The Lanczos Algorithm

Daniel Hernández-Lobato

Universidad Autónoma de Madrid
University of Cambridge

October, 2013

Introduction

The Lanczos algorithm can be used to efficiently find the external eigenvalues (maximum and minimum) of a symmetric matrix $\mathbf{A}$ of size $n \times n$.

Based on computing the following decomposition of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{Q}\mathbf{T}\mathbf{Q}^{\mathsf{T}}.$$

where $\mathbf{Q}$ is an orthonormal basis of vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$ and $\mathbf{T}$ is tri-diagonal

$$\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n], \qquad \mathbf{T} = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & \cdots & 0 \\ \beta_1 & \alpha_2 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \beta_n \\ 0 & \cdots & 0 & \beta_n & \alpha_n \end{bmatrix}$$

The decomposition always exists and is unique given that $\mathbf{q}_1$ has been specified.

Lanczos Iterations

We know that $\mathbf{T} = \mathbf{Q}^\mathsf{T}\mathbf{A}\mathbf{Q}$ which gives $\alpha_k = \mathbf{q}_k^\mathsf{T}\mathbf{A}\mathbf{q}_k$ and $\beta_k = \mathbf{q}_{k+1}^\mathsf{T}\mathbf{A}\mathbf{q}_k$.

The full decomposition is obtained by equating the columns of $\mathbf{AQ} = \mathbf{QT}$.

$$\begin{aligned}\left[\ \mathbf{A}\mathbf{q}_1,\quad \mathbf{A}\mathbf{q}_2,\quad \cdots,\quad \mathbf{A}\mathbf{q}_n\ \right] = \big[\ & \alpha_1\mathbf{q}_1 + \beta_2\mathbf{q}_2,\quad \beta_1\mathbf{q}_1 + \alpha_2\mathbf{q}_2 + \beta_2\mathbf{q}_3,\quad \cdots \\ & \cdots,\quad \beta_{n-1}\mathbf{q}_{n-1} + \alpha_n\mathbf{q}_n\ \big]\end{aligned}$$

This gives the following equations for computing $\mathbf{q}_1, \ldots, \mathbf{q}_n$ iteratively:

$$\begin{aligned}\mathbf{q}_2 &= 1/\beta_1(\mathbf{A} - \alpha_1)\mathbf{q}_1 \\ \mathbf{q}_3 &= 1/\beta_2\left[(\mathbf{A} - \alpha_2)\mathbf{q}_2 - \beta_1\mathbf{q}_1\right] \\ \cdots & \\ \mathbf{q}_k &= 1/\beta_{k-1}\left[(\mathbf{A} - \alpha_{k-1})\mathbf{q}_{k-1} - \beta_{k-2}\mathbf{q}_{k-2}\right] \qquad 2 < k < n\,, \\ 0 &= (\mathbf{A} - \alpha_n)\mathbf{q}_n - \beta_{n-1}\mathbf{q}_{n-1}\end{aligned}$$

Since $\mathbf{q}_{k+1}$ has unit norm $\beta_k = |(\mathbf{A} - \alpha_k)\mathbf{q}_k - \beta_{k-1}\mathbf{q}_{k-1}|$.

Lanczos Algorithm

- $\mathbf{r}_0 = \mathbf{q}_1$; $\mathbf{q}_0 = \mathbf{0}$; $\beta_0 = 1$;
- for $(k = 1, \ldots, n)$
  - if $(\beta_{k-1} = 0)$
    - break;
  - endif
  - $\mathbf{q}_k = \mathbf{r}_{k-1}/\beta_{k-1}$;
  - $\alpha_k = \mathbf{q}_k^\mathsf{T}\mathbf{A}\mathbf{q}_k$;
  - $\mathbf{r}_k = (\mathbf{A} - \alpha_k)\mathbf{q}_k - \beta_{k-1}\mathbf{q}_{k-1}$;
  - $\beta_k = |\mathbf{r}_k|$;
- endfor

$\mathbf{q}_1$ is set randomly and the $\mathbf{q}_k$ are called the Lanczos vectors (orthonormal).

