

THE NONLINEAR GEOMETRY OF LINEAR PROGRAMMING. I AFFINE AND PROJECTIVE SCALING TRAJECTORIES

D. A. BAYER AND J. C. LAGARIAS

ABSTRACT. This series of papers studies a geometric structure underlying Karmarkar's projective scaling algorithm for solving linear programming problems. A basic feature of the projective scaling algorithm is a vector field depending on the objective function which is defined on the interior of the polytope of feasible solutions of the linear program. The geometric structure studied is the set of trajectories obtained by integrating this vector field, which we call P -trajectories. We also study a related vector field, the affine scaling vector field, and its associated trajectories, called A -trajectories. The affine scaling vector field is associated to another linear programming algorithm, the affine scaling algorithm. Affine and projective scaling vector fields are each defined for linear programs of a special form, called strict standard form and canonical form, respectively.

This paper derives basic properties of P -trajectories and A -trajectories. It reviews the projective and affine scaling algorithms, defines the projective and affine scaling vector fields, and gives differential equations for P -trajectories and A -trajectories. It shows that projective transformations map P -trajectories into P -trajectories. It presents Karmarkar's interpretation of A -trajectories as steepest descent paths of the objective function $\langle c, x \rangle$ with respect to the Riemannian geometry $ds^2 = \sum_{i=1}^n dx_i dx_i / x_i^2$ restricted to the relative interior of the polytope of feasible solutions. P -trajectories of a canonical form linear program are radial projections of A -trajectories of an associated standard form linear program. As a consequence there is a polynomial time linear programming algorithm using the affine scaling vector field of this associated linear program: This algorithm is essentially Karmarkar's algorithm.

These trajectories are studied in subsequent papers by two nonlinear changes of variables called Legendre transform coordinates and projective Legendre transform coordinates, respectively. It will be shown that P -trajectories have an algebraic and a geometric interpretation. They are algebraic curves, and they are geodesics (actually distinguished chords) of a geometry isometric to a Hilbert geometry on a polytope combinatorially dual to the polytope of feasible solutions. The A -trajectories of strict standard form linear programs have similar interpretations: They are algebraic curves, and are geodesics of a geometry isometric to Euclidean geometry.

Received by the editors July 28, 1986 and, in revised form, September 28, 1987 and March 21, 1988.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 90C05; Secondary 52A40, 34A34.

Research of the first author partially supported by ONR contract N00014-87-K0214.

© 1989 American Mathematical Society
0002-9947/89 \$1.00 + \$.25 per page

1. INTRODUCTION

In 1984 Narendra Karmarkar [K] introduced a new linear programming algorithm which moves through the relative interior of the polytope of feasible solutions. This algorithm, which we call the *projective scaling algorithm*, takes a series of steps inside this polytope whose direction is specified by a vector field $\mathbf{v}(\mathbf{x})$ which we call the *projective scaling vector field*. This vector field depends on the objective function and is defined at all points inside the feasible solution polytope. Karmarkar proved that the projective scaling algorithm runs in polynomial time in the worst case. He suggested that variants of this algorithm would be competitive with the simplex method on many problems, particularly on large problems having a sparse constraint matrix, and computational experiments are very encouraging [AKRV]. The algorithm has been extended and adapted to fractional linear programming [A] and convex quadratic programming [KV].

In these papers we study the set of trajectories obtained by following the projective scaling vector field exactly. Given an initial point \mathbf{x}_0 one obtains a parametrized curve $\mathbf{x}(t)$ by integrating the projective scaling vector field:

$$(1.1) \quad \begin{cases} d\mathbf{x}/dt = \mathbf{v}(\mathbf{x}), \\ \mathbf{x}(0) = \mathbf{x}_0. \end{cases}$$

A *projective scaling trajectory* (also called a *P-trajectory*) is the point-set covered by a solution to this differential equation extended to the full range of t for which a solution to this differential equation exists.

Our viewpoint is that the set of trajectories is a fundamental mathematical object underlying Karmarkar's algorithm and that the good convergence properties of Karmarkar's algorithm arise from good geometric properties of the set of trajectories.

In these papers we show that the set of all *P-trajectories* has both an algebraic and a geometric structure. Algebraically, all *P-trajectories* are parts of real algebraic curves. Geometrically, there is a metric defined on the relative interior of the polytope of feasible solutions of the linear program such that the *P-trajectories* are geodesics¹ for the geometry induced by this metric. This metric geometry is isometric to Hilbert geometry on the interior of a polytope dual to the feasible solution polytope.

We also study the trajectories of another interior-point linear programming algorithm, the *affine scaling algorithm*, which was originally proposed by Dikin [D1], in 1967, and rediscovered by others [B, VMF] more recently. We call the associated set of trajectories *affine scaling trajectories* or *A-trajectories*. We show that these trajectories also have both an algebraic and geometric structure. Algebraically, they are also parts of real algebraic curves. Geometrically, they make up the complete set of geodesics for a second metric geometry defined on

¹ Actually they are curves of shortest distance (*chords*). In this geometry chords are not always unique; see part III.

the interior of the polytope of feasible solutions. If this polytope is bounded and of dimension n , then this geometry is isometric to Euclidean geometry on \mathbf{R}^n . A -trajectories are also obtainable as \dot{q} -trajectories of a completely integrable Hamiltonian dynamical system, arising from a Lagrangian dynamical system having a simple Lagrangian.

These results for A -trajectories and P -trajectories are proved by nonlinear changes of variable that linearize these trajectories. For A -trajectories we call the associated change of variables *Legendre transform coordinates*. The Legendre transform coordinate mapping is a projection of a gradient of a logarithmic barrier function associated to the linear program's constraints, and is given by rational functions. (We call it the Legendre transform coordinate mapping because it is related to the Legendre transform of a logarithmic barrier function.) For P -trajectories we call the associated change of variable *projective Legendre transform coordinates*. It is also given by rational functions, and is a nonlinearly scaled version of the Legendre transform coordinate mapping. Legendre transform coordinates are introduced in part II of these papers, and the results concerning A -trajectories are proved there. Projective Legendre transform coordinates are introduced by the second author in part III and the results concerning P -trajectories are proved there.

Part I presents elementary facts about affine and projective scaling trajectories and shows that P -trajectories are algebraically related to certain A -trajectories. The affine and projective scaling vector fields have algebraically similar definitions: the affine scaling vector field is defined using rescalings of variables by affine transformations, while the projective scaling vector field is defined using rescalings of variables by projective transformations. This algebraic parallel between the affine and projective scaling vector fields leads to a simple algebraic relation between P -trajectories of a linear program and A -trajectories of a related (homogeneous) linear program, which is given in §6. In particular this result implies that the projective scaling algorithm can be regarded as a special case of the affine scaling algorithm, as described in §7. The contents of part I are summarized in detail in the next section.

The set of P -trajectories for a given linear program differ geometrically from the set of A -trajectories for the same linear program. The metric geometry defined in part II for which A -trajectories are geodesics is Euclidean, hence flat, while the metric geometry defined in part III for which P -trajectories are geodesics behaves in many respects like a geometry of negative curvature. The sets of trajectories also differ in how they behave viewed in the linear program's coordinates (with the usual Euclidean distance). Megiddo and Shub [MS] show that the set of A -trajectories and P -trajectories have qualitatively different behavior. They show that one can find A -trajectories that pass arbitrarily close to all 2^n vertices of the n -cube while P -trajectories for the same linear program do not exhibit this behavior.

In part II we show that the sets of A -trajectories and P -trajectories for a fixed linear program have one trajectory in common, the *central trajectory*. This

trajectory is the trajectory that passes through one particular point in the polytope of feasible solutions called the *center*. This point is defined if the polytope of feasible solutions is bounded (which it always is in Karmarkar's algorithm). It is the unique point \mathbf{x} that maximizes $\prod_{j=1}^m (\langle \mathbf{a}_j, \mathbf{x} \rangle - b_j)$ on the relative interior of the polytope of feasible solutions, where $\{\langle \mathbf{a}_j, \mathbf{x} \rangle \geq b_j; 1 \leq j \leq m\}$ is the set of constraints that are not constant on the set of feasible solutions. This notion of center was introduced and studied by Sonnevend [So1, So2]. The central trajectory has a number of different characterizations, among them that it is the trajectory of a parametrized family of logarithmic barrier functions, in which guise it is studied by Megiddo [M2]. It also has a power-series expansion of a very simple form which is easy to compute, which is given in part II. This leads to interior-point linear programming algorithms that use higher-order power-series expansions, cf. [AKRV, KLSW].

Karmarkar's algorithm may be viewed in the context of nonlinear programming as a path-following method that approximately follows the central trajectory. It is analogous to Euler's method for solving the initial value problem (1.1), cf. Nazareth [N]. Recently there has been rapid development of other interior-point linear programming methods that follow the central trajectory. These include algorithms of Iri and Imai [II], Renegar [Re], Vaidya [Va], Gonzaga [Go], and Kojima, Mizumo and Yoshise [KMY]. The algorithms of Renegar [Re], Vaidya [Va] and Gonzaga [Go] are essentially predictor-corrector methods. Vaidya [Va] and Gonzaga [Go] obtain worst-case running-time bounds that improve on Karmarkar by a factor of \sqrt{m} , where m denotes the number of inequality constraints on the linear program. Megiddo [M2] studies related families of trajectories based on parametrized families of logarithmic barrier functions.

We are indebted to Jim Reeds and Peter Doyle for helpful conversations about convexity and Riemannian geometry, and to Narendra Karmarkar for inclusion of his steepest descent interpretation of A -trajectories. We are also indebted to Mike Todd for references to the discovery of the affine scaling algorithm by Dikin in 1967, and for suggestions that improved the exposition of the paper. The results of parts I and II were presented at MSRI in January 1986.

2. SUMMARY

§3 reviews the affine and projective scaling algorithms. The projective scaling algorithm is defined for linear programs in \mathbf{R}^n of the following *canonical form*:

$$(2.1) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ A\mathbf{x} = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{x} \rangle = n, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

where $\mathbf{e} = (1, 1, \dots, 1)^T$ is feasible. The projective scaling algorithm also requires an objective function $\langle \mathbf{c}, \mathbf{x} \rangle$ that has $\langle \mathbf{c}, \mathbf{x} \rangle \geq 0$ for all feasible \mathbf{x} and $\langle \mathbf{c}, \mathbf{x} \rangle = 0$ for some feasible \mathbf{x} . An objective function with this property is said

to be *normalized*. The affine scaling algorithm is defined for linear programs in \mathbf{R}^n of the following *standard form*:

$$(2.2) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ A\mathbf{x} = \mathbf{b}, \\ \mathbf{x} \geq \mathbf{0}. \end{cases}$$

Such a linear program is in *strict standard form* (or has *strict standard form constraints*) if it has a feasible solution $\mathbf{x} = (x_1, \dots, x_n)$ with all $x_i > 0$. A canonical form linear program is in strict standard form.

§4 defines the affine and projective scaling vector fields and obtains differential equations for *A*-trajectories and *P*-trajectories. The affine scaling vector field is calculated using an affine rescaling of coordinates, and the projective scaling vector field is calculated using a projective rescaling of coordinates. (This motivates our choice of names for these algorithms.) In order to apply these rescaling transformations the linear programs must be of special forms: *strict standard form* for the affine scaling algorithm, and *canonical form* for the projective scaling algorithm. Consequently *A*-trajectories are defined in part I only for strict standard form problems and *P*-trajectories only for canonical form problems. (In part II of this series of papers we extend the definition of *A*-trajectory to other linear programs and in part III we extend the notion of *P*-trajectory similarly.)

In §5 we determine how the projective scaling vector field transforms under projective transformations, and use this to show that a projective transformation maps *P*-trajectories onto *P*-trajectories.

