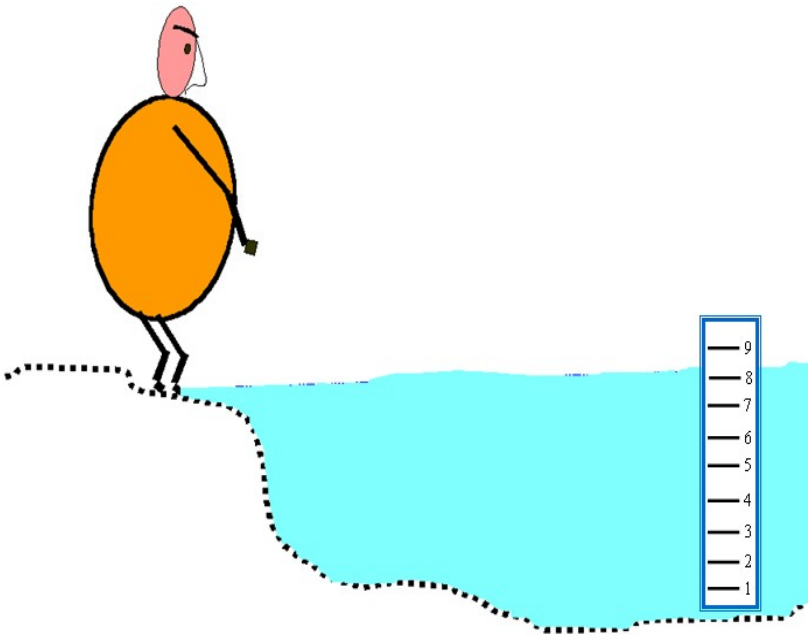


Figure 1: Hotelling (a whale of a man) getting ready to swim in the ocean.

defense of Dantzig saying that the subject will become very important. (For an account of Von Neumann's comments at this conference see Page xxvii of [Dantzig, Thapa, 1997]). The preface in this book contains an excellent account of the early history of LP from the inventor of the most successful method in OR and in the mathematical theory of polyhedra.

Von Neumann's early assessment of the importance of LP (Von Neumann [40]) turned out to be astonishingly correct. Today, the applications of LP in almost all areas of science are so numerous, so well known and recognized that they need no enumeration. Also, LP seems to be the basis for most of the efficient algorithms for many problems in other areas of mathematical programming. Many of the successful approaches in nonlinear programming, discrete optimization, and other branches of optimization are based on LP in their iterations. Also, with the development of duality theory and game theory [Gale, 1960], LP has also assumed a central position in economics.

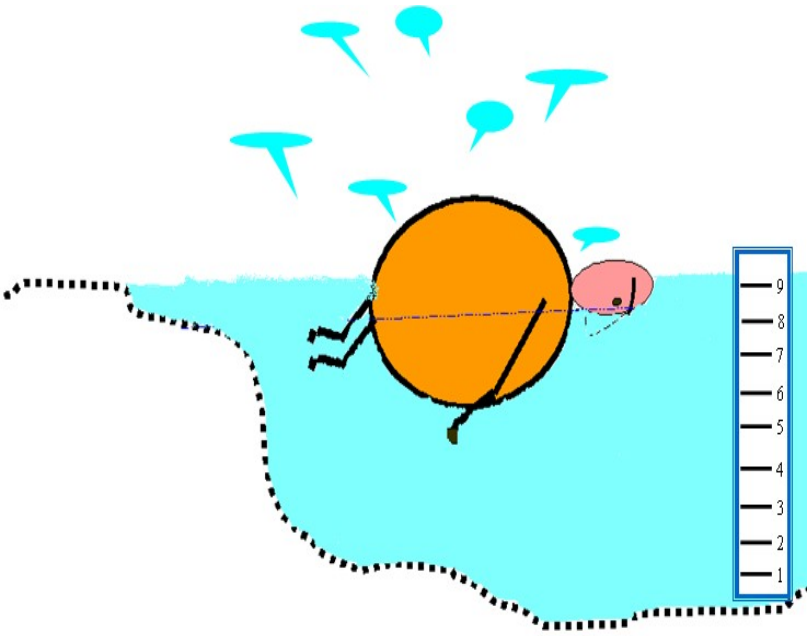


Figure 2: Hotelling swimming in the ocean. Watch the level of the ocean go up.

3 Dantzig's contributions to linear algebra, convex polyhedra, OR, computer science

A lot has been written about Dantzig's contributions. Also, he has a personal assessment of his own contributions in Chapter 1 of his book [Dantzig, 1963]. As someone who started learning LP from his course at Berkeley, I will summarize here some of his contributions that are usually overlooked in other statements (for a brief account of my experiences with Dantzig see [Murty, 2005-3]).

Contributions to OR

The simplex method is the first effective computational algorithm for one of the most versatile mathematical models in OR. Even though LP and also the simplex method for solving it originated much earlier than Dantzig's work as explained in Section 1.6, it started becoming prominent only with Dantzing's work, and OR was just beginning to develop around that time. The success of the simplex method is one of the root causes for the pheonominal development and the maturing of LP, mathematical programming in general, and OR, in the 2nd half of the 20th century.

Contributions to Linear Algebra and Computer Science

Recognizing the Irrelevance of the "RREF" Concept Emphasized in Mathematics Books on Linear Algebra: Dantzig contributed important pedagogic improvements to the teaching of linear algebra. He would state all the algorithmic steps in the GJ elimination method using the fundamental tool of row operations on the detached coefficient tableau for the system with the variable corresponding to each column entered in a top row in every tableau. This makes it easier for young students to see that the essence of this method is to take linear combinations of equations in the original system to get an equivalent but simpler system from which a solution can be read out. In descriptions of the GJ method in most mathematics books on linear algebra, the variables are usually left out.

Also, these books state the termination condition in the GJ elimination method to be that of reaching the RREF (reduced row echelon form, a tableau is defined to be in RREF if it contain a full set of unit vectors in proper order at the left end). Dantzig (and of course a lot of other OR people) realized that it is not important that all unit vectors be at the left end of the tableau (they can be anywhere and can be scattered all over); also it is not important that they be in proper order from left to right. He developed the very simple **data structure** (this phrase means a strategy for storing information generated during the algorithm, and using it to improve the efficiency of that algorithm (perhaps this is the first instance of such a structure in computational algorithms)) of associating the variable corresponding to the rth unit vector in the final tableau as the rth basic variable (or basic variable in the rth row); and storing these basic variables in a column on the tableau as the algorithm progresses. This data structure makes it easier to read the solution directly from the final tableau of the GJ elimination method by making all nonbasic variables = 0; and the rth basic variable = the rth updated RHS constant, for all r. Dantzig called this final tableau the **canonical tableau** to distinguish it from the mathematical concept of RREF. It also opened the possibility of pivot column selection strategies instead of always selecting the leftmost eligible column in this method.

Even today it is sad that in courses on linear algebra in mathematics departments, the RREF is emphasized as the output of the GJ elimination method. For a more realistic statement of the GJ method from an OR perspective see [Murty, 2004].

Evidence (or Certificate) of Infeasibility: A fundamental theorem of linear algebra asserts that a system of linear equations is infeasible iff there is a linear combination of equations in the system which is the **fundamental inconsistent equation** "$0 = a$" (where a is some nonzero number). Mathematically in matrix notation the statement of this theorem is: "Either the system $Ax = b$ has a solution (column) vector x, or there exists a row vector π satisfying $\pi A = 0, \pi b \neq 0$". The coefficient vector π in this linear combination is called an **evidence** (or **certificate) of infeasibility** for the original

system $Ax = b$.

But with the usual descriptions of the GJ elimination method to get an RREF or canonical tableau, this evidence is not available when the infeasibility conclusion is reached. An important contribution of Dantzig, the **revised simplex method**, has very important consequences to the GJ elimination method. When the GJ elimination method is executed in the revised simplex format, pivot computations are not performed on the original system (it remains unchanged throughout the algorithm), but only carried out on an auxiliary matrix set up to accumulate the basis inverse, and all the computations in the algorithm are carried out using this auxiliary matrix and the data from the original system. We will call this auxiliary matrix the **memory matrix**. For solving $Ax = b$ where A is of order $m \times n$, the initial memory matrix is the unit matrix of order m set up by the side of the original system. For details of this implementation of the GJ elimination method see Section 4.11 in [Murty, 2004].

We will illustrate this with a numerical example. At the top in the tableau given below is the original system in detached coefficient form on the right, and the memory matrix on its left. At the bottom we show the final tableau (we show the canonical tableau on the right just for illustration, actually it will not be computed in this implementation). BV = basic variable selected for the row, MM = memory matrix.

BV	MM			Original system				RHS
				x_1	x_2	x_3	x_4	
	1	0	0	1	−1	1	−1	5
	0	1	0	−1	2	2	−2	10
	0	0	1	0	1	3	−3	17
				Canonical tableau				
x_1	2	1	0	1	0	4	−4	20
x_2	1	1	0	0	1	3	−3	15
	−1	−1	1	0	0	0	0	2

The third row in the final tableau represents the inconsistent equation "0 = 2" which shows that the original system is infeasible. The row vector of the memory matrix in this row, $(1, 1, -1)$, is the coefficient vector for the linear combination of equations in the original system that produces this inconsistent equation, it is the certificate of infeasibility for this system.

Contributions to the Mathematical Study of Convex Polyhedra

Dantzig has made fundamental contributions to the mathematical study of convex polyhedra (a classical subject being investigated by mathematicians for more than 2000 years) when he introduced the complete version of the primal simplex method as a computational tool.

We could only see drawings of 2-dimensional polyhedra before this work. Polyhedra in higher dimensions could only be visualized through imagination.

The primal simplex pivot step is the first computational step for actually tracing an edge (either bounded or unbounded) of a convex polyhedron. It opened a revolutionary new computational dimension in the mathematical study of convex polyhedra, and made it possible to visualize and explore higher dimensional polyhedra through computation. At a time when research on convex polyhedra was beginning to stagnate, the simplex method has reignited the spark, and enriched their mathematical study manyfold.

4 Algorithms used for solving LPs today

Now we will summarize the main ideas behind algorithms used for solving LPs today.

4.1 Objective Plane Sliding Geometric Method for 2-Variable LPs

This simple visual geometric method is useful for solving LPs involving only 2 variables by hand. Let $z(x)$ be the linear objective function that we are trying to optimize. First the feasible region is drawn on paper by hand, and then a feasible solution $\bar{x}$ identified in it visually. Then the objective plane (a straight line in R^2) through $\bar{x}$ represented by $z(x) = z(\bar{x})$ is drawn. Changing the RHS constant in the equation for this line (i.e., changing the objective value) is equivalent to moving this straight line parallel to itself. This objective straight line is moved parallelly in the desired direction until it reaches a stage where it is still intersecting the feasible region, but any further move in the desired direction will make it loose contact with the feasible region. The intersection of the objective straight line in this final position with the feasible region is the set of optimum solutions of the problem.

In the fertilizer product mix problem (1) from Section 1.4, we start with the feasible point $\bar{x} = (0, 0)$ with an objective value z_0 of 0. As z_0 is increased from 0, the line $15x_1 + 10x_2 = z_0$ moves up keeping a nonempty intersection with the feasible region, until the line coincides with the dashed line $15x_1+10x_2 = 13,500$ in Figure 3 passing through the point of intersection of the two lines:

$$\begin{aligned} 2x_1 + x_2 &= 1500 \\ x_1 + x_2 &= 1200 \end{aligned}$$

which is $\hat{x} = (300, 900)$. For any value of $z_0 > 13,500$ the line $15x_1 + 10x_2 = z_0$ does not intersect the feasible region. Hence, the optimum objective value in this problem is \$13,500, and the optimum solution of the problem is $\hat{x} = (300, 900)$. Hence the fertilizer maker achieves his maximum daily net profit of \$13,500 by manufacturing 300 tons of Hi-ph and 900 tons of Lo-ph daily.

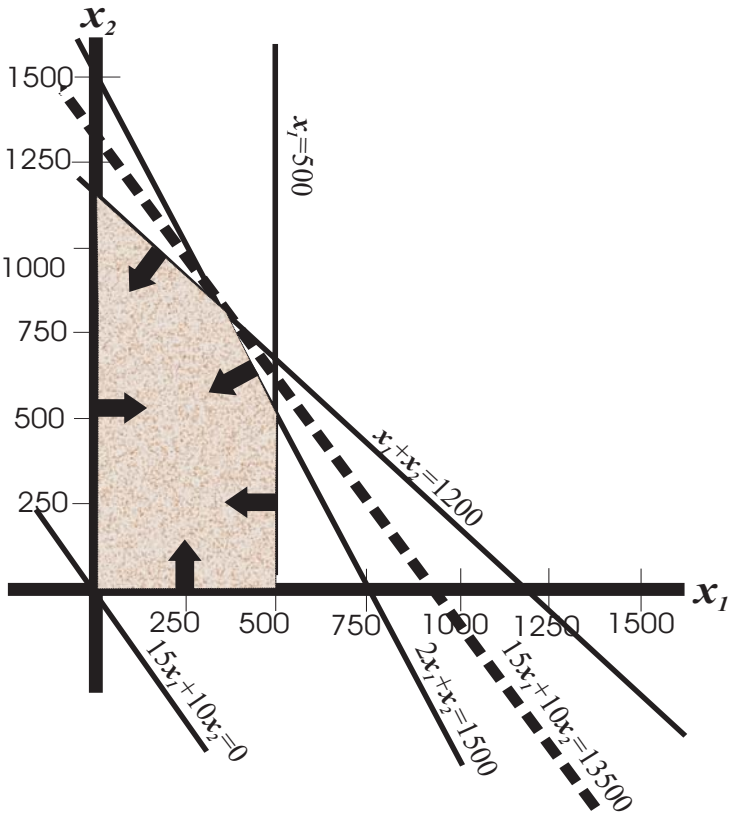


Figure 3: Solution of the Fertilizer product mix problem by the geometric method.

