Modified on Wednesday 11th November, 2020, 11:56:3511:56

Nov 9, 2020

- Factoring matrices: positive and sparse
- Review various decomposition of matrices.
  - LU-decomposition, PLU-decomposition
  - $A = PDP^{-1}$, diagonalizable
  - $A = PDP^T$, P invertible, congruence
  - $A = PDP^T$, P orthogonal, orthogonally diagonalizable

$$A = QDQ^T = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n^T,$$

  - $A = QR$, the QR-decomposition, Q with orthonormal columns, R invertible upper triangular
  - $A = U\Sigma V^T$, SVD, Singular Value Decomposition.

$$A = U\Sigma V^T = U \begin{bmatrix} \Sigma_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} V^T = U_r \Sigma_r V_r^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T.$$

  There are algorithms for decompositions. Are they sufficient?

- Assume that A has rank r. For $k \le r$, define

$$A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T.$$

  Approximating A with low rank matrices. For a given matrix A and any matrix B of rank k,

$$\|A - A_k\| \le \|A - B\|$$

  We considered various norms for matrices.

- Nonnegative matrices are matrices with nonnegative entries.
- A matrix is said to be *sparse*, if a very small portion of its entries are nonzero.
- In some applications, nonnegative matrices, sparse and nonnegative matrices are used.
- Facial Feature Extraction, Text Mining and Document Classification.
- Nonnegative matrices

$$\min \|A - UV\|_F^2 \quad \text{with } U, V \text{ nonnegative}$$

  Is there a unique solution? Is there an efficient algorithm?

- Sparse and Nonnegative matrices

$$\min \|A - UV\|_F^2 + \lambda \|UV\|_N \quad \text{with } U, V \text{ nonnegative}$$

  Is there a unique solution? Is there efficient algorithm?

- Nonnegative Matrix Factorization (NMF)

  NMF. Find *nonnegative matrices U and V* so that $A \approx UV$.

- Find *sparse low rank matrices B and C* so that $A \approx BC$.
- When each column vector of A is a facial image and $A \approx BC$, what do B and C represent?
- When each column vector of A is a document, each row vector represents a word, and $A \approx BC$, what do B and C represent?

- Vector, Matrix and Tensor: $\mathbb{R}^n$, $\mathbb{R}^{m\times n}$, $\mathbb{R}^{m\times n\times p}$, $\mathbb{R}^{m\times n\times p\times q}$
- Examples of tensors. (Tensors in other mathematics may mean different objects.)
  - Color images. A black and white image is a matrix, say of size $m\times n$. A color image, RGB image, is a tensor, say of size $m\times n\times 3$. The pixel (i,j) has three color components, (R,G,B).
  $$(i,j,1)=R,\quad (i,j,2)=G,\quad (i,j,3)=B$$
  - Derivative $\frac{\partial \mathbf{w}}{\partial A}$ of $\mathbf{w}=A\mathbf{v}$ with variables a_{ij}. Assume that $A=(a_{ij})$ is an $m\times n$ matrix and $\mathbf{v}=(v_i)$ is an $n\times 1$ column vector. We regard entries of A as variables. Let $\mathbf{w}=(w_i)$. Then the derivative of A with respect to A is a tensor $T=(T_{ijk})$ where
  $$T_{ijk}=\frac{\partial w_i}{\partial a_{jk}}=v_k\delta_{ij}.$$
  - A tensor induced from a matrix. Let $A=(a_{ij})$ be an $n\times 3$ matrix. Assume that entries of A are positive integers less than or equal to N. Define a tensor $B=(b_{ijk})$ of size $N\times N\times N$ by
  $$b_{ijk}=\text{the number of rows of } A \text{ equal to } \begin{bmatrix} i & j & k\end{bmatrix}.$$
- Tensors can be used in many applications. We can do with tensors almost everything, sometimes more, that can be done with matrices. For example, we can define rank, norm, factorization, and so on, for tensors.
- Numerical linear algebra
- Solve $A\mathbf{x}=\mathbf{b}$.
  - $\mathbf{x}=A^{-1}\mathbf{b}$
  - Gaussian elimination $A\mathbf{x}=\mathbf{b}$: $\frac{2}{3}n^3$ flops for an $n\times n$ matrix A.
  - $A=LU$ $(PA=LU)$, $L\mathbf{y}=\mathbf{b}$, $U\mathbf{x}=\mathbf{y}$
  - $A=QR$, $A^TA\mathbf{x}=A^T\mathbf{b}$, $R\mathbf{x}=Q^T\mathbf{b}$, $\hat{\mathbf{x}}=R^{-1}Q^T\mathbf{b}$
- Computing $A\mathbf{b}$ is fast, especially if A is sparse, while computing A^n is slow for large A.
- **Krylov subspaces and Arnoldi iteration**

