Matrix Methods in Data Mining and Pattern Recognition
Theory Questions

Lars Eldén
http://www.mai.liu.se/~laeld/matrix-methods/

April 13, 2007

Chapter 2. Vectors and Matrices

1. Draw sketches of the unit circle $\| x \| = 1$ in $\mathbb{R}^2$ and $\mathbb{R}^3$ for $\| \cdot \|_1$, $\| \cdot \|_2$, and $\| \cdot \|_\infty$.

Chapter 3. Linear Systems and Least Squares

Chapter 4. Orthogonality

1. (a) Write the definition of the vector norm $\|x\|_2$.

   (b) Show that if Q is an orthogonal matrix, then $\|Qx\|_2 = \|x\|_2$.

2. (a) Compute $\|x\|_2$ for

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

   (b) Let Q be an orthogonal matrix. What is the value of $\|Qx\|_2$? Motivate!

(c) Let U be an $m \times n$ matrix with $m \geq n$, and assume that the columns of U are orthogonal and normalized to have Euclidean length 1. What is the solution of the least squares problem $\min_x \|Ux - z\|_2$? Motivate your answer.

(d) Let A be a matrix of dimension $m \times n$, where $m \geq n$, with QR decomposition

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix} = Q_1 R;$$

(the latter represents the thin QR decomposition). Demonstrate how the least squares problem $\min_x \|Ax - b\|_2$ can be solved using the QR decomposition.

3. (a) Determine the rotation matrix $G \in R^{2\times 2}$ such that

$$G \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \kappa \\ 0 \end{pmatrix}.$$

(b) Let $R \in \mathbb{R}^{n\times n}$ be upper triangular and let $x \in \mathbb{R}^n$. Describe clearly and symbolically how the QR decomposition of the matrix

$$\begin{pmatrix} R \\ x^T \end{pmatrix}.$$

can be computed using n rotations. How can the matrix Q in the QR decomposition be formed?

4. (a) Construct the plane rotation G that makes the transformation

$$G \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} \kappa \\ 0 \end{pmatrix}$$

for some κ (what will that be?).

(b) Let A be a *large* matrix with the structure

$$A = \begin{pmatrix} \times & \times & \cdots & \times \\ \vdots & \vdots & & \vdots \\ \times & \times & \cdots & \times \\ 0 & \times & \cdots & \times \\ \vdots & \vdots & & \vdots \\ 0 & \times & \cdots & \times \end{pmatrix} \in \mathbb{R}^{n\times n},$$

where $\times$ denotes non-zero elements. We want to zero the bottom non-zero element in the first column, which assumed to be in row i. Explain why the following code is unsuitable, even if it is correct.

```
G = eye(n);
G(i-1:i,i-1:i)=G0;    % G0 is a 2 x 2 rotation
A = G * A;
```

Write a better code that does the job.

5. (a) Let $P = I - 2uu^T$ be a Householder transformation, and define

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Compute $\|Px\|_2$. Motivate the answer.

(b) With a Householder matrix one can transform the vector x to a unit vector,

$$Px = \kappa e_1.$$

What is the value of κ?

Chapter 5. QR Decomposition

1. Let H be a tridiagonal matrix, 6×6, say. Describe carefully, using a sequence of "$\times-$matrices", how one can compute the R factor in its QR decomposition by a sequence of rotations. What is the zero-nonzero structure of R?

2. Let A be an $m \times n$ matrix with $m \geq n$, and assume that A has full column rank.

   (a) What is the QR decomposition of A? The *thin* QR decomposition?

   (b) What are the coordinates of column a_j in A in terms of the column vectors of Q?

   (c) Derive the solution of the least squares problem $\min_x \|Ax - b\|_2$ in terms of the QR decomposition of A.

3. QR decomposition with column pivoting.

   (a) Describe (using text and figures) the first two steps of the algorithm. What is the final result?

   (b) Let the matrix A be $m \times n$ with $m \geq n$ and assume that its rank is $r < n$. What is the "shape" of R (zeros-non-zeros)?

   (c) Can the algorithm always be trusted to display the numerical rank of A?

Chapter 6. Singular Value Decomposition

1. Let A be an $m \times n$ matrix.

   (a) What is $\|A\|_F$?

   (b) Let the rank of A be equal to r, where $r < \min(m, n)$. Explain in formulas and in a figure what the SVD of A looks like.

   (c) Let $k < r$. Write the solution of the matrix approximation problem

   $$\min_{\text{rank}(B)=k} \|A - B\|_F.$$

   (d) In Matlab there are two functions for computing the SVD: `svd` and `svds`. What are the main differences in the ways in the way the matrix is used in the algorithms behind the functions? (No details are required).

2. (a) The matrix $A \in R^{m \times n}$ has rank r. Write the SVD of A in "expansion form", i.e. as a sum of rank 1 matrices.

   (b) Write the thin form of the SVD.

   (c) When we compute the tangent distance in digit classification it is quite likely that we need to find the minimum of

   $$\min_x \|Ax - b\|_2,$$

   for a matrix A that does not have full rank. Show how the norm of the residual can be computed using the SVD.

   (d) Explain the concept of "numerical rank".

3. (a) Let the matrices

$$A = (a_1 \; a_2 \; \dots a_n), \qquad B = \begin{pmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_n^T \end{pmatrix},$$

be given, where a_i are column vectors and b_i^T are row vectors. Write the matrix product AB as a sum of rank one matrices.

(b) Let A be a matrix of dimension $m \times n$, where $m \geq n$, with SVD

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T.$$

Express the column vector a_j as a linear combination of the columns of U,

$$a_j = \sum_{i=1}^{n} \beta_i u_i.$$

What are the coordinates β_i?