In §6 we show that *P*-trajectories of a canonical form linear program (2.1) are radial projections of *A*-trajectories of the associated (homogeneous) strict standard form linear program obtained by dropping the inhomogeneous constraint $\langle \mathbf{e}, \mathbf{x} \rangle = n$. This gives an algebraic relation between these *P*-trajectories and *A*-trajectories.

§7 shows that a polynomial time linear programming algorithm for a canonical form linear program having a normalized objective function \mathbf{c}_N results from following the affine scaling vector field of the associated *homogeneous standard form problem*, which is

$$(2.3) \quad \begin{cases} \text{minimize } \langle \mathbf{c}_N, \mathbf{x} \rangle, \\ A\mathbf{x} = \mathbf{0}, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

where \mathbf{e} is feasible, i.e., $A\mathbf{e} = \mathbf{0}$. The piecewise linear steps of the resulting “affine scaling” algorithm radially project onto the piecewise linear steps of Karmarkar’s projective scaling algorithm, so this “affine scaling” algorithm is essentially Karmarkar’s projective scaling algorithm. In fact this “affine scaling” algorithm is not solving the linear program (2.3), but rather is solving the fractional linear program with objective function $\langle \mathbf{c}, \mathbf{x} \rangle / \langle \mathbf{e}, \mathbf{x} \rangle$ subject to homogeneous standard form problem constraints. Thus the results of §7 may be viewed as an interpretation of Karmarkar’s projective scaling algorithm as an

“affine scaling” algorithm for a particular fractional linear programming problem. In this connection see Anstreicher [A].

In §8 we give Karmarkar’s geometric interpretation of A -trajectories for standard form linear programs as steepest descent curves with respect to the Riemannian metric $ds^2 = \sum_{i=1}^n dx_i dx_i / x_i^2$. This Riemannian metric has a rather special property: It is invariant under homogeneous affine transformations taking the positive orthant $\text{Int}(\mathbf{R}_+^n)$ onto itself.

3. AFFINE AND PROJECTIVE SCALING ALGORITHMS

We briefly summarize Karmarkar’s projective scaling algorithm [K] and the affine scaling algorithm [D1, D2, B, VMF].

Karmarkar’s *projective scaling algorithm* is a piecewise linear algorithm which proceeds in steps through the relative interior of the polytope of feasible solutions to the linear programming problem. It has the following main features: an *initial starting point*, a choice of *step direction*, a choice of *step size* at each step, and a *stopping rule*. The algorithm is defined only for linear programming problems whose constraints are of a special form, which we call (*Karmarkar*) *canonical form*, which comes with a particular initial feasible starting point which Karmarkar calls the *center*. Karmarkar’s algorithm also requires that the objective function $z = \langle \mathbf{c}, \mathbf{x} \rangle$ satisfy the special restriction that its value at the optimum point of the linear program is zero. We call such an objective function a *normalized objective function*. In order to obtain a general linear programming algorithm, Karmarkar [K, §5] shows how any linear programming problem may be converted to an associated linear programming problem in canonical form which has a normalized objective function. This conversion is done by combining the primal and dual problems, then adding slack variables and an artificial variable, and as a last step using a projective transformation. An optimal solution of the original linear programming problem can be easily recovered from an optimal solution of the associated linear program constructed in this way. The *step direction* is supplied by a vector field defined on the relative interior $\text{Rel-Int}(P)$ of the polytope of feasible solutions of a canonical form linear program. Karmarkar’s vector field depends on both the constraints and the objective function. It can be defined for any objective function on a canonical form problem, whether or not this objective function is normalized. However Karmarkar only proves good convergence properties for the piecewise linear algorithm he obtains using a normalized objective function. Karmarkar’s vector field is defined implicitly in his paper [K], in which projective transformations serve as a means for its calculation. This is described in §4.

The *step size* in Karmarkar’s algorithm is computed using an auxiliary function $g: \text{Rel-Int}(P) \rightarrow \mathbf{R}$ which he calls a *potential function*. In fact $g: (\mathbf{R}^n)^+ \rightarrow \mathbf{R}$ is defined by

$$g(\mathbf{x}) = n \log \langle \mathbf{c}, \mathbf{x} \rangle - \sum_{i=1}^n \log x_i.$$

It depends on the normalized objective function $\langle \mathbf{c}, \mathbf{x} \rangle$ and approaches $+\infty$ at all nonoptimal points ∂P of the polytope P of feasible solutions, and can be made to approach $-\infty$ approaching any optimal point on the boundary along a suitable curve. It is related to the objective function by the inequality

$$(3.1) \quad g(\mathbf{x}) \geq n \log(\langle \mathbf{c}, \mathbf{x} \rangle).$$

If \mathbf{x}_j is the starting point of the j th step and \mathbf{v} the step direction, then the step size is taken to arrive at that point \mathbf{x}_{j+1} on the ray $\{\mathbf{x} + \lambda \mathbf{v} : \lambda \geq 0\}$ which minimizes $g(\mathbf{x})$ on this ray. If \mathbf{x}_{j+1} is not an optimal point, then \mathbf{x}_{j+1} remains in $\text{Rel-Int}(P)$. Karmarkar proves that

$$(3.2) \quad g(\mathbf{x}_{j+1}) \leq g(\mathbf{x}_j) - \frac{1}{5}$$

provided that $\langle \mathbf{c}, \mathbf{x} \rangle$ is a normalized objective function. Finally, the *stopping rule* is related to the input data and to the bound (3.2) on the potential function. If (3.2) fails to hold at any step, the original linear program was infeasible or unbounded. If we start at the center $\mathbf{x}_0 = \mathbf{e}$ then

$$g(\mathbf{x}_0) = n \log \langle \mathbf{c}, \mathbf{x}_0 \rangle.$$

With (3.1) and (3.2) this implies for a normalized objective function that

$$\langle \mathbf{c}, \mathbf{x}_j \rangle / \langle \mathbf{c}, \mathbf{x}_0 \rangle \leq e^{-j/5}.$$

It is known that there is a bound L easily computable from the input data of a canonical form linear program with normalized objective function and rational data such that

$$\langle \mathbf{c}, \mathbf{w} \rangle \geq 2^{-L}$$

for any nonoptimal vertex \mathbf{w} of the polytope. When $e^{-j/5} \leq 2^{-L}$ the algorithm is stopped, and one locates a vertex \mathbf{w} of P with

$$(3.3) \quad \langle \mathbf{c}, \mathbf{w} \rangle \leq \langle \mathbf{c}, \mathbf{x}_j \rangle,$$

which is then guaranteed to be optimal. In practice one does not wait until the bound $e^{-j/5} \leq 2^{-L}$ is reached; instead every few iterates one derives a solution \mathbf{w} to (3.3) and checks whether or not it is optimal.

The *affine scaling algorithm* is similar to the projective scaling algorithm. It differs in the following respects. The input linear program is required to have *strict standard form constraints*. This form is less restricted than (Karmarkar) canonical form. The *step direction* is given by the affine scaling vector field defined in §4. This vector field is calculated using a scaling transformation based on an affine change of variable; this justifies calling this algorithm the affine scaling algorithm. There are a number of different proposals for calculating the *step size*, one of which is to go a fixed fraction (say 95%) of the way to the boundary along the ray specified by the step direction [VMF]. The *stopping rule* is the same as in Karmarkar's algorithm. The affine scaling algorithm using a suitable step size has been proved (in [D2, B, and VMF]) to converge to an optimum solution under suitable nondegeneracy conditions. The affine scaling

algorithm has not been proved to run in polynomial time in the worst case, and it is likely not a polynomial time algorithm in general.

In §7 we show that a particular special case of the affine scaling algorithm does give a provably polynomial time algorithm for linear programming. This occurs, however, because the resulting algorithm is essentially identical to Karmarkar's projective scaling algorithm.

Surveys of Karmarkar's algorithm and recent developments appear in [Ho, M1].

4. AFFINE AND PROJECTIVE SCALING VECTOR FIELDS AND DIFFERENTIAL EQUATIONS

In this section we define the affine and projective scaling vector fields in terms of scalings of the positive orthant \mathbf{R}_+^n .

A. Affine scaling vector field. The affine scaling vector field is defined for linear programs of a special form called *strict standard form*. A *standard form* linear program is

$$\begin{aligned} (4.1) \quad & \left\{ \begin{array}{l} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ (4.2a) \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \\ (4.2b) \quad \mathbf{x} \geq \mathbf{0}. \end{array} \right. \end{aligned}$$

By eliminating redundant equality constraints one can always reduce to the case in which AA^T is invertible. In that case the projection operator π_{A^\perp} which projects \mathbf{R}^n onto the subspace $A^\perp = \{\mathbf{x}: \mathbf{A}\mathbf{x} = \mathbf{0}\}$ is given by

$$(4.3) \quad \pi_{A^\perp} = I - A^T(AA^T)^{-1}A.$$

In the rest of the paper we assume that AA^T is invertible.

We define *standard form constraints* to be constraints of the form (4.2). A set of linear program constraints is in *strict standard form* if it is a set of standard form constraints that has a feasible solution $\mathbf{x} = (x_1, \dots, x_n)$ such that all $x_i > 0$. A *homogeneous strict standard form problem* is a linear program having strict standard form constraints in which $\mathbf{b} = \mathbf{0}$, and its constraints are *homogeneous strict standard form constraints*.

The notion of a set of strict standard form constraints H is a mathematical convenience introduced to make it easy to describe the relative interior of the polytope P_H of feasible solutions of H , denoted $\text{Rel-Int}(P_H)$, which is then $P_H \cap \text{Int } \mathbf{R}_+^n$, and to give explicit formulae for the effect of affine scaling transformations. A standard form linear program can always be converted to one that is in strict standard form by dropping all variables x_i that are identically zero on P_H .

In defining the affine scaling vector field we first consider a strict standard form linear program having the point $\mathbf{e} = (1, 1, \dots, 1)^T$ as a feasible point. We define the *affine scaling direction* $\mathbf{v}_A(\mathbf{e}; \mathbf{c})$ at the point \mathbf{e} to be the steepest

descent direction for $\langle \mathbf{c}, \mathbf{x} \rangle$ at $\mathbf{x}_0 = \mathbf{e}$, subject to the constraint $A\mathbf{x} = \mathbf{b}$, so that

$$(4.4) \quad \mathbf{v}_A(\mathbf{x}, \mathbf{c}) = -\pi_{A^\perp}(\mathbf{c}).$$

This may be obtained using Lagrange multipliers as a solution to the constrained minimization problem:

$$(4.5) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle - \langle \mathbf{c}, \mathbf{e} \rangle, \\ \langle \mathbf{x} - \mathbf{e}, \mathbf{x} - \mathbf{e} \rangle = \varepsilon, \\ A\mathbf{x} = \mathbf{b}, \end{cases}$$

for any $\varepsilon > 0$.

