ALGORITHM CONSTRUCTION BY DECOMPOSITION

JAN DE LEEUW

ABSTRACT. We discuss and illustrate a general technique, related to augmentation, in which a complicated optimization problem is replaced by a simpler problem with more variables but with the same value and the same solution. Our main example comes from metric unfolding.

1. INTRODUCTION

The following theorem is so simple it's almost embarassing. Nevertheless it seems to have some important applications to algorithm construction.

Theorem 1.1. *Suppose $G : X \otimes Y \Rightarrow Z$ and f is an extended real-valued function. Then*

$$\inf_{z\in\overline{Z}} f(z) = \inf_{x\in X}\inf_{y\in Y} f(G(x,y)),$$

where $\overline{Z} = G(X,Y)$. Moreover if the infimum on the right is attained in $(\hat{x},\hat{y})$, then the infimum on the left is attained in $\hat{z} = G(\hat{x},\hat{y})$.

Proof. In the eating (see below). □

Date: March 29, 2004.

2000 *Mathematics Subject Classification.* 62H25.

Key words and phrases. Multivariate Analysis, Correspondence Analysis.

Here are some quick examples. First take $G(x,y)=x-y$. Then

$$\inf_z f(z)=\inf_{x\geq 0}\inf_{y\geq 0} f(x-y)=$$

$$=\inf_x\inf_y f(x-y).$$

Now take $G(x,\lambda)=\lambda x$, with λ a scalar and x a vector,

$$\inf_z f(z)=\inf_{\lambda\geq 0}\inf_{x'x=1} f(\lambda x)=$$

$$=\inf_\lambda\inf_{x'x=1} f(\lambda x)=$$

$$=\inf_\lambda\inf_x f(\lambda x).$$

If $G(x,\lambda)=\frac{x}{\lambda}$, with $\lambda\neq 0$ and $x'x=1$, then $\overline{Z}$ is the set of all vectors $z\neq 0$. Thus

$$\inf_{z\neq 0} f(z)=\inf_{\lambda\neq 0}\inf_{x'x=1} f(\frac{x}{\lambda}).$$

Somewhat less trivially, for a symmetric matrix argument A,

$$\inf_A f(A)=\inf_{\mathbf{dg}(\Lambda)=\Lambda}\inf_{K'K=I} f(K\Lambda K').$$

Observe we can always interchange the two infimum operations, because $\inf_{x\in X}\inf_{y\in Y}=\inf_{y\in Y}\inf_{x\in X}$. Because f is extended real valued, the infimum always exists, although it may be $-\infty$.

2. APPLICATION

We now discuss an actual example. Consider the problem of minimizing the function

$$f(z) = \frac{(z-b)'A(z-b)+c}{z'z}, \tag{1}$$

over $z \neq 0$, where we make no assumptions on the matrix A, the vector b, and the scalar c. Instead of going the usual route of differentiating and solving the stationary equations, we use the decomposition approach of Theorem 1.1.

Define

$$g(x,\lambda) = \frac{(\lambda x - b)'A(\lambda x - b)+c}{\lambda^2 x'x}, \tag{2a}$$

or, letting $\theta = \lambda^{-1}$,

$$g(x,\theta) = \frac{\theta^2(b'Ab+c) - 2\theta b'Ax + x'Ax}{x'x}, \tag{2b}$$

Then

$$\inf_{z\neq 0} f(z) = \inf_{x'x=1} \inf_{\theta\geq 0} g(x,\theta), \tag{2c}$$

but also

$$\inf_{z\neq 0} f(z) = \inf_{x'x=1} \inf_{\theta} g(x,\theta). \tag{2d}$$

If $x'x=1$ then

$$\inf_\theta g(x,\theta)=\inf_\theta \theta^2(b'Ab+c)-2\theta b'Ax+x'Ax.$$

If $b'Ab+c>0$ the minimum is attained at

$$\hat\theta=\frac{b'Ax}{b'Ab+c}$$

and the minimum is equal to $x'\overline{A}x$, where

$$\overline{A}=A-\frac{Abb'A}{b'Ab+c}.$$

It follows that in this case $\min_z f(z)$ is the smallest eigenvalue of $\overline{A}$, written as $\kappa(\overline{A})$. If $\overline{x}$ is the corresponding unit-length eigenvector, then the minimizer of $f(z)$ is