We cannot draw feasible regions for higher dimensional LPs, so we cannot select an initial feasible solution for them visually (this itself requires solving another LP, a Phase I problem), and we cannot visually check whether the objective plane can be moved further in the desired direction without losing contact with the feasible region. Because this geometric method requires such a high degree of visibility, it has not been generalized to solving LPs of higher dimensions so far. We will show later that the new algorithm discussed in Section 6 is a generalization of this geometric method to higher dimensions made possible computationally through the centering step in it.

4.2 The Simplex Family of Methods (1-Dimensional Boundary Methods)

The simplex method is still the dominant algorithm in use for solving LPs. It exhibits exponential growth in the worst case, but its performance in practice

has been outstanding, and is being improved continually by developments in implementation technologies. There are many variants of the simplex method, the most prominent being the primal simplex method. This method needs an initial feasible basic vector for the primal. If a primal feasible basic vector is not available, the method introduces artificial varibles into the problem and sets up a Phase I problem with a readily available feasible basic vector consisting of artificial basic variables. When this Phase I problem is solved by the same algorithm, at termination it either provides a feasible basic vector for the original primal, or a proof that it is infeasible.

Initiated with a feasible basic vector for the problem, the method goes through a series of GJ pivot steps exchanging one nonbasic variable for a basic variable in each (this type of basic vector change by one variable is the common feature of all variants of the simplex method). In each nondegenerate pivot step the method moves along an edge (a **1-dimensional boundary face or corner**) of the feasible region from one basic feasible solution to an adjacent one, and the objective value strictly improves. We will illustrate with a pivot step carried out for solving the fertilizer problem (1). To solve this problem by the primal simplex method, the constraints are converted into equations by introducing slack variables s_1, s_2, s_3. Here is the original tableau which is also the canonical tableau with respect to the basic vector (s_1, s_2, s_3). BV = basic variable selected in the row, PC = pivot column, PR = pivot row.

Original tableau

BV	x_1	x_2	s_1	s_2	s_3	$-z$	RHS	Ratio	
s_1	2	1	1	0	0	0	1500	1500/2	
s_2	1	1	0	1	0	0	1200	1200/1	
s_3	[1]	0	0	0	1	0	500	500/1	PR
$-z$	15	10	0	0	0	1	0	Min = 500	

All variables ≥ 0, maximize z

Tableau after the pivot step

BV	x_1	x_2	s_1	s_2	s_3	$-z$	RHS
s_1	0	-1	1	0	-2	0	500
s_2	0	1	0	1	-1	0	700
x_1	1	0	0	0	1	0	500
$-z$	0	10	0	0	-15	1	-7500

The initial basic vector (s_1, s_2, s_3) corresponds to the initial BFS $(x_1^1, x_2^1, s_1^1, s_2^1, s_3^1)^T = (0, 0, 1500, 1200, 500)^T$ which corresponds to the point $x^1 = (x_1^1, x_2^1)^T = (0, 0)^T$ in the x_1, x_2-space in Figure 3 of the feasible region for this problem.

A nonbasic variable is eligible to enter this basic vector if its updated objective coefficient (i.e., coefficient in the objective row in the canonical tableau) has the appropriate sign to improve the objective value (positive for maximization, negative for minimization). If there are no nonbasic variables eligible to enter the present feasible basic vector, the present BFS is an optimum solution to the problem, and the method terminates.

In this tableau both nonbasic variables x_1, x_2 are eligible to enter the basic vector, among them we selected x_1 as the entering variable, and its column vector in the present canonical tableau becomes the pivot column for this pivot step. If there are no positive entries among the constraint rows in the pivot column, the objective function is unbounded (unbounded above if the original problem is a maximization problem, or unbounded below if it is a minimization problem) on the feasible region, and again the method terminates.

If unbounded termination did not occur, the dropping basic variable that the entering variable will replace is determined using the primal simplex minimum ratio test to guarantee that the next basic vector will also remain feasible. For this in each row in which the pivot column has a positive entry, the ratio of the updated RHS constant in that row divided by the entry in the pivot column is computed. The smallest of these ratios is called the minimum ratio, and a row in which it occurs is selected as the pivot row for the pivot operation, and the present basic variable in that row is the dropping variable that will be replaced by the entering variable in the next basic vector.

It is s_3 here, hence the row in which s_3 is basic, row 3, is the pivot row for this pivot step. The tableau at the bottom is the canonical tableau with respect to the basic vector $(s_1, s_2, x_1)^T$ obtained after this pivot step. Its BFS corresponds to the extreme point solution $x^2 = (x_1^2, x_2^2)^T = (500, 0)^T$ in the x_1, x_2-space of Figure 3, it is an adjacent extreme point of x^1. Thus in this pivot step the primal simplex method has moved from x^1 to x^2 along the edge of the feasible region joining them, increasing the objective value from 0 to \$7500 in this process. The method continues from x^2 in the same way.

Each step of the simplex method requires the updating of the basis inverse as the basis changes in one column.

Since the method follows a path along the edges (one-dimensional boundary faces or corners) of the set of feasible solutions of the LP, it is classified as a **one-dimensional boundary method**.

4.3 Introduction to Earlier Interior Point Methods for LP

In early 1980s Karmarkar pioneered a new method for LP, an interior point method [Karmarkar, 1984]. Claims were made that this method would be many times faster than simplex method for solving large scale sparse LPs, these claims helped focus researchers attention on it. His work attracted worldwide attention, not only from operations researchers, but also from scientists in other areas. Let me relate a personal experience. When news of his work broke out in world press I was returning from Asia. The person sitting next to me on the flight was a petroleum geologist. When he learned that I am on the OR faculty at Michigan, he asked me excitedly "I understand that an OR scientist from India at Bell labs made a discovery that is going to revolutionize petroleum exploration. Do you know him?!"

In talks on his algorithm that he gave at that time Karmarker repeatedly emphasized the following points: (I) the boundary of a convex polyhedron with its faces of varying dimensions has a highly complex combinatorial structure.

Any method that operates on the boundary or close to the boundary will get caught up in this combinatorial complexity, and there is a limit on improvements we can make to its efficiency, (II) methods that operate in the central portion of the feasible region in the direction of descent of the objective function have the ability to take longer steps towards the optimum before being stopped by the boundary, and hence have the potential of being more efficient than boundary methods for larger problems, (III) from an interior point, one can move in any direction locally without violating feasibility, hence powerful methods of unconstrained optimization can be brought to bear on the problem.

Researchers saw the validity of these arguments, hence his talks stimulated a lot of work on these methods that stay "away" from the boundary. In the tidal wave of research that ensued, many different classes of interior point methods have been developed for LP, and extended to wider classes of problems including convex quadratic programming, monotone linear complementarity problem, and semidefinite programming problems.

Definition of An Interior Feasible Solution, and How to Modify the Problem to Have an Initial Interior Feasible Solution Available

In LP literature, an **interior feasible solution**, (also called **strictly feasible solution**) to an LP model is defined to be a feasible solution at which all inequality constraints including bound restrictions on individual variables in the model are satisfied as strict inequalities, but any equality constraints in the model are satisfied as equations. Most interior point methods need an initial interior feasible solution to start the method. If an interior feasible solution to the model is not available, the problem can be modified by introducing one artificial variable using the big-M strategy into a Phase I problem for which an initial interior feasible solution is readily available. We show these modifications first. Suppose the problem to be solved is in the form:

$$\begin{array}{lll} \text{Minimize} & cx & \\ \text{subject to} & Ax & \geq \quad b \end{array}$$

where A is a matrix of order $m \times n$. For LPs in this form, typically $m \geq n$. Introducing the nonnegative artificial variable x_{n+1}, the Phase I modification of the original problem is:

$$\begin{array}{lll} \text{Minimize} & cx + Mx_{n+1} & \\ \text{subject to} & Ax + ex_{n+1} & \geq \quad b \\ & x_{n+1} & \geq \quad 0 \end{array}$$

where $e = (1, \ldots, 1)^T \in R^m$, and M is a positive number significantly larger than any other number in the problem. Let $x^0_{n+1} > \max\{0, b_1, b_2, \ldots, b_m\}$. Then $(0, \ldots, 0, x^0_{n+1})^T$ is an interior feasible solution of the Phase I modification

which is in the same form as the original problem. If the original problem has an optimum solution, and M is sufficiently large, then the artificial variable x_{n+1} will be 0 at an optimum solution of the Phase I modification.

Now suppose the original problem is in the form:

$$\begin{array}{rrcl} \text{Minimize} & cx & & \\ \text{subject to} & Ax & = & b \\ & x & \geq & 0 \end{array}$$

where A is a matrix of order $m \times n$. For LPs in this form typically $n > m$, and an interior feasible solution is one which is strictly > 0. Select an arbitrary vector $x^0 \in R^n$, $x^0 > 0$; generally one chooses $x^0 = (1, \ldots, 1)^T$, the n-vector of all 1s. If x^0 happens to be feasible to the problem, it is an interior feasible solution, done. Otherwise, let $A_{.n+1} = b - Ax^0$. The Phase I modification including the nonnegative artificial variable x_{n+1} is:

$$\begin{array}{rrcl} \text{Minimize} & cx + Mx_{n+1} & & \\ \text{subject to} & Ax + A_{.n+1}x_{n+1} & = & b \\ & x, x_{n+1} & \geq & 0. \end{array}$$

It is easily confirmed that (x^0, x^0_{n+1}), where $x^0_{n+1} = 1$ is an interior feasible solution of the Phase I problem which is in the same form as the original problem. Again, if the original problem has an optimum solution and M is sufficiently large, then the artificial variable x_{n+1} will be 0 at an optimum solution of the Phase I modification.

Similar modifications can be made to a general LP in any form, to get a Phase I modification in the same form with an interior feasible solution.

The Structure of the General Step in Interior Point Methods

Assume that the problem being solved is a minimization problem. All interior point methods start with a known interior feasible solution x^0 say, and generate a descent sequence of interior feasible solutions $x^0, x^1, \ldots$. Here a descent sequence means a sequence along which either the objective value, or some other measure of optimality strictly decreases. The general step in all the interior point methods has the following structure:

General Step

Substep 1: Let x^r be the current interior feasible solution. Generate a search direction d^r at x^r, a descent direction.

Substep 2: Compute the maximum step length θ_r, the maximum value of λ that keeps $x^r + \lambda d^r$ feasible to the original problem. This is like the minimum

ratio computation in the simplex method. Determine the **step length fraction parameter** α_r, $0 < \alpha_r < 1$ and take $x^{r+1} = x^r + \alpha_r \theta_r d^r$. With x^{r+1} as the next interior feasible solution, go to the next step.

The various methods differ on whether they work on the primal system only, dual system only, or the system consisting of the primal and dual systems together; on the strategy used to select the search direction d^r; and on the choice of the step length fraction parameter.

To give an idea of the main strategies used by interior point methods to select the search directions, we will discuss the two most popular interior point methods.

The first is in fact the first interior point method discussed in the literature, the primal affine scaling method [Dikin, 1967], which predates Karmarkar's work, but did not attract much attention until after Karmarkar made the study of interior point methods popular. This method works on the system of constraints in the original problem (primal) only. To get the search direction at the current interior feasible solution x^r, this method creates an ellipsoid $\bar{E}_r$ centered at x^r inside the feasible region of the original LP. Minimizing the objective function over $\bar{E}_r$ is an easy problem, its optimum solution $\bar{x}^r$ can be computed directly by a formula. The search direction in this method at x^r is then the direction obtained by joining x^r to $\bar{x}^r$.

The second method that we will discuss is a central path-following primal-dual interior point method. It works on the system of constraints of both the primal and dual put together. In this method the search directions used are modified Newton directions for solving the optimality conditions. The class of path-following primal-dual methods evolved out of the work of many authors including [Bayer and Lagarias, 1989]; [Güler, Roos, Terlaky and Vial, 1995]; [Kojima, Mizuno and Yoshishe, 1989]; [McLinden, 1980], [Meggido, 1989], [Mehrotra, 1992], [Mizuno, Todd and Ye, 1993], ; [Monteiro and Adler, 1989]; [Sonnevend, 1985]; [Sonnevend, Stoer, and Zhao, 1989], and others. For a complete list of references to these and other authors see the list of references in [Saigal, 1995], and [Wright, 1997].

4.4 The Primal Affine Scaling Method

This method is due to [Dikin, 1967]. We describe the method when the original LP is in the following standard form:

$$\begin{array}{rrcl} \text{Minimize} & cx & & \\ \text{subject to} & Ax & = & b \\ & x & \geq & 0 \end{array}$$

where A is of order $m \times n$ and rank m. Let x^0 be an available interior feasible solution, i.e., $Ax^0 = b$ and $x^0 > 0$ for initiating the method. The method

generates a series of interior feasible solutions $x^0, x^1, \ldots$. We will discuss the general step.