Now we define the affine scaling vector field $\mathbf{v}_A(\mathbf{d}; \mathbf{c})$ for an arbitrary strict standard form linear program at an arbitrary feasible point $\mathbf{d} = (d_1, \dots, d_n)$ in

$$\text{Int}(\mathbf{R}_+^n) = \{\mathbf{x}: \text{all } x_i > 0\}.$$

Let $D = \text{diag}(d_1, \dots, d_n)$ be the diagonal matrix corresponding to \mathbf{d} , so that $\mathbf{d} = D\mathbf{e}$. Introduce new coordinates by the affine (scaling) transformation

$$\mathbf{y} = \Psi_{D^{-1}}(\mathbf{x}) = D^{-1}\mathbf{x}$$

with inverse transformation

$$\Psi_D(\mathbf{y}) = D\mathbf{y} = \mathbf{x}.$$

Under this change of variables the standard form program (4.1)–(4.2) becomes the following standard form program:

$$(4.6) \quad \begin{cases} \text{minimize } \langle D\mathbf{c}, \mathbf{y} \rangle, \\ AD\mathbf{y} = \mathbf{b}, \\ \mathbf{y} \geq \mathbf{0}. \end{cases}$$

(4.7a)

(4.7b)

Furthermore $\Psi_{D^{-1}}(\mathbf{d}) = \mathbf{e}$. By definition the affine scaling direction for this problem is $-\pi_{(AD)^\perp}(D\mathbf{c})$, and we define the affine scaling vector $\mathbf{v}_A(\mathbf{d}; \mathbf{c})$ to be the image under Ψ_D of this vector, which yields

$$(4.8) \quad \begin{aligned} \mathbf{v}_A(\mathbf{d}; \mathbf{c}) &= (\Psi_D)_* (-\pi_{(AD)^\perp}(D\mathbf{c})) = -D\pi_{(AD)^\perp}(D\mathbf{c}) \\ &= -D(I - DA^T(AD^2A^T)^{-1}AD)D\mathbf{c}. \end{aligned}$$

We check that the affine scaling vector depends only on the component $\pi_{A^\perp}(\mathbf{c})$ of \mathbf{c} in the A^\perp direction, and summarize the discussion so far in the following lemma.

Lemma 4.1. *The affine scaling vector field for a strict standard form problem (4.1)–(4.2) having a feasible solution $\mathbf{x} = (x_1, \dots, x_n)$ with all $x_i > 0$ is*

$$(4.9) \quad \mathbf{v}_A(\mathbf{d}; \mathbf{c}) = -D\pi_{(AD)^\perp}(D\mathbf{c}).$$

It satisfies

$$(4.10) \quad \mathbf{v}_A(\mathbf{d}; \mathbf{c}) = \mathbf{v}_A(\mathbf{d}; \pi_{A^\perp}(\mathbf{c})).$$

Proof. Only (4.10) needs to be demonstrated. Let π_A denote orthogonal projection on the row space of A . Using

$$\pi_A(\mathbf{c}) = A^T(AA^T)^{-1}A\mathbf{c} = A^T\mathbf{w},$$

we find from direct substitution in (4.9) that

$$\mathbf{v}_A(\mathbf{d}; \pi_A(\mathbf{c})) = -D^2 A^T \mathbf{w} + D^2 A^T (AD^2 A^T)^{-1} AD^2 A^T \mathbf{w} = \mathbf{0}.$$

Since $\mathbf{c} = \pi_{A^\perp}(\mathbf{c}) + \pi_A(\mathbf{c})$, (4.10) follows. \square

The affine scaling vector field has no isolated critical points.

Lemma 4.2. *The affine scaling vector field $\mathbf{v}_A(\mathbf{d}; \mathbf{c})$ for a strict standard form problem with constraints given by*

$$\begin{cases} A\mathbf{x} = \mathbf{b}, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

is everywhere nonvanishing if $\pi_{A^\perp}(\mathbf{c}) \neq \mathbf{0}$. It is identically zero if $\pi_{A^\perp}(\mathbf{c}) = \mathbf{0}$.

Proof. Let H denote the constraints and P_H the polytope of feasible solutions. Suppose that $\pi_{A^\perp}(\mathbf{c}) \neq \mathbf{0}$ so that $\langle \mathbf{c}, \mathbf{x} \rangle$ is nonconstant on P_H . For any given \mathbf{d} in $\text{Rel-Int}(P_H)$ the transformed linear program obtained by the affine transformation $\Psi_{D^{-1}}(\mathbf{x}) = D^{-1}\mathbf{x}$ has the polytope of feasible solutions

$$\Psi_{D^{-1}}(P_H) = \{\Psi_{D^{-1}}(\mathbf{x}) : \mathbf{x} \in P_H\}$$

and the transformed objective function $\langle D\mathbf{c}, \mathbf{y} \rangle$ is not constant since

$$\langle D\mathbf{c}, \mathbf{y} \rangle = \langle D\mathbf{c}, D^{-1}\mathbf{x} \rangle = \langle \mathbf{c}, \mathbf{x} \rangle.$$

Since $\Psi_{D^{-1}}(P_H)$ is given explicitly by

$$\begin{cases} AD\mathbf{y} = \mathbf{b}, \\ \mathbf{y} \geq \mathbf{0}, \end{cases}$$

and since $\langle D\mathbf{c}, \mathbf{y} \rangle$ is nonconstant on $\Psi_{D^{-1}}(P_H)$ it follows that

$$\pi_{(AD)^\perp}(D\mathbf{c}) \neq \mathbf{0}.$$

Hence

$$\mathbf{v}_A(\mathbf{d}; \mathbf{c}) = D\pi_{(AD)^\perp}(D\mathbf{c}) \neq \mathbf{0},$$

since D is invertible. Hence $\mathbf{v}_A(\mathbf{d}; \mathbf{c})$ is everywhere nonvanishing.

If $\pi_{A^\perp}(\mathbf{c}) = \mathbf{0}$ then Lemma 4.1 gives

$$\mathbf{v}_A(\mathbf{d}; \mathbf{c}) = \mathbf{v}_A(\mathbf{d}; \pi_{A^\perp}(\mathbf{c})) = \mathbf{v}_A(\mathbf{d}; \mathbf{0}) = \mathbf{0}. \quad \square$$

B. Projective scaling vector field. The projective scaling vector field is defined for linear programs in the following form, which we call *canonical form*:

$$(4.11) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ A\mathbf{x} = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{x} \rangle = n, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

where \mathbf{e} is feasible. A canonical form problem is always in strict standard form. *Canonical form constraints* are constraints of a canonical form linear program.

The projective scaling vector field is more naturally associated with a *canonical form fractional linear program*, which is

$$(4.12) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle / \langle \mathbf{b}, \mathbf{x} \rangle, \\ A\mathbf{x} = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{x} \rangle = n, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

where \mathbf{e} is a feasible solution and the denominator $\mathbf{b} \geq \mathbf{0}$ is scaled so that $\langle \mathbf{b}, \mathbf{e} \rangle = 1$.

We identify a canonical form linear program (4.11) with the fractional linear program having objective function $\langle \mathbf{c}, \mathbf{x} \rangle / \langle \mathbf{e}/n, \mathbf{x} \rangle$. Observe that this FLP objective function agrees with the LP objective function $\langle \mathbf{c}, \mathbf{x} \rangle$ everywhere on the constraint set in view of the constraint $\langle \mathbf{e}, \mathbf{x} \rangle = n$.

The *projective scaling vector* $\mathbf{v}_P(\mathbf{e}; \mathbf{c})$ of a canonical form fractional linear program at \mathbf{e} is the steepest descent direction of the *numerator* $\langle \mathbf{c}, \mathbf{x} \rangle$ of the fractional linear objective function, subject to the constraints $A\mathbf{x} = \mathbf{0}$ and $\langle \mathbf{e}, \mathbf{x} \rangle = n$, which is

$$(4.13) \quad \mathbf{v}_P(\mathbf{e}; \mathbf{c}) = -\pi \left[\begin{smallmatrix} A \\ \mathbf{e}^T \end{smallmatrix} \right]^\perp (\mathbf{c}).$$

The fact that this definition does not take into account the denominator $\langle \mathbf{b}, \mathbf{x} \rangle$ of the FLP objective function may seem rather surprising. We will show however that it gives a reasonable search direction for minimizing a *normalized* objective function.

To define the projective scaling vector field $\mathbf{v}_P(\mathbf{d}; \mathbf{c})$ for a canonical form problem at an arbitrary feasible point \mathbf{d} in $\text{Rel-Int}(S_{n-1}) = \{\mathbf{x}: \langle \mathbf{e}, \mathbf{x} \rangle = n \text{ and } \mathbf{x} > \mathbf{0}\}$, we introduce new variables by the projective transformation

$$(4.14) \quad \mathbf{y} = \Phi_{D^{-1}}(\mathbf{x}) = n \frac{D^{-1}\mathbf{x}}{\mathbf{e}^T D^{-1}\mathbf{x}},$$

which has inverse transformation

$$(4.15) \quad \Phi_D(\mathbf{y}) = n \frac{D\mathbf{y}}{\mathbf{e}^T D\mathbf{y}} = \mathbf{x}.$$

Under this change of variables the canonical form fractional linear program (4.12) with objective function $\langle \mathbf{c}, \mathbf{x} \rangle / \langle \mathbf{e}/n, \mathbf{x} \rangle$ becomes the following canonical form fractional linear program:

$$(4.16) \quad \begin{cases} \text{minimize } \langle D\mathbf{c}, \mathbf{y} \rangle / \langle D\mathbf{e}/n, \mathbf{y} \rangle, \\ AD\mathbf{y} = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{y} \rangle = n, \\ \mathbf{y} \geq \mathbf{0}, \end{cases}$$

where \mathbf{e} is a feasible solution, i.e., $AD\mathbf{e} = \mathbf{0}$. Note that $\Phi_{D^{-1}}(\mathbf{d}) = \mathbf{e}$. By definition the projective scaling direction for this point is

$$(4.17) \quad \mathbf{v}_P(\mathbf{e}; D\mathbf{c}) = -\pi \left[\begin{smallmatrix} AD \\ \mathbf{e}^T \end{smallmatrix} \right]^\perp (D\mathbf{c}).$$

We define the projective scaling vector $\mathbf{v}_P(\mathbf{d}; \mathbf{c})$ to be the image of $\mathbf{v}_P(\mathbf{e}; D\mathbf{c})$ under the inverse map Φ_D acting on the tangent space, i.e.,

$$\mathbf{v}_P(\mathbf{d}; \mathbf{c}) = (\Phi_D)_*(\mathbf{v}_P(\mathbf{e}; D\mathbf{c})).$$

Now Φ_D is a nonlinear map, and a computation gives the formula

$$(\Phi_D)_*(\mathbf{w}) = D\mathbf{w} - \frac{1}{n} \langle D\mathbf{e}, \mathbf{w} \rangle D\mathbf{e}.$$

The last three formulae combine to yield

$$(4.18) \quad \mathbf{v}_P(\mathbf{d}; \mathbf{c}) = -D\pi_{\left[\frac{A\mathbf{d}}{\mathbf{c}}\right]^\perp}(D\mathbf{c}) + \frac{1}{n} \langle D\mathbf{e}, \pi_{\left[\frac{A\mathbf{d}}{\mathbf{c}}\right]^\perp}(D\mathbf{c}) \rangle D\mathbf{e}.$$

One motivation for this definition of the projective scaling direction is that it gives a “good” direction for fractional linear programs having a normalized objective function. To show this we use observations of Anstreicher [A]. Define a *normalized objective function* of an FLP to be one whose value at an optimum point is zero. This property depends only on the numerator $\langle \mathbf{c}, \mathbf{x} \rangle$ of the FLP objective function. The property of being normalized is preserved by the projective change of variable $\mathbf{y} = \Phi_{D^{-1}}(\mathbf{x}) = nD^{-1}\mathbf{x}/\mathbf{e}^T D^{-1}\mathbf{x}$. In fact the FLP (4.12) is normalized if and only if the transformed FLP (4.16) is normalized. Now consider the FLP (4.12) with an arbitrary objective function. Let \mathbf{x}^* denote the optimal solution vector of a fractional linear program of form (4.12), and let $z^* = \langle \mathbf{c}, \mathbf{x}^* \rangle / \langle \mathbf{b}, \mathbf{x}^* \rangle$ be the optimal objective function value. Define the auxiliary linear program with objective function

$$\text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle - z^* \langle \mathbf{b}, \mathbf{x} \rangle$$

and the same constraints as the FLP (4.12). The point \mathbf{x}^* is easily checked to be an optimal solution of this auxiliary linear program, using the fact that $\langle \mathbf{c}, \mathbf{x} \rangle / \langle \mathbf{b}, \mathbf{x} \rangle \geq z^*$ for all feasible \mathbf{x} . In the special case that $z^* = 0$ which arises from a normalized FLP, the steepest descent direction for this auxiliary linear program is just the fractional projective scaling direction (4.13). Since normalization is preserved under the projective transformation $\mathbf{y} = \Phi_{D^{-1}}(\mathbf{x})$ this leads to definition (4.18) of the projective scaling direction $\mathbf{v}_P(\mathbf{d}; \mathbf{c})$ for a canonical form linear program with a normalized objective function.

