Matrix Computations
Chapter 6: Singular value, Singular Vectors, and Singular Value Decomposition
Section 6.2 Matrix Norms

Jie Lu
ShanghaiTech University

Matrix Norms

Definition: A function $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ is a matrix norm if (i) $f(\mathbf{A}) \geq 0$ for all $\mathbf{A}$; (ii) $f(\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{0}$; (iii) $f(\mathbf{A} + \mathbf{B}) \leq f(\mathbf{A}) + f(\mathbf{B})$ for any $\mathbf{A}, \mathbf{B}$; (iv) $f(\alpha \mathbf{A}) = |\alpha| f(\mathbf{A})$ for any $\mathbf{A}$ and any scalar α

- For example, the Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = [\mathrm{tr}(\mathbf{A}^T \mathbf{A})]^{1/2}$ is a norm
- Induced norm or operator norm: the function

  $$f(\mathbf{A}) = \max_{\|\mathbf{x}\|_\beta \leq 1} \|\mathbf{A}\mathbf{x}\|_\alpha$$

  where $\|\cdot\|_\alpha, \|\cdot\|_\beta$ denote any vector norms
- Matrix norms induced by the vector p-norm ($p \geq 1$):

  $$\|\mathbf{A}\|_p = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = \max_{\|\mathbf{x}\|_p \leq 1} \|\mathbf{A}\mathbf{x}\|_p$$

  - $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$
  - $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$
  - $\|\mathbf{A}\|_2 = ?$

Matrix 2-Norm

The Matrix 2-norm or spectral norm is given by

$$\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A})$$

Prove this using SVD

Implication to linear systems: Let $\mathbf{y} = \mathbf{A}\mathbf{x}$ be a linear system. Under the input energy constraint $\|\mathbf{x}\|_2 \leq 1$, the system output energy $\|\mathbf{y}\|_2^2$ is maximized when $\mathbf{x}$ is chosen as the 1st right singular vector

Corollary: $\min_{\|\mathbf{x}\|_2 = 1} \|\mathbf{A}\mathbf{x}\|_2 = \sigma_{\min}(\mathbf{A})$ if $m \geq n$

Properties of Matrix 2-Norm

- $\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$
  - In fact, $\|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{B}\|_p$ for any $p \geq 1$
- $\|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{x}\|_2$
  - A special case of the first property
- $\|\mathbf{QAW}\|_2 = \|\mathbf{A}\|_2$ for any orthogonal $\mathbf{Q}, \mathbf{W}$
  - We also have $\|\mathbf{QAW}\|_F = \|\mathbf{A}\|_F$ for any orthogonal $\mathbf{Q}, \mathbf{W}$
- $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{p}\|\mathbf{A}\|_2$ (here $p = \min\{m, n\}$)

Schatten p-Norm

- The function

$$f(\mathbf{A}) = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{A})^p \right)^{1/p}, \qquad p \geq 1,$$

  is a matrix norm called the Schatten p-norm

- Nuclear norm:

$$\|\mathbf{A}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{A}) = \mathrm{tr}(\sqrt{\mathbf{A}^T\mathbf{A}})$$

  - A special case of the Schatten p-norm
  - A way to prove the nuclear norm is a matrix norm:
    - Show that $f(\mathbf{A}) = \max_{\|\mathbf{B}\|_2 \leq 1} \mathrm{tr}(\mathbf{B}^T\mathbf{A})$ is a norm
    - Show that $f(\mathbf{A}) = \sum_{i=1}^{\min\{m,n\}} \sigma_i$
  - Applications in rank approximation, e.g., for compressive sensing and matrix completion[1]

[1] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

Schatten p-Norm

- $\mathrm{rank}(\mathbf{A})$ is *nonconvex* in $\mathbf{A}$ and is arguably hard to do optimization with it
- **Idea**: The rank function can be expressed as

$$\mathrm{rank}(\mathbf{A}) = \sum_{i=1}^{\min\{m,n\}} \mathbb{1}\{\sigma_i(\mathbf{A}) \neq 0\},$$

  and we may approximate it via

$$f(\mathbf{A}) = \textstyle\sum_{i=1}^{\min\{m,n\}} \varphi(\sigma_i(\mathbf{A}))$$

  for some friendly function φ

- Using $\varphi(z) = z$, $f(\mathbf{A})$ becomes the nuclear norm

$$\|\mathbf{A}\|_* = \textstyle\sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{A})$$

  which is *convex* in $\mathbf{A}$