ON THE BALANCED LEAST-SQUARES TRANSFORMATION

Jan De Leeuw*
Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

An alternative transformation which can be used in nonmetric scaling has been proposed by Alan Yates. In the first part of the paper we outline a general conceptual framework for the transformational approach to scaling. In the second part we analyze the Yates transformation, prove its monotonicity, give explicit (noniterative) formulas, discuss the computation of the transformed values, and discuss some anomalies the transformation has. The conclusion is that this transformation should not be used in Z-space, as Yates does. The use of the transformation in T-space has some conceptual advantages over the use of positive orthant or absolute value methods.

---

*On leave of absence from the University of Leiden, Leiden, The Netherlands.

ON THE BALANCED LEAST-SQUARES TRANSFORMATION

Jan De Leeuw*
Bell Laboratories
Murray Hill, New Jersey 07974

Summary

A nonmetric scaling method proposed by Alan Yates is analyzed in this paper. We give some alternative representations of the loss functions involved, analyze the iterative process proposed by Yates, give some explicit min-max representations, and prove monotonicity of the transformation.

0: ~~Introduction~~ Transformation methods for nonmetric scaling

For our purposes we can define a measurement model as a vector valued function $Z(\theta)$ of a parameter θ varying in a general parameter space Θ. We shall assume a fixed indexing of the n coordinate values of $Z(\theta)$, and we shall call the vector $\{z_i(\theta)\}$ the vector of model values. In this interpretation Z is a mapping of Θ into the measurement space Ω_A, which is n dimensional real space. The data space Δ_A is the set of all n×n matrices $S = \{\sigma_{ij}\}$ satisfying

S_1: σ_{ij} is either 0, 1, or -1

S_2: $\sigma_{ij} = -\sigma_{ji}$

S_3: $\sigma_{ij} = 1$ if $\sigma_{ik} = \sigma_{kj} = 1$ for any k.

An equivalent definition of Δ_A is the set of all partial orderings over $I = \{1,2,...,n\}$. For a given S in the data space, the measurement inequalities are the n^2 inequalities

---

*On leave of absence from the University of Leiden, Leiden The Netherlands.

$$\sigma_{ij}(z_i(\theta)-z_j(\theta)) \geq 0$$

Observe that to require equality of $z_i(\theta)$ and $z_j(\theta)$ in this framework, we have to use a different indexing such that both $z_i(\theta)$ and $z_j(\theta)$ are mapped into the same model value $z_k(\theta)$. Another possibility is to use a second data space Δ_B which contains matrices $A = \{\alpha_{ij}\}$ satisfying

A_1: α_{ij} is either 0 or 1.
A_2: $\alpha_{ij} = \alpha_{ji}$.
A_3: $\alpha_{ij} = 1$ if $\alpha_{ik} = \alpha_{kj} = 1$ for any k.

(or, equivalently, the set of all equivalence relations over I) and a second set of <u>measurement equalities</u>

$$\alpha_{ij}(z_i(\theta)-z_j(\theta)) = 0$$

We shall adopt the first possiblity in this paper. We also define a second measurement space Ω_B, which is an n^2 dimensional real space, and a mapping T of $\Delta_A \times \Theta$ into Ω_B by the rule

$$t_{ij}(S,\theta) = \sigma_{ij}(z_i(\theta)-z_j(\theta))$$

Moreover, we let Ω_A^{S} be the cone of all n-dimensional vectors $\{z_i\}$ satisfying

$$\sigma_{ij}(z_i - z_j) \geq 0 \text{ for all } i,j = 1,\ldots,n$$

and we let Ω_B^+ be the nonnegative orthant of Ω_B, i.e., the cone of all n^2-dimensional vectors $\{t_{ij}\}$ satisfying $t_{ij} \geq 0$ for all $i,j = 1,\ldots,n$. The measurement inequalities can now be written in the two equivalent ways

$$Z(\theta)\epsilon\Omega_A(S)$$

$$T(\theta,S)\epsilon\Omega_B^+$$

Observe that in the first formulation, the cone depends on $S\epsilon\Delta_A$ and the mapping Z is independent of S. In the second formulation, the mapping T depends on S, but the cone is independent of S.

The two different ways of choosing the measurement space define two different approaches to nonmetric scaling. Let ϕ_S be a mapping of Ω_A into $\Omega_A(S)$, and let ψ be a mapping of Ω_B into Ω_B^+ with the property that

$$P_A: \quad z\epsilon\Omega_A(S) \leftrightarrow \phi_S(z) = z$$

$$P_B: \quad t\epsilon\Omega_B^+ \leftrightarrow \psi(t) = t.$$

The mappings ϕ_S and ψ are called the <u>nonmetric</u> transformations. The measurement inequalities are now equivalent with

$$\phi_S(Z(\theta)) = Z(\theta)$$

$$\psi(T(S,\theta)) = T(S,\theta)$$

The class of transformational algorithms in Z-space minimize a suitably normed version of

$$\lambda_A(S,\theta) = \delta_A[Z(\theta), \phi_S\{Z(\theta)\}]$$

and the transformational algorithms in T-space minimize a suitably normed version of

$$\lambda_B(S,\theta) = \delta_B[T(S,\theta),\psi\{T(S,\theta)\}]$$

Here δ_A and δ_B are metrics defined on the respective linear spaces. Both the Kruskal (1964a,b) block almagamation method and the Guttman (1968b) rank image method are transformational methods in Z-space, the positive orthant method (De Leeuw, 1968, 1970), the absolute value method (Guttman, 1968) and the pairwise method (Johnson, 1973) are all transformational methods in T-space. In fact, the positive orthant method has the other two T-space methods as special cases.

A particular class of transformational methods are the projection methods. A projection method in Z-space defines ϕ_S by the rule

$$\hat{z} = \phi_S(Z(\theta)) \text{ iff } \delta_A[Z(\theta),\hat{z}] = \inf_{z \in \Omega_A^{\mathcal{S}}(S)} \delta_A[z(\theta),z]$$

A projection method in T-space defines ψ by

$$\hat{t} = \psi[T(S,\theta)] \text{ iff } \delta_B[T(S,\theta),\hat{t}] = \inf_{t \in \Omega_B^r} \delta_B[T(S,\theta),t]$$

Kruskal's method is a projection method in T-space, but the rank image method is not. The positive orthant method is a projection method, and so are its special cases. In fact, if δ_B is any of the power (or ℓ_p) metrics, we find

$$\hat{t} = \tfrac{1}{2}\{|T(S,\theta)| + T(S,\theta)\} = \max[T(S,\theta),0]$$

and $\lambda(S,\theta)$ is the δ_B-norm of the vector $\max[-T(S,\theta),0]$.