This discussion provides no justification for the claim that the projective scaling direction $\mathbf{v}_P(\mathbf{d}; \mathbf{c})$ given by (4.13) is a good search direction for minimizing a general objective function, and sometimes it is not (see part III). In fact the direction specified by $\mathbf{v}_P(\mathbf{d}; \mathbf{c})$ in the general case does have one reasonable consequence: It leads to the simple relationship between affine scaling trajectories and projective scaling trajectories given in Theorem 7.1.

Now we obtain a simplified formula for the projective scaling direction $\mathbf{v}_P(\mathbf{d}; \mathbf{c})$, and also show that it depends only on the component $\pi_{A^\perp}(\mathbf{c})$ of \mathbf{c} in the A^\perp direction. We summarize the facts in the following lemma.

Lemma 4.3. *The projective scaling vector field for a canonical form linear program (4.11) is given by*

$$(4.19) \quad \mathbf{v}_P(\mathbf{d}; \mathbf{c}) = -D\pi_{(AD)^\perp}(D\mathbf{c}) + \frac{1}{n}\langle D\mathbf{e}, \pi_{(AD)^\perp}(D\mathbf{c}) \rangle D\mathbf{e}.$$

It satisfies

$$(4.20) \quad \mathbf{v}_P(\mathbf{d}; \mathbf{c}) = \mathbf{v}_P(\mathbf{d}; \pi_{A^\perp}(\mathbf{c})).$$

Note that $\mathbf{v}_P(\mathbf{d}; \mathbf{c}) \neq \mathbf{v}_P(\mathbf{d}; \pi_{[\frac{A}{\mathbf{e}^T}]^\perp}(\mathbf{c}))$ in general.

Proof. By construction $\mathbf{v}_P(\mathbf{d}; \mathbf{c})$ lies in $[\frac{A}{\mathbf{e}^T}]^\perp$, so it lies in \mathbf{e}^\perp . Now we simplify (4.18) by observing that the feasibility of \mathbf{d} gives $A\mathbf{d}\mathbf{e} = A\mathbf{d} = \mathbf{0}$. Hence the projections $\pi_{(AD)^\perp}$ and $\pi_{(\mathbf{e}^T)^\perp}$ commute with each other and

$$\pi\left[\frac{AD}{\mathbf{e}^T}\right]^\perp = \pi_{(\mathbf{e}^T)^\perp}\pi_{(AD)^\perp}.$$

Next we observe that $\pi_{(\mathbf{e}^T)^\perp} = I - J/n$ where $J = \mathbf{e}\mathbf{e}^T$ is the matrix with all entries one, and that $J\mathbf{w} = \langle \mathbf{e}, \mathbf{w} \rangle \mathbf{e}$ for all vectors \mathbf{w} . Applying these facts to (4.18) we obtain

$$(4.21) \quad \begin{aligned} \mathbf{v}_P(\mathbf{d}; \mathbf{c}) &= -D\pi_{(\mathbf{e}^T)^\perp}(\pi_{(AD)^\perp}(D\mathbf{c})) + \lambda D\mathbf{e} \\ &= -D\pi_{(AD)^\perp}(D\mathbf{c}) + \frac{1}{n}DJ\pi_{(AD)^\perp}(D\mathbf{c}) + \lambda D\mathbf{e} \\ &= -D\pi_{(AD)^\perp}(D\mathbf{c}) + \mu D\mathbf{e}, \end{aligned}$$

where λ and μ are scalars and

$$(4.22) \quad \mu = \frac{1}{n}\langle D\mathbf{e}, \pi_{[\frac{AD}{\mathbf{e}^T}]^\perp}(D\mathbf{c}) \rangle + \frac{1}{n}\langle \mathbf{e}, \pi_{(AD)^\perp}(D\mathbf{c}) \rangle.$$

Multiplying (4.21) by \mathbf{e}^T , and using the identity $\langle \mathbf{e}, \mathbf{v}_P(\mathbf{d}; \mathbf{c}) \rangle = 0$ we derive an alternate expression for μ which is

$$\mu = \frac{1}{n}\langle D\mathbf{e}, \pi_{(AD)^\perp}(D\mathbf{c}) \rangle,$$

and this proves (4.19).

To prove the remaining formula, start from

$$\pi_A(\mathbf{c}) = A^T(AA^T)^{-1}A\mathbf{c} = A^T\mathbf{w}$$

where we define $\mathbf{w} = (AA^T)^{-1}A\mathbf{c}$. Then

$$\pi_{AD^\perp}(D\pi_A(\mathbf{c})) = -(I - DA^T(AD^2A^T)^{-1}AD)DA^T\mathbf{w} = \mathbf{0}.$$

Substituting this in (4.19) yields $\mathbf{v}_P(\mathbf{d}; \pi_A(\mathbf{c})) = \mathbf{0}$. Since $\mathbf{c} = \pi_{A^\perp}(\mathbf{c}) + \pi_A(\mathbf{c})$ formula (4.20) follows. \square

The projective scaling vector field $\mathbf{v}_P(\mathbf{d}; \mathbf{c})$ depends on the component of \mathbf{c} in the \mathbf{e} -direction. The requirement in Karmarkar's algorithm that the objective function be *normalized* so that it is zero at an optimal solution specifies the component of \mathbf{c} in the \mathbf{e} -direction and removes this ambiguity.

Lemma 4.4. *Given a canonical form linear program and an objective function \mathbf{c} there is a unique normalized objective function \mathbf{c}_N such that*

- (i) \mathbf{c}_N lies in A^\perp .
- (ii) $\pi_{\left[\frac{A}{\mathbf{e}^T}\right]^\perp}(\mathbf{c}) = \pi_{\left[\frac{A}{\mathbf{e}^T}\right]^\perp}(\mathbf{c}_N) = \pi_{(\mathbf{e}^T)^\perp}(\mathbf{c}_N)$.

If $\mathbf{c}^* = \pi_{\left[\frac{A}{\mathbf{e}^T}\right]^\perp}(\mathbf{c})$ and \mathbf{x}_{opt} is an optimal solution for the objective function $\langle \mathbf{c}, \mathbf{x} \rangle$ then \mathbf{c}_N is given by

$$(4.23) \quad \mathbf{c}_N = \mathbf{c}^* - \frac{1}{n} \langle \mathbf{c}^*, \mathbf{x}_{\text{opt}} \rangle \mathbf{e}.$$

Proof. The condition $A\mathbf{e} = \mathbf{0}$ implies that $A^\perp = \left[\frac{A}{\mathbf{e}^T}\right]^\perp \oplus \mathbf{R}(\mathbf{e})$. Hence conditions (i) and (ii) imply that any normalized objective function satisfying (i) and (ii) has $\mathbf{c}_N = \mathbf{c}^* + \mu \mathbf{e}$ for some scalar μ . The normalization condition gives

$$\langle \mathbf{c}_N, \mathbf{x}_{\text{opt}} \rangle = \langle \mathbf{c}^*, \mathbf{x}_{\text{opt}} \rangle - \mu \langle \mathbf{e}, \mathbf{x}_{\text{opt}} \rangle = 0.$$

Since a canonical form problem has $\langle \mathbf{e}, \mathbf{x} \rangle = n$, we have $\langle \mathbf{e}, \mathbf{x}_{\text{opt}} \rangle = n$ so that

$$\mu = \frac{1}{n} \langle \mathbf{c}^*, \mathbf{x}_{\text{opt}} \rangle$$

is unique. \square

Now we study critical points of the projective scaling vector field. It turns out that for some objective functions \mathbf{c} the projective scaling vector field $\mathbf{v}_p(\mathbf{d}; \mathbf{c})$ can have a single isolated critical point, which is either a source or a sink, see part III. We show that for a *normalized* objective function critical points do not occur.

Lemma 4.5. *The projective scaling vector field $\mathbf{v}_p(\mathbf{d}; \mathbf{c})$ for a canonical form problem with constraints given by*

$$\begin{cases} A\mathbf{x} = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{x} \rangle = n, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

having \mathbf{e} as a feasible solution is everywhere nonvanishing if \mathbf{c} is normalized and $\pi_{\left[\frac{A}{\mathbf{e}^T}\right]^\perp}(\mathbf{c}) \neq \mathbf{0}$. It is identically zero if \mathbf{c} is normalized and $\pi_{\left[\frac{A}{\mathbf{e}^T}\right]^\perp}(\mathbf{c}) = \mathbf{0}$.

Proof. Let H denote the constraints and P_H the polytope of feasible solutions, and suppose that \mathbf{c} is normalized, i.e., $\langle \mathbf{c}, \mathbf{x}_{\text{opt}} \rangle = 0$ and $\langle \mathbf{c}, \mathbf{d} \rangle \geq 0$ for all \mathbf{d} in P_H .

Now suppose $\pi_{\left[\frac{A}{\mathbf{e}^T}\right]^\perp}(\mathbf{c}) = \mathbf{c}^* \neq \mathbf{0}$. Then $\langle \mathbf{c}, \mathbf{x} \rangle$ is not constant on P_H so that

$$(4.24) \quad \langle \mathbf{c}, \mathbf{x} \rangle > 0 \quad \text{for all } \mathbf{x} \in \text{Rel-Int}(P_H).$$

Suppose that $\mathbf{d} \in \text{Rel-Int}(P_H)$ is given. Then the canonical form fractional linear program obtained by the projective transformation

$$\mathbf{y} = \Phi_{D^{-1}}(\mathbf{x}) = nD^{-1}\mathbf{x}/\mathbf{e}^T D^{-1}\mathbf{x}$$

has objective function $\langle D\mathbf{c}, \mathbf{y} \rangle / \langle D\mathbf{e}, \mathbf{y} \rangle$, see (4.16). Now for all

$$\mathbf{y} \in \text{Rel-Int}(\Phi_{D^{-1}}(P_H))$$

one has

$$\frac{\langle D\mathbf{c}, \mathbf{y} \rangle}{\langle D\mathbf{e}, \mathbf{y} \rangle} = \frac{\langle \mathbf{c}, \mathbf{x} \rangle}{\langle \mathbf{e}, \mathbf{x} \rangle} = \frac{1}{n} \langle \mathbf{c}, \mathbf{x} \rangle.$$

Choosing $\mathbf{x} = \mathbf{d}$ and $\mathbf{y} = \Phi_{D^{-1}}(\mathbf{d}) = \mathbf{e}$ we have

$$\frac{\langle D\mathbf{c}, \mathbf{e} \rangle}{\langle D\mathbf{e}, \mathbf{e} \rangle} = \frac{1}{n} \langle \mathbf{c}, \mathbf{d} \rangle > 0,$$

using (4.24). Hence $\langle D\mathbf{c}, \mathbf{e} \rangle > 0$ and letting $\mathbf{y}_{\text{opt}} = \Phi_{D^{-1}}(\mathbf{x}_{\text{opt}})$ we know that $\langle D\mathbf{c}, \mathbf{y}_{\text{opt}} \rangle = 0$ since projective transformations take normalized objective functions to normalized objective functions. Hence $\langle D\mathbf{c}, \mathbf{x} \rangle$ is not constant on $\Phi_{D^{-1}}(P_H)$, which is

$$\begin{cases} AD\mathbf{y} = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{y} \rangle = n, \\ \mathbf{y} \geq \mathbf{0}. \end{cases}$$

Consequently the gradient of $\langle D\mathbf{c}, \mathbf{x} \rangle$ in $\Phi_{D^{-1}}(P_H)$ at \mathbf{e} is nonzero, i.e.,

$$\pi_{\left[\begin{smallmatrix} A \\ \mathbf{e}^T \end{smallmatrix}\right]^\perp}(D\mathbf{c}) \neq \mathbf{0}.$$

Then

$$\mathbf{v}_P(\mathbf{d}; \mathbf{c}) = (\Phi_D)_* (-\pi_{\left[\begin{smallmatrix} A \\ \mathbf{e}^T \end{smallmatrix}\right]^\perp}(D\mathbf{c})) \neq \mathbf{0},$$

since $(\Phi_D)_*$ is a vector space isomorphism. This proves that $\mathbf{v}_P(\mathbf{d}; \mathbf{c})$ is an everywhere nonvanishing vector field in this case.

