Nonparametric Multivariate Regression Subject to Constraint

Steven M. Goldman Paul A. Ruud*

May, 1993

Abstract

We review Hildreth's algorithm for computing the least squares regression subject to inequality constraints and Dykstra's generalization. We provide a geometric proof of convergence and several ehancements to the algorithm and generalize the application of the algorithm from convex cones to convex sets.

1. Introduction

Given a design matrix of explanatory variables X and a corresponding vector of dependent variable observations y, there is a functional relationship $z(X) \equiv \mathrm{E}(y \mid X)$. How should we parameterize z and how should we estimate z? Restricting ourselves to the ordinary least squares (OLS) estimation criterion, we focus on the use of prior information for the parameterization of z.

Clearly, if *no* restrictions are placed on the set of possible functions, then the OLS fit, $\hat{z}(X)$, will equal Y at every observation. At unobserved values for the explanatory variables nothing can be said and the function $\hat{z}$ is unconstrained. In contrast, if the set of functions were limited to linear relationships alone, as in the case of linear regression, then $\hat{z}$ would be constrained at *every* possible value of the explanatory variables. We are concerned with an

*We would like to thank Yazhen Wang for bringing Dykstra's article to our attention and William Kahan for helpful discussions.

intermediate case in which $z(X)$ is restricted to a set of functions, say Π. In such cases, the OLS estimation criterion may not yield a unique best fitting function. Instead there is a *subset* of best fitting functions, say, $\Delta \subset \Pi$. For analytical tractability, we will restrict Π to be convex.

A leading example of such regression problems is isotonic regression. In isotonic regression, there is a partial ordering $\succ$ of the observations according to X that implies an ordering of the regression function:

$$\Pi = \{z \mid x_1 \prec x_2 \Rightarrow z(x_1) \leq z(x_2)\}.$$

Monotonic regression is a special case. For a general reference, see Barlow *et al* (1972).

An analytical simplification to computing $\hat{z} \in \Delta$ follows from noting that the best least squares fit only compares the vector $\hat{z}(X)$ to Y and does not directly involve the *function* z. The restrictions on z translate into those for the allowable vectors $z(X)$ as the set $Z = \{z(X) \mid z \in \Pi\}$. As many have noted (e.g., Varian (1984)), viewed in this manner least squares regression becomes a matter of quadratic programming: find the vector which minimizes the Euclidean distance from Y to Z. While conceptually appealing, solving such quadratic programs numerically is infeasible except for small data sets.

Hildreth (1954) attacks this problem quite elegantly employing the concept of duality and the Gauss-Seidel computational algorithm. His analysis is restricted by the requirement that Z be the intersection of *linearly independent half-spaces*. That analysis is extended by Dykstra (1983) to Z that are the intersection of a *finite number of convex cones*. Our analysis builds upon these results by further expanding the possibilities for Z: It can be the intersection of a *finite number of convex sets*. Additionally, we show convergence of the Gauss-Seidel algorithm from an arbitrary starting point. This not only shows the computation to be free of cumulative numerical errors, but it also admits various computational improvements.

2. Basic Notation and Statement of Problem

In particular, consider a data set of N observations on

$$y_n = z(x_n) + u_n$$

$(n = 1, \ldots, N)$ where $E(y_n \mid x_n) = z(x_n)$ is a regression function and x_n is a vector of M explanatory variables for the n^{th} observation. Let y denote the

$N \times 1$ vector of observations $[\,y_n]$ and X denote the $N \times M$ matrix of observed explanatory variables $[x_n']$. We will consider the problem of estimating z by the method of ordinary least squares, minimizing the Euclidean distance between y and z, subject to the restriction that z belong to a convex set of functions of x, $\mathbf{Z}$. Thus, our estimation problem is the solution to the program

$$\min_{z \in \Pi} \|y - z(X)\|^2 \tag{1}$$

where $z(X) \equiv [z(x_n); n = 1, \ldots, N]$. Note that the solution requires only that we find z at the vectors in X. From now on, we will treat z and $z\,(X)$ as synonymous.

