Simple methods for the top generalized eigenvector

Thomas P. Minka

October 29, 2003

1 Introduction

This paper is concerned with finding the vector $\mathbf{x}$ which maximizes the Rayleigh quotient

$$f(\mathbf{x}) = \frac{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}} \tag{1}$$

The matrices $\mathbf{A}, \mathbf{B}$ are assumed to be real, symmetric, and positive definite (denoted $\mathbf{A}, \mathbf{B} > 0$).

2 Power method

The power method repeatedly maximizes a lower bound on the Rayleigh quotient.

Lemma If $\mathbf{A}, y > 0$, the function $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}/y$ is convex in $(\mathbf{x}, y)$.

Proof sketch It is straightforward to show that x^2/y is convex in (x, y) by inspecting the second derivative matrix. From this we can conclude that $\sum_i x_i^2/y = \mathbf{x}^{\mathrm{T}}\mathbf{x}/y$ is convex, therefore so is $(\mathbf{U}\mathbf{x})^{\mathrm{T}}(\mathbf{U}\mathbf{x})/y = \mathbf{x}^{\mathrm{T}}(\mathbf{U}^{\mathrm{T}}\mathbf{U})\mathbf{x}/y$. Finally $\mathbf{A} > 0$ implies $\mathbf{A} = \mathbf{U}^{\mathrm{T}}\mathbf{U}$ for some $\mathbf{U}$.

Convexity implies that the tangent plane at any point $(\mathbf{x}_0, y_0)$ is a lower bound:

$$\frac{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{y} \geq \frac{2\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}_0}{y_0} - \frac{\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0}{y_0^2}y \qquad \text{if } \mathbf{A}, y > 0 \tag{2}$$

Applying this bound to the Rayleigh quotient gives a quadratic function in $\mathbf{x}$:

$$\frac{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}} \geq \frac{2\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}_0 - f(\mathbf{x}_0)\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}}{\mathbf{x}_0^{\mathrm{T}}\mathbf{B}\mathbf{x}_0} \tag{3}$$

This bound is maximum when

$$\mathbf{B}\mathbf{x} = \mathbf{A}\mathbf{x}_0/f(\mathbf{x}_0) \tag{4}$$

Thus, repeatedly solving (4), where $\mathbf{x}_0$ is the old guess for $\mathbf{x}$, will converge to a (local) maximum of f. When $\mathbf{B} = \mathbf{I}$, this is the usual power method for finding the top eigenvector of a matrix. When $\mathbf{B} \neq \mathbf{I}$, the most efficient way to solve (4) is to precompute the Cholesky decomposition of $\mathbf{B}$. However, decomposing $\mathbf{B}$ can be expensive. The next sections consider alternatives to solving (4).

3 Relaxed power method

We don't need to precisely maximize the lower bound at each step for the power method to converge. We only need to take a step uphill on the quadratic surface. In particular, we can go to the maximum of the bound along some direction $\mathbf{u}$. This maximum is

$$\begin{aligned}
\mathbf{x} &= \mathbf{x}_0 - \mathbf{u}\frac{\mathbf{g}^{\mathrm{T}}\mathbf{u}}{\mathbf{u}^{\mathrm{T}}\mathbf{H}\mathbf{u}} && (5)\\
\mathbf{g} &= \mathbf{A}\mathbf{x}_0 - f(\mathbf{x}_0)\mathbf{B}\mathbf{x}_0 && (6)\\
\mathbf{H} &= -f(\mathbf{x}_0)\mathbf{B} && (7)
\end{aligned}$$

One iteration of this algorithm costs $O(n^2)$, compared to $O(n^3)$ for (4). Note that when $\mathbf{B} = \mathbf{I}$, the gradient leads to the true maximum and (5) is equivalent to (4).

The direction for the line search can simply be the gradient $\mathbf{g}$, but a better approach is to use a conjugated gradient based on the last search direction.

4 Analytic line search

Instead of doing a line search on a lower bound, we can do a line search directly on the Rayleigh quotient. It turns out that this line search can be performed analytically.

$$\begin{aligned}
\mathbf{x} &= \mathbf{x}_0 + a\mathbf{u} && (8)\\
f(\mathbf{x}) &= \frac{\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0 + 2a\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{u} + a^2\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u}}{\mathbf{x}_0^{\mathrm{T}}\mathbf{B}\mathbf{x}_0 + 2a\mathbf{x}_0^{\mathrm{T}}\mathbf{B}\mathbf{u} + a^2\mathbf{u}^{\mathrm{T}}\mathbf{B}\mathbf{u}} && (9)\\
&= f(\mathbf{u}) + \frac{(\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{T}}\mathbf{B}\mathbf{x}_0 f(\mathbf{u})) + 2a(\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{u} - \mathbf{x}_0^{\mathrm{T}}\mathbf{B}\mathbf{u} f(\mathbf{u}))}{\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}} && (10)\\
\frac{df(\mathbf{x})}{da} &= \frac{c_0 + ac_1 + a^2c_2}{\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}} && (11)\\
c_0 &= (\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{x}_0)(\mathbf{x}_0^{\mathrm{T}}\mathbf{B}\mathbf{x}_0) - (\mathbf{u}^{\mathrm{T}}\mathbf{B}\mathbf{x}_0)(\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0) && (12)\\
c_1 &= (\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u})(\mathbf{x}_0^{\mathrm{T}}\mathbf{B}\mathbf{x}_0) - (\mathbf{u}^{\mathrm{T}}\mathbf{B}\mathbf{u})(\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0) && (13)\\
c_2 &= (\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u})(\mathbf{u}^{\mathrm{T}}\mathbf{B}\mathbf{x}_0) - (\mathbf{u}^{\mathrm{T}}\mathbf{B}\mathbf{u})(\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{x}_0) && (14)
\end{aligned}$$

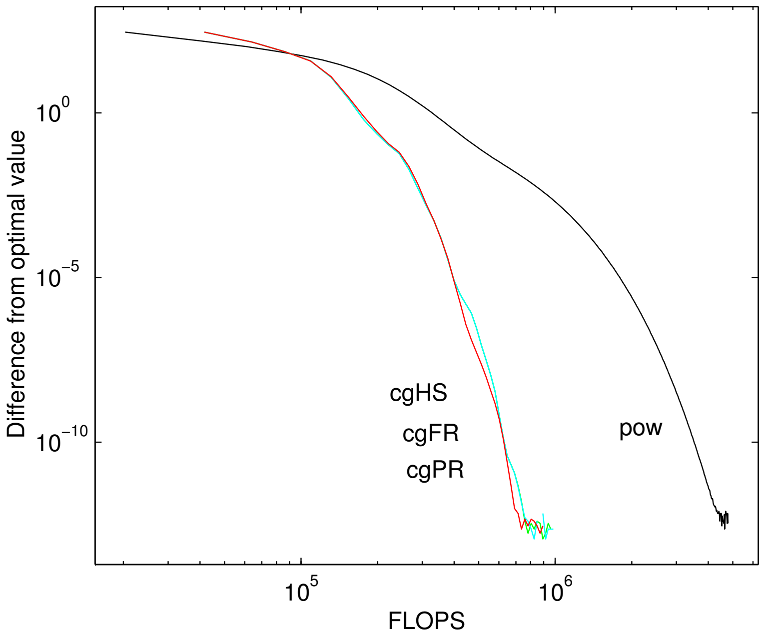
Setting the gradient to zero gives a quadratic equation for a, which is easily solved. The root producing the maximum is

$$a = \frac{-c_1 - \sqrt{c_1^2 - 4c_0c_2}}{2c_2} \qquad (15)$$

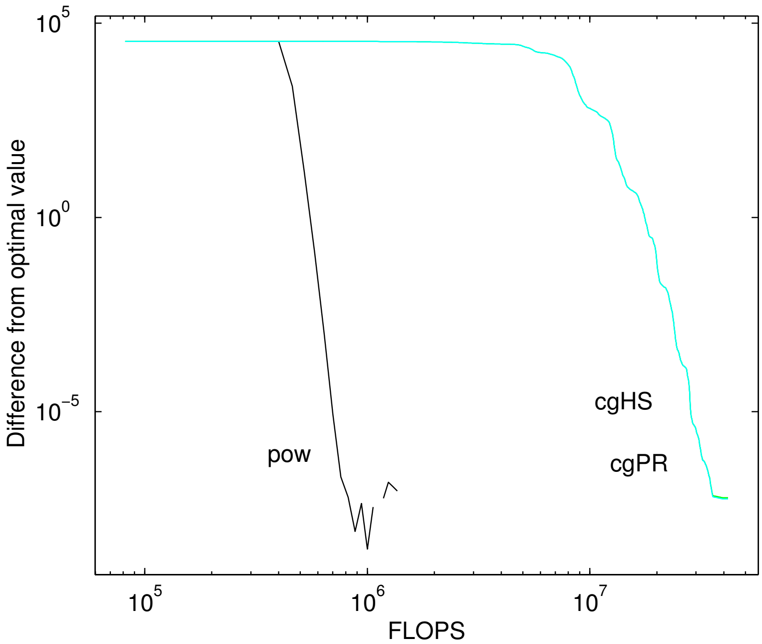
As before, the search direction should be a conjugated gradient.

5 Results

Finding the top eigenvector of a random 100×100 matrix:



Finding the top generalized eigenvector, A and B both random 100×100 matrices:



CG does very well when $\mathbf{B} = \mathbf{I}$, but falls flat in the generalized case. I would be interested if anyone can explain why.