Now suppose that $\pi_{\left[\begin{smallmatrix} A \\ \mathbf{c}^T \end{smallmatrix}\right]^\perp}(\mathbf{c}) = \mathbf{c}^* = \mathbf{0}$. We may suppose without loss of generality that \mathbf{c} is in A^\perp by Lemma 4.3, and then must have $\mathbf{c} = \mu\mathbf{e}$ for some μ . The normalization condition

$$\langle \mathbf{c}, \mathbf{x}_{\text{opt}} \rangle = \mu \langle \mathbf{e}, \mathbf{x}_{\text{opt}} \rangle = 0$$

then forces $\mu = 0$. So $\mathbf{c} = \mathbf{0}$ and $\mathbf{v}_P(\mathbf{d}; \mathbf{c}) = \mathbf{v}_P(\mathbf{d}; \mathbf{0}) = \mathbf{0}$ using (4.19). \square

C. Affine and projective scaling differential equations. The affine and projective scaling trajectories are found by integrating the affine and projective scaling vector fields, respectively.

For the affine scaling case, consider a strict standard form problem

$$\begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ A\mathbf{x} = \mathbf{b}, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

having a feasible solution $\mathbf{x} = (x_1, \dots, x_n)$ with $x_i > 0$. In that case the relative interior $\text{Rel-Int}(P)$ of the polytope P of feasible solutions is

$$\text{Rel-Int}(P) = \{\mathbf{x}: A\mathbf{x} = \mathbf{b} \text{ and } \mathbf{x} > \mathbf{0}\} = P \cap \text{Int}(\mathbf{R}_+^n).$$

Suppose that \mathbf{x}_0 is in $\text{Rel-Int}(P)$. We define the A -trajectory $T_A(\mathbf{x}_0; \mathbf{c}, A, \mathbf{b})$ containing \mathbf{x}_0 to be the point-set given by the integral curve $\mathbf{x}(t)$ of the affine scaling differential equation:

$$(4.25) \quad \begin{cases} d\mathbf{x}/dt = -X\pi_{(AX)^\perp}(X\mathbf{c}), \\ \mathbf{x}(0) = \mathbf{x}_0, \end{cases}$$

in which $X = X(t)$ is the diagonal matrix with diagonal elements $x_1(t), \dots, x_n(t)$, so that $\mathbf{x}(t) = X(t)\mathbf{e}$. This differential equation is obtained from the affine scaling vector field as defined in Lemma 4.1, together with the initial value \mathbf{x}_0 . The integral curve $\mathbf{x}(t)$ is defined for the range $t_1(\mathbf{x}_0; A, \mathbf{c}) < t < t_2(\mathbf{x}_0; \mathbf{c}, A)$ which is chosen to be the maximum interval on which the solution exists. (Here $t_1 = -\infty$ and $t_2 = +\infty$ are allowable values. It turns out that finite values of t_1 or t_2 may occur. Refer to equation (4.30).) An A -trajectory $T(\mathbf{x}_0; \mathbf{c}, A, \mathbf{b})$ lies in $\text{Rel-Int}(P)$ because the vector field in (4.25) is defined only for $\mathbf{x}(t)$ in $\text{Rel-Int}(P)$.

For the projective scaling case, consider a canonical form problem (4.11). In this case

$$\text{Rel-Int}(P) = \{\mathbf{x}: A\mathbf{x} = \mathbf{0}, \langle \mathbf{e}, \mathbf{x} \rangle = n \text{ and } \mathbf{x} > \mathbf{0}\}.$$

Suppose that \mathbf{x}_0 is in $\text{Rel-Int}(P)$. We define the P -trajectory $T_P(\mathbf{x}_0; \mathbf{c}, A)$ containing \mathbf{x}_0 to be the point-set given by the integral curve $\mathbf{x}(t)$ of the projective scaling differential equation:

$$(4.26) \quad \begin{cases} \frac{d\mathbf{x}}{dt} = -X\pi_{(AX)^\perp}(X\mathbf{c}) + \frac{1}{n}\langle X\mathbf{e}, \pi_{(AX)^\perp}(X\mathbf{c}) \rangle X\mathbf{e}, \\ \mathbf{x}(0) = \mathbf{x}_0. \end{cases}$$

This differential equation is obtained from the projective scaling vector field as defined in Lemma 4.3, together with the initial value \mathbf{x}_0 .

We have defined A -trajectories and P -trajectories as point-sets. The solutions to the differential equations (4.25) and (4.26) specify these point-sets as parametrized curves. An arbitrary scaling of the vector fields by an everywhere positive function $\rho(\mathbf{x}, t)$ leads to differential equations whose solutions give the same trajectories with different parametrizations. Conversely, a reparametrization of the curve by a variable $u = \psi(t)$ with $\psi'(t) > 0$ for all t leads to a similar differential equation with a rescaled vector field having $\rho(\mathbf{x}, t) = \psi'(t)$. If $\mathbf{y}(t) = \mathbf{x}(\psi(t))$ and $\mathbf{y}(\mathbf{0}) = \mathbf{x}_0$ and $\mathbf{x}(t)$ satisfies the affine scaling differential equation, then $\mathbf{y}(t)$ satisfies

$$(4.27) \quad \begin{cases} \frac{d\mathbf{y}}{dt} = -\psi'(t)Y\pi_{(AY)^\perp}(Y\mathbf{c}), \\ \mathbf{y}(0) = \mathbf{x}_0. \end{cases}$$

If $\mathbf{x}(t)$ satisfies the projective scaling differential equation instead, then $\mathbf{y}(t)$ satisfies

$$(4.28) \quad \begin{cases} \frac{d\mathbf{y}}{dt} = -\psi'(t)[Y\pi_{(AY)^\perp}(Y\mathbf{c}) - \frac{1}{n}\langle Y\mathbf{e}, \pi_{(AY)^\perp}(Y\mathbf{c}) \rangle Y\mathbf{e}], \\ \mathbf{y}(0) = \mathbf{x}_0. \end{cases}$$

The affine scaling differential equation can be solved in closed form in the special case that the linear program has no equality constraints:

$$\begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ \mathbf{x} \geq \mathbf{0}. \end{cases}$$

The affine scaling differential equation (4.25) becomes in this case

$$\begin{cases} \frac{d\mathbf{x}}{dt} = -X^2 \mathbf{c}, \\ \mathbf{x}(0) = (d_1, \dots, d_n)^T \in \text{Int}(\mathbf{R}_+^n). \end{cases}$$

This is a decoupled set of Riccati equations

$$\begin{cases} \frac{dx_i}{dt} = -c_i x_i^2, \\ x_i(0) = d_i, \end{cases}$$

for $1 \leq i \leq n$. Using the change of variables $y_i = 1/x_i$ we find that

$$\begin{cases} \frac{dy_i}{dt} = c_i, \\ y_i(0) = \frac{1}{d_i}, \end{cases}$$

for $1 \leq i \leq n$. From this we obtain

$$(4.29) \quad \mathbf{x}(t) = \left(\frac{1}{1/d_1 + c_1 t}, \dots, \frac{1}{1/d_n + c_n t} \right).$$

This trajectory is defined for $t_1 < t < t_2$ where

$$(4.30a) \quad t_1 = \max \left\{ -\frac{1}{c_i d_i} : c_i > 0 \right\},$$

$$(4.30b) \quad t_2 = \min \left\{ -\frac{1}{c_i d_i} : c_i < 0 \right\}$$

with the convention that $t_1 = -\infty$ if all $c_i \leq 0$ and $t_2 = \infty$ if all $c_i \geq 0$.

5. PROJECTIVE TRANSFORMATIONS AND PROJECTIVE SCALING TRAJECTORIES

We compute the effect of a projective transformation

$$\Phi_{D^{-1}}(\mathbf{x}) = \frac{nD^{-1}\mathbf{x}}{\langle D^{-1}\mathbf{e}, \mathbf{x} \rangle}$$

on the projective scaling vector field $\mathbf{v}_p(\mathbf{x}; \mathbf{c})$ of a canonical form linear program. The projective scaling vector field $\mathbf{v}_p(\mathbf{x}; \mathbf{c})$ of a canonical form linear program is *not* invariant under projective transformations. The following result shows that instead it transforms at each point by a variable positive scale factor.

Theorem 5.1. *Let $\mathbf{v}_p(\mathbf{x}; \mathbf{c}, A)$ denote a projective scaling vector field for a canonical form problem with feasible polytope P defined by*

$$\begin{cases} A\mathbf{x} = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{x} \rangle = n, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

and let $\mathbf{d} = D\mathbf{e}$ be in $\text{Rel-Int}(P)$. The projective transformation $\Phi_{D^{-1}}$ given by

$$\mathbf{y} = \Phi_{D^{-1}}(\mathbf{x}) = \frac{nD^{-1}\mathbf{x}}{\langle D^{-1}\mathbf{e}, \mathbf{x} \rangle}$$

maps P to the polytope $P^* = \Phi_{D^{-1}}(P)$ defined by

$$\begin{cases} AD\mathbf{y} = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{y} \rangle = n, \\ \mathbf{y} \geq \mathbf{0}. \end{cases}$$

Then

$$(5.1) \quad (\Phi_{D^{-1}})_*(\mathbf{v}_P(\mathbf{x}; \mathbf{c}, A)) = \frac{1}{n} \langle D^{-1}\mathbf{e}, \mathbf{x} \rangle \mathbf{v}_P(\Phi_{D^{-1}}(\mathbf{x}); D\mathbf{c}, AD).$$

Proof. Before beginning the proof, we observe that

$$(5.2) \quad \mathbf{v}_P(\mathbf{x}; \lambda\mathbf{c}, A) = \lambda\mathbf{v}_P(\mathbf{x}; \mathbf{c}, A)$$

for all $\lambda \neq 0$ and

$$(5.3) \quad \mathbf{v}_P(\mathbf{x}; \mathbf{c}, \lambda A) = \mathbf{v}_P(\mathbf{x}; \mathbf{c}, A)$$

if $\lambda > 0$. Now let $\mathbf{x} = X\mathbf{e}$ and

$$\mathbf{y} = \Phi_{D^{-1}}(\mathbf{x}) = \frac{nD^{-1}X\mathbf{e}}{\langle D^{-1}\mathbf{e}, X\mathbf{e} \rangle}.$$