$$\hat z=\frac{b'Ab+c}{b'A'\overline{x}}\,\overline{x}.$$

If $b'Ab+c<0$ the minimum is not attained and $\inf_\theta g(x,\theta)=-\infty$ for each x. Thus $\inf_z f(z)=-\infty$ as well.

If $b'Ab+c=0$ then we must distinguish two sub-cases. If $b'Ax=0$ then $\min_\theta g(x,\theta)=x'Ax$. If $b'Ax\neq 0$ then $\inf_\theta g(x,\theta)=-\infty$ again. Thus if $b'Ab+c=0$ we have $\inf_z f(z)=-\infty$, unless both $c=0$ and $Ab=0$, in which sub-case we have $\min_z f(z)$ equal to $\kappa(A)$, the smallest eigenvalue of A and the minimizer equal to any corresponding eigenvector.

Of course if $Ab = 0$ we have $\overline{A} = A$. Thus

$$\inf_z f(z) = \begin{cases} \kappa(\overline{A}) & \text{if } (b'Ab + c > 0) \vee (Ab = 0 \wedge c = 0), \\ -\infty & \text{otherwise.} \end{cases}$$

Now start with the alternative decomposition

$$\min_{z \neq 0} f(z) = \min_{x'x=1} \min_{\theta \geq 0} \theta^2 (b'Ab + c) - 2\theta b'Ax + x'Ax.$$

We want to show that although the intermediate calculations are different, the result is the same.

If $b'Ab + c > 0$ and $b'Ax \geq 0$ then $\min_\theta g(x, \theta) = x'\overline{A}x$, as before. But if If $b'Ab + c > 0$ and $b'Ax < 0$ the minimum is attained at $\hat{\theta} = 0$, and $\min_\theta g(x, \theta) = x'Ax$. Because $\kappa(\overline{A})$ is less than or equal to $\kappa(A)$, we still have $\min_z f(z)$ equal to the smallest eigenvalue of $\overline{A}$.

If $b'Ab + c < 0$ we stiil have $\inf_z f(z) = -\infty$.

If $b'Ab+c = 0$ we distinguish three sub-cases. If $b'Ax = 0$ then $\min_\theta g(x, \theta) = x'Ax$, as before. If $b'Ax$ 0 then $\inf_\theta g(x, \theta) = -\infty$. And if $b'Ax < 0$ the minimum is attained at $\hat{\theta} = 0$ and equal to $x'Ax$. Again we have $\inf_z f(z) = -\infty$, unless both $c = 0$ and $Ab = 0$, when $\min_z f(z)$ is equal to $\kappa(A)$.

We have solved the problem by using the decompositions (2c) and (2d). But we can also interchange the order of the infimums and use

$$\inf_{z\neq 0} f(z) = \inf_{\theta\geq 0}\inf_{x'x=1} g(x,\theta), \tag{3a}$$

or

$$\inf_{z\neq 0} f(z) = \inf_{\theta}\inf_{x'x=1} g(x,\theta). \tag{3b}$$

Let's look at the problem $\min_{x'x=1} g(x,\theta)$. For a minimum we must have $x = \theta(A-\mu I)^{-1}Ab$, where the Lagrange multiplier μ is chosen such that $\theta^2 b'A(A-\mu I)^{-2}Ab = 1$. At the minimum