At iteration k the algorithm generates intermediate matrices $\mathbf{Q}_k$ and $\mathbf{T}_k$:

$$\mathbf{Q}_k = [\ \mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_k\ ],$$

$$\mathbf{T}_k = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & \cdots & 0 \\ \beta_1 & \alpha_2 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & \beta_k \\ 0 & \cdots & 0 & \beta_k & \alpha_k \end{bmatrix}$$

that satisfy $\mathbf{T}_k = \mathbf{Q}_k^\mathsf{T}\mathbf{A}\mathbf{Q}_k$ and have important properties.

$\mathbf{A}$ is not modified and it is only required to provide a procedure to compute matrix-vector products involving $\mathbf{A}$. No other matrices are generated.

Properties of $\mathbf{q}_k$ and $\mathbf{T}_k$ and Low-Rank Approximation of $\mathbf{A}$

At iteration k the k-th Lanczos vector, $\mathbf{q}_k$, is proven to maximize the l.h.s. of

$$\max_{|\mathbf{y}|=1} \frac{\mathbf{y}^\mathsf{T} \left(\mathbf{Q}_k^\mathsf{T} \mathbf{A} \mathbf{Q}_k\right) \mathbf{y}}{\mathbf{y}^\mathsf{T} \mathbf{y}} = \max_{|\mathbf{y}|=1} \frac{\mathbf{y}^\mathsf{T} \mathbf{T}_k \mathbf{y}}{\mathbf{y}^\mathsf{T} \mathbf{y}} = \lambda_1(\mathbf{T}_k) \leq \lambda_1(\mathbf{A}) = \lambda_1(\mathbf{T}),$$

and to simultaneously minimize the l.h.s. of

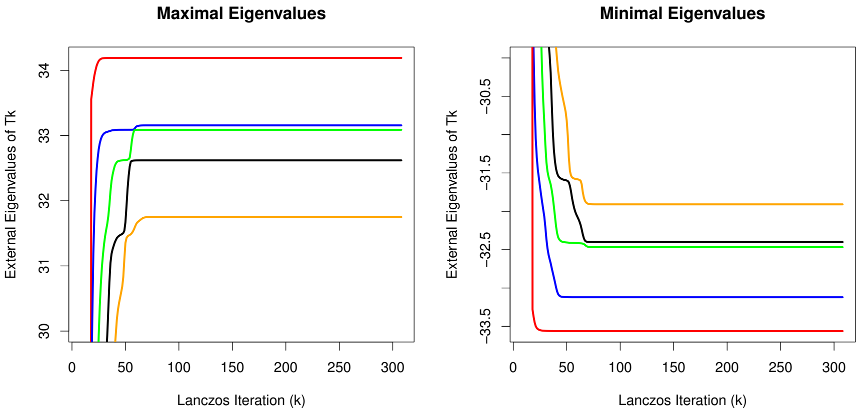
$$\min_{|\mathbf{y}|=1} \frac{\mathbf{y}^\mathsf{T} \left(\mathbf{Q}_k^\mathsf{T} \mathbf{A} \mathbf{Q}_k\right) \mathbf{y}}{\mathbf{y}^\mathsf{T} \mathbf{y}} = \min_{|\mathbf{y}|=1} \frac{\mathbf{y}^\mathsf{T} \mathbf{T}_k \mathbf{y}}{\mathbf{y}^\mathsf{T} \mathbf{y}} = \lambda_n(\mathbf{T}_k) \geq \lambda_n(\mathbf{A}) = \lambda_n(\mathbf{T}),$$

where $\lambda_1(\mathbf{A})$ and $\lambda_1(\mathbf{A})$ are the maximum and the minimum eigenvalue of $\mathbf{A}$.

The external eigenvalues of $\mathbf{T}_k$ progressively become more similar to the ones of $\mathbf{A}$. This information emerges long before the tri-diagonalization is complete!

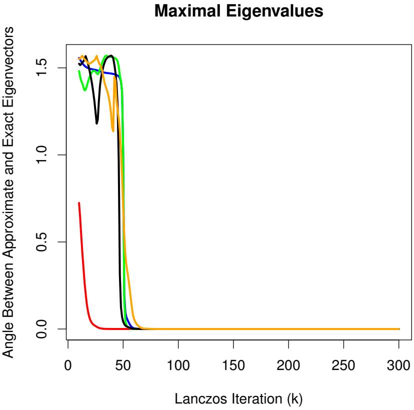
Let $\mathbf{T}_k = \mathbf{H}_k \mathbf{\Delta}_k \mathbf{H}_k^\mathsf{T}$ be the diagonalization of $\mathbf{T}_k$, computed in $\mathcal{O}(k^2)$ steps. One expects that $\mathbf{Q}_k \mathbf{H}_k$ approximates the eigenvectors of $\mathbf{A}$. A rank $k < n$ approximation of $\mathbf{A}$ is hence $\mathbf{A} \approx \mathbf{Q}_k \mathbf{H}_k \mathbf{\Delta}_k \mathbf{H}_k^\mathsf{T} \mathbf{Q}_k^\mathsf{T} = \mathbf{V}_k \mathbf{\Delta}_k \mathbf{V}_k^\mathsf{T}$.

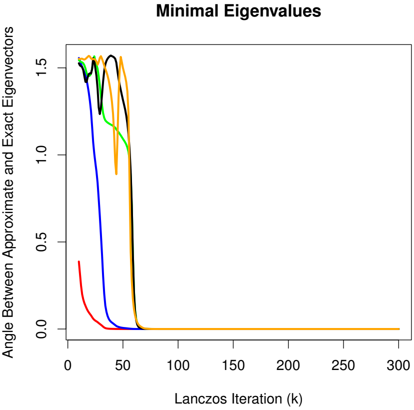
Convergence of external eigenvalues Δ_k



Convergence is observed after only a few of iterations of the Lanczos algorithm.

Convergence of external eigenvectors $\mathbf{V}_k = \mathbf{Q}_k \mathbf{H}_k$





Convergence is observed after only a few of iterations of the Lanczos algorithm.

Summary

- The Lanczos algorithm can be used to compute the external eigenvalues of a symmetric matrix **A**.
- It is very simple to implement. It only requires matrix-vector multiplications with respect to **A**.
- The matrix **A** is not modified during the algorithm. Very useful if **A** has a sparse form. No other generated matrices besides $\mathbf{Q}_k$ and $\mathbf{T}_k$.
- Can be used to efficiently find a low-rank approximation of **A**. This could be useful for approximate Bayesian inference (Seeger *et al.*).
- The algorithm is very sensitive to round-off problems. The Lanczos vectors $\mathbf{q}_k$ loose orthogonality. External eigenvalues are duplicated.
- The vectors have to be re-orthogonalized (complete or selective). Selective re-orthogonalization does not increase the cost too much.

References

- G. Golub and C. Van Loan. Matrix Computations. Johns Hopkins University Press, 3rd edition, 1996.
- J. W. Demmel, Applied Numerical Linear Algebra, SIAM, Philadelphia, 1997.
- M. Seeger, H. Nickisch, R. Pohmann & B. Schölkopf. Bayesian experimental design of magnetic resonance imaging sequences. Advances in Neural Information Processing Systems 21 (pp. 1441-1448), 2009.
- Dhillon, I., Parlett, B.: Multiple Representations to Compute Orthogonal Eigenvectors of Symmetric Tridiagonal Matrices. Linear Algebra Appl. 387, 128 (2004)

Thank you for your attention!