Strategy of the General Step

Let $x^r = (x_1^r, \ldots, x_n^r)^T$ be the current interior feasible solution. The method creats an ellpsoid with x^r as center inside the feasible region of the original LP. It does this by replacing the nonegativity restrictions "$x \geq 0$" by "$x \in E_r = \{x : \sum_{i=1}^n ((x_i - x_i^r)/(x_i^r))^2 \leq 1\}$". E_r is an ellipsoid in R^n with its center at x^r. The ellipsoidal approximating problem is then

$$\begin{aligned} \text{Minimize} \quad & cx & & \\ \text{subject to} \quad & Ax & = & b \\ & \sum_{i=1}^{n} ((x_i - x_i^r)/(x_i^r))^2 & \leq & 1 \end{aligned}$$

It can be shown that $E_r \subset \{x : x \geq 0\}$. The intersection of E_r with the affine space defined by the system of equality constraints $Ax = b$ is an ellipsoid $\bar{E}_r$ with center x^r inside the feasible region of the original LP. The ellipsoidal approximating problem given above is the problem of minimizing the objective function cx over this ellipsoid $\bar{E}_r$. Its optimum solution $\bar{x}^r = (\bar{x}_j^r)$ can be computed by the formula:

$$\bar{x}^r = x^r - [X_r P_r X_r c^T]/(||P_r X_r c^T||) = x^r - [X_r^2 s^r]/(||X_r s^r||)$$

where $||.||$ indicates the Euclidean norm, and

$X_r =$ diag$(x_1^r, \ldots, x_n^r)$, the diagonal matrix of order n with diagonal entries $x_1^r, \ldots, x_n^r$ and off-diagonal entries 0

$I =$ unit matrix of order n

$P_r = (I - X_r A^T (A X_r^2 A^T)^{-1} A X_r)$, a projection matrix

$y^r = (A X_r^2 A^T)^{-1} A X_r^2 c^T$, known as the **tentative dual solution** corresponding to the current interior feasible solution x^r

$s^r = c^T - A^T y^r$, **tentative dual slack vector** corresponding to x^r

It can be shown that if $\bar{x}_j^r = 0$ for at least one j, then $\bar{x}^r$ is an optimum solution of the original LP, and the method terminates. Also, if the tentative dual slack vector s^r is ≤ 0, then the objective value is unbounded below in the original LP, and the method terminates. If these termination conditions are not satisfied, then the search direction at x^r is

$$d^r = \bar{x}^r - x^r = -(X_r^2 s^r)/(||X_r s^r||)$$

is known as the **primal affine scaling direction** at the primal interior feasible solution x^r. Since both $x^r, \bar{x}^r$ are feasible to the original problem, we have

$Ax^r = A\bar{x}^r = b$, hence $Ad^r = 0$. So, d^r is a descent feasible direction for the primal along which the primal objective value decreases. The maximum step length θ_r that we can move from x^r in the direction d^r is the maximum value of λ that keeps $x^r_j + \lambda d^r_j \geq 0$ for all j. It can be verified that this is

$$\theta_r = \begin{array}{l} \infty \quad \text{if } s^r \leq 0 \text{ (this leads to the unboundedness condition stated above);} \\ \quad \text{and if } s^r \not\leq 0 \text{ it is equal to} \\ \quad \min\{(||X_r s^r||)/(x^r_j s^r_j) \text{: over } j \text{ such that } s^r_j > 0\}. \end{array}$$

It can be verified that $\theta_r = 1$ if $\bar{x}^r_j = 0$ for some j (in this case $\bar{x}^r$ is an optimum solution of the original LP as discussed above). Otherwise $\theta_r > 1$. In this case the method takes the next iterate to be $x^{r+1} = x^r + \alpha\theta_r d^r$ for some $0 < \alpha < 1$. Typically $\alpha = 0.95$ in implementations of this method. This α is the **step length fraction parameter**. Then the method moves to the next step with x^{r+1} as the current interior feasible solution. Here is a summary statement of the general step in this method.

General Step

Substep 1: Let $x^r = (x^r_1, \ldots, x^r_n)^T$ be the current interior feasible solution of the problem. Let $X_r = \text{diag}(x^r_1, \ldots, x^r_n)$.

Substep 2: Compute the tentative dual solution $y^r = (AX^2_r A^T)^{-1} AX^2_r c^T$, the tentative dual slack $s^r = c^t - A^T y^r$, and the primal affine scaling search direction at x^r which is $d^r = -(X^2_r s^r)/(||X_r s^r||)$.

If $s^r \leq 0$, $\{x^r + \lambda d^r : \lambda \geq 0\}$ is a feasible half-line for the original problem along which the objective function $cx \to -\infty$ as $\lambda \to +\infty$, terminate.

Substep 3: If $s^r \not\leq 0$, compute the maximum step length that we can move from x^r in the direction d^r, this is the maximum value of λ that keeps $x^r_j + \lambda d^r_j \geq 0$ for all j. It is $\theta_r = \min\{(||X_r s^r||)/(x^r_j s^r_j)$: over j such that $s^r_j > 0\}$. If $\theta_r = 1$, $x^r + d^r$ is an optimum solution of the original LP, terminate.

Otherwise let $x^{r+1} = x^r + \alpha\theta_r d^r$ for some $0 < \alpha < 1$ (typically $\alpha = 0.95$). With x^{r+1} as the current interior feasible solution, go to the next step.

Under some minor conditions it can be proved that if the original problem has an optimum solution, then the sequence of iterates x^r converges to a strictly complementary optimum solution, and that the objective value cx^r converges at a linear or better rate. Also if the step length fraction parameter α is $< 2/3$, then the tentative dual sequence y^r converges to the analytic center of the optimum dual solution set. For proofs of these results and a complete discussion of the convergence properties of this method see [Saigal, 1995]. So far this method has not been shown to be a polynomial time method.

Versions of this method have been developed for LPs in more general forms like the bounded variable form, and the form in which the LP consists of some unrestricted variables too. When the original LP has unrestricted variables, instead of an ellipsoid, the method creates a hyper-cylinder with an elliptical cross section inside the feasible region centered at the current interior feasible solution. The point minimizing the objective function over this hyper-cylinder can also be computed directly by a formula, and other features of the method remain essentially similar to the above.

A version of this method that works on the constraints in the dual problem only (instead of those of the primal) has also been developed, this version is called the **dual affine scaling method**. There is also a **primal-dual affine scaling method** that works on the system consisting of both the primal and dual constraints together, search directions used in this version are based on Newton directions for the system consisting of the complementary slackness conditions.

4.5 Primal-Dual Interior Point Methods for LP

The central path following primal-dual interior point methods are some of the most popular methods for LP. They consider the primal LP:

minimize $c^T x$, subject to $Ax = b, x \geq 0$;
and its dual in which the constraints are: $A^T y + s = c, s \geq 0$;

where A is a matrix of order $m \times n$ and rank m. The system of primal and dual constraints put together is:

$$\begin{aligned} Ax &= b \\ A^T y + s &= c \\ (x, s) &\geq 0 \end{aligned} \tag{4}$$

A feasible solution (x, y, s) to (4) is called an **interior feasible solution** if $(x, s) > 0$. Let $\mathcal{F}$ denote the set of all feasible solutions of (4), and $\mathcal{F}^0$ the set of all interior feasible solutions. For any $(x, y, s) \in \mathcal{F}^0$ define $X = \text{diag}(x_1, \ldots, x_n)$, the square diagonal matrix of order n with diagonal entries $x_1, \ldots, x_n$; and $S = \text{diag}(s_1, \ldots, s_n)$.

For each $j = 1$ to n, the pair (x_j, s_j) is known as the *j-th complementary pair of variables* in these primal, dual pair of problems. The *complementary slackness conditions* for optimality in this pair of problems are: the product $x_j s_j = 0$ for each $j = 1$ to n; i.e., $XSe = 0$ where e is a vector of all 1s. Since each product is ≥ 0, these conditions are equivalent to $x^T s = 0$.

The Central Path

The central path, $\mathcal{C}$ for this family of primal-dual path-following methods is a curve in $\mathcal{F}^0$ parametrized by a positive parameter $\tau > 0$. For each $\tau > 0$, the point $(x^\tau, y^\tau, s^\tau) \in \mathcal{C}$ satisfies: $(x^\tau, s^\tau) > 0$ and

$$\begin{aligned} A^T y^\tau + s^\tau &= c^T \\ Ax^\tau &= b \\ x_j^\tau s_j^\tau &= \tau, \quad j = 1, \ldots, n \end{aligned}$$

If $\tau = 0$, the above equations define the optimality conditions for the LP. For each $\tau > 0$, the solution (x^τ, y^τ, s^τ) is unique, and as τ decreases to 0 the central path converges to the center of the optimum face of the primal, dual pair of LPs.

Optimality Conditions

From optimality conditions, solving the LP is equivalent to finding a solution (x, y, s) satisfying $(x, s) \geq 0$, to the following system of $2n + m$ equations in $2n + m$ unknowns:

$$F(x, y, s) = \begin{bmatrix} A^T y + s - c \\ Ax - b \\ XSe \end{bmatrix} = 0 \tag{5}$$

This is a nonlinear system of equations because of the last equation.

Selecting the Directions to Move

Let the current interior feasible solution be $(\bar{x}, \bar{y}, \bar{s})$. So, $(\bar{x}, \bar{s}) > 0$. Also, the variables in y are unrestricted in sign in the problem.

Primal-dual path-following methods try to follow the central path $\mathcal{C}$ with τ decreasing to 0. For points on $\mathcal{C}$, the value of τ is a measure of closeness to optimality, when it decreases to 0 we are done. Following $\mathcal{C}$ with τ decreasing to 0 keeps all the complementary pair products $x_j s_j$ equal and decreasing to 0 at the same rate.

However there are two difficulties for following $\mathcal{C}$. One is that it is difficult to get an initial point on $\mathcal{C}$ with all the $x_j s_j$ equal to each other, the second is that $\mathcal{C}$ is a nonlinear curve. At a general solution $(x, y, s) \in \mathcal{F}^0$, the products $x_j s_j$ will not be equal to each other, hence the parameter $\mu = (\sum_{j=1}^n x_j s_j)/n = x^T s/n$, the *average complementary slackness violation measure*, is used as a measure of optimality for them. Since path-following methods cannot exactly follow $\mathcal{C}$, they stay within a loose but well defined neighborhood of $\mathcal{C}$ while steadily reducing the optimality measure μ to 0.

Staying explicitly within a neighborhood of $\mathcal{C}$ serves the purpose of excluding points (x, y, s) that are too close to the boundary of $\{(x, y, s) : x \geq 0, s \geq 0\}$, to make sure that the lengths of steps towards optimality remain long.

To define a neighborhood of the central path, we need a measure of deviation from centrality, this is obtained by comparing a measure of deviation of the various $x_j s_j$ from their average μ to μ itself. This leads to the measure

$$(||(x_1 s_1, \ldots, x_n s_n)^T - \mu e||)/\mu = (||XSe - \mu e||)/\mu$$

where $||.||$ is some norm. Different methods use neighborhoods defined by different norms.

The parameter θ is used as a bound for this measure when using the Euclidean norm. A commonly used neighborhood based on the Euclidean norm $||.||_2$, called the 2-norm neighborhood, defined by

$$\mathcal{N}_2(\theta) = \{(x, y, s) \in \mathcal{F}^0 : \quad ||XSe - \mu e||_2 \leq \theta\mu\}$$

for some $\theta \in (0, 1)$. Another commonly used neighborhood based on the ∞-norm is the $\mathcal{N}_{-\infty}(\gamma)$ defined by

$$\mathcal{N}_{-\infty}(\gamma) = \{(x, y, s) \in \mathcal{F}^0 : \quad x_j s_j \geq \gamma\mu, \quad j = 1, \ldots, n\}$$

parametrized by the parameter $\gamma \in (0, 1)$. This is a one-sided neighborhood that restricts each product $x_j s_j$ to be at least some small multiple γ of their average μ. Typical values used for these parameters are $\theta = 0.5$, and $\gamma = 0.001$. By keeping all iterates inside one or the other of these neighborhoods, path-following methods reduce all $x_j s_j$ to 0 at about the same rates.

Since the width of these neighborhoods for a given μ depends on μ, these neighborhoods are conical (like a *horn*), they are wider for larger values of μ, and become narrow as $\mu \to 0$.

Once the direction to move from the current point $(\bar{x}, \bar{y}, \bar{s})$ is computed, we may move from it only a small step length in that direction, and since $(\bar{x}, \bar{s}) > 0$ such a move in any direction will take us to a point that will continue satisfying $(x, s) > 0$. So, in computing the direction to move at the current point, the nonnegativity constraints $(x, s) \geq 0$ can be ignored. The only remaining conditions to be satisfied for attaining optimality are the equality conditions (5). So the direction finding routine concentrates only on trying to satisfy (5) more closely.

Ignoring the inactive inequality constraints in determining the direction to move at the current point is the main feature of **barrier methods** in nonlinear programming, hence these methods are also known as barrier methods.