We use the fact that $\mathbf{v}_P(\mathbf{x}; \mathbf{c}, A)$ and $\mathbf{v}_P(\mathbf{y}; D\mathbf{c}, AD)$ are both defined as pushforwards (by different projective transformations) of $\mathbf{v}_P(\mathbf{e}; X\mathbf{c}, AX)$. By definition

$$\mathbf{v}_P(\mathbf{x}; \mathbf{c}, A) = (\Phi_X)_*(\mathbf{v}_P(\mathbf{e}; X\mathbf{c}, AX)),$$

so

$$(5.4) \quad (\Phi_{X^{-1}})_*(\mathbf{v}_P(\mathbf{x}; \mathbf{c}, A)) = \mathbf{v}_P(\mathbf{e}; X\mathbf{c}, AX).$$

Next, using (5.2) and (5.3) we have

$$\begin{aligned} (5.5) \quad \mathbf{v}_P(\mathbf{y}; D\mathbf{c}, AD) &= (\Phi_Y)_*(\mathbf{v}_P(\mathbf{e}; YD\mathbf{c}, ADY)) \\ &= (\Phi_{D^{-1}X})_*(n\langle D^{-1}\mathbf{e}, X\mathbf{e} \rangle^{-1} \mathbf{v}_P(\mathbf{e}; X\mathbf{c}, AX)) \\ &= n\langle D^{-1}\mathbf{e}, X\mathbf{e} \rangle^{-1} (\Phi_{D^{-1}X})_*(\mathbf{v}_P(\mathbf{e}; X\mathbf{c}, AX)). \end{aligned}$$

Now we compute using (5.4) and (5.5) that

$$\begin{aligned} (\Phi_{D^{-1}})_*(\mathbf{v}_P(\mathbf{x}; \mathbf{c}, A)) &= (\Phi_{D^{-1}} \circ \Phi_X \circ \Phi_{X^{-1}})_*(\mathbf{v}_P(\mathbf{x}; \mathbf{c}, A)) \\ &= (\Phi_{D^{-1}X})_* \circ ((\Phi_{X^{-1}})_*(\mathbf{v}_P(\mathbf{x}; \mathbf{c}, A))) \\ &= (\Phi_{D^{-1}X})_*(\mathbf{v}_P(\mathbf{e}; X\mathbf{c}, AX)) \\ &= \frac{1}{n} \langle D^{-1}\mathbf{e}, X\mathbf{e} \rangle \mathbf{v}_P(\mathbf{y}; D\mathbf{c}, AD), \end{aligned}$$

proving the theorem. \square

An immediate consequence of this result is that a projective transformation maps P -trajectories to P -trajectories.

Corollary 5.1a. *Let a canonical form linear program with feasible polytope P be determined by the constraints*

$$\begin{cases} Ax = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{x} \rangle = n, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

and let $\mathbf{d} = D\mathbf{e}$ be in $\text{Rel-Int}(P)$. Then the projective transformation

$$\Phi_{D^{-1}} = \frac{D^{-1}\mathbf{x}}{\langle D^{-1}\mathbf{e}, \mathbf{x} \rangle}$$

maps the P -trajectory $T_P(\mathbf{x}; \mathbf{c}, A)$ to the P -trajectory $T_P(\Phi_{D^{-1}}(\mathbf{x}); D\mathbf{c}, AD)$.

Proof. The trajectory $T_P(\mathbf{x}; \mathbf{c}, A)$ is given by the differential equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}_P(\mathbf{x}; \mathbf{c}, A), \quad \mathbf{x}(0) = \mathbf{x}.$$

By Theorem 5.1 the curve $\mathbf{y} = \Phi_{D^{-1}}(\mathbf{x})$ satisfies the differential equation

$$\frac{d\mathbf{y}}{dt} = (\Phi_{D^{-1}})_*(\mathbf{x}) = \frac{1}{n} \langle D^{-1}\mathbf{e}, \mathbf{x} \rangle \mathbf{v}_P(\mathbf{y}; D\mathbf{c}, AD), \quad \mathbf{y}(0) = \Phi_{D^{-1}}(\mathbf{x}).$$

Since $\langle D^{-1}\mathbf{e}, \mathbf{x} \rangle > 0$ everywhere on $\text{Rel-Int}(P)$, this is a positive rescaling of the projective scaling differential equation on $\Phi_{D^{-1}}(P)$ so it gives exactly the P -trajectory $T_P(\Phi_{D^{-1}}(\mathbf{x}); D\mathbf{c}, AD)$. \square

6. RELATIONS BETWEEN P -TRAJECTORIES AND A -TRAJECTORIES

There is a simple relationship between the P -trajectories of the canonical form linear program:

$$(6.1) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ Ax = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{x} \rangle = n, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

and the A -trajectories of the associated homogeneous strict standard form linear program:

$$(6.2) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ Ax = \mathbf{0}, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

where $A\mathbf{e} = \mathbf{0}$ in both linear programs, so that \mathbf{e} is feasible. It is as follows.

Theorem 6.1. *If $T_A(\mathbf{x}_0; \mathbf{c}, A, \mathbf{0})$ is an A -trajectory of the homogeneous strict standard form problem (6.2) then its radial projection*

$$(6.3) \quad T = \{n\mathbf{x}/\langle \mathbf{e}, \mathbf{x} \rangle : \mathbf{x} \in T_A(\mathbf{x}_0; \mathbf{c}, A, \mathbf{0})\}$$

is a P -trajectory of the associated canonical form linear program, which is given by

$$(6.4) \quad T = T_P(n\mathbf{x}_0/\langle \mathbf{e}, \mathbf{x}_0 \rangle; \mathbf{c}, A).$$

Proof. Geometrically the radial projection produces the radial component in the projective scaling vector field evident on comparing Lemmas 4.1 and 4.3. The trajectory $T_A(\mathbf{x}_0; \mathbf{c}, A, \mathbf{0})$ is parametrized by a solution $\mathbf{x}(t)$ of the differential equation

$$(6.5) \quad \begin{cases} \frac{d\mathbf{x}}{dt} = -X\pi_{(AX)^\perp}(X\mathbf{c}), \\ \mathbf{x}(0) = \mathbf{x}_0. \end{cases}$$

Now define

$$\mathbf{y}(t) = \frac{n\mathbf{x}(t)}{\langle \mathbf{e}, \mathbf{x}(t) \rangle}.$$

We verify directly that $\mathbf{y}(t)$ satisfies a (scaled) version of the projective scaling differential equation.

Let $Y(t) = \text{diag}(y_1(t), \dots, y_n(t))$ and note that $Y(t) = n\langle \mathbf{e}, \mathbf{x}(t) \rangle^{-1} X(t)$ so that

$$X\pi_{(AX)^\perp}(X\mathbf{c}) = n^{-2} \langle \mathbf{e}, \mathbf{x}(t) \rangle^2 Y\pi_{(AY)^\perp}(Y\mathbf{c}).$$

Using this fact and $Y\mathbf{e} = n\langle \mathbf{e}, \mathbf{x}(t) \rangle^{-1} \mathbf{x}$ we obtain

$$\begin{aligned} \frac{d\mathbf{y}}{dt} &= n\langle \mathbf{e}, \mathbf{x}(t) \rangle^{-1} \frac{d\mathbf{x}}{dt} - n\langle \mathbf{e}, \mathbf{x}(t) \rangle^{-2} \left\langle \mathbf{e}, \frac{d\mathbf{x}}{dt} \right\rangle \mathbf{x} \\ &= -n\langle \mathbf{e}, \mathbf{x}(t) \rangle^{-1} (n^{-2} \langle \mathbf{e}, \mathbf{x}(t) \rangle^2 Y\pi_{(AY)^\perp}(Y\mathbf{c})) \\ &\quad - n^{-3} \langle \mathbf{e}, \mathbf{x}(t) \rangle^2 \langle \mathbf{e}, Y\pi_{(AY)^\perp}(Y\mathbf{c}) Y\mathbf{e} \rangle \\ &= \frac{1}{n} \langle \mathbf{e}, \mathbf{x}(t) \rangle \left(-Y\pi_{(AY)^\perp}(Y\mathbf{c}) + \frac{1}{n} \langle Y\mathbf{e}, \pi_{(AY)^\perp}(Y\mathbf{c}) \rangle Y\mathbf{e} \right) \\ &= \frac{1}{n} \langle \mathbf{e}, \mathbf{x}(t) \rangle \mathbf{v}_P(\mathbf{y}; \mathbf{c}). \end{aligned}$$

Since $\psi'(t; \mathbf{x}_0) = \langle \mathbf{e}, \mathbf{x}(t) \rangle / n > 0$ for $\mathbf{x}(t) \in \text{Int}(\mathbf{R}_+^n)$ this is a version of the projective scaling differential equation (4.28). This proves (6.4) holds. \square

As an example we apply Theorem 6.1 to the canonical form linear program with no extra equality constraints:

$$\begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ \langle \mathbf{e}, \mathbf{x} \rangle = n, \\ \mathbf{x} \geq \mathbf{0}. \end{cases}$$

The feasible solutions to this problem form a regular simplex S_{n-1} . In this case the associated homogeneous standard form problem has no equality constraints:

$$\begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ \mathbf{x} \geq \mathbf{0}. \end{cases}$$

Formula (4.29) parametrizing the affine scaling trajectories for the problem gives

$$T_A(\mathbf{d}; \mathbf{c}, \phi, \phi) = \left\{ \left(\frac{1}{1/d_1 + c_1 t}, \dots, \frac{1}{1/d_n + c_n t} \right) : t_1 < t < t_2 \right\}.$$

Hence Theorem 6.1 implies that the projective scaling trajectories are

$$(6.6) \quad T_P(\mathbf{d}; \mathbf{c}, \phi) = \left\{ \frac{n}{\sum_{i=1}^n (1/d_i + c_i t)^{-1}} \left(\frac{1}{1/d_1 + c_1 t}, \dots, \frac{1}{1/d_n + c_n t} \right) : t_1 < t < t_2 \right\},$$

where t_1 and t_2 are given by (4.30).

7. THE HOMOGENEOUS AFFINE SCALING ALGORITHM

Consider the homogeneous standard form linear program:

$$(7.1) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ A\mathbf{x} = \mathbf{0}, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

where $A\mathbf{e} = \mathbf{0}$. We define the *homogeneous affine scaling algorithm* to be a piecewise linear algorithm in which the *starting value* is given by $\mathbf{x}_0 = \mathbf{e}$, the *step direction* is specified by the affine scaling vector field associated with (7.1) and the *step size* is chosen to minimize Karmarkar's "potential function"

$$g(\mathbf{x}) = \sum_{i=1}^n \log \left(\frac{\langle \mathbf{c}, \mathbf{x} \rangle}{x_i} \right)$$

along the line segment inside the feasible solution polytope specified by the step direction. Let $\mathbf{x}_0, \dots, \mathbf{x}_n$ denote the resulting sequence of interior points obtained using this algorithm. Consider the associated canonical form problem:

$$(7.2) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ A\mathbf{x} = \mathbf{0}, \\ \langle \mathbf{e}, \mathbf{x} \rangle = n, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

where $A\mathbf{e} = \mathbf{0}$. We have the following result.

Theorem 7.1. *If $\{\mathbf{x}^{(k)} : 0 \leq k < \infty\}$ are the homogeneous affine scaling algorithm iterates associated with the linear program (7.1) and if $\mathbf{y}^{(k)}$ are defined by*

$$(7.3) \quad \mathbf{y}^{(k)} = \frac{n\mathbf{x}^{(k)}}{\langle \mathbf{e}, \mathbf{x}^{(k)} \rangle},$$

then $\{\mathbf{y}^{(k)} : 0 \leq k < \infty\}$ are the projective scaling algorithm iterates of the canonical form problem (7.2).