The distinction made by Guttman (1968b), Johnson (1973) and Roskam (1969b) which classifies algorithms as nontransformational or transformational is confusing. They maintain that the positive orthant method is nontransformational, while our discussion shows that it has a transformational (even projectional) interpretation in a different space. The fact that the transformation is very easy to compute, and that it can be substituted into the formula for $\lambda(S,\theta)$ right away, is not essential. The distinction between projection algorithms and other nonprojectional transformational algorithms is much more important, because for a suitable choice of the metrics δ the projection algorithms have a number of advantages like differentiability (De Leeuw, 1973).

The Z-space and T-space methods do not exhaust the possible nonmetric scaling methods. In fact, a more general description is possible, in which the two methods turn out to be the opposite extremes of a large class of methods. First we define a <u>nonmetric data structure</u> as a finite set of matrices $S_1,\ldots,S_m \varepsilon \Delta_A$. The measurement inequalities read

$$\sigma_{ij}^k (z_i(\theta)-z_j(\theta)) \geq 0.$$

For each $k = 1,\ldots,m$ the appropriate subset of these inequalities defines a cone Ω_k in Z-space. The <u>loss function</u> is a suitably scaled version of

$$\lambda[S_1,\ldots,S_m;\theta] = \sum_{k=1}^{m} \delta_k[Z(\theta),\ \phi_k\{Z(\theta)\}]$$

with $\phi_k : Z \to \Omega_k$ defined in such a way that $\phi_k(z) = z \leftrightarrow z \varepsilon \Omega_k$. From a computational point of view, it is important if we norm each of the δ_k separately before we add them, or if we norm the sum of the "raw" δ_k (cf. Kruskal & Carroll, 1969).

Our previous T-space methods correspond with the case in which for each k we have $\sigma^k_{ij} = \pm 1$ for at most one pair i,j, and the Z-space methods correspond with the case in which we have only a single matrix S_k. This general case has also been discussed by Roskam (1969b), but in his paper the metrics δ_k, and the transformations ϕ_k are much more specific, and the emphasis is more on the norming of the loss functions.

In this paper we investigate an interesting approach suggested by Yates which combines aspects of T-space and Z-space approaches, and which combines projection and nonprojection aspects.

1: <u>The Balanced Least-Squares Transformation</u>

Because all formulas in this paper are developed for a fixed $\theta \varepsilon \Theta$ and a fixed $S \varepsilon \Delta_A$, we do not use these indices any more. We generalize Yates' definition of the balanced least-squares (BLS) transformation in two respects. In the first place we use a matrix of known weights in the definition of the loss function; in the second place we generalize the analysis to partial orders. We start by decomposing a weighted version of the T-space squared error loss function λ_B into n components

$$\lambda_i = \frac{1}{4} \sum_{j=1}^{n} w_{ij} [|t_{ij}| - t_{ij}]^2$$

with

$$t_{ij} = \sigma_{ij}(z_i - z_j)$$

and $\{w_{ij}\}$ a matrix of positive weights (with some minor modifications all of the results of this paper also hold for <u>nonnegative</u> weights).

Define the function $\lambda_i(y)$ of the single real variable y by

$$\lambda_i(y) = \frac{1}{4}\sum_{j=1}^{n} w_{ij}[|t_{ij}(y)| - t_{ij}(y)]^2$$

with

$$t_{ij}(y) = \sigma_{ij}(y - z_j).$$

Clearly $\lambda_i(z_i) = \lambda_i$ and $\lambda_i(y)$ indicates, in some sense, how large loss function component i would have been if z_i had been equal to y. We now define the BLS-transform $\tilde{z}_i$ of z_i by

$$\lambda_i(\tilde{z}_i) = \inf_y \lambda_i(y)$$

Questions of existence and uniqueness are considered at a later stage. An alternative definition of $\lambda_i(y)$ is also useful for some purposes. If

$$\delta_{ij}(y) = \begin{cases} 1 \text{ if } t_{ij}(y) < 0 \\ 0 \text{ if } t_{ij}(y) \geq 0 \end{cases}$$

then

$$\lambda_i(y) = \sum_{j=1}^{n} w_{ij}\delta_{ij}(y)(y-z_j)^2$$

Observe then this representation remains valid if we define

$$\delta_{ij}(y) = \begin{cases} 1 \text{ if } t_{ij}(y) < 0 \\ 0 \text{ if } t_{ij}(y) > 0 \\ 0 \text{ if } \sigma_{ij} = 0 \text{ and } y \neq z_j \\ \mu \text{ if } \sigma_{ij} \neq 0 \text{ and } y = z_j \end{cases}$$

with μ completely arbitrary.

Some of the definitions are illustrated in the following simple example with σ_{ij} = sign (i-j) and $w_{ij} = 1$ for all i,j. The z_i-values are given in the second column of Table 1. The third column has the monotone regression values $\hat{z}_i$; the fourth column the rank images z_i^*. In Fig. 1:1 we have plotted the functions $t_{3j}(y)$, and in 1:2 we have plotted $t_{4j}(y)$. In Fig. 2:1 the function $\lambda_3(y)$ is given, and 2:2 shows $\lambda_4(y)$,

2: Existence and Uniqueness

It is clear that a third way of writing $\lambda_i(y)$ is

$$\lambda_i(y) = \inf_{\hat{t}_{ij} \geq 0} \frac{1}{4}\sum_{j=1}^{n} w_{ij}[t_{ij}(y)-\hat{t}_{ij}]^2$$

It follows directly from this representation that $\lambda_i(y)$ is a convex and continuously differentiable function of y (cf Rockafellar, 1970, corollary 26-3-2). In fact, $\lambda_i(y)$ is strictly convex outside the interval in which it is zero. If this interval consists of at most a single point, then the minimizing value $\tilde{z}_i$ is unique. A necessary and sufficient condition for $\lambda_i(y) = 0$ is that $\delta_{ij}(y) = 0$ for all i,j. If we define

$$L_i = \{j | \sigma_{ij} = +1\}$$
$$U_i = \{j | \sigma_{ij} = -1\}.$$

Then this condition is equivalent to

$$\max_{j \varepsilon L_i} z_j \leq y \leq \min_{j \varepsilon U_i} z_j$$

Throughout this paper we adopt the convention that the max (or sup) over an empty set is $-\infty$, and the min (or inf) over an empty set is $+\infty$.

3: Iterative Algorithms

The derivative of $\lambda_i(y)$ is simply

$$g_i(y) = 2 \sum_{j=1}^{n} w_{ij} \delta_{ij}(y)(y - z_j)$$

This made Yates suggest the iterative process

$$y^{+} = \frac{\sum_{j=1}^{n} w_{ij}\delta_{ij}(y)z_j}{\sum_{j=1}^{n} w_{ij}\delta_{ij}(y)}$$

if

$$\varepsilon_i(y) = \sum_{j=1}^{n} w_{ij}\delta_{ij}(y) \neq 0$$

and

$$y^{+} = y$$

otherwise. Here y^{+} stands for the successor of y in the iterations.

The algorithm has some peculiarities, which deserve attention. Consider our previous example. The iterations for $i = 3$ are represented graphically in Fig. 3:1, and the iterations for $i = 4$ in Fig. 3:2. It is clear from this figure that if we want to compute $\tilde{z}_4$, then the algorithm converges to the correct value 5.5 iff we start in the open interval (5,6). Otherwise the iterations will cycle, i.e., there is an s such that $y^t = y^{t+2} = 5$ and $y^{t+1} = y^{t+3} = 6$ for all $t > s$. One possible way out is to use the more general definition of $\delta_{ij}(y)$. If $\mu > 0$, then this cycling does not occur, and the algorithm converges in a finite number of steps. The dots in 3:1 and 3:2 indicate where iterations with $\mu = 1$ differ from those with $\mu = 0$.

Observe that a gradient algorithm for this problem has the form

$$y^+ = y - \kappa \sum_{j=1}^{n} w_{ij}\sigma_{ij}(y)(y-z_j)$$

for some $\kappa > 0$. If we denote the Yates iterate by y^* again, this can be rewritten as

$$y^+ = (1-\kappa\varepsilon_i(y))y + \kappa\varepsilon_i(y)y^*$$

Thus the Yates process is a gradient process with stepsize $\kappa = [\varepsilon_i(y)]^{-1}$. This seems to suggest that we take large steps if we are close to the minimum, and the cycling is an example that this may go wrong. Moreover, it is clear that for all $0 \leq \kappa \leq \varepsilon_i(y)$, we have

$$\lambda_i(y^+) \leq \kappa\varepsilon_i(y)\lambda_i(y^*) + (1-\kappa\varepsilon_i(y))\lambda_i(y)$$

Moreover if $g_i(y) \neq 0$, there is a real number $\omega_i(y)$ such that $\lambda_i(y^+) < \lambda_i(y)$ for all $0 < \kappa < \omega_i(y)$. It is consequently not difficult to construct a finite gradient algorithm. We start with stepsize $[\varepsilon_i(y)]^{-1}$. If this does not decrease $\lambda_i(y)$, we half it, and so on.

4: Some Examples

In Table 1 we have collected the $\tilde{z}_i$ values for our previous example in the fifth and sixth columns. Only $\tilde{z}_4$ is unique, all others are only restricted to lie in the interval $I[\tilde{z}_i]$ given in column five. We make the choice of $\tilde{z}_i$ unique by

$$(z_i-\tilde{z}_i)^2 = \inf_{y\in I[\tilde{z}_i]} (z_i-y)^2$$

This makes $\tilde{z}_i$ equal to the endpoint of $I[\tilde{z}_i]$ that is closer to the data point z_i. If $I[\tilde{z}_i]$ only contains one point, this does not lead to inconsistencies. Another consequence is that $\tilde{z}_1$ is equal to the smallest of the z_i and $\tilde{z}_n$ to the largest of the z_i (or: the z_i and the $\tilde{z}_i$ have equal range).

A second example is taken from Roskam (1969a, p. 11). He uses it to illustrate the importance of the different possibilities of handling ties (in fact, his example tends to overemphasize this). Each of the transformational principles has a strong and a weak version. In the example there are three tie-blocks of three different index values. The implied partial order is (1,2,3) < (4,5,6) < (7,8,9). We interpret this to mean that σ_{ij} within tie-blocks is zero. Clearly this already implies that $I[\tilde{z}_i]$ = $I[\tilde{z}_j]$ if i and j are in the same tie-block. We can now define a weak version of BLS by taking $\tilde{z}_i$ from the interval in the usual way, we can also define a strong version by setting all $\tilde{z}_i$ corresponding with tie-block J equal to the constant $\tilde{z}(J)$ define by

$$\sum_{i \varepsilon J} (z_i - \tilde{z}(J))^2 = \inf_{y \varepsilon I[\tilde{z}_i]} \sum_{i \varepsilon J} (z_i - y)^2;$$

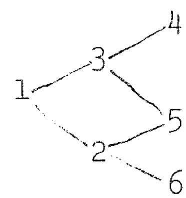
i.e., $\tilde{z}(J)$ equals the endpoint of $I[\tilde{z}_i]$ that is closer to the block average. Observe that λ_A for BLS is larger than λ_A for the other approaches, observe also that the $\tilde{z}_i$ do not add up to the same value as the z_i. The last property is of some importance. The $\hat{z}_i$ are averages of disjoint blocks of z_i values, the z_i^* are permutations of the z_i values, which also can be considered as averages of disjoint blocks of size one. This observation can be used to construct compromise transformations between the two extremes z_i^* and $\hat{z}_i$ (Young, 1973). In the case of linear orders, it can also be used to apply a familiar theorem of Hardy, Littlewood, and Polya (1929).

The most important consequence of this theorem for our purposes is that

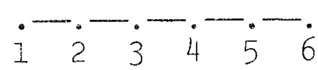
$$z_1^* + z_2^* + \ldots + z_k^* \leq \hat{z}_1 + \hat{z}_2 + \ldots + \hat{z}_k$$

for all $k = 1, \ldots, n$. There is no way to derive relationships like this between $\hat{z}_i$ and $\tilde{z}_i$.

The final example in Table 3a illustrates the BLS transformation for the more complicated partial order.



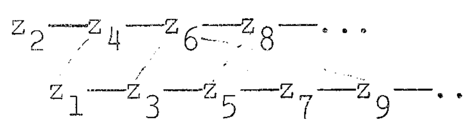
We need the generalized definition of isotone regression, and the generalized definition of rank images (the permutation of the data values that satisfies the inequalities and minimizes λ_A). In 3b the same data values are used for the linear order



Observe that the λ_A value for $\tilde{z}_i$ is lower for the linear order than for the partial order. This indicates that λ_A is not really a good measure to compare the different transformations in Z-space with the BLS transformation. Of course, it is easy to verify that in the BLS case we also have $\lambda_A = 0$ iff z_i satisfies the order restrictions. Observe that the natural loss-function suggested by the construction of the $\tilde{z}_i$ is

$$\lambda(\tilde{z}) = \sum_i \lambda_i(\tilde{z}_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}\delta_{ij}(\tilde{z}_i)(\tilde{z}_i - z_j)^2$$

If the z_i satisfy the order restrictions, then $\lambda(\tilde{z}) = 0$, but if $\lambda(\tilde{z}) = 0$ then it does not necessarily follow that $\{z_i\}$ satisfies the restrictions. In the case of a linear order, for example, with $w_{ij} = 1$ for all i,j, it just follows then the z_i satisfy the partial order



The main reason for this is the curious fact that $\tilde{z}_i$ is independent of the value of z_i, although it depends on the value of the z_j with $j \neq i$ (in fact, it is a piecewise linear function of those values).

5: <u>An Explicit Formula</u>

Define the index sets

$$L_i(y) = \{j \mid y < z_j \wedge \sigma_{ij} = +1\}$$
$$U_i(y) = \{j \mid y > z_j \wedge \sigma_{ij} = -1\}$$

Obviously

$$L_i(x) \cap U_i(y) = \phi$$

for all real x,y. Moreover, we can decompose the unions $L_i(x) \cup U_i(y)$ in a very useful way. For all real x,y we can write

$$L_i(x) \cup U_i(y) = \{[L_i(x) \cup U_i(x)] \cup [U_i(y)-U_i(x)]\}-[U_i(x)-U_i(y)]$$
$$= \{[L_i(y) \cup U_i(y)] \cup [L_i(x)-L_i(y)]\}-[L_i(y)-L_i(x)]$$

We now define for each subset I of N = {1,2,...,n}

$$S_i(I) = \sum\{w_{ij}z_j | j\varepsilon I\}$$
$$M_i(I) = \sum\{w_{ij} | j\varepsilon I\}$$

By the usual conventions $S_i(\phi) = M_i(\phi) = 0$. If $M_i(I) \neq 0$, then we also define

$$A_i(I) = S_i(I)/M_i(I).$$

It is clear from the definitions given so far that

$$j\varepsilon U_i(y) - U_i(x) \text{ iff } \sigma_{ij} = -1 \wedge x \leq z_j < y,$$
$$j\varepsilon U_i(x) - U_i(y) \text{ iff } \sigma_{ij} = -1 \wedge y \leq z_j < x,$$
$$j\varepsilon L_i(x) - L_i(y) \text{ iff } \sigma_{ij} = +1 \wedge x < z_j \leq y,$$
$$j\varepsilon L_i(y) - L_i(x) \text{ iff } \sigma_{ij} = +1 \wedge y < z_j \leq x.$$

By performing weighted summations over the indices in the subset, we conclude

$$M_i[U_i(y)-U_i(x)]x \le S_i[U_i(y)-U_i(x)] \le M_i[U_i(y)-U_i(x)]y,$$
$$M_i[U_i(x)-U_i(y)]y \le S_i[U_i(x)-U_i(y)] \le M_i[U_i(x)-U_i(y)]x,$$
$$M_i[L_i(x)-L_i(y)]x \le S_i[L_i(x)-L_i(y)] \le M_i[L_i(x)-L_i(y)]y,$$
$$M_i[L_i(y)-L_i(x)]y \le S_i[L_i(y)-L_i(x)] \le M_i[L_i(y)-L_i(x)]x.$$

From the set theoretical decomposition of $L_i(x) \cup U_i(y)$ we find

$$S_i[L_i(x) \cup U_i(y)] = S_i[L_i(x) \cup U_i(x)] + S_i[U_i(y)-U_i(x)]$$
$$- S_i[U_i(x)-U_i(y)] = S_i[L_i(y) \cup U_i(y)] + S_i[L_i(x)-L_i(y)] - S_i[L_i(y)-L_i(x)]$$

Together with the inequalities derived above, this gives

$$S_i[L_i(x) \cup U_i(x)] + D_i(x,y)x$$
$$\le S_i[L_i(x) \cup U_i(y)] \le S_i[L_i(x) \cup U_i(x)] + D_i(x,y)y,$$

and

$$S_i[L_i(y) \cup U_i(y)] + E_i(x,y)x$$
$$\le S_i[L_i(x) \cup U_i(y)] \le S_i[L_i(y)+U_i(y)] + E_i(x,y)y,$$

with

$$D_i(x,y) = M_i[U_i(y)-U_i(x)] - M_i[U_i(x)-U_i(y)],$$
$$E_i(x,y) = M_i[L_i(x)-L_i(y)] - M_i[L_i(y)-L_i(x)].$$

For any $\tilde{z}$ which minimizes $\lambda_i(y)$ we have the relationship

$$M_i[L_i(\tilde{z}) \cup U_i(\tilde{z})]\tilde{z} = S_i[L_i(\tilde{z}) \cup U_i(\tilde{z})].$$

If we substitute $x = \tilde{z}$ in the first double inequality, we find

$$M_i[L_i(\tilde{z}) \cup U_i(y)]\tilde{z} \leq S_i[L_i(\tilde{z}) \cup U_i(y)]$$
$$\leq M_i[L_i(\tilde{z}) \cup U_i(\tilde{z})]\tilde{z} + D_i(\tilde{z},y)y,$$

and if we substitute $y = \tilde{z}$ in the second one, we find

$$M_i[L_i(\tilde{z}) \cup U_i(\tilde{z})]\tilde{z} + E_i(x,\tilde{z})x$$
$$\leq S_i[L_i(x) \cup U_i(\tilde{z})] \leq M_i[L_i(x) \cup U_i(\tilde{z})]\tilde{z}$$

The most interesting consequence of these inequalities is that, if we assume that $L_i(\tilde{z})$ and $U_i(\tilde{z})$ are both nonempty,

$$A_i[L_i(x) \cup U_i(\tilde{z})] \leq \tilde{z} \leq A_i[L_i(\tilde{z}) \cup U_i(y)]$$

for all real x and y. This is equivalent with

$$\tilde{z} = \max_x \min_y A_i[L_i(x) \cup U_i(y)] = \min_y \max_x A_i[L_i(x) \cup U_i(y)],$$

where we do not consider pairs x,y such that $L_i(x) \cup U_i(y) = \phi$.

If only one of $L_i(\tilde{z})$ and $U_i(\tilde{z})$ is nonempty, then this result remains valid. If both $L_i(\tilde{z})$ and $U_i(\tilde{z})$ are empty then

$$\min_y \max_x A_i[L_i(x) \cup U_i(y)] \leq \tilde{z} \leq \max_x \min_y A_i[L_i(x) \cup U_i(y)]$$

because these values give exactly the endpoints of the interval we have used in the previous section (again: if there are no lower sets $L_i(x)$ except ϕ then the lower bound is $-\infty$ by our previous conventions, if there are no upper sets $U_i(y)$ except ϕ then the upper bound is $+\infty$). The formulas are illustrated in Table 4a and 4b where we recomputed $\tilde{z}_3$ and $\tilde{z}_4$ from our first example, and in Table 4c where we computed $\tilde{z}_3$ for the partial order. The lower sets correspond with the rows of the tables, the upper sets correspond with columns.

6: An Alternative Representation

In this section we prove an explicit formula which is considerably less economical than the previous one, but somewhat easier to handle for theoretical purposes. Let $\mathcal{L}_i$ be the set of all subsets of L_i and $\mathcal{U}_i$ be the set of all subsets of U_i. Suppose that for some $L \epsilon \mathcal{L}_i$, we find $\underline{U} \epsilon \mathcal{U}_i$ such that

$$A_i[L \cup \underline{U}] = \min_{U \epsilon \mathcal{U}_i} A_i[L \cup U]$$

We prove that $\underline{U}$ is of the form $U_i(y)$ for some real y by distinguishing some special cases. The case $L = \underline{U} = \phi$ is excluded again. Moreover we assume (without loss of generality) that $\underline{U}$ does not contain any indices k such that $z_k = A_i[L \cup \underline{U}]$. This can be done because adding these values to $\underline{U}$ would not change the average anyway.

Case 1: $L = \phi$, $\underline{U} \neq \phi$. Now $A_i[L \cup U] = A_i[U]$, and consequently $\underline{U}$ is the single element set consisting of z_k, the smallest of the z_j for which $\sigma_{ij} = -1$. If z_ℓ is the next smallest of the z_j with $\sigma_{ij} = -1$, then $\underline{U} = U_i(y)$ for all $z_k < y < z_\ell$. Nonuniqueness of z_k and nonexistence of z_ℓ do not cause problems.

Case 2: $L \neq \phi$, $\underline{U} = \phi$. Now we can set $\underline{U} = U_i(y)$ with $y < z_k = \min \{z_j | \sigma_{ij} = -1\}$.

Case 3: $L \neq \phi$; $\underline{U} = \underline{U}_i$. Then $\underline{U} = U_i(y)$ with $y > z_k = \max \{z_j | \sigma_{ij} = -1\}$.

Case 4: $L \neq \phi$; $\underline{U} \neq \phi$; $\underline{U} \neq \underline{U}_i$. In this case we can delete an index from $\underline{U}$. Suppose we delete k. Then by definition

$$A_i[L \cup \underline{U} - \{k\}] \geq A_i[L \cup \underline{U}]$$

or

$$z_k \leq A_i[L \cup \underline{U}].$$

In the same way we can add an index ℓ with $\sigma_{ij} = -1$ to $\underline{U}$ and this gives

$$z_\ell \geq A_i[L \cup \underline{U}]$$

We remember that we excluded all elements with $z_k = A_i[L \cup \underline{U}]$ from $\underline{U}$, and set $\underline{U} = U_i(y)$ with $y = A_i[L \cup \underline{U}]$. This is the last case and in the same way we can prove that if $\underline{L}$ maximizes $A_i[L \cup U]$ over $L \varepsilon \mathcal{L}_i$ for fixed $U \varepsilon \mathcal{U}_i$, then $\underline{L}$ is of the form $L_i(x)$. Consequently,

$$\min_{L \varepsilon \mathcal{L}_i} \max_{U \varepsilon \mathcal{U}_i} A_i[L \cup U] = \max_x \min_y A_i[L_i(x) \cup U_i(y)]$$

$$\min_{U \varepsilon \mathcal{U}_i} \max_{L \varepsilon \mathcal{L}_i} A_i[L \cup U] = \min_y \max_x A_i[L_i(x) \cup U_i(y)]$$

This new representation helps in proving one of the main results of the paper. Suppose $\sigma_{ik} = +1$. Then $j \varepsilon L \varepsilon \mathcal{L}_k$ implies $\sigma_{kj} = +1$ implies $\sigma_{ij} = +1$. Thus $\mathcal{L}_k \subseteq \mathcal{L}_i$. In the same way $\mathcal{U}_i \subseteq \mathcal{U}_k$. Consequently both

$$\max_{L \varepsilon \mathcal{L}_i} \min_{U \varepsilon \mathcal{U}_i} A_i[L \cup U] \geq \max_{L \varepsilon \mathcal{L}_k} \min_{U \varepsilon \mathcal{U}_k} A_k[L \cup U]$$

and

$$\min_{U \varepsilon \mathcal{U}_i} \max_{L \varepsilon \mathcal{L}_i} A_i[L \cup U] \geq \min_{U \varepsilon \mathcal{U}_k} \max_{L \varepsilon \mathcal{L}_k} A_k[L \cup U]$$

If $\tilde{z}_i$ and $\tilde{z}_k$ are unique, this implies directly that

$$\sigma_{ik}(\tilde{z}_i-\tilde{z}_k) \geq 0.$$

If $\tilde{z}_i$ is uniquely determined and $\tilde{z}_k$ is not, then

$$\sigma_{ik}(\tilde{z}_i-y) \geq 0$$

for all $y \varepsilon I(\tilde{z}_k)$. If $\tilde{z}_k$ is uniquely determined and $\tilde{z}_i$ is not, then

$$\sigma_{ik}(y-\tilde{z}_k) \geq 0$$

for all y in the interval $I[\tilde{z}_i]$. If both $\tilde{z}_i$ and $\tilde{z}_k$ are not uniquely determined, then

$$\tilde{z}_i^- \equiv \max_{j\varepsilon L_i} z_j \leq \tilde{z}_i \leq \min_{j\varepsilon U_i} z_j \equiv \tilde{z}_i^+$$

$$\tilde{z}_k^- \equiv \min_{j\varepsilon L_k} z_j \leq \tilde{z}_k \leq \min_{j\varepsilon U_k} z_j \equiv \tilde{z}_k^+$$

If in addition $L_i \cap U_k \neq \phi$, i.e., if there is an ℓ such that $\sigma_{i\ell} = \sigma_{\ell k} = +1$ then even $z_k^+ \leq z_i^-$ and $I[\tilde{z}_i] \cap I[\tilde{z}_k]$ contains at most one point. If $L_i \cap U_k = \phi$ and $z_i \leq z_k$ then again $z_k^+ \leq z_i^-$. Only if $L_i \cap U_k = \phi$ and $z_k < z_i$ the intervals may overlap. By checking the possible cases, it is easy to see that our choice of $\tilde{z}_i$ and $\tilde{z}_k$ from the intervals again makes

$$\sigma_{ik}(\tilde{z}_i-\tilde{z}_k) \geq 0.$$

7: Additional Comments

The most important theoretical contributions of this paper are the explicit min-max formulas. They provide us with alternative algorithms and they make it easy to prove the conjecture of Yates and Young that BLS always gives a monotonic transformation. Another valuable contribution is the proof that Yates' algorithm does not necessarily converge, and the modification which makes it convergent. These contributions, however, do not imply in any sense that the BLS transformation is either better or worse than the existing ones. The only obvious advantage of BLS is that it is relatively easy to compute, even for complicated partial orders. A disadvantage is the fact that $\lambda(\tilde{z}) = 0$ does not necessarily imply $\sigma_{ij}(z_i-z_j) \geq 0$ for all i,j. Going back to the original definition, we find that $\lambda(\tilde{z}) = 0$ iff $t_{ij}(\tilde{z}_i) \geq 0$ for all i,j iff $\sigma_{ij}(\tilde{z}_i-z_j) \geq 0$, for all i,j iff the intervals $I[\tilde{z}_i]$ are nonempty for all i, iff $\max_{j\epsilon L_i} z_j \leq \min_{j\epsilon U_i} z_j$. An obvious counter example to the idea that $\lambda(\tilde{z})$ can be made equal to zero iff z is monotonic is given in Table 5a. Several proposals are possible which redefine BLS in such a way that this disadvantage disappears. They shold be based, in a natural way, on the fact that $\tilde{z}_i$ is not influenced by the value of z_i or, to put it differently, that we do not need z_i to compute $\tilde{z}_i$. The obvious proposal is to compute the BLS for all midpoints M(i,j). In the example of Table 5a, this gives the results in Table 5b. We compute the fit over midpoints. This gives $\lambda(\tilde{z}^M) = 10.5$ (for the six values which are shown in the table). Clearly we now have $\lambda(\tilde{z}^M) = 0$ iff z_i is monotonic. After our computations, we construct the $\tilde{z}_i$ as the set of numbers which lie in the interval defined by the midpoints $\tilde{z}_i^M$ and which minimize λ_A under these conditions. For the partial order on page 15, this last step is a bit more complicated. In Table 6a we investigated the midpoints, in Table 6b we constructed the intervals and computed the $\tilde{z}_i$. The nice thing about this modification is that our results on monotonicity and our explicit formulas still apply.

8: A Scaled Loss Function

Define

$$\lambda = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}[|t_{ij}(\tilde{z}_i,\theta)|-t_{ij}(\tilde{z}_i,\theta)]^2}{4\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}t_{ij}^2(\tilde{z}_i,\theta)}$$

with

$$t_{ij}(\tilde{z}_i,\theta) = \sigma_{ij}(\tilde{z}_i - z_j(\theta))$$

and $\{\tilde{z}_i\}$ the BLS-transformation. A new scaling technique, which can be used for any measurement model, is to minimize this function over $\theta \varepsilon \Theta$.

ACKNOWLEDGMENTS

I had some useful discussions with Forrest Young on the subject of transformations in general. As a matter of fact, he introduced me to BLS and the (unpublished) work of Alan Yates. The fact that they both did not have a proof that the transformation is actually monotone inspired me to do this work.

REFERENCES

De Leeuw, J. Smoothness properties of nonmetric loss functions. Unpublished paper, Bell Laboratories, 1973.

De Leeuw, J. Nonmetric multidimensional scaling. Department of Data Theory, University of Leiden, RN010-68.

De Leeuw, J. The positive orthant method for nonmetric multidimensional scaling. Department of Data Theory, University of Leiden, RN001-70.

Guttman, L. A. Smallest space analysis based on the absolute value principle. Paper presented at the XIX International Congress of Psychology, London, 1968. (a)

Guttman, L. A. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 1968, 33, 469-506. (b)

Hardy, G. H., Littlewood, J. E., & Poya, G. Some simple inequalities satisfied by convex functions. Messenger of Mathematics, 1929, 58, 145-152.

Johnson, R. A. Pairwise nonmetric multidimensional scaling. Psychometrika, 1973, 38, 11-18.

Kruskal, J. B., & Carroll, J. D. Geometrical models and badness-of-fit functions. In: P. R. Krishnaiah (Ed.), Multivariate analysis II. New York: Academic Press, 1969.

Kruskal, J. B. Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-28. (a)

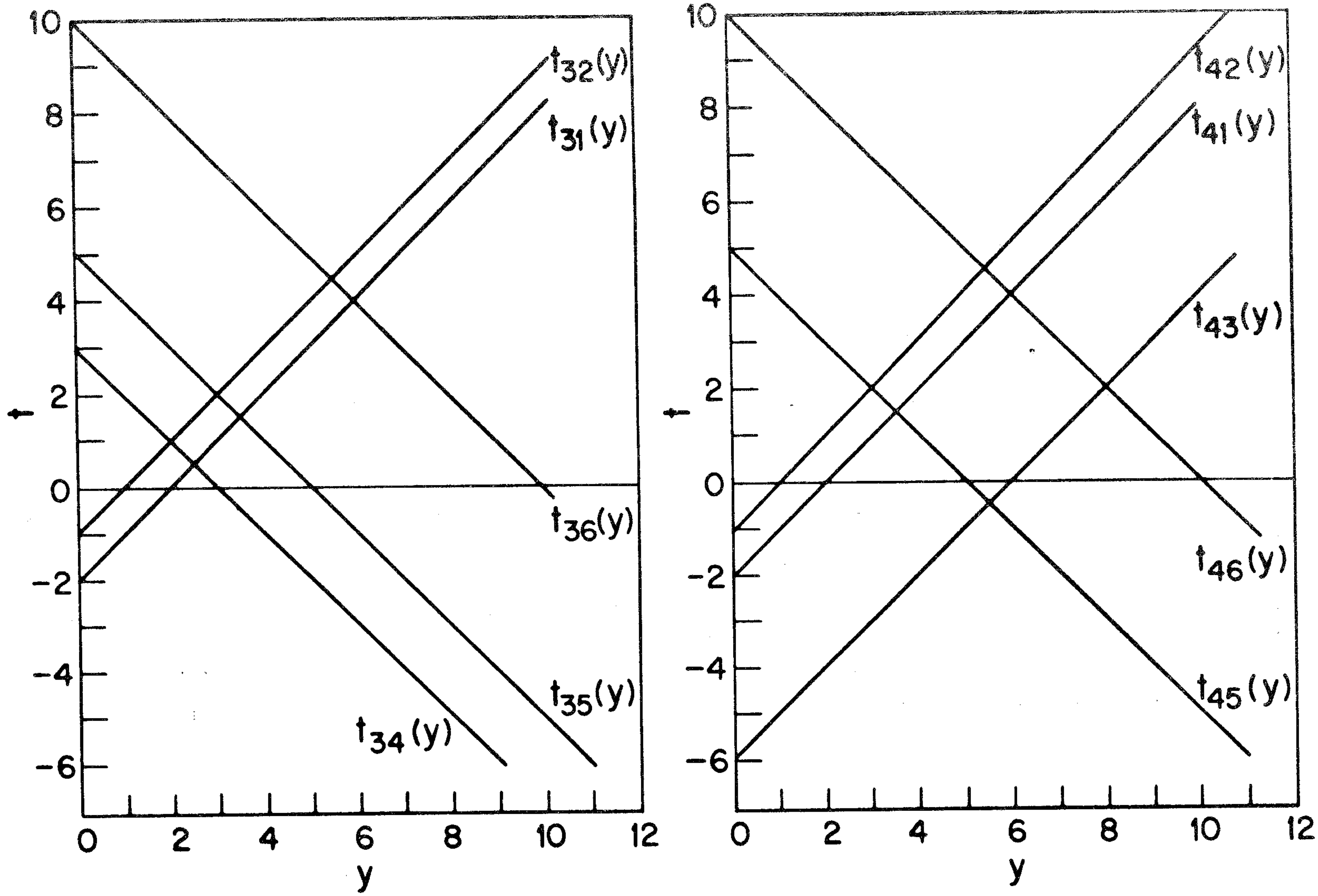
Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964, 29, 115-129. (b)

Rockafellar, R. T. Convex analysis. Princeton, N. J.: Princeton University Press, 1970.

Roskam, E. E. A comparison of principles for algorithm construction in nonmetric scaling. Department of Psychology, University of Michigan, MMPP69-2, 1969. (a)

Roskam, E. E. Data theory and algorithms for nonmetric scaling. Department of Psychology, University of Nijmegen, Mimeo, 1969. (b)

Young, F. W. Conjoint scaling. LL. Thurstone Laboratory, University of North Carolina, No. 118, 1973.



10
8
6
4
2
0
-2
-4
-6
t
0
2
4
6
8
10
12
y
$t_{32}(y)$
$t_{31}(y)$
$t_{36}(y)$
$t_{35}(y)$
$t_{34}(y)$
$t_{42}(y)$
$t_{41}(y)$
$t_{43}(y)$
$t_{46}(y)$
$t_{45}(y)$

Figure 1

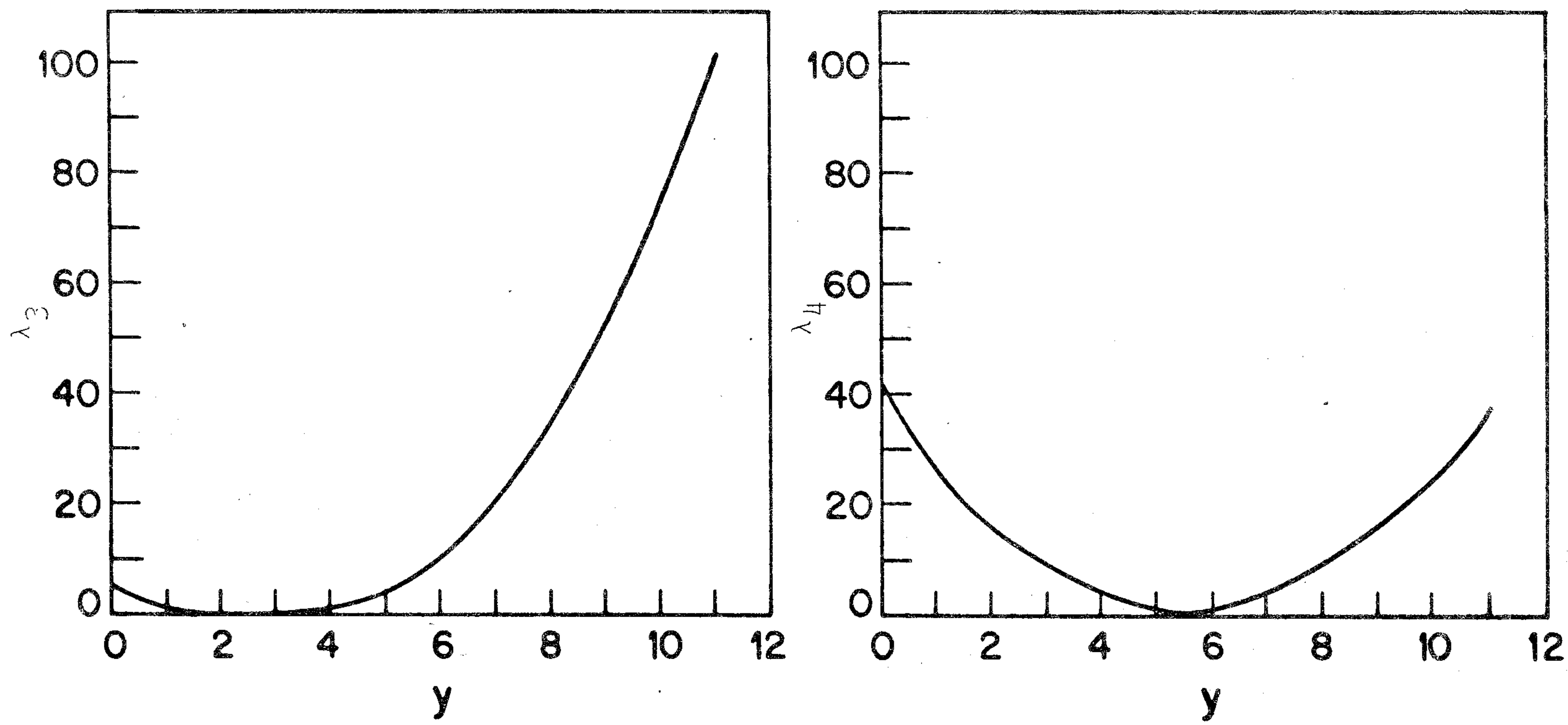


Figure 2

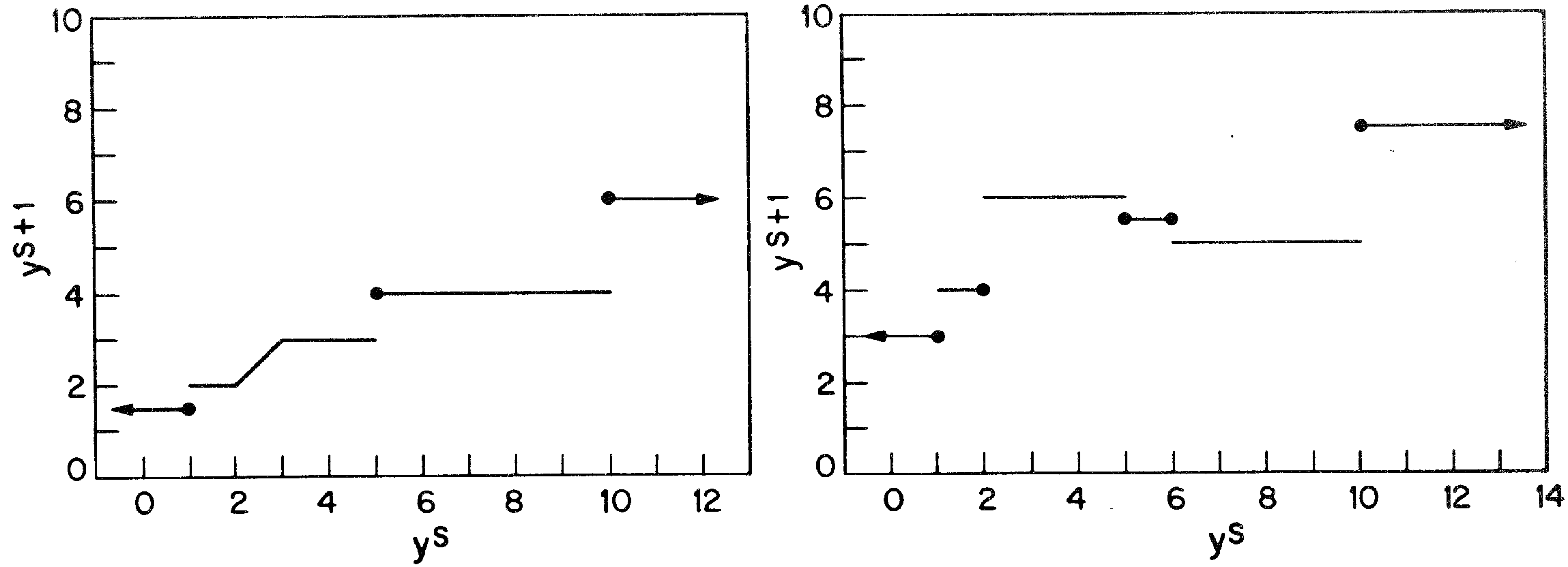


Figure 3

Table 1

i	z_i	$\hat{z}_i$	z_i^*	$I[\tilde{z}_i]$	$\tilde{z}_i$
1	2	1.5	1	$[-\infty,1]$	1.00
2	1	1.5	2	[2,3]	2.00
3	6	4.5	3	[2,3]	3.00
4	3	4.5	5	5.5	5.50
5	5	5.0	6	[6,10]	6.00
6	10	10.0	10	$[10,\infty]$	10.00
λ_A		5.0	16		18.25

Table 2

i	z_i	${}_w\hat{z}_i$	${}_s\hat{z}_i$	${}_wz^*_i$	${}_sz^*_i$	$I[\tilde{z}_i]$	${}_w\tilde{z}_i$	${}_s\tilde{z}_i$
1	3	3.00	2.000	2	1.66..	$(-\infty,2]$	2.00	2.00
2	1	1.00	2.000	1	1.66..	$(-\infty,2]$	1.00	2.00
3	2	2.00	2.000	2	1.66..	$(-\infty,2]$	2.00	2.00
4	7	4.25	4.833..	4	3.33..	2.5	2.50	2.50
5	5	4.25	4.833..	3	3.33..	2.5	2.50	2.50
6	4	4	4.833..	3	3.33..	2.5	2.50	2.50
7	3	4.25	4.833..	7	6.66..	$[7,+\infty)$	7.00	7.00
8	8	8.00	4.833..	8	6.66..	$[7,+\infty)$	8.00	7.00
9	2	4.25	4.833..	5	6.66..	$[7,+\infty)$	7.00	7.00
λ_A		14.75	28.833..	40	56.00		70.75	72.75

Table 3

i	z_i	$\hat{z}_i$	z_i^*	$I[\tilde{z}_i]$	$\tilde{z}_i$
1	2	1.5	1	$(-\infty,1]$	1
2	1	1.5	2	2	2
3	5	3.5	4	2	2
4	4	4.0	5	$[5,\infty)$	5
5	2	3.5	4	$[5,\infty)$	5
6	4	4.0	2	$[2,\infty)$	4
λ_A		5.0	12		21

(a)

i	z_i	$\hat{z}_i$	z_i^*	$I(\tilde{z}_i)$	$\tilde{z}_i$
1	2	3/2	1	$(-\infty,1]$	1.0
2	1	3/2	2	2.0	2.0
3	5	11/3	2	2.0	2.0
4	4	11/3	4	3.5	3.5
5	2	11/3	4	4.5	4.5
6	4	4	5	$[5,\infty)$	5.0
λ_A		31/6	16		18.5

(b)

<u>Table 4</u>

	φ	{4}	{4,5}	{4,5,6}	min
φ	x	3	4	6	3
{1}	2	5/2	10/3	5	2
{1,2}	3/2	2	11/3	21/5	2
max	2	3	4	6	[2,3]

(a)

	φ	{5}	{5,6}	min
φ	x	5	15/2	5
{3}	6	11/2	7	11/2
{1,3}	4	13/3	23/4	4
{1,2,3}	3	14/4	24/5	3
max	6	11/2	15/2	11/2

(b)

	φ	{5}	{4,5}	min
φ	x	2	3	2
{1}	2	2	8/3	2
max	2	2	3	2

(c)

Table 5

i	z_i	$\hat{z}_i$	z_i^*	$I[z_i]$	$\tilde{z}_i$
1	2	1.5	1	$(-\infty,1]$	1
2	1	1.5	2	$[2,3]$	2
3	4	8.5	3	$[2,3]$	3
4	3	3.5	4	$[4,5]$	4
5	6	5.5	5	$[4,5]$	5
6	5	5.5	6	$[6,7]$	6
⋮	⋮	⋮	⋮	⋮	⋮

(a)

i	z_i	$\tilde{z}_M$	$I[\tilde{z}_i]$	$\tilde{z}_i$
1	2	?	$[-\infty,1.5]$	1
M(1,2)	?	1.5	?	?
2	1	?	$[1.5,2.5]$	2
M(2,3)	?	2.5	?	?
3	4	?	$[2.5,3.5]$	3
M(3,4)	?	3.5	?	?
4	3	?	$[3.5,4.5]$	4
M(4,5)	?	4.5	?	?
5	6	?	$[4.5,5.5]$	5
M(5,6)	?	5.5	?	?
6	5	?	$[5.5,6.5]$	6
⋮	⋮	⋮	⋮	⋮

(b)

Table 6

	$\tilde{z}$
M(1,3)	2
M(1,2)	1.5
M(3,4)	4.5
M(3,5)	2
M(2,5)	2
M(2,6)	[2,4]

(a)

i	$\tilde{z}_i$	$I[\tilde{z}_i]$	$\tilde{z}_i$
1	2	$(-\infty,1.5]$	1.5
2	1	[1.5,2]	1.5
3	5	2	2
4	4	$[4.5,+\infty)$	4.5
5	2	$[2,+\infty)$	2
6	4	$[2,+\infty)$	4

(b)