  Consider $A\mathbf{x}=\mathbf{b}$. We may assume that A is an $m\times m$ invertible matrix. Let $\mathcal{K}_n$ denote the subspace of $\mathbb{R}^m$ spanned by a basis $\{\mathbf{b},A\mathbf{b},\ldots,A^{n-1}\mathbf{b}\}$. This subspace is called a Krylov subspace. We want to find a better basis for $\mathcal{K}_n$, in fact, an orthogonal basis $\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_n$.
- **Arnoldi iteration** (Modified Gram-Schmidt process on a Krylov subspace)
  1. $\mathbf{q}_1=\frac{\mathbf{b}}{\|\mathbf{b}\|},\mathbf{q}_2,\ldots,\mathbf{q}_k$ are known.
  2. $\mathbf{v}=A\mathbf{q}_k$
  3. for $j=1$ to k

     $h_{jk}=\mathbf{q}_j^T\mathbf{v}$

     $\mathbf{v}=\mathbf{v}-h_{jk}\mathbf{q}_j$
  4. $h_{k+1,k}=\|\mathbf{v}\|$
  5. $\mathbf{q}_{k+1}=\frac{\mathbf{v}}{h_{k+1,k}}$

From the above algorithm, we obtain

$$h_{k+1,k}\mathbf{q}_{k+1} = A\mathbf{q}_k - \sum_{j=1}^{k} h_{jk}\mathbf{q}_j,$$

$$A\mathbf{q}_k = \sum_{j=1}^{k+1} h_{jk}\mathbf{q}_j$$

In matrix notation,

$$AQ_k = A\begin{bmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_k\end{bmatrix} = \begin{bmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_k & \mathbf{q}_{k+1}\end{bmatrix}\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & h_{2k} \\ 0 & h_{32} & \ddots & \vdots \\ \vdots & \vdots & \ddots & h_{k,k} \\ 0 & 0 & \cdots & h_{k+1,k}\end{bmatrix} = Q_{k+1}H_{k+1,k}$$

Please check the above identity carefully.

$$AQ_k = Q_{k+1}H_{k+1,k}$$

$$Q_k^T AQ_k = Q_k^T Q_{k+1}H_{k+1,k} = \begin{bmatrix} I_{k\times k} & \mathbf{0}_{k\times 1}\end{bmatrix}\begin{bmatrix} H_k \\ \cdot \end{bmatrix} = H_k$$

- If $k = m$, $H_m = Q_m^T AQ_m$. H_m and A are similar. They have the same eigenvalues. If we stop at $k < m$, then eigenvalues of H_k have some relations with the eigenvalues of A.
- Linear systems by Arnoldi and GMRES(Generalized Minimum RESidual).
- $A\mathbf{x} = \mathbf{b}$: Find the vectors $\mathbf{x}_k \in \mathcal{K}_k$ solving $\min \|\mathbf{b} - A\mathbf{x}_k\|$.
- GMRES with Arnoldi's basis $\mathbf{q}_1, \ldots, \mathbf{q}_k$

  Find $\mathbf{y}_k$ solving $\min \|H_{k+1,k}\mathbf{y}_k - (\|\mathbf{b}\|, 0, \ldots, 0)^T\|$.

  Then $\mathbf{x}_k = Q_k\mathbf{y}_k$.
- Note the following justification of the above.

$$\|\mathbf{b} - A\mathbf{x}_k\| = \|\mathbf{b} - AQ_k\mathbf{y}\| = \|\mathbf{b} - Q_{k+1}H_{k+1,k}\mathbf{y}\| = \|Q_{k+1}^T\mathbf{b} - H_{k+1,k}\mathbf{y}\|,$$

$$Q_{k+1}^T\mathbf{b} = \|\mathbf{b}\|\begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix}^T$$

- When k increases, $\mathbf{x}_k$ approaches the exact solution $A^{-1}\mathbf{b}$.