Non-Negative
Matrix Approximation

Joel A. Tropp

<jtropp@ices.utexas.edu>

Institute for Computational Engineering and Sciences

The University of Texas at Austin

Austin, TX 78712

Multi-DSP

EE 381K

30 April 2003

INTRODUCTION

- Suppose we have a collection of N non-negative, d-dimensional data. Form a $d \times N$ matrix $\boldsymbol{A}$ whose columns are the data points.
- What data? Images, document word counts, ...
- We would like to find $r \ll N$ representative vectors. We hope that non-negative linear combinations of the representatives will approximate the data well.
- Problem can be expressed as approximating

  $$\boldsymbol{A} \approx \boldsymbol{V}\boldsymbol{H},$$

  where $\boldsymbol{V}^{d \times r}$ and $\boldsymbol{H}^{r \times N}$ are non-negative.
- Quantify the error using a measure that reflects knowledge about the problem domain
- Applications: Dimensionality reduction, feature extraction, compression, sparse matrix approximation

ALTERNATING LEAST SQUARES (ALS)

Problem

- Measuring the error with the Frobenius norm yields the minimization

$$\min_{\boldsymbol{V},\boldsymbol{H}\geq 0} \|\boldsymbol{A} - \boldsymbol{V}\boldsymbol{H}\|_F.$$

Algorithm

- Make an initial guess $\boldsymbol{H}_0$.
- For $j \geq 0$, alternate between the least-squares problems

$$\boldsymbol{V}_{j+1} \in \arg\min_{\boldsymbol{V}\geq 0} \|\boldsymbol{A} - \boldsymbol{V}\boldsymbol{H}_j\|_F$$
$$\boldsymbol{H}_{j+1} \in \arg\min_{\boldsymbol{H}\geq 0} \|\boldsymbol{A} - \boldsymbol{V}_{j+1}\boldsymbol{H}\|_F.$$

- The subproblems can be solved using a finite algorithm of Lawson and Hanson (1973).
- This approach was first considered by Paatero and Tapper (1994). No meaningful convergence proofs were provided.

Convergence of ALS

- Assume that $\boldsymbol{V}_j, \boldsymbol{H}_j$ have full rank r at each step j.
- Then each of the minimizations has a unique solution.
- The minimizers are continuous functions of the fixed variable.
- It follows from a theorem of R. Meyer (1976) that
  1. $\|\boldsymbol{V}_{j+1} - \boldsymbol{V}_j\|_F \longrightarrow 0$ and
     $\|\boldsymbol{H}_{j+1} - \boldsymbol{H}_j\|_F \longrightarrow 0$.
  2. So the sequences converge or have a continuum of accumulation points.
  3. Every accumulation point is a fixed point of the algorithm.
- Weaker convergence results are possible without the rank assumption.
- Convergence may be sublinear, so hybrid algorithms make more sense.

Extensions

Other Error Metrics

- Might try to minimize the information divergence of the approximation

$$\min_{\boldsymbol{V},\boldsymbol{H}\geq 0} D(\boldsymbol{A} \,\|\, \boldsymbol{V}\boldsymbol{H})$$

- Might try to minimize another Bregman divergence
- Alternating algorithms apply; similar convergence proofs
- Solving the subproblems may be much more difficult.

Other Constraints

- Write approximation as $\boldsymbol{A} \approx \boldsymbol{V}\boldsymbol{D}\boldsymbol{H}$.
- If $\boldsymbol{V}$, $\boldsymbol{H}$ are unconstrained, ALS yields the SVD approximation.
- If $\boldsymbol{V}$, $\boldsymbol{H}$ are constrained to have 0-1 entries, problem amounts to clustering.