Chapter 11. Classification of Handwritten Digits

1. In the classification of handwritten digits we deal with bases for handwritten digits. Describe how this is done. In principle, how is the step from digits to matrices made? What is the purpose of using orthogonal bases? How are the orthogonal bases computed? How is the classification done? Why is it possible to classify unknown digits with this method? Discuss the number of basis vectors, will the results become steadily better as the number of basis vectors is increased?

Chapter 11. Text Mining

1. (6p) Given a term-document matrix A.

   (a) What is the significance of the the matrix element a_{ij}?

(b) Describe the k-means algorithm for clustering.

(c) Given a centroid matrix C, how can we compute the coordinates of the documents in terms of the centroid basis?

2. (a) Describe the basic ideas of the vector space model for text mining.

   (b) What is a dictionary? How is it related to A?

   (c) Explain the concepts *stop words* and *stemming*!

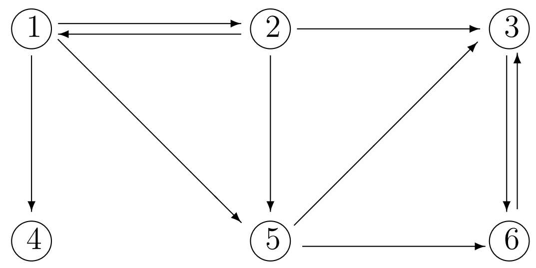
   (d) Explain the concepts "Precision" and "Recall".

   (e) What is the cosine distance measure? How can it be computed?

   (f) Describe latent semantic indexing (LSI). Explain why in some cases it can improve retrieval performance.

Chapter 12. Pagerank

1. Let the following link graph be given, and assume that we want to compute pagerank corresponding to the graph.



   (a) Write the corresponding link graph matrix.

   (b) Is the graph strongly connected? Motivate your answer.

   (c) Describe what modifications are done to such a graph for making the pagerank computation well-defined.

   (d) What is the mathematical problem that defines the pagerank vector for this graph?

2. (a) In Google's pagerank concept, a page i is considered as more important the more inlinks it has. Discuss why pagerank cannot be based on this statement alone. What is the alternative formulation?
   (b) Describe the random surfer model. How can one handle pages without outlinks? Subgraphs of the internet that have no outlinks?
   (c) What is the mathematical problem that gives the pagerank vector? Mathematical properties?
   (d) Which numerical method is used to compute the pagerank vector?

3. Explain the connection between the concepts *irreducile matrix* and *strongly connected graph*! (Give an axample of a graph that is not strongly connected and show that the cooresponding matrix is reducible).

Chapter 12. Automatic Key Word and Key Sentence Extraction

1. The assignment of saliency scores for extraction of key words and key sentences is made based on the *mutual reinforcement principle*:

   > A term should have a high saliency score if it appears in many sentences with high saliency scores. A sentence should have a high saliency score if it contains many words with high saliency scores.

   Describe how this principle can be translated to a linear algebra problem. What is the matrix? What are the vectors involved? How can one compute the saliency scores?

Chapter 15. Computing Eigenvalues and Singular Values

1. (a) Describe symbolically (use a sequence of 'x-matrices' with explaining text) how a non-symmetric matrix can be transformed to up-

per Hessenberg form.

$$A \to A_1 = P_1 A P_1^T \to A_2 = P_2 A_1 P_2^T \to \cdots \to H,$$

where P_i are Householder transformations. How many Householder transformations are needed? If A has dimension n, the computation of H requires Cn^y flops approximately, for some C. What is the value of y?

(b) Show that the eigenvalues of A and H are the same.

(c) What is the structure of H if A is symmetric?

2. Assume that T is tridiagonal.

   (a) Describe how the QR decomposition $T = QR$ can be computed using a sequence of Givens rotations. What is the structure of R?

   (b) Show that the matrix RQ is tridiagonal (cf. the QR algorithm).

3. (a) Assume that we use the power method to compute the largest eigenvalue of a matrix. If the two largest (in magnitude) are 1 and 0.9, how many iterations do we need to perform to (approximately) reduce the error in the initial approximation by factor of 10^{-3}? It is sufficient to demostrate how the number of iterations can be computed.

   (b) Describe inverse iteration (it is good if you can write a pseudocode). In which case can it be quite or even very fast?

4. (a) The computation of a QR decomposition of a matrix A, which, for simplicity, we assume is $n \times n$, requires Cn^y flops. What is y?

   (b) Order the following decompositions by the amount of work needed to compute them (for a matrix of dimension n): SVD, LU, QR.

   (c) Assume that the matrix A is large and sparse. Explain what "sparseness" means.

   (d) The Google matrix is large and sparse. Approximately how large? Explain why one cannot compute its eigenvalues by the Matlab command `eig`.

5. In the computation of pagerank one uses the power method for computing the eigenvector r_1 that corresponds to the largest eigenvalue, i.e. $Ar_1 = \lambda_1 r_1$. Assume that the second largest eigenvalue in modulus is $\lambda_2 = 0.85$. Further assume that A has dimension n.

   (a) What is λ_1?

   (b) Describe the power method. What is computed in each step? What are the memory requirements?

   (c) Let $r^{(0)}$ denote the initial approximation, and assume that $r^{(0)} = c_1 r_1 + c_2 r_2 + \cdots + c_n r_n$, where r_i are eigenvectors of A, with $c_1 \neq 0$. Give an expression for $r^{(k)}$, and explain why the error in the eigenvector approximation decays as 0.85^k for large k.

   (d) How many iterations are needed to reduce the error in the initial approximation by a factor of 10^{-6}?