(5) is a square system of nonlinear equations ($2n + m$ equations in $2n + m$ unknowns, it is nonlinear because the third condition in (5) is nonlinear). Experience in nonlinear programming indicates that the best directions to move in algorithms for solving nonlinear equations are either the Newton direction, or some modified Newton direction. So, this method uses a modified Newton direction to move. To define that, a centering parameter $\sigma \in [0, 1]$ is used. Then the direction for the move denoted by $(\triangle x, \triangle y, \triangle s)$ is the solution to the following system of linear equations

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{pmatrix} \begin{pmatrix} \triangle x \\ \triangle y \\ \triangle s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -XSe + \sigma\mu e \end{pmatrix} \tag{6}$$

where 0 in each place indicates the appropriate matrix or vector of zeros, I the unit matrix of order n, and e indicates the column vector of order n consisting of all 1s.

If $\sigma = 1$, the direction obtained will be a centering direction, which is a Newton direction towards the point (x^μ, y^μ, s^μ) on $\mathcal{C}$ at which the products $x_j s_j$ of all complementary pairs in this primal, dual pair of problems are $= \mu$. Moving in the centering direction helps to move the point towards $\mathcal{C}$, but may make little progress in reducing the optimality measure μ. But in the next iteration this may help to take a relatively long step to reduce μ. At the other end the value $\sigma = 0$ gives the standard Newton direction for solving (5). Many algorithms choose σ from the open interval $(0, 1)$ to trade off between twin goals of reducing μ and improving centrality.

We will now describe two popular path-following methods.

The Long-Step Path-Following Algorithm (LPF)

LPF generates a sequence of iterates in the neighborhood $\mathcal{N}_{-\infty}(\gamma)$, which for small values of γ (for example $\gamma = 0.001$) includes most of the set of interior feasible solutions $\mathcal{F}^0$. The method is initiated with an $(x^0, y^0, s^0) \in \mathcal{F}^0$. In each step the method chooses the centering parameter σ between two selected limits $\sigma_{\min}, \sigma_{\max}$ where $0 < \sigma_{\min} < \sigma_{\max} < 1$. The neighborhood defining parameter γ is selected from $(0, 1)$. Here is the general step in this algorithm.

General Step k

Let (x^k, y^k, s^k) be the current interior feasible solution, and $\mu_k = (x^k)^T s^k / n$ the current value of the optimality measure corresponding to it. Choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$. Find the direction $(\triangle x^k, \triangle y^k, \triangle s^k)$ by solving

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S^k & 0 & X^k \end{pmatrix} \begin{pmatrix} \triangle x^k \\ \triangle y^k \\ \triangle s^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X^k S^k e + \sigma_k \mu_k e \end{pmatrix} \tag{7}$$

Find α_k = the largest value of $\alpha \in [0, 1]$ such that $(x^k, y^k, s^k) + \alpha(\triangle x^k, \triangle y^k, \triangle s^k) \in \mathcal{N}_{-\infty}(\gamma)$.

Setting $(x^{k+1}, y^{k+1}, s^{k+1}) = (x^k, y^k, s^k) + \alpha_k(\triangle x^k, \triangle y^k, \triangle s^k)$ as the new current interior feasible solution, go to the next step.

The Predictor-Corrector Path-Following Method (PC)

Path-following methods have two goals, one to improve centrality (closeness to the central path while keeping optimality measure unchanged), and the other to decrease the optimality measure μ. The PC method takes two different steps alternately to achieve each of these twin goals. The PC uses two $\mathcal{N}_2$ neighborhoods nested one inside the other. They are $\mathcal{N}_2(\theta_1), \mathcal{N}_2(\theta_2)$ for selected $0 < \theta_1 < \theta_2 < 1$. For example $\theta_1 = 0.25, \theta_2 = 0.5$. In some versions of this method values of θ larger than 1 are also used successfully.

Every second step in this method is a "predictor" step, its starting point will be in the inner neighborhood. The direction to move in this step is computed by solving the system (7) corresponding to the current solution with the value of $\sigma = 0$. The step length in this step is the largest value of α that keeps the next point within the outer neighborhood. The gap between the inner and outer neighborhoods is wide enough to allow this step to make significant progress in reducing μ.

The step taken after each predictor step is a "corrector" step, its starting point will be in the outer neighborhood. The direction to move in this step is computed by solving the system (7) corresponding to the current solution with the value of $\sigma = 1$. The step length in this step is $\alpha = 1$, which takes it back inside the inner neighborhood to prepare for the next predictor step.

It has been shown that the sequence of interior feasible solutions obtained in this method converges to a point in the optimum face. All these path-following methods have been shown to be polynomial time algorithms.

Each step of these interior point methods requires a full matrix inversion, a fairly complex task in solving large scale problems, this involves much more work than a step of the simplex method. But the number of steps required by these interior point methods is smaller than the number of steps needed by the simplex method.

5 Gravitational Methods With Small Balls (Higher Dimensional Boundary Methods)

[Chang, 1988], pointed out that the path taken by the simplex algorithm to solve an LP can itself be interpreted as the path of a point ball falling under the influence of a gravitational force inside a thin tubular network of the one dimensional skeleton of the feasible region in which each vertex is open to all the edges incident at it. See Figure 4 for a 2-dimensional illustration.

[Murty, 1986, 1988] introduced newer methods for LP based on the principle of the gravitational force, [Chang and Murty, 1989] extended this further. They consider an LP in the form

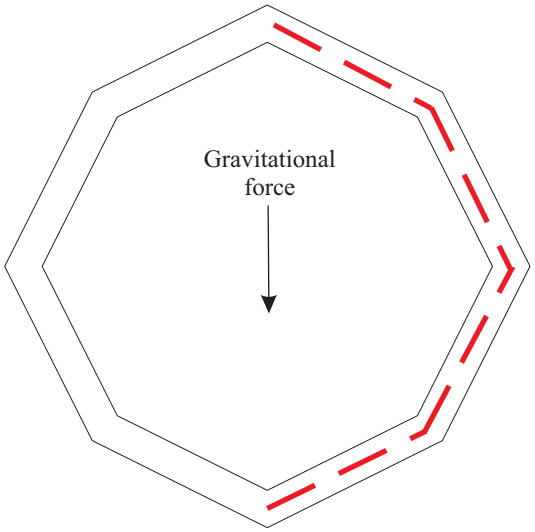


Figure 4: The gravitational interpretation of the simplex method. The dashed lines indicate the path taken by a point ball beginning at the top vertex inside a tubular network for the edges of the feasible region of an LP under the gravitational force pulling it towards the optimum.

$$\begin{aligned} &\text{maximize} \quad \pi b \\ &\text{subject to} \quad \pi A \quad = \quad c, \quad \pi \geq 0 \end{aligned} \tag{8}$$

where A is a matrix of order $m \times n$, $\pi \in R^m$ is the row vector of primal variables. As explained in Section 1, for problems in this form typically $n \leq m$. Its dual is

$$\begin{aligned} &\text{minimize} \quad z(x) \qquad = cx \\ &\text{subject to} \quad Ax \quad \geq \quad b \end{aligned} \tag{9}$$

where $x \in R^n$ is the column vector of dual variables.

We use the symbols $A_{i.}, A_{.j}$ to denote the i-th row vector, j-th column vector of the matrix A. We assume that the rows of A have all been normalized so that $||A_{i.}|| = 1$ for all i, where $||.||$ is the Euclidean norm. We also assume that $c \neq 0$ and that it is normalized so that $||c|| = 1$.

The method is applied on (9). We denote its feasible region $\{x : Ax \geqq b\}$ by K, and its interior $\{x : Ax > b\}$ by K^0. The method needs an initial *interior point* $x^0 \in K^0$. It introduces a spherical drop (we will refer to it as the *drop* or the *ball*) of small radius with center x^0 lying completely in the interior of K, and traces the path of its center as the drop falls under a gravitational force pulling

everything in the direction $-c^T$. The drop cannot cross the boundary of K, so after an initial move in the direction $-c^T$ it will be blocked by the face of K that it touches; after which it will start rolling down along the faces of K of varying dimensions. Hence the center of the drop will follow a piecewise linear descent path completely contained in the interior of K, but since the drop's radius is small, the center remains very close to the boundary of K after the first change in direction in its path. Therefore the method is essentially a boundary method. However, unlike the simplex method which follows a path strictly along the one dimensional boundary of K, this method is a **higher dimensional boundary**

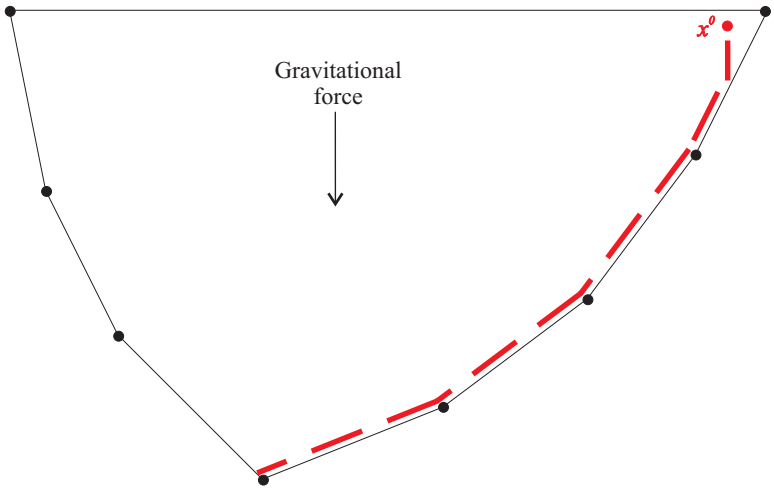


Figure 5: A 2-dimensional polytope and its faces on which the ball rolls down (dashed path) to the optimum.

method in which the path followed remains very close to faces of K of varying dimensions. See Figures 5, 6, for 2-, 3-dimensional illustrations.

After a finite number of changes in the direction of movement, the drop will reach the lowest possible point in the direction $-c^T$ that it can reach within K and then halt. If the radius of the drop is sufficiently small, the touching constraints (i.e., those whose corresponding facets of K are touching the ball) in (9) at this final halting position will determine an actual optimum solution of the LP (8). If its radius is not small enough, the direction finding step in the method at the final halting position with center x^* yields a feasible solution $\tilde{\pi}$ of (8) and the optimum objective value in (8) lies in the interval $[\tilde{\pi}b, cx^*]$. Then the radius of the drop is reduced and the method continues the same way. In [3]

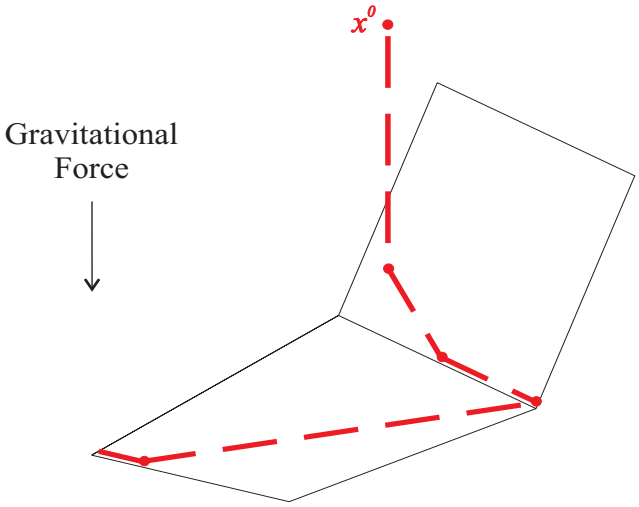


Figure 6: The ball rolling (dashed path, with dots indicating where its direction changes) inside a 3-dimensional polyhedron. Only the faces along which it rolls to the optimum are shown.

finite termination of the method to find an optimum solution has been proved.

The algorithm consists of one or more stages. In each stage the diameter of the ball remains unchanged, and consists of a series of iterations. Each iteration consists of two steps, a step that computes the gravitational direction for moving the entire ball, and a step in which the step length for the move is computed and the ball moved. The stage ends when the ball cannot move any further and halts. In the very first iteration of each stage the ball will be strictly in the interior of K without touching any of the facets of K. In subsequent iterations it will always be touching one or more facets of K. We will now describe a general stage.

A Stage in the Gravitational Method

First Iteration: Let x^0 be the present interior feasible solution. The largest sphere we can construct within K with x^0 as center has radius $= \min\{A_{i.}x^0 - b_i : i = 1 \text{ to } m\}$. Let $B(x^0, \epsilon) = \{x : ||x - x^0|| \leq \epsilon\}$ be the present ball. In this iteration we will have $0 < \epsilon < \min\{A_{i.}x^0 - b_i : i = 1 \text{ to } m\}$, so $B(x^0, \epsilon)$ is not touching any of the facets of K.

In this iteration the entire ball is moved in the direction $-c^T$. The step length is the maximum value of λ satisfying $A_{i.}(x^0 - \lambda c^T) - b_i \geq \epsilon$ for all i. So, it is

$$\gamma = \begin{cases} \infty \quad \text{if } A_{i.}c^T \leq 0 \text{ for all } i \\ \min\{\frac{A_{i.}x^0 - b_i - \epsilon}{A_{i.}c^T} : \quad i \text{ such that } A_{i.}c^T > 0\} \text{ otherwise.} \end{cases}$$

If $\gamma = \infty$, the objective function in (9) is unbounded below on its feasible set, and (8) is infeasible, terminate. Otherwise move the center of the ball from x^0 to $x^1 = x^0 - \gamma c^T$. With the new position $B(x^1, \epsilon)$ of the ball go to the next iteration.