Such estimation problems have a long and broad history. The classic example may be monotonic regression, where

$$\mathbf{Z} = \left\{ z \in \mathbf{R}^N \mid \forall i,j : x_i \leq x_j \Rightarrow z(x_i) \leq z(x_j) \right\}.$$

One step more difficult are such second-order restrictions as concavity in z:

$$\mathbf{Z} = \left\{ z \in \mathbf{R}^N \mid \forall \alpha \in \mathbf{R}_+^{N-1}, \forall n : \alpha' X_{\text{-}n} = x_n \Rightarrow \alpha' z(X_{\text{-}n}) \geq z(x_n) \right\}.$$

For example, economists are interested in estimating regressions functions with both monotonicity and concavity because the costs of efficient firms should exhibit these properties with respect to the prices of inputs and outputs. Both monotonic and concave z exhibit $\mathbf{Z}$ that are convex cones. Still more difficult are problems for which $\mathbf{Z}$ is merely convex. This occurs in regression subject to Lipschitz smoothness restrictions:

$$\mathbf{Z} = \left\{ z \in \mathbf{R}^N \mid |z(x_i) - z(x_j)| \leq \|x_i - x_j\|\, M \right\}.$$

3. Hildreth's Procedure

The estimation problem (1) is a form of restricted least squares. This observation is not enough to make computation attractive.[1] Fortunately, Hildreth (1954) proposed a computational procedure that can be applied to this

[1] Linear regression with a small number of inequality restrictions has been previously studied by Judge and Takayama (1966) and Liew (1976).

problem. Recently, Dykstra (1983) has extended Hildreth's procedure; see the next section. Here, we offer a new and simple description of the algorithm as a combination of two classic ideas: the duality theorem of Kuhn and Tucker and the numerical algorithm called Gauss-Seidel. After describing the algorithm, we give a geometric proof of its convergence property.

Hildreth begins with the primal problem (1) written as

$$\min_{Az \leq 0} \|y - z\|^2 \tag{2}$$

where A is an $R \times N$ matrix whose rows are the elements of the set $\mathcal{A}$. The algorithm breaks down (2) into a sequence of sub-problems which are easy to solve. Hildreth's procedure rests on the following duality theorem ((Kuhn & Tucker, 1951, pp. 487, 491-2) (Hildreth, 1954, p. 604) (Luenberger, 1969, pp. 299-300)).

Theorem 1 (Kuhn-Tucker). *The* primal *constrained quadratic minimization problem (1) is equivalent to the* dual *problem*

$$\min_{\lambda \geq 0} \|y - A'\lambda\|^2 \tag{3}$$

If we denote the solution to the dual problem by $\hat{\lambda}$, then the solution to the primal problem is $\hat{z} = y - A'\hat{\lambda}$.

The elements of the vector λ are the Lagrange multipliers associated with each of the convexity constraints. The dual problem has constraints with a convenient form: they are univariate inequalities. This property makes the application of the Gauss-Seidel method natural and this is essentially what Hildreth did to derive the sequence of simple sub-problems of the form[2]

$$\min_{\lambda_r \geq 0} (g_r - p_{rr}\lambda_r)^2 \tag{4}$$

given g_r and p_{rr}. Given a feasible $\lambda \geq 0$, at each step of the process we focus on a single element of λ, λ_r. The dual objective function is a simple

[2]The Gauss-Seidel method solves a system of n equations of the form $x_i = f_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ $(i = 1, \ldots, n)$ by mapping x^{k-1} into x^k through the steps $x_i^k = f_i(x_1^k, \ldots, x_{i-1}^k, x_{i+1}^{k-1}, \ldots, x_n^{k-1})$ on the k^{th} iteration. In optimization problems, these equations are first-order conditions. See Quandt (1983) for a brief survey of the applications of the Gauss-Seidel method by econometricians.

quadratic in λ_r and the constrained optimization for this element is easily solved: If the unconstrained solution yields a negative value for λ_r then we set $\lambda_r = 0$. Treating one iteration as a cycle through all of the elements of λ, the i^{th} iteration can be expressed in terms of the previous iteration as

$$\lambda_r^i = \max(0, q_r^i) \tag{5}$$

where

$$q_r^i = -\frac{1}{p_{rr}} \left(d_r + \sum_{j=1}^{r-1} p_{rj} \lambda_j^i + \sum_{j=r+1}^{R} p_{rj} \lambda_j^{i-1} \right) \tag{6}$$

$P \equiv [p_{ij}] = AA'$, $d = Ay$, and R is the number of constraints imposed by convexity.

In general, the Gauss-Seidel method is not guaranteed to converge. It is sufficient for convergence, however, for the mapping from λ^{i-1} to λ^i to be continuous and a contraction mapping on a compact set (Quandt, 1983, p. 726). In Hildreth's particular application of Gauss-Seidel, these conditions are satisfied for the sequence $\{z^i = y - A'\lambda^i\}$. Continuity follows when there is a one-to-one relationship between z and λ in the equation $z = y - A'\lambda$: That is, A must be full row rank so that there are no redundancies among the inequality restrictions. Given the continuity in the sequence $\{z^i = y - A'\lambda^i\}$ produced by the Hildreth procedure, the contraction property follows immediately: By construction, $\|z^i\| < \|z^{i-1}\|$ if $\lambda^i \neq \lambda^{i-1}$.