Proof. We observe that Karmarkar's "potential function" is constant on rays through the origin:

$$(7.4) \quad g(\lambda \mathbf{x}) = g(\mathbf{x}) \quad \text{if } \lambda > 0.$$

Now we prove the theorem by induction on the iteration number k . It is true by definition for $k = 0$. Suppose it is true for a given k . Then the proof of

Theorem 6.1 shows that the nonradial component of the affine scaling vector field agrees with the projective scaling vector field. Hence the radial projection of the homogeneous affine scaling step direction line segment inside \mathbf{R}_+^n is the projective scaling step direction line segment inside \mathbf{R}_+^n . Since Karmarkar's potential function is constant on rays, the step size criterion for the homogeneous affine scaling algorithm causes (7.3) to hold for $k+1$, completing the induction step. \square

Theorem 7.1 proves that the iterates of the homogeneous affine scaling algorithm and the projective scaling algorithms correspond for *any* objective function. Karmarkar [K] proves that the projective scaling algorithm converges in polynomial time provided that the objective function \mathbf{c} is *normalized* so that $\langle \mathbf{c}, \mathbf{x} \rangle \geq 0$ on the polytope of feasible solutions to (7.2) and $\langle \mathbf{c}, \mathbf{x} \rangle = 0$ for at least one feasible \mathbf{x} . Theorem 7.1 allows us to infer that the homogeneous affine scaling algorithm also converges in polynomial time for *normalized* objective functions. These results do not hold for general objective functions; in fact the projective scaling algorithm for a general objective function may not converge to an optimal point, see part III.

The homogeneous affine scaling algorithm may be regarded as an algorithm for solving the *fractional linear program* with objective function $\langle \mathbf{c}, \mathbf{x} \rangle / \langle \mathbf{e}, \mathbf{x} \rangle$. The condition that an objective function be normalized is that $\langle \mathbf{c}, \mathbf{x} \rangle / \langle \mathbf{e}, \mathbf{x} \rangle \geq 0$ on the polytope P of feasible solutions to the homogeneous standard form problem (7.1), with equality for at least one feasible \mathbf{x} . If Karmarkar's stopping rule is used one obtains a polynomial time algorithm for solving this fractional linear program.

8. THE AFFINE SCALING VECTOR FIELD AS A STEEPEST DESCENT VECTOR FIELD

The affine scaling vector field of a strict standard form linear program has an interpretation as a steepest descent vector field of the objective function $\langle \mathbf{c}, \mathbf{x} \rangle$ with respect to a particular Riemannian metric ds^2 defined on the relative interior of the polytope of feasible solutions of the linear program.

We first review the definition of a steepest descent direction with respect to a Riemannian metric. Let

$$(8.1) \quad ds^2 = \sum_{i=1}^n \sum_{j=1}^n g_{ij}(\mathbf{x}) dx_i dx_j$$

be a Riemannian metric defined on an open subset Ω of \mathbf{R}^n , i.e., we require that the matrix

$$(8.2) \quad G(\mathbf{x}) = [g_{ij}(\mathbf{x})]$$

be a positive-definite symmetric matrix for all $\mathbf{x} \in \Omega$. Let

$$(8.3) \quad f: \Omega \rightarrow \mathbf{R}$$

be a differentiable function. The *differential* $df_{\mathbf{x}}$ at \mathbf{x} is a linear map on the tangent space \mathbf{R}^n at \mathbf{x} ,

$$(8.4) \quad df_{\mathbf{x}}: \mathbf{R}^n \rightarrow \mathbf{R}$$

given by

$$(8.5) \quad f(\mathbf{x} + \varepsilon \mathbf{v}) = f(\mathbf{x}) + \varepsilon df_{\mathbf{x}}(\mathbf{v}) + O(\varepsilon^2)$$

as $\varepsilon \rightarrow 0$ and $\mathbf{v} \in \mathbf{R}^n$. The Riemannian metric ds^2 permits us to define the gradient vector field $\nabla_G f: \Omega \rightarrow \mathbf{R}^n$ with respect to $G(\mathbf{x})$ by letting $\nabla_G f(\mathbf{x})$ be that tangent direction such that f increases most steeply with respect to ds^2 at \mathbf{x} . This is the direction of the minimum of $f(\mathbf{x})$ on an infinitesimal unit ball of ds^2 (which is an ellipsoid) centered at \mathbf{x} . Formally we have

$$(8.6) \quad \nabla_G f(\mathbf{x}) = G(\mathbf{x})^{-1} \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{bmatrix}.$$

Note that if $ds^2 = \sum_{i=1}^n (dx_i)^2$ is the Euclidean metric then $\nabla_G f$ is the usual gradient ∇f . (See [Fl, p. 43].)

There is an analogous definition for the gradient vector field $\nabla_G f|_F$ of a function f restricted to a k -dimensional flat F in \mathbf{R}^n . Let the flat F be $\mathbf{x}_0 + V$ where V is an $(n-m)$ -dimensional subspace of \mathbf{R}^n given by $V = \{\mathbf{x}: A\mathbf{x} = \mathbf{0}\}$, in which A is an $m \times n$ matrix of full row rank m . Geometrically the steepest descent direction $\nabla_G f(\mathbf{x}_0)|_F$ is that direction in F that maximizes $f(\mathbf{x})$ on an infinitesimal unit ball centered at \mathbf{x}_0 of the metric $ds^2|_F$ restricted to F . A computation with Lagrange multipliers given in the Appendix shows that

$$(8.7) \quad \nabla_G f(\mathbf{x}_0)|_F = (G^{-1} - G^{-1}A^T(AG^{-1}A^T)^{-1}AG^{-1}) \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \end{bmatrix}$$

where ds^2 has coefficient matrix $G = G(\mathbf{x}_0)$ at \mathbf{x}_0 .

Now we consider a linear programming problem given in strict standard form:

$$(8.8) \quad \begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ A\mathbf{x} = \mathbf{b}, \\ \mathbf{x} \geq \mathbf{0}, \end{cases}$$

having a feasible solution \mathbf{x} with all $x_i > 0$. Karmarkar's steepest descent interpretation of the affine scaling vector field is as follows.

Theorem 8.1 (Karmarkar). *The affine scaling vector field $\mathbf{v}_A(\mathbf{d}; \mathbf{c})$ of a strict standard form problem is the steepest descent vector $-\nabla_G(\langle \mathbf{c}, \mathbf{x}_0 \rangle)|_F$ at $\mathbf{x}_0 = \mathbf{d}$ with respect to the Riemannian metric obtained by restricting the metric*

$$(8.9) \quad ds^2 = \sum_{i=1}^n \frac{dx_i dx_i}{x_i^2}$$

defined on $\text{Int}(\mathbf{R}_+^n)$ to the flat $F = \{\mathbf{x}: A\mathbf{x} = \mathbf{b}\}$.

Before proving this result we discuss the metric (8.9). It may be characterized as the unique Riemannian metric (up to a positive constant factor) on $\text{Int}(\mathbf{R}_+^n)$ which is invariant under the *scaling transformations* $\Phi_D: \mathbf{R}_+^n \rightarrow \mathbf{R}_+^n$ given by

$$x_i \rightarrow d_i x_i \quad \text{for } 1 \leq i \leq n,$$

with all $d_i > 0$, and under the *inverting transformations*

$$I_i((x_1, \dots, x_i, \dots, x_n)) = (x_1, \dots, 1/x_i, \dots, x_n) \quad \text{for } 1 \leq i \leq n,$$

and under all permutations $\sigma((x_1, \dots, x_n)) = (x_{\sigma(1)}, \dots, x_{\sigma(n)})$. The geometry induced by ds^2 on $\text{Int}(\mathbf{R}_+^n)$ is isometric to Euclidean geometry on \mathbf{R}^n under the change of variables $y_i = \log x_i$ for $1 \leq i \leq n$. (These facts are verified by simple calculations which we omit.)

Proof of Theorem 8.1. The metric $ds^2 = \sum_{i=1}^n (dx_i)^2/x_i^2$ induces a unique Riemannian metric $ds^2|_F$ on the region

$$\text{Rel-Int}(\mathbf{P}) = \{\mathbf{x}: A\mathbf{x} = \mathbf{b} \text{ and } \mathbf{x} > \mathbf{0}\}$$

inside the flat $F = \{\mathbf{x}: A\mathbf{x} = \mathbf{b}\}$. The matrix $\overline{G}(\mathbf{x})$ associated with ds^2 is the diagonal matrix

$$\overline{G}(\mathbf{x}) = \text{diag}(1/x_1^2, \dots, 1/x_n^2) = X^{-2},$$

where $X = \text{diag}(x_1, \dots, x_n)$. Using definition (8.7) applied to the function $l_c(\mathbf{x}) = \langle \mathbf{c}, \mathbf{x} \rangle$ we obtain

$$\nabla_{\overline{G}}(l_c(\mathbf{x}))|_F = X(I - XA(AX^2A^T)^{-1}AX)X\mathbf{c}.$$

The right side of this equation is $-\mathbf{v}_A(\mathbf{x}; \mathbf{c})$ by Lemma 4.1. \square

These steepest descent curves are not geodesics of the metric $ds^2|_F$ even in the simplest case. To show this, we consider the strict standard form problem with no equality constraints:

$$\begin{cases} \text{minimize } \langle \mathbf{c}, \mathbf{x} \rangle, \\ \mathbf{x} \geq \mathbf{0}. \end{cases}$$

The A -trajectories for this problem are given by

$$\mathbf{x}(t) = \left(\frac{1}{1/d_1 + c_1 t}, \dots, \frac{1}{1/d_n + c_n t} \right)$$

where $(d_1, \dots, d_n) \in \mathbf{R}_+^n$, by (4.29). On the other hand, the geodesics of $ds^2 = \sum_{i=1}^n (dx_i)^2 / x_i^2$ are

$$\gamma(t) = (e^{a_1 t + b_1}, \dots, e^{a_n t + b_n}),$$

where $\sum_{i=1}^n a_i^2 = 1$ and $-\infty < t < \infty$. To see this, we use the change of variables $y_i = \log x_i$ which transforms the metric to the Euclidean metric $\sum_{i=1}^n (dy_i)^2$, whose geodesics are $(a_1 t + b_1, \dots, a_n t + b_n)$ with $\sum_{i=1}^n a_i^2 = 1$.

It is easy to see that for $n \geq 2$ the point-sets covered by the geodesics $\gamma(t)$ do not coincide with those covered by the curves $\mathbf{x}(t)$, because the coordinates of $\mathbf{x}(t)$ have algebraic dependencies while those of $\gamma(t)$ do not.

APPENDIX. STEEPEST DESCENT DIRECTION WITH RESPECT TO A RIEMANNIAN METRIC

We compute the steepest descent direction $-\nabla_G f(\mathbf{x}_0)|_F$ of a function $f(\mathbf{x})$ defined on a flat $F = \mathbf{x}_0 + \{\mathbf{x}: A\mathbf{x} = \mathbf{0}\}$ with respect to a Riemannian metric $ds^2 = \sum_{i=1}^n \sum_{j=1}^n g_{ij}(x) dx_i dx_j$ at \mathbf{x}_0 . We may suppose without loss of generality that $\mathbf{x}_0 = \mathbf{0}$, and set $G = [g_{ij}(\mathbf{0})]$.

The gradient direction is found by maximizing the linear functional

$$(A.1) \quad \langle df_0, \mathbf{v} \rangle = \left(\frac{\partial f}{\partial x_1}(\mathbf{0}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{0}) \right) \mathbf{v}$$

on the ellipsoid

$$(A.2) \quad \sum_{i=1}^n \sum_{j=1}^n g_{ij} \mathbf{v}_i \mathbf{v}_j = \varepsilon^2,$$

subject to the constraints

$$(A.3) \quad A\mathbf{v} = \mathbf{0}.$$

The direction obtained will be independent of ε .