General Iteration $r \geq 1$

Let x^{r-1} be the current interior feasible solution and $B(x^{r-1}, \epsilon)$ the present ball. Let

$J(x^{r-1}, \epsilon) =$ $\{i : A_{i.}x^{r-1} = b_i + \epsilon\}$, the index set of touching constraints for $B(x^{r-1}, \epsilon)$

$Q =$ the matrix consisting of rows $A_{i.}$ for $i \in J(x^{r-1}, \epsilon)$

$G(x^{r-1}, \epsilon) =$ $\{y : cy < 0,\ A_{i.}y \geqq 0 \text{ for all } i \in J(x^{r-1}, \epsilon)\}$, the set of descent feasible directions for the ball $B(x^{r-1}, \epsilon)$.

Step 1: Selecting the Gravitational Direction at x^{r-1} for moving the entire current ball $B(x^{r-1}, \epsilon)$

The SDGM (Steepest Descent Gravitational Method) developed in [3] takes this direction to be the steepest direction among all those in $G(x^{r-1}, \epsilon)$. This direction, called the SDGD (steepest descent gravitational direction) at x^{r-1} is the optimum solution of

$$\begin{aligned} \text{Minimize} \quad & cy \\ \text{subject to} \quad & Qy \geqq 0 \\ & 1 - y^T y \geqq 0 \end{aligned} \tag{10}$$

This problem is equivalent to

$$\begin{aligned} \text{Minimize} \quad & (c - \eta Q)(c - \eta Q)^T \\ \text{subject to} \quad & \eta \geqq 0 \end{aligned} \tag{11}$$

which is the same as that of finding the nearest point by Euclidean distance to c in the cone $\text{Rpos}(Q)$ = the nonnegative hull of row vectors of Q. This is a quadratic program, but is expected to be small as the number of variables in it is equal to the number of touching constraints at x^{r-1} which is likely to be small.

Also, this is a special quadratic program of finding the nearest point to c in a cone expressed as the nonnegative hull of row vectors of a matrix, for which efficient geometric methods are available [Wolfe, 1974, 1976]; [Wilhelmsen, 1976]; [Murty and Fathi, 1982].

If $\bar{\eta}$ is an optimum solution of (11), let

$$\bar{y}^{r-1} = \begin{cases} 0 \text{ if } \bar{\xi} = (c - \bar{\eta}Q) = 0 \\ -\bar{\xi}^T/||\bar{\xi}|| \text{ otherwise} \end{cases}$$

then $\bar{y}^{r-1}$ is an optimum solution of (10).

If $\bar{\xi} = \bar{y}^{r-1} = 0$, then $G(x^{r-1}, \epsilon) = \emptyset$, this implies that the drop $B(x^{r-1}, \epsilon)$ cannot move any further in gravitational descent with gravity pulling everything in the direction of $-c^T$, hence it halts in the present position, and the method moves to the final step in this stage.

If $\bar{y}^{r-1} \neq 0$, it is selected as the gravitational direction for the ball $B(x^{r-1}, \epsilon)$ to move, and the method goes to Step 2 in this iteration.

Reference [3] also discusses simpler methods for choosing the gravitational direction for the ball $B(x^{r-1}, \epsilon)$ to move, by solving the nearest point problem (11) approximately rather than exactly based on efficient geometric procedures discussed in [Murty and Fathi, 1982].

Step 2: Step length determination and moving the ball: The maximum step length that the ball $B(x^{r-1}, \epsilon)$ can move in the direction $\bar{y}^{r-1}$ is the maximum value of λ that keeps $A_{i.}(x^{r-1} + \lambda\bar{y}^{r-1}) \geq b_i + \epsilon$ for all $i = 1$ to m. It is

$$\gamma_{r-1} = \begin{cases} \infty \text{ if } A_{i.}\bar{y}^{r-1} \geq 0 \text{ for all } i \\ \min\{\frac{A_{i.}x^{r-1} - b_i - \epsilon}{-A_{i.}y^{r-1}} : i \text{ such that } A_{i.}y^{r-1} < 0\} \text{ otherwise.} \end{cases}$$

If $\gamma_{r-1} = \infty$, the algorithm terminates with the conclusion that the objective function is unbounded below in (9) (in fact the half-line $\{x^{r-1} + \lambda y^{r-1} : \lambda \geq 0\}$ is a feasible half-line in K along which $z \to -\infty$), and (8) is infeasible. If γ^{r-1} is finite, the center of the drop is moved from x^{r-1} to $x^r = x^{r-1} + \gamma_{r-1}y^{r-1}$. With the ball in the new position $B(x^r, \epsilon)$ the method now moves to the next iteration.

The Final Step in a Stage

Suppose the ball halts in some iteration r with the ball in position $B(x^{r-1}, \epsilon)$. $J(x^{r-1}, \epsilon)$ is the index set of touching constraints in this iteration, and let $\bar{\eta}^{r-1}$ be the optimum solution of (11). Then it can be verified that if we define

$$\bar{\pi}_i = \begin{cases} \bar{\eta}_i^{r-1} & \text{for } i \in J(x^{r-1}, \epsilon) \\ 0 & \text{otherwise} \end{cases}$$

then $\bar{\pi} = (\bar{\pi}_i)$ is a feasible solution to (8). In this case both (8) and (9) have optimum solutions and the optimum objective value z^* in them satisfies $\bar{\pi}b \leq z^* \leq cx^{r-1}$. If the difference $cx^{r-1} - \bar{\pi}b$ is sufficiently small there are several results in LP theory to obtain an optimum solution to (8) from $\bar{\pi}$ that require a small number of pivot steps. Also, let $F = \{i : \bar{\pi}_i > 0\}$, and $E \subset F$ such that $\{A_{i.} : i \in E\}$ is a maximal linearly independent subset of $\{A_{i.} : i \in F\}$, and $d = (b_i : i \in E)$. Let $\hat{x} = x^{r-1} + E^T(EE^T)^{-1}(d - Ex^{r-1})$, the orthogonal projection of x^{r-1} on the flat $\{x : A_{i.}x = b_i, i \in E\}$. If $\hat{x}$ is feasible to (9), then it is optimal to (9), and $\bar{\pi}$ is optimal to (8), terminate the algorithm.

Suppose $\hat{x}$ is not feasible to (9), then reduce the radius of the ball to half its present value, and with $B(x^{r-1}, \epsilon/2)$ go to the next stage.

In [3] finite convergence of this algorithm has been proved. In a computational experiment on LPs with up to 200 variables, an experimental code for this method performed up to 6 times faster than versions of simplex method professional software available at that time.

In the simplex method and all the interior point methods discussed earlier, all the constraints in the problem including any redundant constraints play a role in the computations (i.e., pivot steps or matrix inversions) in every step. One of the biggest advantages of the gravitational methods is that in each step only a small locally defined set of constraints (these are the touching constraints in that step) play a role in the major computation, and in particular redundant constraints can never enter the touching set, therefore the computational effort in each iteration is significantly less than in other methods.

The radius of the ball is kept small, and after the first move in the direction $-c^T$ the ball keeps rolling on the boundary faces of K of various dimensions, hence as explained earlier this method can be classified as a higher dimensional boundary method. The worst case complexity of this method when the ball has positive radius that chages over the algorithm has not been established, but [Morin, Prabhu, Zhang, 2001] showed that the version of the method with a point ball having 0 radius or any fixed radius has exponential complexity in the worst case.

6 A New Predictor-Corrector Type Interior Point Method Based on a New Simpler Centering Strategy that Can be Implemented Without Matrix Inversions

We will now discuss a new interior point method developed recently in [Murty, 2005, 2006]. We have seen that in the gravitational methods discussed in Section 5 using balls of small radius, the path traced by the center of the ball, even

though it is strictly in the interior of the set of feasible solutions of the LP, essentially rolls very close to the boundary, hence making the method behave like a boundary method rather than a truely interior point method.

To make the gravitational method follow a path truely in the central part of the feasible region and benefit from the long steps towards optimality possible under it, this new method modifies it by using balls of the highest possible radius obtained through a special centering strategy.

In the gravitational methods of Section 5 majority of work goes into computing the descent directions for the ball to move. In the new method however, much of the work is in centering steps. The method considers LPs in the form

$$\begin{aligned} \text{Minimize} \quad & z(x) = cx \\ \text{subject to} \quad & Ax \geq b \end{aligned} \tag{12}$$

where A is a matrix of order $m \times n$. In this form typically $m \geq n$. We let K denote the set of feasible solutions of this LP and $K^0 = \{x : Ax > b\}$ its interior. The method needs an initial interior feasible solution $x^0 \in K^0$ to start, if such a solution is not available, the problem can be modified using an artificial variable and the big-M augmentation technique into another one for which an initial interior feasible solution is readily available as explained in Section 4.3. We assume $c \neq 0$ as otherwise x^0 is already an optimum solution of this LP and 0 is the optimum solution of its dual. We normalize so that $||c|| = ||A_{i.}|| = 1$ for all i, here $A_{i.}$ is the i-th row vector of A.

The method consists of a series of iterations; each consisting of two steps, a centering step and a descent step. The first iteration begins with the initial interior feasible solution x^0, subsequent iterations begin with the interior feasible solution obtained at the end of the previous iteration. For any interior feasible solution x, the radius of the largest ball with center at x that can be constructed within K is denoted by:

$$\delta(x) = \quad \text{minimum } \{A_{i.}x - b_i : i = 1 \text{ to } m\}$$

Also, in this method ϵ denotes a small positive tolerance number for "interiorness" (i.e., for $\delta(x)$) for the feasible solution x to be considered an interior feasible solution. We will now describe the steps in a general iteration.

General Iteration $r + 1$

Step 1: Centering: Let x^r be the current interior feasible solution for initiating this iteration. With x^r as center, the largest ball we can construct within K has radius $\delta(x^r)$, which may be too small. To construct a larger ball inside K this step tries to move the center of the ball from x^r to a better interior feasible solution while keeping the objective value unchanged. So, starting with x^r it tries to find a new position x for the center of the ball in $K^0 \cap H$ where

$H = \{x : cx = cx^r\}$ is the objective plane through x^r, to maximize $\delta(x)$. The model for this choice is

$$\begin{array}{rrcl} \text{Maximize} & \delta & & \\ \text{subject to} & \delta & \leq & A_{i.}x - b_i, \quad i = 1 \text{ to } m \\ & cx & = & cx^r \end{array} \tag{13}$$

This is another LP with variables (δ, x). It may have alternate optimum solutions with different x-vectors, but the optimum value of δ will be unique. If $(\bar{x}^r, \bar{\delta}^r)$ is an optimum solution for it, $\bar{x}^r$ is taken as the new center for the drop, and $\bar{\delta}^r = \delta(\bar{x}^r)$ is the maximum radius for the drop within K^0 subject to the constraint that its center lie on $K^0 \cap H$.

But this itself is another LP, this type of model may have to be solved several times before we get a solution for our original LP, so solving this model (13) exactly will be counterproductive. But (13) has a very special structure, using it we discuss procedures to get an approximate solution for it later on.

Step 2: Descent Move Following Centering: Let $\bar{x}^r$ denote the center of the ball selected in Step 1. The ball is $B(\bar{x}^r, \delta(\bar{x}^r))$. Unlike the gravitational methods discussed in Section 5 in which the entire ball is moved, this method does not move the ball $B(\bar{x}^r, \delta(\bar{x}^r))$ at all, but only uses the center $\bar{x}^r$ and its property of being close to the center of $K^0 \cap H$. It takes a step of maximum possible length from $\bar{x}^r$ in a descent direction for cx.

If $r = 0$ (i.e., this is the first iteration in the method), the only descent direction that we have readily available at this time is $-c^T$, and we use that as the direction to move from $\bar{x}^0$.

If $r \geq 1$, besides $-c^T$ we have another descent direction for cx namely the direction of the **path of centers** (the path of the center of the drop in its descent to the optimum face of (12) in this algorithm) at the current center $\bar{x}^r$, which can be approximated by $\bar{x}^r - \bar{x}^{r-1}$ where $\bar{x}^{r-1}$ was the center of the drop in the previous iteration. See Figure 7.

If $d \in \{-c^T, \bar{x}^r - \bar{x}^{r-1}\}$ is the direction selected for moving from $\bar{x}^r$, we will move in this direction the maximum distance possible while still remaining inside K^0 which is:

$$\gamma \quad = \quad \min\{\frac{-A_{i.}\bar{x}^r + b_i + \epsilon}{A_{i.}d} : i \text{ such that } \ A_{i.}d < 0\}$$

If $\gamma = \infty$, the objective function is unbounded below in (12), and its dual is infeasible, terminate the algorithm.

If γ is finite, the decrease in the objective value in this move is $|\gamma cd|$. Select the direction d from $\{-c^T, \bar{x}^r - \bar{x}^{r-1}\}$ to be the one which yields the maximum decrease in the objective value in this move. With the point obtained after the move, $\quad x^{r+1} = \bar{x}^r + \gamma d \quad$ go to the next iteration.

