Beginning of Lecture 40

Singular Value Decomposition of a matrix

Suppose $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is defined as:

$$Tx = Ax, \quad \forall\, x \in \mathbb{R}^n,$$

where $A \in \mathbb{R}^{m,n}$. So in particular, $\mathcal{M}(T) = A$ in the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, and we can identify T with the matrix A. We know that the singular value decomposition of T is:

$$Tx = \sum_{k=1}^{r} \sigma_k \langle x, e_k \rangle f_k$$

where $\sigma_1, \ldots, \sigma_1$ are the nonzero singular values of T, $e_1, \ldots, e_r \in \mathbb{R}^n$ are orthonormal, and $f_1, \ldots, f_r \in \mathbb{R}^m$ are orthonormal. We can rewrite the SVD in terms of matrices, which gives the SVD decomposition of the matrix A. Consider any $x \in \mathbb{R}^n$ as an $n \times 1$ vector (similarly $y \in \mathbb{R}^m$ is an $m \times 1$ vector). Define:

$$\begin{aligned}
\widetilde{\Sigma} &= \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r,r} \\
\widetilde{B} &= (e_1, \ldots, e_r) \in \mathbb{R}^{n,r} \\
\widetilde{C} &= (f_1, \ldots, f_r) \in \mathbb{R}^{m,r}
\end{aligned}$$

Then:

$$Ax = Tx = \sum_{k=1}^{r} \sigma_k \langle x, e_k \rangle f_k \Longleftrightarrow A = \widetilde{C}\widetilde{\Sigma}\widetilde{B}^\dagger$$

The representation $A = \widetilde{C}\widetilde{\Sigma}\widetilde{B}^\dagger$ is called the <u>compact SVD</u> for A.

We can also compute a (standard) SVD representation of A as:

$$A = C\Sigma B^\dagger$$

where $\Sigma \in \mathbb{R}^{m,n}$, $B \in \mathbb{R}^{n,n}$ and $C \in \mathbb{R}^{m,m}$ with $S \in \mathcal{L}(\mathbb{R}^n)$, $Sx := Bx$ and $R \in \mathcal{L}(\mathbb{R}^m)$, $Ry := Cy$, both being isometries. The matrix Σ is simply the "diagonal" extension of $\widetilde{\Sigma}$:

$$\Sigma_{j,k} = \begin{cases} \sigma_k & j = k \leq r \\ 0 & \text{otherwise} \end{cases}$$

To define B, extend $e_1, \ldots, e_r$ to an ONB $e_1, \ldots, e_n$ for $\mathbb{R}^n$. Define B as:

$$B = (e_1, \ldots, e_n) \in \mathbb{R}^{n,n}$$

Similarly, extend $f_1, \ldots, f_r$ to an ONB $f_1, \ldots, f_m$ for $\mathbb{R}^m$ and define C as:

$$C = (f_1, \ldots, f_m) \in \mathbb{R}^{m,m}$$

Note that by definition the columns of B and C or orthonormal bases, and $\mathcal{M}(S) = B$ and $\mathcal{M}(R) = C$, so by Theorem 7.42 in the book, S and R are isometries.

Application 3: Condition number of a matrix

Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined as:

$$Tx = Ax, \quad \forall\, x \in \mathbb{R}^n,$$

where $A \in \mathbb{R}^{n,n}$. So in particular, $\mathcal{M}(T) = A$ in the standard basis, and we can identify T with the matrix A. Suppose additionally that A (and hence T) is invertible. Now suppose that we want to solve:

$$Ax = b$$

for some $b \in \mathbb{R}^n$. The solution is clearly:

$$x = A^{-1}b$$

However, as happens in "real life", we may only know the data approximately or round off errors during computations on a computer may occur, which distort the data. We consider a model in which b is only approximately known, so instead of having $Ax = b$ we are solving:

$$A\tilde{x} = b + \Delta b,$$

where Δb is a small perturbation of b, i.e.,

$$\|\Delta b\| < \epsilon \|b\|, \quad \epsilon \ll 1$$

The solution $\tilde{x}$ is approximately x; indeed:

$$\tilde{x} = A^{-1}b + A^{-1}\Delta b = x + \Delta x, \quad \text{where } \Delta x = A^{-1}\Delta b$$

We want to know how big the relative error $\|\Delta x\|/\|x\|$ in the solution $\tilde{x}$ is in comparison with the relative error $\|\Delta b\|/\|b\|$ of the initial data. Note that:

$$\begin{aligned}
\frac{\|\Delta x\|}{\|x\|} &= \frac{\|A^{-1}\Delta b\|}{\|x\|} \\
&= \frac{\|A^{-1}\Delta b\|}{\|b\|}\frac{\|b\|}{\|x\|} \\
&= \frac{\|A^{-1}\Delta b\|}{\|b\|}\frac{\|Ax\|}{\|x\|} \\
&\leq \frac{\|A^{-1}\|\|\Delta b\|}{\|b\|}\frac{\|A\|\|x\|}{\|x\|} \\
&\leq \|A\|\|A^{-1}\|\frac{\|\Delta b\|}{\|b\|}
\end{aligned}$$

The quantity $\|A\|\|A^{-1}\|$ is the condition number of A. It estimates how the relative error in the solution x depends on the relative error of the initial data b.

We can relate the condition number of A to its singular values. Let $\sigma_1, \ldots, \sigma_n$ be the singular values of A. Assume they are ordered so that:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$$

Note that $\sigma_n > 0$ since A is invertible and:

$$A = C\Sigma B^\dagger$$

where B and C are isometries and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$. Thus B and C are invertible with

$$B^{-1} = B^\dagger, \;\; C^{-1} = C^\dagger \quad [\text{recall } \mathcal{M}(T^*) = \mathcal{M}(T)^\dagger]$$

Thus $\Sigma = C^{-1}A(B^\dagger)^{-1} = C^\dagger AB$ must also be invertible and:

$$A^{-1} = (C\Sigma B^\dagger)^{-1} = (B^\dagger)^{-1}\Sigma^{-1}C^{-1} = B\Sigma^{-1}C^\dagger$$

Note that $\Sigma^{-1} = \mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_n)$ and so the singular values of A^{-1} are

$$\frac{1}{\sigma_n} \geq \frac{1}{\sigma_{n-1}} \geq \cdots \geq \frac{1}{\sigma_1} > 0$$

We know that $\|A\| = \sigma_1$ and by the calculation we just completed $\|A^{-1}\| = 1/\sigma_n$. Therefore the condition number of A is:

$$\|A\|\|A^{-1}\| = \frac{\sigma_1}{\sigma_n}$$

A matrix is well conditioned if its condition number is not too large (the closer to one the better) or ill conditioned if its condition number is too large.

End of Lecture 40