LOW-RANK APPROXIMATION OF SYMMETRIC MATRICES USING MAJORIZATION ALGORITHMS

JAN DE LEEUW

ABSTRACT. Meet the abstract. This is the abstract.

1. INTRODUCTION

Suppose A is a symmetric matrix of order n, not necessarily positive semi-definite. Our problem is to minimize

$$\sigma(C) = \mathbf{tr}\ (A-C)^2$$

over the set $\mathscr{C}_r$ of all positive semi-definite matrices C with $\mathbf{rank}(C) \leq r$. For the projection of A on $\mathscr{C}_r$ we write $\Pi_r(A)$. Thus

$$\mathbf{tr}\ (A-\Pi_r(A))^2 = \min_{C\in\mathscr{C}_r} \mathbf{tr}\ (A-C)^2.$$

Suppose A_+ is the projection of A on the cone of positive semi-definite matrices. Then $A_- = A_+ - A$ is positive semi-definite, and $\mathbf{tr}\ A_+A_- = 0$. Thus

$$\begin{aligned}\sigma(C) = \mathbf{tr}\ ((A_+ - C) - A_-)^2 = \\ = \mathbf{tr}\ (A_+ - C)^2 + \mathbf{tr}\ A_-^2 + 2\mathbf{tr}\ CA_-.\end{aligned}$$

If we choose $\hat{X}$ such that $\hat{X}\hat{X}'$ is the best rank-p approximation to A_+, then $\mathbf{tr}\ \hat{X}\hat{X}'A_- = 0$ and thus

$$\min_X \sigma(X) = \min_X \mathbf{tr}\ (A_+ - XX')^2 + \mathbf{tr}\ A_-^2.$$

The solution can be easily described in terms of the spectral decomposition $A = K\Lambda K'$. Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Number the eigenvectors accordingly. Let

$$\hat{\lambda}_s = \begin{cases}\lambda_s & \text{if } \lambda_s \geq 0 \text{ and } s \leq p, \\ 0 & \text{otherwise.}\end{cases}$$

Date: January 13, 2008 — 9h 39min — Typeset in TIMES ROMAN.

2000 *Mathematics Subject Classification.* 00A00.

Key words and phrases. Binomials, Normals, LaTeX.

Then $\hat{x}_s = k_s\hat{\lambda}_s$ for $s = 1, \cdots, p$. Thus $\textbf{rank}(\hat{X}) = \textbf{rank}(\hat{A}_+)$.

It may be impractical, or inefficient to compute an complete spectral decomposition. In some cases $r \ll n$ and we only need a small number eigenvectors. In other cases, such as within other iterative algorithms, we may not need the exact approximation, but we only need to improve the current approximation by a small step.

2. Basic Majorization

If A is positive semi-definite, then by convexity

$$\textbf{tr}\ X'AX \geq \textbf{tr}\ Y'AY + 2\textbf{tr}\ Y'A(X-Y) = 2\textbf{tr}\ Y'AX - \textbf{tr}\ Y'AY.$$

Thus

$$\sigma(X) \leq \eta(X,Y) \stackrel{\Delta}{=} \textbf{tr}\ A^2 - 4\textbf{tr}\ X'AY + \textbf{tr}\ (X'X)^2 + 2\textbf{tr}\ Y'AY.$$

The general theory of majorization methods now tells us that the update

$$X^{(k+1)} \in \underset{X}{\textbf{argmin}}\, \eta(X, X^{(k)})$$

will generate a convergent algorithm. We solve

$$\frac{\partial\eta(X,Y)}{\partial X} = -4(AY - X(X'X)) = 0,$$

which gives $(X'X)^3 = Y'A^2Y$, and thus

$$X^{(k+1)} = AX^{(k)}((X^{(k)})'A^2X^{(k)})^{-\frac{1}{3}}.$$

The algorithm can be implemented by using $U^{(k)} = AU^{(k-1)} = A^kU^{(0)}$ and then

$$C^{(k+1)} = X^{(k+1)}(X^{(k+1)})' = U^{(k)}((U^{(k)})'U^{(k)})^{-\frac{2}{3}}(U^{(k)})'.$$

Then $\sigma(X^{(k+1)}) = \textbf{tr}\ (A - C^{(k+1)})^2$, and the sequence $C^{(k)}$ will converge to $\Pi_r(A)$.