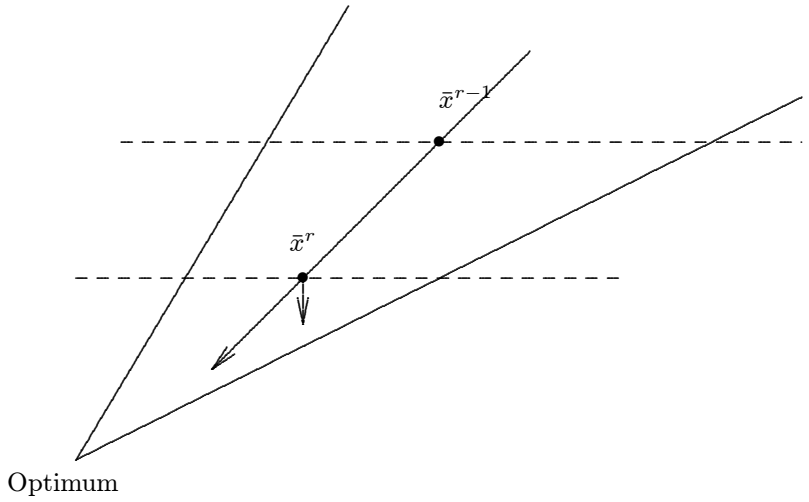


Figure 7: The two descent directions to move in Step 2 when the center is at $\bar{x}^r$ in an iteration. One is $\bar{x}^r - \bar{x}^{r-1}$ where $\bar{x}^{r-1}$ is the center in the previous iteration, another is $-c^T$ (here shown as pointing downsouth). The dashed lines are the objective planes in the two iterations.

Other Descent Directions

Suppose r iterations have been carried out so far. Then $\bar{x}^q - \bar{x}^p$ is a descent direction for the objective function in (12) for all $1 \leq p < q \leq r$. Among all these descent directions, the ones obtained using recent pairs of centers may have useful information about the shape of the feasible region between the objective value at present and at its optimum. So, using a weighted average of these descent directions as the direction to move next (instead of using either $-c^T$ or $\bar{x}^r - \bar{x}^{r-1}$ as discussed above) may help in maximizing the improvement in the objective value in this move. The best weighted average to use for maximum practical effectiveness can be determined using computational experiments.

Convergence Results

We will summarize the main convergence results on this algorithm under the

assumption that centering is carried to optimality in each iteration. Proofs are not given, for them see [Murty, 2006].

Here t is a parameter denoting the objective value cx. $t_{\min}, t_{\max}$ denote the minimum and maximum values of cx over K. For any t between $t_{\min}$ and $t_{\max}$, $\delta[t]$ denotes the maximum value of $\delta(x)$ over $x \in K^0 \cap \{x : cx = t\}$, it is the radius of the largest sphere that can be constructed within K with its center restricted to $K^0 \cap \{x : cx = t\}$, it is the optimum value of δ in the LP

$$\begin{aligned} \delta[t] = \text{Maximum value of } & \delta \\ \text{subject to } \delta - A_{i.}x &\leq -b_i, \quad i = 1, \ldots, n \\ cx &= t \end{aligned} \tag{14}$$

The **set of touching constraints at** t is the set of all inequality constraints in (14) satisfied as equations by any of the optimum solutions of (14).

The **essential touching constraint index set at** t is the set $J(t) = \{i : A_{i.}x = b_i + \delta[t]\}$ for every optimum solution $(\delta[t], x)$ of (14)$\}$. The i-th constraint in (12), (14) is said to be in the set of essential touching constraints at t if $i \in J(t)$.

We assume that the center selected in the centering strategy is an $x(t)$ satisfying the property that the facets of K touching the ball $B(x(t), \delta[t])$ (the ball with $x(t)$ as center and $\delta[t] = \delta(x(t))$ as radius) are those corresponding to the essential touching constraint set $J(t)$.

The Path of Centers $\mathcal{P}$

In primal-dual path following interior point algorithms discussed in Section 4.5, we defined the central path $\mathcal{C}$ in the space of primal, dual variables, parametrized by the parameter τ (the common complementary slackness violation parameter, for points on the central path this violation is equal in all complementary pairs in this primal, dual pair of LPs). Analogous to that we have the path $\{x(t) : t_{\max} \geq t \geq t_{\min}\}$ in the space of the variables in the original LP (12) being solved in this algorithm, parametrized by the parameter t denoting the objective function value. We will call this the **path of centers** in this method, and denote it by $\mathcal{P}$. We also have the associated path $\{\delta[t] : t_{\max} \geq t \geq t_{\min}\}$ of the radii of the balls, which is piecewise linear concave (see Theorem 2 next). Notice the differences. The point on the central path $\mathcal{C}$ is unique for each positive value of the parameter τ. The point $x(t)$ on the path of centers $\mathcal{P}$ however may not be unique.

Theorem 2: $\delta[t]$ is a piecewise linear concave function defined over $t_{\min} \leq t \leq t_{\max}$.

Let t^* = the value of t where $\delta[t]$ attains its maximum value. So, $\delta[t]$ is

monotonic increasing as t increases from $t_{\min}$ to t^*, and from t^* it is monotonic decreasing as t increases on to $t_{\max}$.

Theorem 3: If $J(t)$ remains the same for all $t_1 \leq t \leq t_2$, then $\delta[t]$ is linear in this interval.

Theorem 4: For t in the interval $t_{\min}$ to t^*, $x(t)$, an optimum solution of (14), is also an optimum solution of

$$\begin{array}{rcl} \text{minimize} & cx & \\ \text{subject to} & Ax & \geq \quad b + e\delta[t] \end{array}$$

where e is the column vector of all 1s of appropriate dimension. And for t in the interval t^* to $t_{\max}$, $x(t)$ is also an optimum solution of

$$\begin{array}{rcl} \text{maximize} & cx & \\ \text{subject to} & Ax & \geq \quad b + e\delta[t] \end{array}$$

Theorem 5: Suppose for $t_1 \geq t \geq t_2$ the index set of essential touching constraints $J(t)$ does not change. Then the method will descend from objective value t_1 to t_2 in no more than three iterations.

Theorem 6: As t, the value of cx, decreases to $t_{\min}$, the set of essential touching constraints can change at most $2m$ times.

Theorems 5, 6 together show that this algorithm is a strongly polynomial algorithm in terms of the number of centering steps, if centering is carried out exactly. So, if the centering steps are carried to good accuracy, these results indicate that this method will have superior computational performance.

Procedures for Getting Approximate Solutions to Centering Steps Efficiently

Consider the centering step in iteration $r+1$ of the method when x^r is the interior feasible solution at the start of this iteration. We discuss three procedures for solving this step approximately. Procedures 1, 2 use a series of line searches on $K^0 \cap \{x : cx = cx^r\}$. Each line search involves only solving a 2-variable linear programming problem, so can be solved very efficiently without complicated matrix inversions. So, these searches generate a sequence of points which we denote by $\hat{x}^1, \hat{x}^2, \ldots$ in $K^0 \cap \{x : cx = cx^r\}$ beginning with $\hat{x}^1 = x^r$, along which $\delta(\hat{x}^s)$ is strictly increasing.

Let $\hat{x}^s$ be the current point in this sequence. Let $T(\hat{x}^s) = \{q : q \text{ ties for the minimum in } \{A_{i.}\hat{x}^s - b_i : i = 1 \text{ to } m\}\}$. In optimization literature, when considering a line search at $\hat{x}^s$ in the direction P, only moves of positive step length α leading to the point $\hat{x}^s + \alpha P$ are considered. Here our step length α can be either positive or negative, so even though we mention P as the direction of movement, the actual direction for the move may be either P or $-P$. With $\hat{x}^s + \alpha P$ as the center, the maximum radius of a ball inside K has radius

$$f(\alpha) = \min\{A_{i.}(\hat{x}^s + \alpha P) - b_i : i = 1, \ldots, m\}$$

Since we want the largest ball inside K with its center in $K^0 \cap \{x : cx = cx^r\}$, we will only consider directions P satisfying $cP = 0$, and call such a direction P to be a

profitable direction to move at $\hat{x}^s$ if $f(\alpha)$ increases as α changes from 0 to positive or negative values (i.e., $\max\{f(\alpha) \text{ over } \alpha\}$ is attained at some $\alpha \neq 0$).

unprofitable direction to move at $\hat{x}^s$ if $\max\{f(\alpha) \text{ over } \alpha\}$ is attained at $\alpha = 0$.

We have the following results.

Result 1: $\hat{x}^s$ is an optimum solution for the centering problem (14) iff 0 is the unique feasible solution for the following system in P

$$\begin{aligned} A_{i.}P &\geq 0 \quad \text{for all } \ i \in T(\hat{x}^s) \\ cP &= 0 \end{aligned} \tag{15}$$

Any nonzero solution to this system is a profitable direction to move at $\hat{x}^s$ for this centering step. Hence a direction P is a profitable direction to move at $\hat{x}^s$ if $cP = 0$ and all $A_{i.}P$ for $i \in T(\hat{x}^s)$ have the same sign.

Result 2: Suppose P is a profitable direction to move at $\hat{x}^s$, and let $\bar{\alpha}$ denote the value of α that maximizes $f(\alpha)$, and $\bar{\theta} = f(\bar{\alpha})$. Then $(\bar{\theta}, \bar{\alpha})$ is an optimum solution of the following 2-variable LP in which the variables are θ, α.

$$\begin{aligned} &\text{Maximize} \quad \theta \\ \text{subject to} \quad &\theta - \alpha A_{i.}P \leqq A_{i.}\hat{x}^s - b_i \quad 1 = 1, \ldots, m \\ &\theta \geqq 0, \quad \alpha \qquad \text{unrestricted in sign.} \end{aligned} \tag{16}$$

The optimum solution of (16) can be found by applying the simplex algorithm. Transform (16) into standard form. Let $u_1, \ldots, u_m$ denote the slack variables corresponding to the constraints in (16) in this order. Then $(u_1, \ldots, u_{q-1}, \theta, u_{q+1}, \ldots, u_m)$ is a feasible basic vector for this standard form for $q \in T(\hat{x}^s)$.

The BFS corresponding to this basic vector for the standard form corresponds to the extreme point $(\delta(\hat{x}^s), 0)$ of (16) in the (θ, α)-space. Starting from this feasible basic vector, the optimum solution of (16) can be found efficiently by the primal simplex algorithm with at most $O(m)$ effort. It may be possible to develop even more efficient ways for finding the optimum value of α in (16), that value is the optimum step length for the move at $\hat{x}^s$ in the profitable direction P.

Using these results, we discuss two procedures for approximating the centering problem (16).

Procedure 1 for Getting an Approximate Solution to the Centering Step: Since our goal is to increase the minimum distance of x to each of the facetal hyperplanes of K, this procedure considers only moves in directions

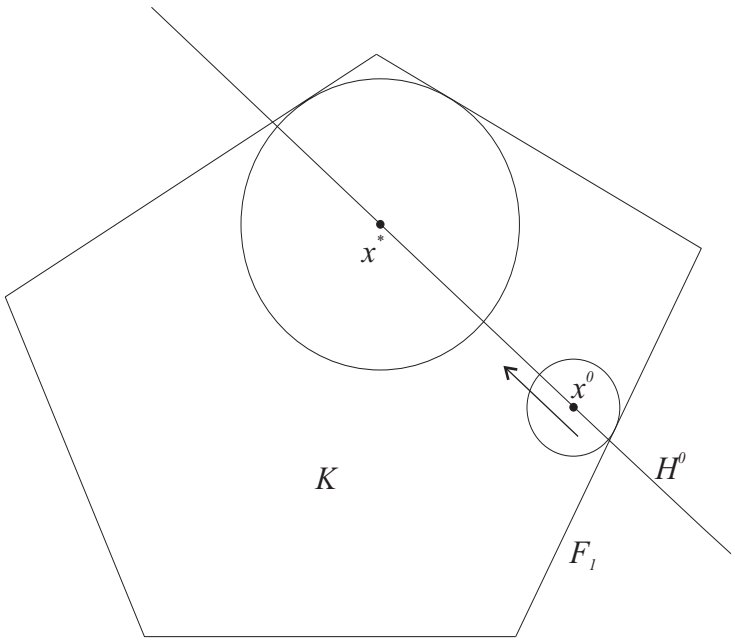


Figure 8: Moving the center from x^0 along the direction $P_{.1}$ to x^*, leads to a larger ball inside K.

perpendicular to the facetal hyperplanes of K, these are the directions $A_{i.}^T$ for $i = 1$ to m. Let $P_{.i} = (I - c^T c)A_{i.}^T$ (where I is the unit matrix of order n), it is the orthogonal projection of $A_{i.}^T$ on $\{x : cx = 0\}$.

This procedure looks for profitable directions to move at current point $\hat{x}^s$

only among the set $\{P_{.1}, \ldots, P_{.m}\}$. If a profitable direction P in this set is found, it finds the optimum solution $(\bar{\theta}, \bar{\alpha})$ of (16) with this P, takes $\hat{x}^{s+1} = \hat{x}^s + \bar{\alpha}P$ if $\bar{\alpha}$ is finite, and continues the same way with $\hat{x}^{s+1}$ as the new point in the sequence. See Figure 8.

If $\bar{\alpha} = \infty$, then the objective value in the original LP (12) is unbounded below and its dual infeasible, and so the whole method terminates.

This procedure stops when there are no profitable directions in the set $\{P_{.1}, \ldots, P_{.m}\}$, or when the improvement in the radius of the ball becomes small.