Theorem 2 (Hildreth). *If A is full row rank, then the iterations in (5) and (6) converge to $\hat{\lambda}$.*

3.1. The Geometry of Hildreth's Procedure

Hildreth's procedure has a useful geometric interpretation that is explained by Dykstra (1983).[3] The basic Kuhn-Tucker duality between (2) and (3) is illustrated in Figure 1 for a case in which $N = 2$ and $M = 2$. The two-dimensional cone $K = \{z | Az \leq 0\}$ describes the intersection of two inequality constraints on z and the point y is the data. The normal vectors α_1 and α_2 determine the directions of the inequalities, pointing into the interior of K.

[3] Dykstra cites Barlow and Brunk (1972) for the geometrical interpretation of the quadratic program. They were studying isotonic regression by least squares. Barlow and Brunk cite Sinden (1962).

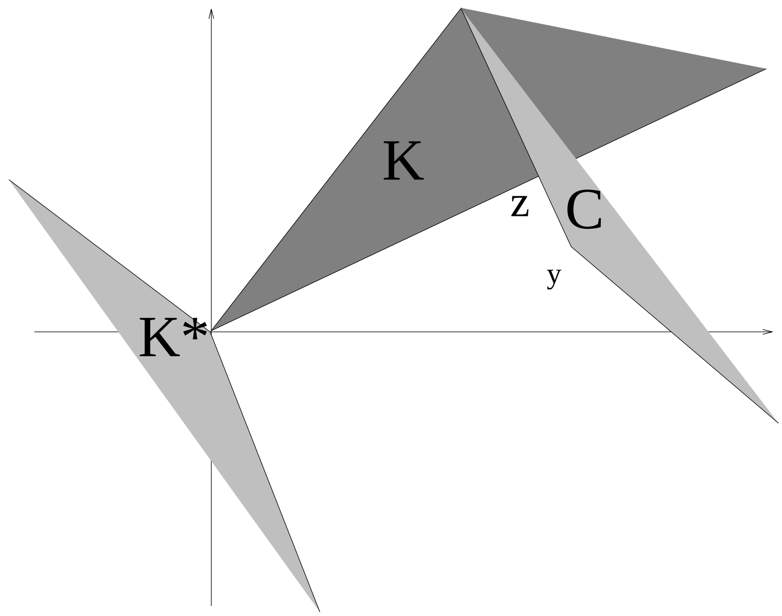

Figure 1: Duality in the Gauss-Seidel Method

The primal quadratic programming problem is to find the point z in K that is closest to y and the solution $\hat{z}$ is found at the orthogonal projection of y onto the nearest "facet" of K. The constraints in the dual problem are also a convex cone which we will call K^*. The cone K^* is the *dual cone* of K.

Definition 1. *The dual cone to any convex cone K, denoted K^*, is given by $K^* = \{x \mid \forall k \in K, k'x \leq 0\}$.*

The geometry of Figure 1 shows that $y - \hat{z} = A'\hat{\lambda}$ is the solution to the dual problem of finding the point in K^* that is closest to y.

We can also view the dual problem as a minimum distance problem in the same parameter space as the primal problem

$$\begin{cases} \min_z \|z\|^2 \\ \text{subject to } z \in \mathcal{C} = \{x | x = y - A'\lambda,\ \lambda \geq 0\} \end{cases} \tag{7}$$

The set $\mathcal{C}$ is a rotation of the dual cone with its vertex translated to y. Figure 1 shows that $\hat{z}$ also solves the reparameterized dual problem.