Computationally the algorithm requires an eigenvalue routine to compute the fractional matrix power.

3. Doing Better

Since $C^{(k+1)} = U^{(k)}S^{(k)}(U^{(k)})'$ with $S^{(k)} = ((U^{(k)})'U^{(k)})^{-\frac{2}{3}}$ we see that

$$\min_S \textbf{tr}\ (A - U^{(k)}S(U^{(k)})')^2 \leq \sigma(X^{(k+1)}).$$

The minimum over S is attained at

$$\tilde{S}^{(k)} = ((U^{(k)})'U^{(k)})^{-1}(U^{(k)})'AU^{(k)}((U^{(k)})'U^{(k)})^{-1}.$$

This suggest the algorithm in which we define $P^{(k)} = U^{(k)}((U^{(k)})'U^{(k)})^{-1}(U^{(k)})'$ and then

$$\tilde{C}^{(k+1)} = P^{(k)}AP^{(k)}.$$

It follows that $\mathbf{tr}\ (A-\tilde{C}^{(k+1)})^2 \leq \mathbf{tr}\ (A-C^{(k+1)})^2$, and $\tilde{C}^{(k)}$ also converges to $\Pi_r(A)$.

The projectors can be computed by QR-decomposition of $U^{(k)}$, which is not too demanding.

4. Even Better

Both algorithms we have discussed so far are of the form $X^{(k+1)} = U^{(k)}T^{(k)}$,

5. Projection and Majorization

Rewrite the problem as minimizing $\sigma(K,T) = \|A-KTK'\|^2$ with K of size $n\times r$ and orthonormal, and with T psd of order r. Now

$$\sigma(K,T) = \mathbf{tr}\ A^2 - 2\mathbf{tr}\ TK'AK + \mathbf{tr}\ T^2,$$

and thus

$$\min_T \sigma(K,T) = \mathbf{tr}\ A^2 - \mathbf{tr}\ (K'AK)^2.$$

This means we have to maximize $\mathbf{tr}\ (K'AK)^2$ over K. Suppose $\tilde{K}$ is the current best solution. Then

$$\mathbf{tr}\ (K'AK)^2 \geq \frac{1}{\mathbf{tr}\ (\tilde{K}'A\tilde{K})^2}\{\mathbf{tr}\ K'A\tilde{K}\}^2$$

Thus the majorization step maximizes $\mathbf{tr}\ K'A\tilde{K}$ over $K'K=I$, i.e. $K = \mathbf{procrus}(A\tilde{K})$. Because $\min_T \sigma(K,T)$ is invariant under rotations of K we get the same sequence of function values by taking $K = \mathbf{gram}(A\tilde{K})$.

If A is not positive semi-definite, then

$$\min_{T\gtrsim 0} \sigma(K,T) = \mathbf{tr}\ A^2 - \mathbf{tr}\ (K'AK)_+^2.$$

6. Iterating with A^2

7. Asymmetric Iterations

Minimize

$$\sigma(X,Y) = \mathbf{tr}\ (A - XY')'(A - XY')$$

Algorithm

$$Y^{(k)} = AX^{(k)}((X^{(k)})'X^{(k)})^{-1},$$
$$X^{(k+1)} = AY^{(k)}((Y^{(k)})'Y^{(k)})^{-1},$$

or

$$X^{(k+1)} = A^2X^{(k)}((X^{(k)})'A^2X^{(k)})^{-1}(X^{(k)})'X^{(k)}.$$

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`