When there are several profitable directions to move at the current point $\hat{x}^s$ in the set $\{P_{.1}, \ldots, P_{.m}\}$ in this procedure, efficient selection criteria to choose the best among them can be developed. In fact the best may be among the $P_{.i}$ that correspond to i that tie for the minimum in $\delta(\hat{x}^s) = \min\{A_{i.}\hat{x}^s - b_i : i = 1 \text{ to } m\}$, or a weighted average of these directions (even though this direction is not included in our list of directions to pursue).

As can be seen, the procedure used in this centering strategy does not need any matrix inversion, and only solves a series of 2-variable LPs which can be solved very efficiently.

Procedure 2 for Getting an Approximate Solution to the Centering Step: We noticed that at the beginning of solving this centering step, $T(\hat{x}^s)$ for small s has small cardinality and usually the set of row vectors $\{c, A_{i.} \text{ for } i \in T(\hat{x}^s)\}$ tends to be linearly independent. Whenever this set of row vectors is linearly independent, a profitable direction to move at $\hat{x}^s$ can be obtained by solving the following system of linear equations in P

$$\begin{aligned} A_{i.}P &= 1 \quad \text{for each } \ i \in T(\hat{x}^s) \\ cP &= 0 \end{aligned}$$

This system has a solution because the coefficient matrix has full row rank. Finding a solution to this system of course requires one matrix inversion operation.

Using a solution P of this system as the profitable direction to move has the advantage that if the next point in the sequence is $\hat{x}^{s+1}$ then the corresponding set $T(\hat{x}^{s+1}) \supset T(\hat{x}^s)$. The same process can be continued if $\{c, A_{i.} \text{ for } i \in T(\hat{x}^{s+1})\}$ is again linearly independent. This process can be continued until we reach a point $\hat{x}^u$ for which $\{c, A_{i.} \text{ for } i \in T(\hat{x}^u)\}$ is linearly dependent. At that stage this procedure shifts to Procedure 1 and continues as in Procedure 1.

Procedure 3 for Getting an Approximate Solution to the Centering Step: Suppose the value of the objective function at the current interior feasible solution is t. Then the centering step at it is to

$$\text{maximize } \delta(x) = \min\{A_{i.}x - b_i : \ i = 1 \text{ to } m\} \quad \text{subject to} \quad cx = t.$$

This is a nonsmooth optimization problem, efficient schemes for solving such max-min problems have been developed in non-smooth convex minimization literature. One good example is [Nestrov, 2005], which can be used to solve it. Also, the effectiveness of Procedure 1 can be improved by including in it some of the line search directions used in these methods.

Numerical example: We apply one iteration of this method on the fertilizer product mix problem (1) of Section 1.4 to illustrate the method, both numerically and with a figure. We will use Procedure 1 for the centering step. Here is the problem in minimization form.

$$\begin{aligned} \text{Minimize} \quad & z = -15x_1 - 10x_2 \\ \text{subject to} \quad & 1500 - 2x_1 - x_2 \geq 0 \\ & 1200 - x_1 - x_2 \geq 0 \\ & 500 - x_1 \geq 0 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

Normalizing the coefficient vectors of all the constraints and the objective function to Euclidean norm 1, here it is again:

$$\begin{aligned} \text{Minimize} \quad & z = -0.832x_1 - 0.555x_2 \\ \text{subject to} \quad & 670.820 - 0.894x_1 - 0.447x_2 \geq 0 \\ & 848.530 - 0.707x_1 - 0.707x_2 \geq 0 \\ & 500 - x_1 \geq 0 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned} \tag{17}$$

The Centering Step

Let K denote the set of feasible solutions, and let $x^0 = (10, 1)^T$ be the initial interior feasible solution. When we plug in x^0 in the constraints in (17), the left hand side expressions have values 661.433, 840.753, 490, 10, 1 respectively. So, the radius of the largest ball inside K with x^0 as center is $\delta^0 = \min\{661.433, 840.753, 490, 10, 1\} = 1$.

The objective plane through x^0 is the straight line in R^2 defined by $-0.832x_1 - 0.555x_2 = -8.875$. This is the straight line joining $(10.667, 0)^T$ and $(0, 15.991)^T$ in the x_1, x_2-plane. So, the only direction on it is $P_{.1} = (10.667, -15.991)^T$. Moving from x^0 in the direction of $P_{.1}$ a step length α leads to the new point $(10 + 10.667\alpha, 1 - 15.991\alpha)^T$. Finding the optimum step length α leads to the following 2-variable LP in variables θ, α:

θ	α		
1	2.388	$\leqq$	661.433
1	-3.765	$\leqq$	840.753
1	10.667	$\leqq$	490
1	-10.667	$\leqq$	10
1	15.991	$\leqq$	1
1	0		Maximize

$\theta \geqq 0$, α unrestricted

Since the minimum RHS constant in this problem occurs in only one row, from Result 1 we know that the optimum value of α in this problem will be nonzero. Actually the optimum solution of this problem is $(\bar{\theta}, \bar{\alpha})^T = (6.4, -0.338)^T$. See Figure 9. The new position for the center is $\hat{x}^1 = x^0 - 0.338P_{.1} = (10, 1)^T - 0.338(10.667, -15.991)^T = (6.4, 6.4)^T$, and the maximum radius ball with it as center has radius 6.4. Since $P_{.1}$ is the only direction in $K \cap \{x : cx = cx^0\}$ in this case, this ball is the maximum radius ball inside K with center on the objective plane through x^0.

If we try to get a larger ball by moving from x^1 in the direction $P_{.1}$ a step length of α, it can be verified that in the 2-variable LP to find the optimum step length α, the entries in the RHS vector are: 662.238, 839.48, 493.6, 6.4, 6.4; and the coefficient vector of α remains the same as in the above table. In this problem the minimum RHS constant occurs in both rows 4 and 5; and the coefficients of α in these two rows have opposite signs, indicating by Result 1 that the optimum value for step length α will be 0. This indicates that $\hat{x}^1$ is the best position for the center of the ball on the objective plane through x^0 in this problem, which in the algorithm is denoted by $\bar{x}^0$.

Descent Move Following Centering

The current center is $\bar{x}^0 = (6.4, 6.4)^T$. In this initial iteration, the only descent direction we have available at $\bar{x}^0$ is $-c^T = (0.832, 0.555)^T$. Moving from $\bar{x}^0$ a step length γ in the direction $-c^T$ leads to the point $(6.4 + 0.832\gamma, 6.4 + 0.555\gamma)^T$. Taking the tolerance $\epsilon = 1$, we see that the maximum step length is $\gamma = \min\{666.571, 854.72, 592.067\} = 592.067$. Fixing $\gamma = 592.067$, we get the new interior feasible solution $x^1 = (499, 335)^T$.

With x^1, we need to go to the next iteration and continue in the same way. Figure 9 illustrates both the centering step carried out beginning with the initial interior feasible solution x^0, and the descent move carried out here.

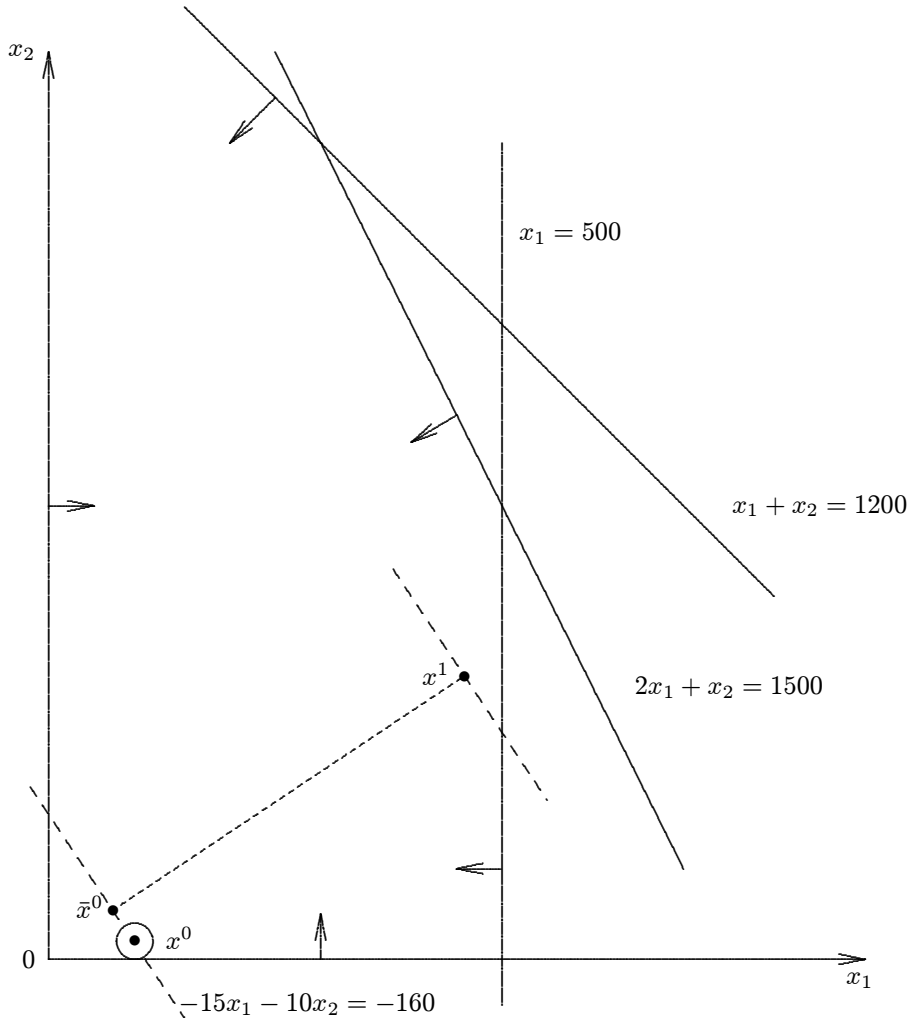


Figure 9: Figure (not drawn to scale) shows feasible region K with 5 facets, each has an arrow pointing its feasible side. Only a small sphere of radius 1 can be drawn inside K with initial point x^0 as center. Dashed line through x^0 is the objective plane, centering strategy moves point to $\bar{x}^0 = (6.4, 6.4)^T$ on this plane. With $\bar{x}^0$ as center, a sphere of radius 6.4 can be inscribed inside K. The descent move from $\bar{x}^0$ in Step 2 in direction $-c^T$ (dotted line) leads to $x^1 = (499, 335)^T$ with objective value -10835. The dashed line through x^1 is the objective plane $\{x : -15x_1 - 10x_2 = -10835\}$. Another iteration begins with x^1.

Some Advantages of This Method

Redundant constraints in a linear program can effect the efficiency for solving it by the simplex method, or the earlier interior point methods. In fact in [Deza, Nematollahi, Peyghami, Terlaky, 2004] they show that when redundant constraints are added to the Klee-Minty problem over the n-dimensional cube, the central path in these methods takes $2^n - 2$ turns as it passes through the neighborhood of all the vertices of the cube before converging to the optimum solution.

Since gravitational methods and this method operate only with the touching constraints, their performance is not affected by redundant constraints. Also, redundant constraints in (12) do not correspond to facets of K. So, in the centering step, having redundant constraints in (12) just adds some additional directions $P_{.i}$ in the set of directions used in the centering Procedure 1. Programming tricks can be developed for efficiently selecting promising directions in this set to search for improving the value of $f(\alpha)$ in this procedure, and keep this centering procedure and this method efficient.

Also, since this method needs no matrix inversions when Procedure 1 is used for centering, it can be used even when A is dense.

Interpretation as a Predictor-Corrector Path-Following Interior Point Method

This method is a path-following interior point method that tries to follow the path of centers $\mathcal{P}$ defined above, just as the methods discussed in Section 4.5 try to follow the central path $\mathcal{C}$ defined there. This method is like the predictor-corrector path-following method PC discussed in Section 4.5. In each iteration of this method, Step 1 (the centering step) is like a corrector step, it tries to move the current interior feasible solution towards the path of centers $\mathcal{P}$ while keeping the objective value constant, using line searches based on solving 2-variable LP models if Procedure 1 is employed. Step 2 (the descent step) is like a predictor step moving the longest possible step in a descent direction.

The central path of Section 4.5 depends on the algebraic representation of the set of feasible solutions through the constraints in the problem being solved , and may become very long and crooked if there are many redundant constraints in the model. The path of centers $\mathcal{P}$ followed by this algorithm however, is unaffected by redundant constraints in the model and only depends on the set of feasible solutions K of the problem as a geometric set.

Relation to the Geometric Method of Section 4.1

We will now show that this method can be viewed as computationally duplicating the geometric algorithm for solving 2-variable LPs discussed in Section 4.1. In that method, the graph of the feasible region K is drawn on paper, a

point $x^0 \in K$ is selected visually, and the straight line $z(x) = cx = cx^0$ (objective plane through x^0) is drawn. Looking at the picture of the feasible region, the objective plane is moved parallel to itself in the desirable direction as far as possible until any further move will make the line loose contact with the feasible region K. The intersection of K with the final position of the line is the set of optimum solutions of the LP.

Due to lack of visibility in higher dimensional spaces to check if the objective plane can be moved further in the desirable direction while still keeping its contact with the feasible region, this simple geometric method could not be generalized to dimensions ≥ 3. In this method, the centering step guarantees that in the descent step, the objective plane through the center $\bar{x}^r$ of the current ball $B(\bar{x}^r, \delta(\bar{x}^r))$ can move a distance of $\delta(\bar{x}^r)$ in the descent direction and still keep its contact with the feasible region. Thus this method can be viewed as a generalization of the objective plane moving step in the geometric method for two dimensional LPs.