$$\min_{x'x=1} g(x,\theta) =$$

$$\theta^2[(b'Ab+c)-2b'A(A-\mu I)^{-1}Ab+b'A(A-\mu I)^{-1}A(A-\mu I)^{-1}Ab]$$

3. Least-squares-squared Unfolding

Now a data analysis example. In least-squares-squared metric unfolding (LSSMU) we must minimize

$$\sigma(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}(\delta_{ij}^2 - [x_i'x_i + y_j'y_j - 2x_i'y_j])^2. \tag{4}$$

over the $n\times p$ and $m\times p$ configuration matrices X and Y. This has been typically handled by block decomposition. The $(n+m)p$ unknowns are partitioned into a number of subsets. Block relaxation algorithms then cycle

through the subsets, minimizing over the parameters in the subset while keeping all parameters fixed at their current values. One cycle through the subsets is one iteration of the algorithm.

In ALSCAL [Takane et al., 1977] coordinate descent is used, which means that the blocks consist of a single coordinate. There are $(n+m)p$ blocks. Solving for the optimal coordinate, with all other fixed, means minimizing a quartic, which in turn means finding the roots of a cubic. The algorithm converges to a stationary point which is a global minimum with respect to each coordinate separately. An alternative algorithm, proposed by Greenacre and Browne [1986], uses the $n+m$ points as blocks. Each substep is again an easy unidimensional minimization. Their algorithm converges to a stationary point which is a global minimum with respect to each point. Generally it is considered to be desirable to have fewer blocks, both to increase the speed of convergence and to restrict the class of local minima we can converge to.

Let us use our basic theorem to construct a four-block algorithm for LSSMU. Minimizing (4) is the same as minimizing

$$\sigma(X,Y,\alpha,\beta)=\sum_{i=1}^n\sum_{j=1}^m w_{ij}(\delta_{ij}^2-[\alpha_i^2+\beta_j^2-2\alpha_i\beta_j x_i'y_j])^2 \tag{5}$$

over α, β, X, and Y, where the configuration matrices X and Y are constrained by $\mathbf{diag}(XX')=I$ and $\mathbf{diag}(YY')=I$.

The algorithm starts with values $\Theta^{(0)} = (\alpha^{(0)}, \beta^{(0)}, X^{(0)}, Y^{(0)})$ satisfying the constraints. Suppose we have arrived at $\Theta^{(k)}$. We then update

$$\begin{aligned}
&\text{(6a)} \qquad \alpha^{(k+1)} = \mathop{\textbf{argmin}}_{\alpha} \ \sigma(X^{(k)}, Y^{(k)}, \alpha, \beta^{(k)}),\\
&\text{(6b)} \qquad \beta^{(k+1)} = \mathop{\textbf{argmin}}_{\beta} \ \sigma(X^{(k)}, Y^{(k)}, \alpha^{(k+1)}, \beta),\\
&\text{(6c)} \qquad X^{(k+1)} = \mathop{\textbf{argmin}}_{\textbf{diag}(XX')=I} \ \sigma(X, Y^{(k)}, \alpha^{(k+1)}, \beta^{(k+1)}),\\
&\text{(6d)} \qquad Y^{(k+1)} = \mathop{\textbf{argmin}}_{\textbf{diag}(YY')=I} \ \sigma(X^{(k+1)}, Y, \alpha^{(k+1)}, \beta^{(k+1)}).
\end{aligned}$$

This gives $\Theta^{(k+1)}$. It is understood that in each of the four substeps of (6) we compute the global minimum, and if the global minimum happens to be nonunique we select any of them. We also remark that, as with any block relaxation method having more than two blocks, there are many variations on this basic scheme. We can travel through the substeps in a different order, we can change the order in each cycle, we can pass through the substeps in random order, we can cycle through the first two substeps a number of times before going to the third and fourth, and so on. Each of these strategies has its own overall convergence rate, and further research would be needed to determine what is best.