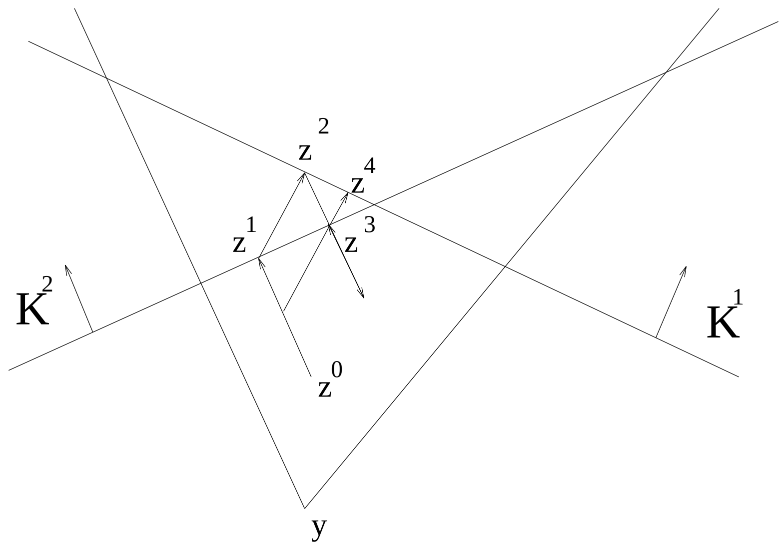

Figure 2: The Iteration Process

The cycles of Hildreth's procedure are pictured in Figure 2 for a case in which $N = 3$ and $M = 2$. In the previous example in Figure 1, the path is trivial. In Figure 2, the starting point for the procedure is simply y so that $\lambda = 0$ initially. The first iteration of the algorithm chooses the best value for λ_1, holding the value of λ_2 fixed at zero. Thus, the first step z_1^1 is the orthogonal projection of y onto K_1. In the second step, λ_1 is held fixed at its new value and the algorithm chooses the best value for λ_2. This results in a move to z_1^2, the orthogonal projection of $y - \lambda_1\alpha_1$ onto K_2. This completes a single cycle. The figure also shows the second cycle so that the zig-zag pattern of convergence is evident.

The discussion below justifies Theorems 1 and 2. While the former is widely understood, its explanation bears a brief repetition in this context. We give a new, geometric proof of the latter theorem that we will also apply to Dykstra's generalization of Hildreth's procedure. The strategy of the argument is as follows: We show

i. the necessary and sufficient conditions for the unique optimum;

ii. that the solution to the minimization problem 1 is the closest point to the origin of the form $y - w$ for $w \in \mathcal{Z}^*$;

iii. that the algorithm defines a function $f : \mathcal{R}^N \rightarrow \mathcal{R}^N$ which maps a closed bounded subset into itself;

iv. that the function f is a contraction mapping with a fixed point on the closed bounded subset described in (iii) and the fixed point satisfies the conditions described in (i).

Items (i) and (ii) comprise Theorem 1 and (iii) and (iv) justify Theorem 2.

3.2. Necessary and Sufficient Conditions for the Unique Optimum

We noted previously that $\hat{z}$ exists and is unique. The necessary and sufficient conditions that $\hat{z}$ must satisfy are

$$\hat{z} \in \mathcal{Z} \text{ and } \hat{z}'(y - \hat{z}) = 0. \tag{8}$$

and, for every other ray in the cone, say $k \in \mathcal{Z}$,

$$k'(y - \hat{z}) < 0 \tag{9}$$

(that is, $y - \hat{z} \in \mathcal{Z}^*$). These are necessary because otherwise, by simple geometry, the convexity of $\mathcal{Z}$ implies that there would be a closer point to y than $\hat{z}$ on the line segment between k and $\hat{z}$.

The sufficiency of (8) and (9) follows from

$$\|y - w\|^2 = \|y - \hat{z}\|^2 - 2w'(y - \hat{z}) + \|w - \hat{z}\|^2 > \|y - \hat{z}\|^2 \tag{10}$$

for all $w \in \mathcal{Z}$, $w \neq \hat{z}$.

3.3. The Dual Problem: Proximity to the Origin

According to (9), $\hat{z}$ can be expressed as $y - \hat{w}$ for some $\hat{w} \in \mathcal{Z}^*$, where $\mathcal{Z}^*$is the dual cone of $\mathcal{Z}$.[4] Define $\mathcal{C} = \{z \mid z = y - w, w \in Z^*\}$. We will now show that if y is external to $\mathcal{Z}$, $\hat{z}$, the solution to (8) and (9), is also the

[4]This result is cited by Dykstra (1983) from Barlow and Brunk (1972).

closest point in $\mathcal{C}$ to the origin. Consider any other $x \in \mathcal{C}$. Then we can write $x = \hat{z} - w + \hat{w}$. Then

$$\|x\|^2 = \|\hat{z}\|^2 + 2(\hat{w} - w)'\hat{z} + \| \hat{w} - w\|^2.$$

Now according to (8), $\hat{w}'\hat{z} = 0$, so that

$$\|x\|^2 = \|\hat{z}\|^2 - 2w'\hat{z} + \|\hat{w} - w\|^2 \geq \|\hat{z}\|^2.$$

since $w'\hat{z} \leq 0$ for all $w \in Z^*$.

Thus, if $\hat{z}$ solves the original problem (1) then $y - \hat{z}$ solves the dual problem

$$\min_{w \in \mathcal{Z}^*} \|y - w\|^2 \tag{11}$$

We can apply this general duality to obtain Theorem 1: When $\mathcal{Z} = \{z | Az \leq 0\}$, then $\mathcal{Z}^* = \{w | w'z \leq 0 \ \forall z \in \mathcal{Z}\} = \{w | w = A'\lambda, \lambda \geq 0\}$.

3.4. The Iterative Process

Now consider the iterative process in Hildreth's procedure. It begins with a point $z^0 \in \mathcal{C}$. Over the i^{th} cycle through (5) and (6), we obtain a sequence of tentative solutions to the primal problem in

$$z_r^i = y - \sum_{j=1}^{r} \alpha_j \lambda_j^i - \sum_{j=r+1}^{R} \alpha_j \lambda_j^{i-1} = y - \sum_{j=1}^{r} w_j^i - \sum_{j=r+1}^{R} w_j^{i-1} \tag{12}$$

where $w_j^i = \alpha_j \lambda_j^i$. Every point z_r^i in the sequence is also a member of $\mathcal{C}$. To see this, note that according to the Gauss-Seidel method, z_r^i is chosen to be the closest element in K_j to $z_r^{i-1} + w_r^{i-1}$. Thus, $w_r^i = z_r^{i-1} - z_r^i + w_r^{i-1} \in K_j^*$ by (9). The process begins with $w_r^0 \in K_r^*$ so that all $w_r^i \in K_r^*$ and $z^i \in \mathcal{C}$. Therefore, the algorithm is a mapping from $\mathcal{C}$ onto itself.

In addition, every z_r^i is at least as close to the origin as z_r^{i-1}. By (10),

$$\left\|z^{i-1}\right\|^2 = \left\|z^i\right\|^2 - 2w_j^{i-1\prime} z^i + \left\|w_j^i - w_j^{i-1}\right\|^2 \geq \left\|z^i\right\|^2 \tag{13}$$

because $z_r^i \in K_r$ and $w_r^{i-1} \in K_r^*$. Therefore all z_r^i are elements of the closed sphere $S(z^0) = \{z | \, \|z\| \leq \|z^0\|\}$. The intersection of $S(z^0)$ and $\mathcal{C}$ is a closed and bounded subset of $\mathcal{C}$.

For the moment, we will assume that any point $z \in \mathcal{C}$ may be written $z = y - \sum_{r=1}^{R} w_r$ where the $w_r \in K_r^*$ are unique. This can always be made to occur, by reducing the K_r to a non redundant set of restrictions. Let us define a function $f : C \to C$ obtained from passing through a complete cycle in the $K_r, r = 1, \ldots, R$: Let z be any z_1^{i-1} of the algorithm and let $f(z) = z_1^i$. Then f is continuous. We have continuity if we look at the mapping of $\{w_1^{i-1}, \ldots, w_R^{i-1}\} \to \{w_1^i, \ldots, w_R^i\}$. The continuity of f follows from the uniqueness of w for every $z \in \mathcal{C}$.

3.5. The Contraction Process

By (13), $\|f(z)\| \leq \|z\|$. The inequality is strict if $w^0 \neq w^R$, so that a stationary length for z requires stationary w_r's; and stationary w_r's imply a fixed point of $f(z)$. We can show that the only fixed point of f is $\hat{z}$. By construction, if (z, w) is a stationary point, then for each r, $z'w_r = 0$. Then, denoting $\sum_{n=1}^{R} w_r$ as w, $z'w = 0$ with $z \in \mathcal{Z}$ and $w \in \mathcal{Z}^*$ and satisfies the necessary and sufficient conditions for the unique minimum. Therefore the fixed point is the closest point we seek.

We now show the process described in Section (3.4) converges to the $\hat{z}$ and $\hat{w}$ described by Sections (3.2) and (3.3). Take the intersection of $\mathcal{C}$ with the space of points at least as near to the origin as the original point, y. Then for any point $z \in \mathcal{C}$, $f(z)$ is also in that intersection. Since $g(z) \equiv \|f(z)\| - \|z\|$ is strictly negative on $\mathcal{C}$ except at the unique $\hat{z}$ and $g(z)$ is continuous on $\mathcal{C}$, the process $f \cdot f \cdot \ldots \cdot f(z)$ must converge to $\hat{z}$. This establishes convergence to $\hat{z}$ from an arbitrary starting point in the intersection of $\mathcal{C}$ and the ball of radius $\|y\|$. Since y itself is a member of this intersection, we may simply start with y.

