Efficient Implementation of the Tensor Power Method

Y.Cem Sübakan

October 14, 2014

1 Introduction

Tensor Power method is used to extract model parameters out of a moment tensor [1]. A naive implementation of this method would explicit store the full tensor. This may be limiting for observations with large dimensionality. In this write-up we describe how to perform tensor power iterations without explicitly forming the moment tensor.

2 Eigenvectors of a tensor

First of all let's define the operation on the tensor $A \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$ we will use extensively:

$$[A(V_1, V_2, \ldots, V_p)]_{i_1, i_2, \ldots, i_p} = \sum_{j_1, \ldots, j_p} A_{j_1, \ldots, j_p} [V_1]_{j_1, i_1} [V_2]_{j_2, i_2} \ldots [V_p]_{j_p, i_p} \tag{1}$$

where, $V_k \in \mathbb{R}^{m_k \times n_k}$ and consequently $A(V_1, V_2, \ldots, V_p) \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_p}$. Examples:

- Matrix - matrix multiplication: $A(V_1, V_2) = V_1^\top A V_2$, where $A \in \mathbb{R}^{n_1 \times n_2}$.
- Matrix - vector multiplication: $[A(I, v)]_i = \sum_{j_1, j_2} A_{j_1, j_2} I_{j_1, i} v_{j_2} = \sum_{j_2} A_{i, j_2} v_{j_2}$, where $A \in \mathbb{R}^{n_1 \times n_2}, I \in [0, 1]^{n_2 \times n_2}$ is an identity matrix and $v \in \mathbb{R}^{n_2}$ is a vector.

Now, let's consider the mapping $u \to A(I, u, u)$:

$$[A(I, u, u)]_i = \sum_{j_1, j_2, j_3} A_{j_1, j_2, j_3} I_{j_1, i} u_{j_2} u_{j_3} = \sum_{j_2, j_3} A_{i, j_2, j_3} u_{j_2} u_{j_3} = \sum_{j_2, j_3} A_{i, j_2, j_3} (e_{j_2}^\top u)(e_{j_3}^\top u) \tag{2}$$

where, $e_{1:K}$ is the canonical basis. We say that the vector $u \in \mathbb{R}^n$ is an eigenvector with eigenvalue $\lambda \in \mathbb{R}$ if we have,

$$A(I, u, u) = \lambda u. \tag{3}$$

Now, if the tensor has the special structure $A = \sum_{k=1}^K w_k \; \mu_k \otimes \mu_k \otimes \mu_k$, we have the following output:

$$[A(I,u,u)]_i = \sum_{j_1,j_2,j_3} \sum_{k=1}^{K} w_k \; I_{j_1,i} \mu_{k,j_1} \mu_{k,j_2} \mu_{k,j_3} \; u_{j_2} u_{j_3} \tag{4}$$

$$= \sum_{j_2,j_3} \sum_{k=1}^{K} w_k \; \mu_{k,i} \, (u_{j_2} \mu_{k,j_2}) \, (u_{j_3} \mu_{k,j_3})$$

$$= \sum_{k=1}^{K} w_k \; \mu_{k,i} \left(u^\top \mu_k\right) \left(u^\top \mu_k\right) = \sum_{k=1}^{K} w_k \; \mu_{i,j_1} \left(u^\top \mu_k\right)^2$$

Further, if the parameter vectors $\mu_{1:K}$ are orthonormal to each other, we have

$$A(I, \mu_{k'}, \mu_{k'}) = w_{k'} \mu_{k'}. \tag{5}$$

So, $\mu_{k'}$ is an eigenvector with corresponding eigenvalue $w_{k'}$ of the orthogonal tensor A.

3 Efficient Implementation of Tensor Power Iterations

Tensor power iterations to find the eigenvectors are essentially about successive application of the operation $A(I,u,u)$, which we have introduced in the previous section. The most straightforward implementation would be to form the tensor A explicitly, and apply the operation until convergence. However, since the tensor is composed of sum of rank-1 tensors, we can avoid forming the tensor explicitly which would result in a much more efficient implementation.

The empirical moment tensor we are interested in is:

$$M_3 = \sum_{k=1}^{K} w_k \; \mu_k \otimes \mu_k \otimes \mu_k \approx \frac{1}{N} \sum_{n=1}^{N} x_n \otimes x_n \otimes x_n \tag{6}$$

But this tensor is not orthogonal. We orthogonalize it with a whitening matrix W. So, the orthogonal tensor we are interested in decomposing is,

$$[\widetilde{M}_3]_{i_1,i_2,i_3} = [M_3(W,W,W)]_{i_1,i_2,i_3} \approx \frac{1}{N} \sum_{j_1,j_2,j_3} \sum_{n=1}^{N} x_{n,j_1} x_{n,j_2} x_{n,j_3} W_{j_1,i_1} W_{j_2,i_2} W_{j_3,i_3} \tag{7}$$

$$= \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{j_1} W_{j_1,i_1} x_{n,j_1} \right) \left(\sum_{j_2} W_{j_2,i_2} x_{n,j_2} \right) \left(\sum_{j_3} W_{j_3,i_3} x_{n,j_3} \right)$$

So,

$$\widetilde{M}_3 = \frac{1}{N} \sum_{n=1}^{N} \left(W^\top x_n\right) \otimes \left(W^\top x_n\right) \otimes \left(W^\top x_n\right) \tag{8}$$

Now let's look at $\widetilde{M}_3(I,u,u)$:

$$[\widetilde{M}_3(I,u,u)]_i = \frac{1}{N} \sum_{j_1,j_2,j_3} \sum_{n=1}^{N} I_{j_1,i} \left(W^\top x_n\right)_{j_1} u_{j_2} \left(W^\top x_n\right)_{j_2} u_{j_3} \left(W^\top x_n\right)_{j_3} \tag{9}$$

so,

$$\widetilde{M}_3(I,u,u) = \frac{1}{N} \sum_{n=1}^{N} \left(W^\top x_n\right) \left(u^\top W^\top x_n\right)^2 \tag{10}$$

Therefore, the operation $\widetilde{M}_3(I,u,u)$ can be computed without explicitly forming the three dimensional tensor $\widetilde{M}_3$. Finally note that,

$$\widetilde{M}_3(u,u,u) = \frac{1}{N} \sum_{n=1}^{N} \left(u^\top W^\top x_n\right)^3 \tag{11}$$

which is the generalized Rayleigh quotient and is equal to the eigenvalues of $\widetilde{M}_3$.

4 Dealing with Nuisance Terms in GMM

In GMM learning unfortunately there are nuisance terms in GMM third order moment to make the life a little bit more difficult:

$$\mathbb{E}[x \otimes x \otimes x] = \sum_{k=1}^{K} w_k\, \mu_k \otimes \mu_k \otimes \mu_k + \sigma^2 \left(\sum_{l=1}^{L} m \otimes e_l \otimes e_l + \sum_{l=1}^{L} e_k \otimes m \otimes e_l + \sum_{l=1}^{L} e_l \otimes e_l \otimes m \right) \tag{12}$$

So,

$$\begin{aligned} M_3 =& \mathbb{E}[x \otimes x \otimes x] - \sigma^2 \left(\sum_{l=1}^{L} m \otimes e_l \otimes e_l + \sum_{l=1}^{L} e_k \otimes m \otimes e_l + \sum_{l=1}^{L} e_l \otimes e_l \otimes m \right) \qquad (13) \\ \approx& \frac{1}{N} \sum_{n=1}^{N} x_n \otimes x_n \otimes x_n - \sigma^2 \left(\sum_{l=1}^{L} m \otimes e_l \otimes e_l + \sum_{l=1}^{L} e_k \otimes m \otimes e_l + \sum_{l=1}^{L} e_l \otimes e_l \otimes m \right) \end{aligned}$$

And,

$$\begin{aligned} \widetilde{M}_3 = M_3(W,W,W) =& \frac{1}{N} \sum_{n=1}^{N} W^\top x_n \otimes W^\top x_n \otimes W^\top x_n \qquad (14) \\ &- \sigma^2 \left(\sum_{l=1}^{L} W^\top m \otimes W^\top e_l \otimes W^\top e_l + \sum_{l=1}^{L} W^\top e_k \otimes W^\top m \otimes W^\top e_l + \sum_{l=1}^{L} W^\top e_l \otimes W^\top e_l \otimes W^\top m \right) \end{aligned}$$

So,

$$
\begin{aligned}
\widetilde{M}_3(I,u,u) &= \frac{1}{N}\sum_{n=1}^{N} W^\top x_n \left(u^\top W^\top x_n\right)^2 \qquad (15)\\
&- \sigma^2 \left(\sum_{l=1}^{L} W^\top m \otimes u^\top W^\top e_l \otimes u^\top W^\top e_l + \sum_{l=1}^{L} W^\top e_l \otimes u^\top W^\top m \otimes u^\top W^\top e_l + \sum_{l=1}^{L} W^\top e_l \otimes u^\top W^\top e_l \otimes u^\top W^\top m\right)\\
&= \frac{1}{N}\sum_{n=1}^{N} W^\top x_n \left(u^\top W^\top x_n\right)^2\\
&- \sigma^2 \left(W^\top m \sum_{l=1}^{L}\left(u^\top W^\top e_l\right)^2 + 2u^\top W^\top m \sum_{l=1}^{L} W^\top e_l (u^\top W^\top e_l) + \cancel{\sum_{l=1}^{L} W^\top e_l (u^\top W^\top e_l)(u^\top W^\top m)}\right)\\
&= \frac{1}{N}\sum_{n=1}^{N} W^\top x_n \left(u^\top W^\top x_n\right)^2 - \sigma^2 \left(W^\top m \sum_{l=1}^{L}\left(u^\top W^\top e_l\right)^2 + 2u^\top W^\top m \left(W^\top W u\right)\right)
\end{aligned}
$$

Finally,

$$
\widetilde{M}_3(u,u,u) = \frac{1}{N}\sum_{n=1}^{N}\left(u^\top W^\top x_n\right)^3 - 3\sigma^2\left(u^\top W^\top m \sum_{l=1}^{L}\left(u^\top W^\top e_l\right)^2\right) \qquad (16)
$$

5 Simultaneous Iterations

Instead of the deflation method in the original paper, we can simultaneously apply the operation $\widetilde{M}_3(I,u_k,u_k)$ for all vectors $u_{1:K}$. The problem with that approach is that some vectors may converge to the same eigenvector. The remedy for this is to orthogonalize the vectors at every iteration. This eigenvector algorithm is known as the *Simultaneous Iterations* [2] in the numerical linear algebra literature.

References

[1] Anandkumar, A., R. Ge, D. Hsu, S. Kakade and M. Telgarsky, "Tensor Decompositions for Learning Latent Variable Models", *arXiv:1210.7559v2*, 2012.

[2] Trefethen, L. N. and I. David Bau, *Numerical Linear Algebra*, SIAM, 1997.

Algorithm 1 Simultaneous Iterations for Tensor Eigenvectors

Input: Data matrix X, Whitening matrix W.
Output: Estimated model parameters $\widehat{\mu}_{1:K}$ and $\widehat{w}_{1:K}$.

Initialize $u_{1:K}^{(0)}$ with an orthonormal set of vectors.
for $\tau = 1 : maxiter$ **do**
  for $k = 1 : K$ **do**
    $v_k = \widetilde{M}_3(I, u_k^{(\tau-1)}, u_k^{(\tau-1)})$ (Apply Equation 10)
    $v_k = v_k / \|v_k\|_2$ (Normalize)
  end for
  $u_{1:K}^{(\tau)} R = v_{1:K}$ (Orthogonalize using QR factorization)
end for

for $k = 1 : K$ **do**
  $\lambda_k = \widetilde{M}_3(u_k^{(\tau-1)}, u_k^{(\tau-1)}, u_k^{(\tau-1)})$
  $\widehat{\mu}_k = \lambda_k (W^\top)^\dagger u_k^{(\tau)}, \quad \widehat{w}_k = 1/\sqrt{\lambda_k}$
end for