Let us look at the subproblems a bit more in detail to see how they can be best solved. Expanding (5) and organizing terms by powers of α gives

$$\begin{aligned}\sigma(X, Y, \alpha, \beta) = &\sum_{i=1}^{n} \alpha_i^4 \sum_{j=1}^{m} w_{ij} + \\ &- \sum_{i=1}^{n} \alpha_i^3 \sum_{j=1}^{m} w_{ij}\beta_j c_{ij} + \\ &+ \sum_{i=1}^{n} \alpha_i^2 \sum_{j=1}^{m} w_{ij}(4\beta_j^2 c_{ij}^2 + 2\beta_j^2 - 2\delta_{ij}^2) + \\ &- \sum_{i=1}^{n} \alpha_i^2 \sum_{j=1}^{m} 4w_{ij}\beta_j^3 c_{ij} + \\ &+ \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}(\delta_{ij}^4 + \beta_j^4 + 4\delta_{ij}^2 c_{ij} - 2\delta_{ij}^2\beta_j^2),\end{aligned}$$

where $c_{ij} = x_i' y_j$. This is a sum of n univariate quartic polynomials, which can be minimized separately to give the global minimum over α. Obviously the same applies to minimization over β.

For minimization over X and Y we define

$$r_{ij} = \frac{\delta_{ij}^2 - [\alpha_i^2 + \beta_j^2]}{2\alpha_i\beta_j},$$

$$w_{ij} = 4\alpha_i^2\beta_j^2 w_{ij}.$$

Then

$$\sigma(X, Y, \alpha, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}[r_{ij} - x_i' y_j]^2.$$

Expanding and collecting terms gives

$$\sigma(X,Y,\alpha,\beta)=\sum_{i=1}^n\psi_i(x_i)$$

with

$$\psi_i(x_i)=f_i-2x_i'g_i+x_i'H_ix_i)$$

and

$$\begin{aligned}f_i&=\sum_{j=1}^m w_{ij}r_{ij}^2,\\g_i&=\sum_{j=1}^m w_{ij}r_{ij}y_j,\\H_i&=\sum_{j=1}^m w_{ij}y_jy_j'.\end{aligned}$$

Again this is the sum of n separate functions ψ_i, quadratics in this case, which can be minimized separately for each x_i. By symmetry, we have the same strategy to minimize over Y.

Mimizing over x_i, under the constraint $x_i'x_i=1$, leads to the secular equation problem discussed in the Appendix. Since typically p is two or at most three, the subproblems are very small indeed and can be solved efficiently.

APPENDIX A. SECULAR EQUATIONS

Consider the problem of finding the stationary values of $f(x) = x'Ax - 2b'x + c$ over all x satisfying $x'x = d$. The stationary equations are

$$(A - \mu I)x = b, \tag{7a}$$

where the Lagrange multiplier μ must satisfy

$$b'(A - \mu I)^{-2}b = d \tag{7b}$$

Equations (7) occur in many different contexts, and have been studied very thoroughly. Some examples are rank-one modified eigen problems, oblique Procrustus rotation, least squares with a quadratic constraint, fitting smoothing splines, trust region optimization methods, and the Greenacre-Browne unfolding algorithm. The key references are Forsythe and Golub [1965]; Spjøtvoll [1972]; Gander [1981]. It may be of interest to statisticians that Forsythe and Golub state in their basic paper that the problem was suggested to them by C.R. Rao.