The above argument is easily modified to allow for redundancy in the set of restrictions. Even if an element of $\mathcal{C} \cap S(z^0)$, say w, does not have a unique representation in terms of the w_r's, it would suffice if the set

$$W(z^0) \equiv \{(w_1, ..., w_R) | \sum_{r=1}^{R} w_r = w, w \in \mathcal{C} \cap S(z^0)\}$$

were bounded. Then we could be assured that the contraction was bounded away from zero (for w's bounded away from $y - \hat{z}$) by simply looking at the smallest contraction on the closed set of w's within that bound.

An additional argument is required if $\mathcal{C}$ contains a subspace, because in such cases the set $W(z^0)$ is not bounded. Nevertheless, the contraction is still bounded away from zero. Since the K_r's are half spaces, every z^r must lie on a closed line segment between z^{r-1} and K_r no matter which $w \in W(z^0)$ one chooses. Thus z^R is contained in a closed set $Z^R(z^0)$ that depends only on z^0; this closed set is the range of f over all possible w. Furthermore, this set does not contain z^0 unless $z^0 = \hat{z}$. The closest point to z^0 in $Z^R(z^0)$ is the smallest, non-zero, contraction.

4. Dykstra's Procedure for Convex Cones

In this section we describe Dykstra's generalization of Hildreth's procedure. Hildreth focused on convexity constraints in linear regression which, as we have seen, are linear inequalities that correspond to polyhedral convex cones. Dykstra considers constraints that imply more general convex cones than finite ones. Such cones hold our interest as well because they arise with the inclusion of additional smoothness constraints, as we will explain below. In this section, we extend our interpretation of Hildreth's procedure to that of Dykstra. In so doing, we provide an alternative, geometric proof of the convergence of this procedure.

Theorem 3 (Dykstra). [5]*Let $y \in \mathcal{R}^N$ and $K_r, r = 1, \ldots, R$ be R closed convex cones. A convergent, iterative solution to the restricted least squares problem*

$$\min_{z \in \cap_{r=1}^{R} K_r} \|y - z\|^2 \tag{14}$$

is to begin with a point $z_0 = y - \sum_{r=1}^{R} w_r^0$ where $w_r^0 \in K_r^*$ and on the r^{th} iteration of the i^{th} cycle choose z_r^i to be the point in K_r closest to $z_{r-1}^i + w_r^{i-1}$, where $w_r^i \equiv z_{r-1}^i - z_r^i + w_r^{i-1}$.

By direct substitution, $z_r^i = y - \sum_{j=1}^{r} w_r^i - \sum_{j=r+1}^{R} w_r^{i-1}$ and we formally reproduce (12), an iteration of Hildreth. Thus, Dykstra's procedure has the same interpretation as Hildreth's: they are applications of Gauss-Seidel to the dual problem. As Dykstra points out, the dual to (14) is

$$\min_{w \in K_1^* \oplus K_2^* \oplus \ldots \oplus K_R^*} \|y - w\|^2 \tag{15}$$

[5]Dykstra (1983, Theorem 3.1, p. 839).

and the r^{th} iteration of a cycle is the r^{th} iteration of a cycle of Gauss-Seidel applied to each of the K_r^* in succession:

$$\min_{w \in K_r^*} \|y - w\|^2 \text{ is dual to } \min_{z \in K_r} \|y - z\|^2$$

For this reason, $w_r^i \in K_r^*$ for all $i, r > 0$ and every point z_r^i in the sequence is also a member of $\mathcal{C}$, as defined in Section 3.3.

Dykstra gives a formal proof of this theorem. We can extend the arguments in Sections 3.4 and 3.5 to provide an alternative, geometric proof. Such an extension is straightforward for polyhedral convex cones. Even though the set of possible z_r that can follow a z_{r-1} is no longer a simple line segment, this set is still closed so that by induction $Z^R(z_0)$ remains closed. Our difficulties arise in the case of smooth cones where such sets can be open. This openness causes no problems, however, because the boundary points involve the fastest rates of contraction.

Lemma 1. *Let $P(z|K)$ be the point in the convex cone K closest to z. Then*

$$\|P(z|K)\| \geq \|P(z+w|K)\|$$

for all $w \in K^*$.

Proof. According to duality, $P(z|K)$ is the shortest vector in the set $C = \{z' | z' = z - v, v \in K^*\}$. Similarly, $P(z+w|K)$ is the shortest vector in the set $C' = \{z' | z' = z + w - v, v \in K^*\}$. Since $C \subset C'$, the result follows.□

Lemma 2. *Let $Z^K(z)$ be the range of $P(z+w|K)$ for all $w \in K^*$. If $z \notin K$, then every member of the closure of $Z^K(z)$ has a length strictly less than $\|z\|$.*

Proof. If $z \notin K$ then $z \notin Z^K(z)$ and $\|P(z|K)\| < \|z\|$. From lemma 1, $\|z\| > \|P(z|K)\| \geq \|P(z+w|K)\|$ for all $w \in K^*$. Since $\|P(z+w|K)\|$ is continuous in w, the result follows immediately. □

Let $W(z_0) = \{w | w \in K_1^* \times K_2^* \times ... \times K_R^*, y - \sum_{r=1}^R w_r = z_0\}$ denote the feasible w which could generate z_0.

Define $P^r(z_0, w)$ inductively:

$$P^1(z_0, w) = P(z_0 + w_1 | K^1),$$

$$P^r(z_0, w) = P(P^{r-1}(z_0, w) + w_r | K_r).$$

Let $Z^r(z_0) = \{z | \exists w^0 \in W(z_0), z = P^r(z_0, w^0)\}$ denote the set of projections on the r^{th} cone which are feasible from an initial z^0, and let $\overline{Z}^r(z_0)$ denote the closure of $Z^r(z_0)$.

Lemma 3. *Either*

1. $z_0 \notin Z^R(z_0)$ and $\forall z \in \overline{Z}^R(z_0), \|z\| < \|z_0\|$, or
2. $z_0 \in Z^R(z_0)$ and $z_0 = \hat{z}$.

Proof. If $z_0 \notin K^r$ then the result is true by Lemma 2.

If $z_0 \in K^r$ then $w_r^0 = 0$ implies $z_0 \in Z^r(z_0)$, a contradiction, so $w_r^0 \neq 0$. But then either there is a contraction and the result is true, or $P(z_0 + w_r^0 | K^r) = z_0$ is on the boundary of K^r in which case $z_0 \in Z^r(z_0)$ and there is a contradiction. So if $z_0 \notin Z^R(z_0)$, then $\forall z \in \overline{Z}^R(z_0)$, $\|z\| < \|z_0\|$. Alternatively, if $z_0 \in Z^R(z_0)$ then $\exists w \in W(z_0)$ such that $\forall r, z_0 \in Z^r(z_0)$ and $w_r' z_0 = 0$. But this is simply the sufficiency condition (8) and $z_0 = \hat{z}$. □

We are now ready to prove the theorem, namely that for $z_0 \in \mathcal{C} \cap S(y)$, $\lim_{i \to \infty} z_R^i = \hat{z}$.

Proof. **(Theorem 3)** It immediately follows from Lemma 3 that if $\|z_0\| - \|\hat{z}\| \geq \epsilon$, then $\min_{z \in \overline{Z}^R(z_0)} (\|z_0\| - \|z\|) > 0$. Denote $D \equiv \{z_0 \in \mathcal{C} \cap S(y), \|z_0\| - \|\hat{z}\| \geq \epsilon\}$. Since $Z \equiv \bigcup_{z_0 \in D} \overline{Z}^R(z_0)$ is closed and bounded, then $\exists \delta > 0$, such that $\forall z \in Z$, $\|z_0\| - \|z\| > \delta$. Thus the contraction proceeds, unimpeded towards $\hat{z}$. □

Drawing this analogy between the procedures of Hildreth and Dykstra gives a key insight into the latter: One can start the latter anywhere in $\mathcal{C}$, not just at y as Dykstra does. As a result of this insight, one can potentially begin from better starting values, guard against accumulated numerical errors, and switch between various parameterizations of K as an intersection of K_r's. The last technique can substantially accelerate the rate of convergence because it permits attempts at the final solution based on the active constraints.

5. Closing Remarks

The programming problem can be generalized from convex cones to convex sets. Hildreth and Dykstra both restricted themselves to cones, but this is not necessary. For example, Hildreth's primal problem (2) can be generalized to

$$\min_{Az \le b} \|y - z\|^2 \tag{16}$$

where b is an arbitrary vector in $\mathbf{R}^N$. This program can be transformed into one with conic constraints by taking

$$u \equiv \begin{pmatrix} z - y \\ t \end{pmatrix}, \; B = \begin{pmatrix} A & -b \end{pmatrix}$$

so that (16) can be written as

$$\min_{\substack{Bu \le 0, \\ t=1}} \|u\|^2 . \tag{17}$$

Given the solution to the conic problem

$$\begin{pmatrix} z^* - y \\ t^* \end{pmatrix} \equiv u^* = \arg \min_{Bu \le 0} \|w - u\|^2 \quad \text{where} \quad w \equiv \begin{pmatrix} 0_N \\ 1 \end{pmatrix}$$