7 An Iterative Method for LP

The name **iterative method** usually refers to a method that generates a sequence of points using a simple formula for computing the $(r+1)$th point in the sequence as an explicit function of the rth point: like $\xi^{r+1} = f(\xi^r)$. An iterative method begins with an initial point ξ^0 (often chosen arbitrarily, or subject to simple constraints that are specified, such as $\xi^0 \geq 0$), and generates the sequence $\xi^0, \xi^1, \xi^2, \ldots$ using the above formula.

Their advantage is that they are extremely simple and easy to program (much more so than the methods discussed so far) and hence may be preferred for tackling very large problems lacking special structure. A variety of iterative methods have been developed for LP and shown to converge to an optimum solution in the limit under some assumptions. But so far these methods have not been popular because in practice then convergence rate has been observed to be very slow.

As an example, we discuss an iterative method known as the **saddle point algorithm** recently developed by [Yi, Choi, Saigal, Zhu and Troutt, 1999] (see also [Kallio and Rosa, 1999], and [Choi, 2001]) that shows promise. They consider:

$$\text{the primal LP:} \quad \text{minimixe } z = cx, \quad \text{subject to} \quad Ax = b,\ x \geq 0$$
$$\text{and the dual:} \quad \text{maximize } b^T y, \quad \text{subject to} \quad A^T y \leq c^T$$

where A is a matrix of order $m \times n$. The Lagrangian function for this primal, dual pair of LPs is $L(x, y) = cx - (Ax - b)^T y$ defined over $x \in R^n_+, y \in R^m$.

Starting with an arbitrary (x^0, y^0) satisfying $x^0 \geq 0$ and $y^0 \in R^m$, this algorithm generates a sequence of points (x^r, y^r) always satisfying $x^r \geq 0$, $r = 0, 1, \ldots$. For $r = 0, 1, \ldots$ we define corresponding to (x^r, y^r)

the dual slack vector $s^r = c^T - A^T y^r = \nabla_x L(x^r, y^r)$, and the primal constriaint violation vector $v^r = b - Ax^r = \nabla_y L(x^r, y^r)$.

In (x^r, y^r), even though $x^r \geq 0$, v^r may be nonzero, and s^r may not be nonnegative, so x^r may not be primal feasible and y^r may not be dual feasible.

The pair $(\bar{x}, \bar{y})$ is said to be a *saddle point* for this primal, dual pair of LPs if

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \quad \text{for all } x \geq 0, \text{ and for all } y.$$

In LP theory these conditions are called *saddle point optimality conditions*, if they are satisfied $(\bar{x}, \bar{y})$ is called a *saddle point* for this primal, dual pair of LPs, and then $\bar{x}$ is an optimum solution for the primal, and $\bar{y}$ is an optimum solution for the dual. The aim of this algorithm is to generate a sequence converging to a saddle point.

For any real number γ define $\gamma^+ = \text{maximum}\{\gamma, 0\}$. For any vector $\xi = (\xi_j)$ define $\xi^+ = (\xi_j^+)$. We will now describe the general iteration in this algorithm. $\alpha > 0, \beta > 0$ are two step length parameters used in the iterative formula, typical values for them are: α (step length parameter in the x-space), β (step legth parametr in the y-space) both equal to 10.

General Iteration $r + 1$

Let (x^r, y^r) be the current point in the sequence. Compute $x_I^r = (x^r - \alpha s^r)^+$, $y_I^r = y^r + \beta v^r$, $\ell_x^r = L(x^r, y^r) - L(x_I^r, y^r)$, $\ell_y^r = L(x^r, y_I^r) - L(x^r, y^r)$, $\ell_r = \ell_x^r + \ell_y^r$.

It can be shown that ℓ_x^r, ℓ_y^r are both ≥ 0. If $\ell_r = 0$, then (x^r, y^r) is a saddle point, terminate the algorithm.

If $\ell_r > 0$, then compute $s_I^r = c^T - A^T y_I^r$, $v_I^r = b - Ax_I^r$, $\rho_r = \ell_r / (||s_I^r||^2 + ||v_I^r||^2)$, where $||.||$ denotes the Euclidean norm. Let $x^{r+1} = (x^r + \rho_r s_I^r)^+$, $y^{r+1} = y^r + \rho_r v_I^r$. With (x^{r+1}, y^{r+1}) as the new current pair, go to the next iteration.

Under the assumption that both the primal and dual have feasible solutions, this algorithm has been proved to generate a sequence converging to a saddle point. In implementing this algorithm, instead of keeping the step length parameters α, β fixed, their values can be chosen by line searches to optimize $L(x, y)$ (minimize with respect to x, maximize with respect to y).

8 Summary/Conclusion

We traced the history of mathematical models involving systems of linear constraints including linear inequalities, and linear programs; and algorithms for

solving them. All existing methods in use for solving them need complicated matrix inversion operations, and are suitable for solving large scale problems only when the data is very sparse. These methods encounter difficulties for solving large scale dense problems, or even those that only have some important dense columns. We also discussed in Section 6 a new efficient descent method that does not need matrix inversion operations, that shows great promise for solving large scale problems fast.

9 References

[1] D. A. Bayer and J. C. Lagarias, 1989. The Nonlinear Geometry of Linear Programming, I. Affine and Projective Scaling Trajectories, II. Legendre Transform Coordinates and Central Trajectories, III. Projective Legendre Transform Coordinates and Hilbert Geometry". *Transactions of the American Mathematical Society* 314:499-581, 1989.

[2] S. Y. Chang, 1988. The Steepest Descent Gravitational Method for Linear Programming. Ph. D. dissertation, University of Michigan, Ann Arbor, MI.

[3] S. Y. Chang and K. G. Murty, 1989. The Steepest Descent Gravitational method for Linear Programming. *Discrete Applied Mathematics* 25:211-239.

[4] B. Choi, 2001. *Theory and Algorithms for Semidefinite Programming*. Ph.D. Dissertation, University of Michigan, Ann Arbor.

[5] G. B. Dantzig, 1963. *Linear Programming and Extensions.* Princeton University Press, NJ.

[6] G. B. Dantzig and M. N. Thapa, 1997. *Linear Programming, 1. Introduction.* Springer-Verlag New York.

[7] A. Deza, E. Nematollahi, R. Peyghami, T. Terlaky, 2004. The Central Path Visits All the Vertices of the Klee-Minty Cube. AdvOL-Report No. 2004/11, McMaster University, Hamilton, Ontario, Canada.

[8] I. I. Dikin, 1967. Iterative Solution of Problems of Linear and Quadratic Programming. *Soviet Mathematics Doklady*, 8:674-675.

[9] J. Farkas, 1895. Über die Anwendungen des mechanischen Princips von Fourier. *Mathematische und natur wissenschaftliche Berichte aus Ungarn*, 12:263-281.

[10] D. Gale, 1960. *The Theory of Linear Economic Models.* McGraw-Hill, NY.

[11] P. Gordan, 1873. Ueber die Auflösung linearer Gleichungen mit reellen Coefficienten. *Mathematische Annalen*, 6:23-28.

[12] O. Güler, C. Roos, T. Terlaky, and J. -P. Vial, 1995. A survey of the implications of the behavior of the central path for the duality theory of linear programming. *Management Science*, 41:1922-1934.

[13] M. Kallio and C. Rosa, 1999. Large Scale Convex Optimization via Saddle Point Computation. *Operations Research*, 47:373-395.

[14] S. Kangshen, John N. Crossley, and Anthony W. C. Lun, 1999. *The Nine Chapters on the Mathematical Art: Companion and Commentary.* xiv + 596 pp., Oxford University Press (ISBN 0-19- 853936-3), Oxford, and Science Press. Beijing.

[15] L. V. Kantorovich, 1960. *The Mathematical Method of Production Planning and Organization.* 1939 in Russian. Translated and published in *Management Science*, 6, no. 4:363-422, July 1960.

[16] N. Karmarkar, 1984, A new Polynomial-Time Algorithm for Linear Programming", Combinatorica, 4, 373-395.

[17] M. Kojima, S. Mizuno, and A. Yoshise, 1989. A primal-dual interior point algorithm for linear programming. *Progress in Mathematical Programming: Interior Point and Related Methods*, N. Meggiddo, ed., Springer-Verlag, New York, ch. 2, pp 29-47.

[18] V. Lakshmikantham and S. Leela, 2000, *The Origin of Mathematics*, 104pp, University Press of America, Inc., Lanham, MD (visit www.univpress.com for information, also a summary of this book can be seen in its book review at: http://www.tlca.com/adults/origin-math.html).

[19] L. Mclinden, 1980. The analogue of Moreau's proximation theorem, with applications to the nonlinear complementarity problem. *Pacific Journal of Mathematics* 88:101-161.

[20] N. Meggiddo, 1989. Pathways to the optimal set in linear programming. *Progress in Mathematical Programming: Interior Point and Related Methods*, N. Meggiddo, ed., Springer-Verlag, New York, ch. 8, pp 131-158.

[21] S. Mehrotra, 1992. On the implementation of a primal-dual interior point method. *SIAM Journal on Optimization*, 2:575-601.

[22] H. Minkowski, 1896. *Geometrie der Zahlen (Erste Lieferung)*, Teubner, Leipzig, Germany.

[23] S. Mizuno, M. Todd, and Y. Ye, 1993. On adaptive step primal-dual interior point algorithms for linear programming. *Mathematics of Operations Research* 18:964-981.

[24] R. D. C. Monteiro and I. Adler, 1989. Interior path-following primal-dual algorithms, Part I: Linear programming. *Mathematical Programming* 44:27-41.

[25] T. L. Morin, N. Prabhu, and Z. Zhang, 2001, "Complexity of the Gravitational Method for Linear Programming", *Journal of Optimization Theory and Applications* 108, 633-658.

[26] K. G. Murty, 1983, *Linear Programming*, Wiley.

[27] K. G. Murty, 1986, "The Gravitational method for Linear Programming", *Opsearch* 23, 206-214.

[28] K. G. Murty, 1988, *Linear Complementarity, Linear and Nonlinear Programming*, Helderman Verlag, Berlin, can be accessed on the web from Murty's webpage at: http://www-personal.engin.umich.edu/˜ murty/

[29] K. G. Murty, 2004, *Computational and Algorithmic Linear Algebra and n-dimensional Geometry*, Freshman-Sophomore level linear algebra book available as a download for a requested small contribution at: http://ioe.engin.umich.edu/people/fac/books/murty/algorithmic_linear_algebra/

[30] K. G. Murty, 2005-1. "A Gravitational Interior Point Method for LP", *Opsearch*, 42, No. 1, 28-36.

[31] K. G. Murty, 2005-2. *Optimization Models for Decision Making: Vol. 1 (Junior Level)*, Available as a download for a requested small contribution at http://ioe.engin.umich.edu/people/fac/books/murty/opti_model/

[32] K. G. Murty, 2005-3. "My Experiences with George Dantzig", see: http://www.informs.org/History/dantzig/rem_murty.htm

[33] K. G. Murty, 2006. A new Practically Efficient Interior Point Method for LP, *Algorithmic Operations Research*, 1, 3-19; paper can be seen at the website: http://journals.hil.unb.ca/index.php/AOR/index.

[34] K. G. Murty and Y. Fathi, 1982. "A Critical Index Algorithm for Nearest Point Problems on Simplicial Cones", *Mathematical Programming*, 23, 206-215.

[35] Y. Nestrov, 2005. "Smooth Minimization of Non-Smooth Functions", Mathematical Programming, Series A, 103, 127-152.

[36] R. Saigal, 1995. *Linear Programming A Modern Integrated Analysis*, Kluwer Academic Publishers, Boston.

[37] A. Schrijver, 1986. *Theory of Linear and Integer Programming*, Wiley-Interscience, NY.

[38] G. Sonnevend, 1985. "An Analytic Center for Polyhedrons and New Classes of Global Algorithms for Linear (Smooth, Convex) Programming", *Proc. 12th Conf. on System Modelling and Optimization*, Springer Verlag, NY.

[39] G. Sonnevend, J. Stoer, and G. Zhao, 1989. On the complexity of following the central path of linear programming by linear extrapolation. *Mathematics of Operations Research* 62:19-31.

[40] J. Von Neumann. Discussion of a maximum problem. *John von Neumann, Collected Works, Vol VI*, A. H. Taub, ed. Pergamon Press, Oxford, England, 1963, pp 89-95. Also see "Communication on the Borel notes" pp 27-28.

[41] D. R. Wilhelmsen, 1976, "A Nearest Point Algorithm for Convex Polyhedral Cones and Applications to Positive Linear Approximation", *Mathematics of Computation*, 30, 48-57.

[42] P. Wolfe, 1974, "Algorithm for a Least Distance Programming Problem", *Mathematical Programming Study 1*, 190-205.

[43] P. Wolfe, 1976. "Finding the Nearest Point in a Polytope", *Mathematical Programming*, 11, 128-149.

[44] S. J. Wright, 1997. *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia.

[45] Y. Ye, 1997. *Interior Point Algorithms, Theory and Analysis*, Wiley-Interscience.

[46] S. Yi, B. Choi, R. Saigal, W. Zhu, and M. Troutt, 1999. "Convergence of a gradient based algorithm for linear programming that computes a saddle point", Technical Report, Univ of Michigan, Ann Arbor.