We set this problem up as a Lagrange multiplier problem. Let

$$\mathbf{d} = \left(\frac{\partial f}{\partial x_1}(\mathbf{0}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{0}) \right)^T.$$

We wish to find a stationary point of

$$(A.4) \quad L = \langle \mathbf{d}, \mathbf{v} \rangle - \lambda^T A\mathbf{v} - \mu(\mathbf{v}^T G \mathbf{v} - \varepsilon^2).$$

The stationarity conditions are

$$(A.5) \quad \partial L / \partial \mathbf{v} = \mathbf{d} - A^T \lambda - \mu(G + G^T) \mathbf{v} = \mathbf{0},$$

$$(A.6) \quad \partial L / \partial \lambda = -A\mathbf{v} = \mathbf{0},$$

$$(A.7) \quad \partial L / \partial \mu = \mathbf{v}^T G \mathbf{v} - \varepsilon^2 = 0.$$

Using (A.5) and $G = G^T$ we find that

$$(A.8) \quad \mathbf{v} = \frac{1}{2\mu} G^{-1} (\mathbf{d} - A^T \lambda).$$

Substituting this into (A.6) yields

$$AG^{-1}A^T\lambda = AG^{-1}\mathbf{d}.$$

Hence

$$(A.9) \quad \lambda = (AG^{-1}A^T)^{-1}AG^{-1}\mathbf{d}.$$

Substituting this into (A.8) yields the stationary point

$$(A.10) \quad \mathbf{v} = \frac{1}{2\mu}(G^{-1} - G^{-1}A^T(AG^{-1}A^T)^{-1}AG^{-1})\mathbf{d}.$$

We show that the tangent vector

$$(A.11) \quad \mathbf{w} = (G^{-1} - G^{-1}A^T(AG^{-1}A^T)^{-1}AG^{-1})\mathbf{d}$$

points in the maximizing direction. To show this, it suffices to show that $\langle \mathbf{d}, \mathbf{w} \rangle \geq 0$. Recall that any positive-definite symmetric matrix G has a unique positive-definite symmetric square root $G^{1/2}$. Using this fact we obtain

$$\begin{aligned} \langle \mathbf{d}, \mathbf{w} \rangle &= \mathbf{d}^T G^{-1} \mathbf{d} - \mathbf{d}^T G^{-1} A^T (AG^{-1}A^T)^{-1} AG^{-1} \mathbf{d} \\ &= (\mathbf{d}^T G^{-1/2})(I - G^{-1/2}A^T(AG^{-1}A^T)^{-1}AG^{-1/2})G^{-1/2}\mathbf{d}. \end{aligned}$$

Now $\pi_w = I - G^{-1/2}A^T(AG^{-1}A^T)^{-1}AG^{-1/2}$ is a projection operator onto the subspace $W = \{\mathbf{x}: AG^{-1/2}\mathbf{x} = \mathbf{0}\}$, so that

$$\begin{aligned} \langle \mathbf{d}, \mathbf{w} \rangle &= (G^{-1/2}\mathbf{d})^T \pi_w (G^{-1/2}\mathbf{d}) \\ &= \|\pi_w(G^{-1/2}\mathbf{d})\|^2 \geq 0, \end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm. Note that there are two special cases where $\langle \mathbf{d}, \mathbf{w} \rangle = 0$. The first is where $\mathbf{d} = \mathbf{0}$, which corresponds to $\mathbf{0}$ being a stationary point of f , and the second is where $\mathbf{d} \neq \mathbf{0}$ but $\langle \mathbf{d}, \mathbf{w} \rangle = 0$, in which case the linear functional $\langle df_0, \mathbf{v} \rangle = \langle \mathbf{d}, \mathbf{v} \rangle$ is constant on the flat F .

The vector (A.11) is the gradient vector field with respect to G . We obtain the analogue of a *unit gradient field* by using the Lagrange multiplier μ to scale the length of \mathbf{v} . Substituting (A.10) into (A.7) yields

$$4\mu^2\varepsilon^2 = \mathbf{d}^T G^{-1} \mathbf{d} - \mathbf{d}^T G^{-1} A^T (AG^{-1}A^T)^{-1} AG^{-1} \mathbf{d},$$

so that

$$\mu = \frac{\pm 1}{2\varepsilon}(\mathbf{d}^T G^{-1} \mathbf{d} - \mathbf{d}^T G^{-1} A^T (AG^{-1}A^T)^{-1} AG^{-1} \mathbf{d})^{1/2}.$$

Choosing the plus sign (for maximization) we obtain from (A.10) that

$$(A.12) \quad \lim_{\varepsilon \rightarrow 0} \frac{\mathbf{v}}{\varepsilon} = \theta(G, \mathbf{d})(G^{-1} - G^{-1}A^T(AG^{-1}A^T)^{-1}AG^{-1})\mathbf{d},$$

where $\theta(G, \mathbf{d})$ is the scaling factor

$$\theta(G, \mathbf{d}) = (\mathbf{d}^T G^{-1} \mathbf{d} - \mathbf{d}^T G^{-1} A^T (AGA^T)^{-1} AG^{-1} \mathbf{d})^{-1/2}.$$

Here $\theta(G, \mathbf{d})$ measures the length of the tangent vector \mathbf{w} with respect to the metric ds^2 . (As a check, note that for the Euclidean metric and $F = \mathbf{R}^n$ formula (A.11) for \mathbf{w} gives the ordinary gradient and (A.12) gives the unit gradient.)

REFERENCES

- [AKRV] I. Adler, N. Karmarkar, G. Resende and S. Veiga, *An implementation of Karmarkar's algorithm for linear programming*, preprint, Univ. of California, Berkeley, 1986.
- [A] K. Anstreicher, *A monotonic projective algorithm for fractional linear programming*, *Algorithmica* **1** (1986), 483–498.
- [Ar] V. I. Arnold, *Mathematical methods of classical mechanics*, Springer-Verlag, New York, 1978.
- [B] E. R. Barnes, *A variation on Karmarkar's algorithm for solving linear programming problems*, *Math. Programming* **36** (1986), 174–182.
- [BL2] D. A. Bayer and J. C. Lagarias, *The nonlinear geometry of linear programming. II. Legendre transform coordinates, and central trajectories*, *Trans. Amer. Math. Soc.* (to appear).
- [BP] H. Busemann and B. B. Phadke, *Beltrami's theorem in the large*, *Pacific J. Math.* **115** (1984), 299–315.
- [Bu] H. Busemann, *The geometry of geodesics*, Academic Press, New York, 1955.
- [Bu2] —, *Spaces with distinguished shortest joins*, *A Spectrum of Mathematics*, Auckland, 1971, pp. 108–120.
- [CH] R. Courant and D. Hilbert, *Methods of mathematical physics*, Vols. I, II, Wiley, New York, 1962.
- [D1] I. I. Dikin, *Iterative solution of problems of linear and quadratic programming*, *Dokl. Akad. Nauk SSSR* **173** (1967), 747–748. (English transl., *Soviet Math. Dokl.* **8** (1967), 674–675.)
- [D2] —, *About the convergence of an iterative process*, *Controllable Systems IM IK SO AN SSR* 1974, No. 12, 54–60. (Russian)
- [F] W. Fenchel, *On conjugate convex functions*, *Canad. J. Math.* **1** (1949), 73–77.
- [FM] A. V. Fiacco and G. W. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*, Wiley, New York, 1968.
- [Fl] W. H. Fleming, *Functions of several variables*, Addison-Wesley, Reading, Mass., 1965.
- [GZ] C. B. Garcia and W. I. Zangwill, *Pathways to solutions, fixed points and equilibria*, Prentice-Hall, Englewood Cliffs, N. J., 1981.
- [Go] C. Gonzaga, *An algorithm for solving linear programming problems in $O(m^3L)$ operations*, *Progress in Mathematical Programming, Interior-Point and Related Methods* (N. Megiddo, Ed.), Springer-Verlag, New York, 1989, pp. 1–28.
- [H] D. Hilbert, *Grundlagen der Geometrie*, 7th ed., Leipzig, 1930. (English transl., *Foundations of geometry*.)
- [Ho] J. Hooker, *The projective linear programming algorithm*, *Interfaces* **16** (1986), 75–90.
- [II] M. Iri and H. Imai, *A multiplicative barrier function method for linear programming*, *Algorithmica* **1** (1986), 455–482.
- [K] N. Karmarkar, *A new polynomial time algorithm for linear programming*, *Combinatorica* **4** (1984), 373–395.
- [KLSW] N. Karmarkar, J. C. Lagarias, L. Slutsman and P. Wang, *Poser series variants of Karmarkar-type algorithms*, A.T. & T. Technical J. (to appear).
- [KV] S. Kapoor and P. M. Vaidya, *Fast algorithms for convex quadratic programming and multi-commodity flows*, *Proc. 18th ACM Sympos. on Theory of Computing*, 1986, pp. 147–159.
- [KMY] M. Kojima, S. Mizuno, and A. Yoshise, *A primal-dual interior point method for linear programming*, *Progress in Mathematical Programming, Interior-Point and Related Methods* (N. Megiddo, Ed.), Springer-Verlag, New York, 1989, pp. 29–48.
- [L3] J. C. Lagarias, *The nonlinear geometry of linear programming. III, Projective Legendre transform coordinates and Hilbert geometry*, *Trans. Amer. Math. Soc.* (to appear).
- [Ln] C. Lanczos, *The variational principles of mechanics*, Univ. of Toronto Press, Toronto, 1949.
- [M1] N. Megiddo, *On the complexity of linear programming*, *Advances in Economic Theory* (T. Bewley, Ed.), Cambridge Univ. Press, 1986.

- [M2] —, *Pathways to the optimal set in linear programming*, Progress in Mathematical Programming, Interior-Point and Related Methods (N. Megiddo, Ed.), Springer-Verlag, New York, 1989, pp. 131–158.
- [MS] N. Megiddo and M. Shub, *Boundary behavior of interior point algorithms in linear programming*, IBM Research Report RJ-5319, Sept. 1986.
- [N] J. L. Nazareth, *Homotopy techniques in linear programming*, *Algorithmica* **1** (1986), 529–535.
- [Re] J. Renegar, *A polynomial-time algorithm, based on Newton's method for linear programming*, *Math. Programming* **40** (1988), 59–94.
- [R1] R. T. Rockafellar, *Conjugates and Legendre transforms of convex functions*, *Canad. J. Math.* **19** (1967), 200–205.
- [R2] —, *Convex analysis*, Princeton Univ. Press, Princeton, N. J., 1970.
- [Sh] M. Shub, *On the asymptotic behavior of the projective scaling algorithm for linear programming*, IBM Tech. Report R5 12522, Feb. 1987.
- [So1] Gy. Sonnevend, *An "analytical centre" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming*, Proc. 12th IFIP Conf. System Modelling, Budapest, 1985, Lecture Notes in Computer Science, 1986.
- [So2] —, *A new method for solving a set of linear (convex) inequalities and its application for identification and optimization*, Proc. Sympos. on Dynamic Modelling, IFAC-IFORS, Budapest, June 1986.
- [SW] J. Stoer and C. Witzgall, *Convexity and optimization in finite dimensions. I*, Springer-Verlag, New York, 1970.
- [Va] P. Vaidya, *An algorithm for linear programming which requires $O(((m+n)n^2 + (m+n)^{1.5}n)L)$ arithmetic operations*, Proc. 19th ACM Sympos. on Theory of Computing, 1987, pp. 29–38.
- [VMF] R. J. Vanderbei, M. J. Meketton, and B. A. Freedman, *A modification of Karmarkar's linear programming algorithm*, *Algorithmica* **1** (1986), 395–407.

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY, NEW YORK, NEW YORK 10027

AT & T BELL LABORATORIES, MURRAY HILL, NEW JERSEY 07974