and 0_N is the N-dimensional zero vector, $z^{**} = z^*/t^*$ is the solution to (17). To see this note that

$$\forall u : t = t^*, \; \|t^* w - u\|^2 = \|w - u\|^2 - (1 - t^*)^2$$

so that

$$z^* \equiv \arg \min_{\substack{z: Bu \le 0, \\ t=t^*}} \|t^* w - u\|^2 \quad \Rightarrow \quad z^{**} \equiv z^*/t^* = \arg \min_{\substack{z: Bu \le 0, \\ t=1}} \|w - u\|^2$$

by a simple rescaling. The relationship among the points is illustrated by Figure 3.

As an example, an iterative solution to the smooth regression problem

$$\begin{cases} \min \|y - z\|^2 \\ \text{subject to } z_m - z_n \le M_{mn}, \; n, m = 1, \ldots, N \end{cases}$$

sets

$$\begin{pmatrix} z_m^{i+1} \\ z_n^{i+1} \\ t^{i+1} \end{pmatrix} = \frac{1}{2 + M_{mn}^2} \begin{pmatrix} (1 + M_{mn}^2) z_m^i + z_n^i + M_{mn} t^i \\ (1 + M_{mn}^2) z_n^i + z_m^i - M_{mn} t^i \\ M_{mn} z_m^i - M_{mn} z_n^i + 2t^i \end{pmatrix} .$$

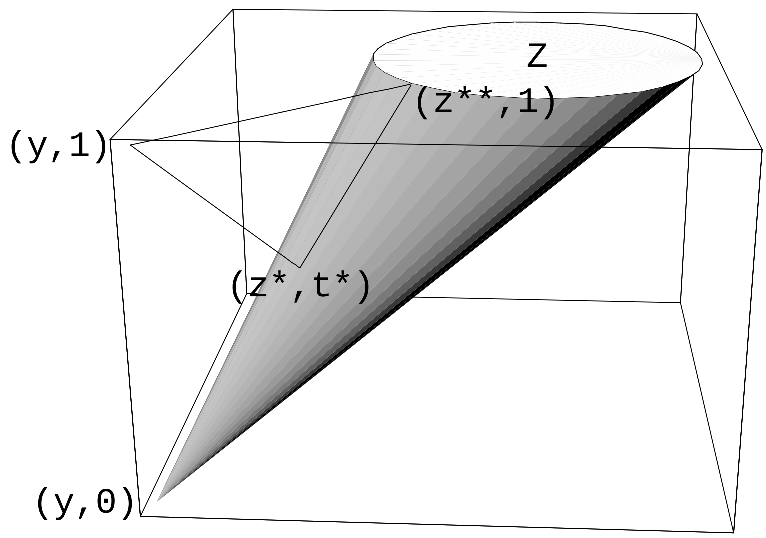

Figure 3: The Case of a Convex Constraint Set

References

Barlow, R. E., & Brunk, H. D. 1972. The Isotonic Regression Problem and Its Dual. *Journal of the American Statistical Association*, **67**(337), 140–147.

Barlow, R. E., Bartholomew, D. J., Bremner, J., & Brunk, H. D. 1972. *Statistical Inference Under Order Restrictions*. New York: Wiley.

Dykstra, Richard L. 1983. An Algorithm for Restricted Least Squares. *Journal of the American Statistical Association*, **78**(364), 837–842.

Hildreth, Clifford. 1954. Point Estimates of Ordinates of Concave Functions. *Journal of the American Statistical Association*, **49**, 598–619.

Judge, George G., & Takayama, T. 1966. Inequality Restrictions in Regression Analysis. *Journal of the American Statistical Association*, **61**, 166–181.

Kuhn, H. W., & Tucker, A. W. 1951. Nonlinear Programming. *In:* Neyman, Jerzy (ed), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley: University of California Press.

Liew, Chong Kiew. 1976. Inequality Constrained Least-Squares Estimation. *Journal of the American Statistical Association*, **71**, 746–751.

Luenberger, David G. 1969. *Optimization by Vector Space Methods.* New York: Wiley.

Quandt, Richard E. 1983. Handbook of Econometrics. vol. I. Amsterdam: North-Holland.

Sinden, F. W. 1962. A Geometric Representation for Pairs of Dual Quadratic or Linear Programs. *Journal of Mathematical Analysis and Applications*, **5**(2), 378–402.

Varian, Hal R. 1984. The Nonparametric Approach to Production Analysis. *Econometrica*, **52**(2), 579–597.