Fast algorithms for sparse principal component analysis based on Rayleigh quotient iteration

Volodymyr Kuleshov

Department of Computer Science
Stanford University

June 18, 2013

Sparse principal component analysis

Seeks principal components that maximize variance subject to a sparsity constraint:

Sparse PCA	PCA
$$\begin{aligned} \max \quad & \frac{1}{2} x^T \Sigma x \\ \text{s.t.} \quad & \|x\|_2 \leq 1 \\ & \|x\|_0 \leq k \end{aligned}$$	$$\begin{aligned} \max \quad & \frac{1}{2} x^T \Sigma x \\ \text{s.t.} \quad & \|x\|_2 \leq 1 \end{aligned}$$

where $\Sigma \in \mathbb{R}^{n \times n}$, $\Sigma = \Sigma^T$ and $k > 0$.

Current state-of-the-art

The most successful methods are variations of the *generalized power method*.

Algorithm 1 GPM(Σ, x_0, γ, ϵ)

$j \leftarrow 0$
repeat
  $y \leftarrow \Sigma x^{(j)}$
  $x^{(j+1)} \leftarrow \mathrm{SparsifyAndScale}_\gamma(y)$ // New relative to pow. met.
  $j \leftarrow j + 1$
until $||x^{(j)} - x^{(j-1)}|| < \epsilon$

return $x^{(j)}$

The sparsification step typically consists of soft thresholding and scaling to a norm of one.

Rayleigh quotient iteration

A more sophisticated algorithm for computing eigenvalues than the power method.

Algorithm 2 RayleighQuotientIteration(Σ, x_0, ϵ)

$j \leftarrow 0$
repeat
  $\mu \leftarrow \dfrac{(x^{(j)})^T \Sigma x^{(j)}}{(x^{(j)})^T x^{(j)}}$ // Rayleigh quotient

  $x^{(j+1)} \leftarrow \dfrac{(\Sigma - \mu I)^{-1} x^{(j)}}{||(\Sigma - \mu I)^{-1} x^{(j)}||}$

  $j \leftarrow j + 1$
until $||x^{(j)} - x^{(j-1)}|| < \epsilon$
return $x^{(j)}$

Generalized Rayleigh quotient iteration

Algorithm 3 GRQI(Σ, x_0, k, J, ϵ)

$j \leftarrow 0$
repeat
  $\mathcal{W} \leftarrow \{i | x_i^{(j)} \neq 0\}$
  $x_{\mathcal{W}}^{(j)} \leftarrow \mathrm{RQIStep}(x_{\mathcal{W}}^{(j)}, \Sigma_{\mathcal{W}})$ // Rayleigh quotient update

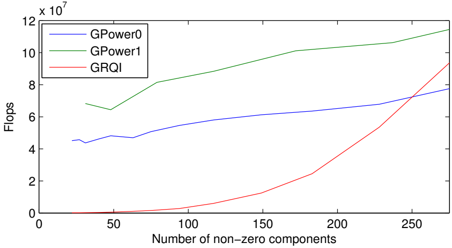
  if $j < J$ **then**
    $x^{(j)} \leftarrow \Sigma x^{(j)} / ||\Sigma x^{(j)}||_2$ // Power met. update
  end if

  $x^{(j+1)} \leftarrow \mathrm{Project}_k\,(x_{\mathrm{new}})$ // Project on $l_0 \cap l_2$ ball.
  $j \leftarrow j + 1$
until $||x^{(j)} - x^{(j-1)}|| < \epsilon$
return $x^{(j)}$

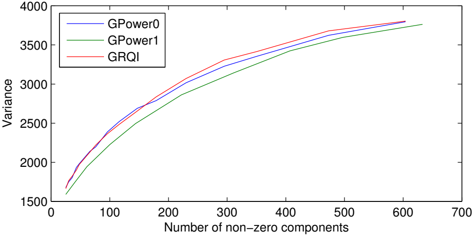
Comparison

Gen. power method	**Gen. Rayleigh quotient iter.**
Extends power method	Extends Rayleigh quotient iter.
Form of gradient descent	A second-order method
Linear convergence	Cubic convergence
$O(nk + n^2)$ flops per iter.	$O(nk + k^3)$ flops per iter.
Converges in about 100 iter.	Converges in about 10 iter.

Comparison



(a) Flops to compute eigenvector as a function of sparsity ($\mathbb{R}^{1000\times1000}$)



(b) Variance/sparsity tradeoff (random matrices in $\mathbb{R}^{1000\times1000}$)

Summary

New algorithms for sparse PCA that

- Use 10-100x fewer flops than the best current methods;
- Find sparse components which are as good or better than ones from existing algorithms;
- Generalize Rayleigh quotient iteration.

This motivates further research into second-order methods for doing matrix factorizations.