We review the basic details. If $A = K\Lambda K'$ is an eigen decomposition of A, and $\gamma = K'b$ and we let

$$\phi(\mu) = \sum_{s=1}^{p} \frac{\gamma_s^2}{(\lambda_s - \mu)^2}$$

we have $\phi(\mu) = b'(A - \mu I)^{-2}b$ and thus (7b) becomes $\phi(\mu) = 1$. This is one form of the "secular equation" discussed by Golub [1973]. To solve it, we first observe that the function ϕ is positive. It goes to $+\infty$ if μ approaches one of the λ_s from either direction. It is convex between any two consecutive eigenvalues λ_s and λ_{s+1}. For μ largest than the largest eigenvalue ϕ is convex and decreasing from $+\infty$ to zero, for μ smallest than the smallest eigenvalue $\phi$is convex and increasing from zero to $+\infty$. It follows from these qualitative considerations that we are interested in the unique root of $\phi(\mu) = 1$ which is smaller than the smallest eigenvalue of H_i. Figure 1 shows ϕ with eigenvalues 1,2, and 3 and with the gamma's equal to 0.9, 0.3, and 2.5.

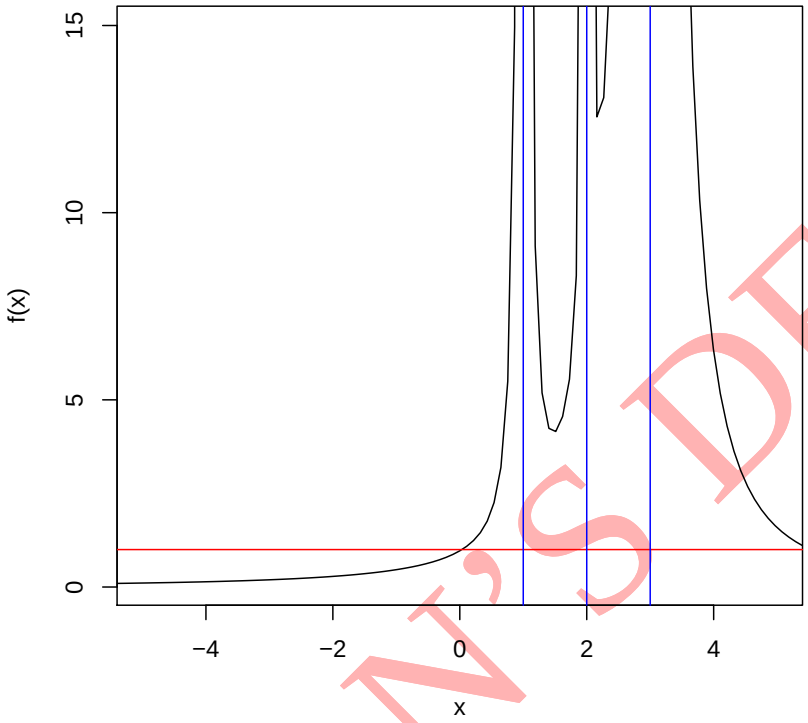

FIGURE 1. Secular Function

Efficient ways of finding this root are given by Gander et al. [1989] and by Melman [1997].

REFERENCES

G.E. Forsythe and G.H. Golub. On the Stationary Values of a Second Degree Polynomial on the Unit Sphere. *Journal of the Society for Industrial and Applied Mathematics*, 13:1050–1068, 1965.

W. Gander. Least Squares with a Quadratic Constraint. *Numerische Mathematik*, 36:291–307, 1981.

W. Gander, G.H. Golub, and U. von Matt. A Constrained Eigenvalue Problem. *Linear Algebra and Its Applications*, 114/115:815–839, 1989.

G.H. Golub. Some Modified Matrix Eigenvalue Problems. *SIAM Review*, 15:318–334, 1973.

M.J. Greenacre and M.W. Browne. An Efficient Alternating Least-Squares Algorithm to Perform Multidimensional Unfolding. *Psychometrika*, 51: 241–250, 1986.

A. Melman. A Unifying Convergence Analysis of Second-Order Methods for Secular Equations. *Mathematics of Computation*, 66:333–344, 1997.

E. Spjøtvoll. A Note on a Theorem of Forsythe and Golub. *SIAM Journal on Applied Mathematics*, 23:307–311, 1972.

Y. Takane, F.W. Young, and J. De Leeuw. Nonmetric Individual Differences in Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 42:7–67, 1977.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu