OPTIMAL SCALING BY ALTERNATING LENGTH-CONSTRAINED NONNEGATIVE LEAST SQUARES: AN APPLICATION TO DISTANCE-BASED PRINCIPAL COMPONENTS ANALYSIS

Patrick J.F. Groenen
Bart-Jan van Os
Jacqueline J. Meulman

Data Theory Group
Leiden University

RR-98-01

Optimal Scaling by Alternating Length-Constrained Nonnegative Least Squares: An Application to Distance-Based Principal Components Analysis

Patrick J.F. Groenen*, Bart-Jan van Os, and Jacqueline J. Meulman

February 6, 1998

*Supported by The Netherlands Organization for Scientific Research (NWO) by grant nr. 575-67-053 for the 'PIONEER' project 'Subject Oriented Multivariate Analysis' to the third author. Requests for reprints should be sent to Patrick Groenen, Department of Education, Data Theory Group, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands (e-mail: groenen@rulfsw.fsw.leidenuniv.nl).

1 Introduction

The problem of nonnegative least squares (NNLS) arises in several applications. It is defined as the minimization of a quadratic function over the space of nonnegative elements. In this paper, we study the *length-constrained* NNLS problem (LC-NNLS), which amounts to the maximization of a linear function over the space of nonnegative elements under a length constraint. Such problems with explicit length constraint come up in the area of optimal scaling in multidimensional scaling (see, e.g., De Leeuw & Heiser, 1977), smoothed monotone regression (Heiser, 1985), multivariate analysis with optimal scaling (see, e.g., Young, De Leeuw, & Takane, 1976; Breiman & Friedman, 1985; Gifi, 1990), and distance-based multivariate analysis with optimal scaling (Meulman, 1986, 1992). In these examples, the length-constrained NNLS problem appears as a subproblem in a larger iterative process that solves the overall minimization problem. For now, we will concentrate on distance-based principal components analysis (DB-PCA) as our overall minimization problem.

In this paper, we discuss two strategies to LC-NNLS. The first strategy consists of computing a solution to the NNLS problem without length constraint by the method of Lawson and Hanson (1974, p. 161) followed by proper normalization to impose the length constraint. The second strategy uses a new iterative method for LC-NNLS that never increases loss and never violates the constraints. There are several reasons for introducing an iterative method. First, if good starting values are available, then a few iterations of the new strategy might result in the optimal solution with little effort. Fortunately, such good start solutions typically are available if the NNLS problem appears as an inner optimization problem in a larger iterative scheme (as we will show in a moment). Then the new strategy may be initialized by the solution obtained from the previous iteration. Second, in such a context, the quality of the solution to exterior problem may be improved and obtained faster when the solution of the NNLS problem is approximated and not determined exactly.

The purpose of this paper is to investigate whether obtaining an approximate solution of the LC-NNLS problem during the overall minimization process results in a better quality for the final DB-PCA solution. It is our expectation that the approximate solution of the LC-NNLS problems in the initial stages of the exterior algorithm for DB-PCA directs the exterior algorithm to better quality solutions. At convergence of the exterior algorithm, the LC-NNLS problem will be solved as well. However, we also expect that using the approximate update will cause the exterior algorithm to converge more slowly.

expressed as the linear sum $\mathbf{q}_k = \mathbf{S}_k\mathbf{b}_k$, where $\mathbf{b}_k$ is a vector of nonnegative weights to be estimated and the $n \times p_k$ matrix $\mathbf{S}_k$ is an integrated spline basis depending only on the original variable $\mathbf{z}_k$, the degree of the spline, and a given knot sequence that defines the interval of the pieces. For more details on monotone splines we refer to De Boor (1978) and Ramsay (1988). Note that in principal components analysis Ramsay (1988) uses a range restriction instead of a length-constraint on the $\mathbf{q}_k$'s, thereby losing the property of equal sum of squares of the $\mathbf{q}_k$'s as in classical principal component analysis.

We prefer monotone spline transformations instead of the usual least squares monotonic transformations as emphasized in Gifi (1990), because such monotonic transformations seem less suitable for variables measured on a continuous scale, and often yields a nonsmooth step function, whereas monotonic splines result in a smooth transformation and estimate only a few parameters. Moreover, monotonic splines are easily embedded in a least-squares framework, and the least-squares monotonic transformation can be considered a special case of a monotone spline transformation.

Meulman (1992) proposed an convergent algorithm in which $\mathbf{X}$ and $\mathbf{Q}$ are updated alternatingly. For ease of notation, we drop in the sequel the subscript k when referring to a variable $\mathbf{q}$. For each variable k, an unconstrained update $\overline{\mathbf{q}}$ is computed, and next the nonnegative least-squares problem with respect to $\mathbf{b}$ has to be solved, i.e.,

$$\begin{aligned} f(\mathbf{b}) &= ||\overline{\mathbf{q}} - \mathbf{S}\mathbf{b}||^2 \\ &= \overline{\mathbf{q}}'\overline{\mathbf{q}} + \mathbf{b}'\mathbf{G}\mathbf{b} - 2\mathbf{b}'\mathbf{h} \text{ subject to } \mathbf{b} \geq \mathbf{0} \text{ and } \mathbf{b}'\mathbf{G}\mathbf{b} = n, \end{aligned} \quad (3)$$

where $\mathbf{G} = \mathbf{S}'\mathbf{S}$ and $\mathbf{h} = \mathbf{S}'\overline{\mathbf{q}}$. We assume that $\mathbf{S}$ is of full rank, which is generally true for I-spline bases. Thus, in each iteration, m LC-NNLS problems have to be solved. To retain convergence, it is not necessary to solve (3) completely, but to find an update of $\mathbf{b}$ for which (3) is not larger than the previous iteration.

3 Length-Constrained NNLS (LC-NNLS)

LC-NNLS fixes the length of $\mathbf{b}$ to $\mathbf{b}'\mathbf{G}\mathbf{b} = n$, so that (3) becomes

$$f(\mathbf{b}) = c + n - 2\mathbf{b}'\mathbf{h} \qquad \text{subject to } \mathbf{b} \geq \mathbf{0} \text{ and } \mathbf{b}'\mathbf{G}\mathbf{b} = n, \quad (4)$$

where $c = ||\overline{\mathbf{q}}||^2$. Below we analyze (4) using convex analysis, and show that two cases of LC-NNLS can be distinguished; we also show under which conditions (4) has a unique solution.

(b) there exists at least one positive h_i.

For case (a) a direct solution exists, but for case (b) a special algorithm to minimize (4) is needed. We now show that the function

$$f_2(\mathbf{b}) \quad = \quad c + n - 2\mathbf{b}'\mathbf{h} \qquad \text{with } \mathbf{b} \in Y \cap P \tag{9}$$

reaches it minimum at one of the vertices of the polyhedron defined by $Y \cap P$, that is, for $b_i = (n/g_{ii})^{1/2}$ and all other $b_{j \neq i} = 0$. The vertex that should be nonzero, is the one for which $f_2(\mathbf{b})$ is minimal, that is,

$$\text{argmin}_i - h_i(n/g_{ii})^{1/2}. \tag{10}$$

Since $Y \cap P \supset Z$, this solution also minimizes $f(\mathbf{b})$ in (4). Note that $Y \cap P$ defines a polyhedral set. Rockafellar (1970, Corollary 32.3.4, p. 345) states that if $f(\mathbf{b})$ is bounded below by a polyhedral set, then the minimum is obtained at an extremal point. For negative $\mathbf{h}$, boundedness from below is implied, because $-\mathbf{h}'\mathbf{b}$ is positive for all feasible $\mathbf{b}$. Checking the minimum values of $f(\mathbf{b})$ at the extremal points as in (10) gives the minimum.

4 Algorithms for LC-NNLS

In this section we discuss two computational methods that provide a solution for case (b) of LC-NNLS. The first method is to compute an unconstrained minimum (i.e., without length constraint) and then apply the normalization constraint such that $\mathbf{b}'\mathbf{G}\mathbf{b} = n$ (Kruskal & Carroll, 1969; De Leeuw, 1977; Gifi, 1990). For NNLS this unconstrained minimum can be found by the method of Lawson and Hanson (1974), which gives an exact solution in a finite number of steps. We refer to this approach as LH-NNLS. The second method is an improvement of the approach used in Meulman (1992). The latter approach consists of cyclically updating one $b_i \geq 0$ at a time disregarding the length constraint, evaluating (3) or (4) after $\mathbf{b}$ is updated for all variables, and continuing with a subsequent round of updates until the value of (3) or (4) does not increase. At that point the length constraint can be applied. The present method solves for the nonnegativity constraints and the length constraint simultaneously. We call this algorithm Alternating Length-Constrained Non-Negative Least Squares (ALC-NNLS) and it is described below.

4.1 ALC-NNLS

In this section we first propose the ALC-NNLS algorithm. To show why it works, we reformulate the LC-NNLS problem, and then prove that $f(\mathbf{b})$ in provided $\mathbf{b}'\mathbf{h} \geq 0$. Under this condition, LC-NNLS is equivalent to the maximization of (11).

Suppose that we can find an update $\mathbf{b}^{(\ell)}$ at iteration ℓ for which $-g(\mathbf{b}^{(\ell)}) \leq -g(\mathbf{b}^{(\ell-1)})$. Then, we must have

$$\begin{aligned} -\frac{\mathbf{h}'\mathbf{b}^{(\ell)}}{\mathbf{b}^{(\ell)\prime}\mathbf{G}\mathbf{b}^{(\ell)}} &\leq -\frac{\mathbf{h}'\mathbf{b}^{(\ell-1)}}{\mathbf{b}^{(\ell-1)\prime}\mathbf{G}\mathbf{b}^{(\ell-1)}}, \text{ so that} \\ \frac{\mathbf{h}'\mathbf{b}^{(\ell-1)}}{\mathbf{b}^{(\ell-1)\prime}\mathbf{G}\mathbf{b}^{(\ell-1)}} - \frac{\mathbf{h}'\mathbf{b}^{(\ell)}}{\mathbf{b}^{(\ell)\prime}\mathbf{G}\mathbf{b}^{(\ell)}} &\leq 0. \end{aligned} \tag{13}$$

Multiplying both sides by the positive value $\mathbf{b}^{(\ell)\prime}\mathbf{G}\mathbf{b}^{(\ell)}$ gives

$$\begin{aligned} \frac{\mathbf{h}'\mathbf{b}^{(\ell-1)}}{\mathbf{b}^{(\ell-1)\prime}\mathbf{G}\mathbf{b}^{(\ell-1)}}\mathbf{b}^{(\ell)\prime}\mathbf{G}\mathbf{b}^{(\ell)} - \mathbf{h}'\mathbf{b}^{(\ell)} &\leq 0, \\ a\mathbf{b}^{(\ell)\prime}\mathbf{G}\mathbf{b}^{(\ell)} - \mathbf{h}'\mathbf{b}^{(\ell)} &\leq 0, \end{aligned} \tag{14}$$

where $a = g(\mathbf{b}^{(\ell-1)})$. Note that if $\mathbf{b}^{(\ell)} = \mathbf{b}^{(\ell-1)}$ then (14) becomes a strict equality. Thus, a monotone nonincreasing sequence of $-g(\mathbf{b}^{(\ell)})$ can be obtained by minimizing (14).

It remains to be proven that $\mathbf{h}'\mathbf{b}^{(\ell)} > 0$ in our algorithm so that maximizing $g(\mathbf{b})$ and $g^{1/2}(\mathbf{b})$ are equivalent. Suppose that $\mathbf{b}^{(0)}$ is such that $\mathbf{h}'\mathbf{b}^{(0)} > 0$. (Our algorithm satisfies this condition in Step 1.) Also suppose that (14) is satisfied in every iteration, then the sequence

$$-\frac{\mathbf{h}'\mathbf{b}^{(\ell)}}{\mathbf{b}^{(\ell)\prime}\mathbf{G}\mathbf{b}^{(\ell)}} \leq -\frac{\mathbf{h}'\mathbf{b}^{(\ell-1)}}{\mathbf{b}^{(\ell-1)\prime}\mathbf{G}\mathbf{b}^{(\ell-1)}} < 0,$$

implies that $\mathbf{h}'\mathbf{b}^{(\ell)} > 0$, since $\mathbf{b}^{(\ell)\prime}\mathbf{G}\mathbf{b}^{(\ell)} > 0$ by assumption. Therefore, the condition $\mathbf{h}'\mathbf{b}^{(\ell)} > 0$ is satisfied in our algorithm in every iteration ℓ.

The core of the ALC-NNLS algorithm is to satisfy (14) by minimizing

$$f_2(\mathbf{b}) = a\mathbf{b}'\mathbf{G}\mathbf{b} - 2\mathbf{b}'\mathbf{h} \qquad \text{subject to } \mathbf{b} \geq \mathbf{0} \tag{15}$$

over one b_i at a time subject to $b_i > 0$. This leads to Step 6 of the algorithm. Of course, one could also repeat Step 6 several times, so that a better value of $f_2(\mathbf{b})$ in (15) is obtained. In fact, some experimentation of ALC-NNLS in DB-PCA showed that setting $\ell_{\max} = 1$ and repeating Step 6 one or more times yields better quality solutions than doing Step 6 only once with $\ell_{\max} \geq 1$.

The derivation of this algorithm uses elements of Dinkelbach's (1967) approach to minimize a ratio of functions. The iterative majorization approach of Kiers (1995) for maximizing a ratio of quadratic functions can be shown to lead to the same algorithm, although by a different argument.

strategies (LH-NNLS, ALC-NNLS-1, -2, -3, -5, -10, -50) constituted the first within-subject factor.

We varied several design factors that might influence the importance of the LC-NNLS problem in minimizing Stress. These factors are:

- number of objects ($n = 50, 100$),
- number of variables ($m = 5, 10, 20$),
- number of interior knots in the spline transformations ($k = 3, 9$), and
- error added to the data ($\varepsilon = 10\%, 25\%, 40\%$).

The number of objects controls for the size of the problem of estimating $\mathbf{X}$ and $\mathbf{Q}$, which is of different order of magnitude compared to the LC-NNLS problems. The number of variables determines the number of LC-NNLS problems to be solved at each major iteration. The number of knots used in the spline transformations influences the size of the LC-NNLS problem to be solved. To ensure that transformations are required, we added random noise to the variables in each dataset. Below we describe how the error perturbed datasets were constructed.

The four non-strategy factors lead to an overall design, where seven strategies are tested in 36 cells. Because the variability in Stress values due to the design factors is not known a priori, a statistical decision on the number of replications needed in each cell can not easily be made in advance. In a trial set of 50 replications, the main-effects of the within-subjects factors proved to be relatively small. We decided to have 150 replications in each cell to be on the safe side. In total, 2700 datasets ($2 \times 3 \times 3 \times 150$) were analyzed using seven different strategies under the two transformation conditions (number of interior knots).

Error Perturbed Data

The error perturbed datasets $\mathbf{Z}$ were constructed as follows. We first compute a 100×2 matrix $\mathbf{H}$ with elements $h_{ij} = N(0,1)$, where $N(0,1)$ denotes the normal distribution with mean zero and variance one. Then, $\mathbf{H}$ was rotated and expanded to m dimensions using a random rotation-expansion matrix $\mathbf{A}$ with rows on the unit circle, i.e, $\mathbf{A}'\mathbf{A} = c\mathbf{I}, \text{Diag}(\mathbf{A}\mathbf{A}') = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. These restrictions ensure that the distances between the rows of $\mathbf{H}$ are the same as those of $\mathbf{H}\mathbf{A}'$ up to a constant factor, and the unit circle constraint additionally ensures that the columns of $\mathbf{H}\mathbf{A}'$ have approximately the same sum of squares. For the design factor $n = 50$, only the first 50 rows of $\mathbf{H}$ were used, and for $n = 100$ all rows were used. To

Table 1: ANOVA decomposition of Stress values: average univariate within-subjects effects. The factor error level is represented by ε, number of objects by n, number of variables by m, and number of knots by k.

	Sum of Squares	df	Mean Square	F	Sig.
Strategy	.005978	6	.0009963	627.59	.000
Strategy $\times$ ε	.001032	12	.0000860	54.17	.000
Strategy $\times$ n	.000031	6	.0000051	3.21	.004
Strategy $\times$ m	.000712	12	.0000593	37.36	.000
Error(Strategy)	.025659	16164	.0000016		
Strategy $\times$ k	.000079	6	.0000132	9.01	.000
Error(Strategy $\times$ k)	.023743	16164	.0000015		
Total sum of squares	30.737754	37799	.0008132		

the spline-transformation constitutes a larger part of the overall decrease in Stress.

In these analyses, the absolute difference in Stress between strategies for the same dataset are still very small in a large number of cases. Therefore, we report in Table 3 the proportion of runs of a strategy with difference in Stress with the lowest Stress smaller than 10^{-4} (10 times the convergence criterion). On average, in 75% of the runs, the ALC-NNLS-2 and 3 strategies end up in or close to the lowest minimum, whereas LH-NNLS does so in 68%. Note that ALC-NNLS-50 behaves almost identically to LH-NNLS, and so does ALC-NNLS-10 to a smaller extent. In the 40% error condition, LH-NNLS yields more often the lowest minimum compared to ALC-NNLS-2 and 3. It is evident that ALC-NNLS-1 performs the worst of all strategies. For the 9-knots and 20 variables conditions, the ALC-NNLS-2 strategy was best, ending in 74% and 89% of the runs close to the lowest minimum.

To investigate the convergence efficiency conditional on the quality of the solution, the same ANOVA model as above was used with the number of iterations as dependent variable correcting for the Stress value. This model did not find a significant main effect of the strategies, indicating that the average number of iterations needed by the strategies did not differ when corrected for the quality of the solution.

To exemplify the effect of the strategy-quality interaction on the average number of DB-PCA iterations, we compare in Table 4 the average number of iterations of the LH-NNLS and ALC-NNLS-2 strategies, the latter being the best of the ALC-NNLS strategies. We selected three groups of runs based on the difference between the two strategies in terms of Stress: (a) a group with differences smaller than $2 \cdot 10^{-5}$, (b) a group with differences larger than 10^{-4} in favor of LH-NNLS, and (c) a group with differences larger than 10^{-4}

Table 4: Average number of DB-PCA iterations using the LH-NNLS or the LC-ANNLS-2 strategy.

		Average # iterations		
		LH-NNLS	ALC-NNLS-2	# Repl
(a)	No difference	129	130	849
(b)	LH-NNLS better	171	153	853
(c)	ALC-NNLS-2 better	133	148	1303

in favor of ALC-NNLS-2. For equally good solutions, the average number of iterations did not differ substantially between LH-NNLS and ALC-NNLS-2. For group (b), where NNLS-LH performed better, its average number of iterations is larger, and the reverse situation occurs in group (c), where ALC-NNLS-2 performed better. The conclusion is that the number of iterations is determined by the quality of the solution and the design factors, but not by the the strategies.

6 Discussion and Conclusions

We have studied the length-constrained nonnegative least squares (LC-NNLS) problem in some detail. We identified the cases in which an explicit solution is available. The other cases can be solved by the method of Lawson and Hanson (1974) followed by proper normalization, or by a new algorithm, called alternating length-constrained nonnegative least squares (ALC-NNLS), for solving the length-constrained problem in an iterative manner. It was hypothesized that the quality of the minimum would improve by not completely solving the LC-NNLS problem, but only taking one or more steps in the right direction. However, if the LC-NNLS problem is not embedded in a larger iterative scheme but appears on its own, we recommend the strategy of Lawson and Hanson (1974) followed by proper normalization, since it yields an exact solution.

To test the above mentioned hypothesis, we investigated the LC-NNLS problem occurring in distance-based principal components (DB-PCA) analysis using I-spline transformations. A Monte Carlo study compared ALC-NNLS strategies to the Lawson and Hanson (1974) approach. With respect to the quality of the DB-PCA solutions, ALC-NNLS with two or three inner iterations is the most preferable alternative, doing only one inner iteration is least preferred. This result might warrant revision of the ALS-approach advocated in Gifi (1990), which is using only one inner iteration. In the majority of the cases, ALC-NNLS with two or three inner iterations performs better than the strategy of Lawson and Hanson, except in high error

Gower, J. C., & Hand, D. J. (1996). *Biplots.* London: Chapman and Hall.

Groenen, P. J. F., Commandeur, J. J. F., & Meulman, J. J. (1997). PIONEER: A program for distance-based multivariate analysis. In W. Bandilla & F. Faulbaum (Eds.), *Softstat'97: Advances in statistical software* (pp. 83–90). Stuttgart: Lucius & Lucius.

Heiser, W. J. (1985). *Multidimensional scaling by optimizing goodness-of-fit to a smooth hypothesis* (Tech. Rep. No. RR-85-07). Leiden, The Netherlands: Department of Data Theory, Leiden University.

Jones, L. K., & Seaton, L. (Eds.). (1994). *Math77: Reference manual.* Sterling, VA: Language Systems Coorporation.

Kiers, H. A. L. (1995). Maximization of sums of quotients of quadratic forms and some generalizations. *Psychometrika, 60*, 221–245.

Kruskal, J. B., & Carroll, J. D. (1969). Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. 2, pp. 639–671). New York: Academic Press.

Lawson, C. L., & Hanson, R. J. (1974). *Solving least squares problems.* Englewood Cliffs, NJ: Prentice Hall.

Meulman, J. J. (1986). *A distance approach to nonlinear multivariate analysis.* Leiden, The Netherlands: DSWO Press.

Meulman, J. J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations. *Psychometrika, 57*, 539–565.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science, 3*(4), 425–461.

Rockafellar, R. T. (1970). *Convex analysis.* Princeton, NJ: Princeton University Press.

Verboon, P. (1994). *A robust approach to nonlinear multivariate analysis.* Leiden, The Netherlands: DSWO Press.

Young, F. W., De Leeuw, J., & Takane, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares approach. *Psychometrika, 41*, 505–529.