3.2. **Other norms for matrices.**

3.2.1. *The Frobenius norm (the Hilbert-Schmidt norm).* This is defined as

$$\|M\|_F = \sqrt{Tr(M^*M)} = \sqrt{\sum_{k=1}^{r} \sigma_k^2}$$

It is the same as the Euclidian norm when M is viewed as a vector with mn components:

$$\|M\|_F = \sqrt{\sum_{i,j} |M_{ij}|^2}$$

(see also (13)).

The Frobenius norm is easier to calculate than the operator norm, and it is invariant under unitary transformations (i.e. under changes of orthonormal bases), since $\|M\|_F = \|UMV^*\|_F$ if U, V are unitary (because the matrices M and UMV^* have the same singular values).

The Frobenius norm is compatible to matrix multiplication, as relation (12) can be checked by direct calculation:

$$\|MN\|_F^2 = \sum_{ij} |(MN)_{ij}|^2 = \sum_{ij} \Big|\sum_{k} M_{ik}N_{kj}\Big|^2$$

and, using the Cauchy-Schwartz inequality,

$$\le \sum_{ij} (\sum_{k} |M_{ik}|^2)(\sum_{l} |N_{lj}|^2) = \|M\|_F^2 \, \|N\|_F^2$$

Lower rank approximations. Suppose that M is an $m \times n$ matrix of rank r, with singular values $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_r$ and singular value decomposition $M = U\Sigma V^*$. Then among all $m \times n$ matrices of lower rank $k \le r$ the best approximation is $X_k = U\Sigma_k V^*$ where Σ_k is the diagonal matrix with singular values $\sigma_1, \sigma_2, \ldots, \sigma_k$, in the sense that

$$\|M - X_k\|_F = \min\{\|M - X\|_F \,;\, X \in \mathcal{M}_{m,n},\ \operatorname{rank} X = k\}$$

(this theorem is due to Eckart and Young, 1936).

Proof. Since $\|M - X\|_F = \|U\Sigma V^* - X\|_F = \|\Sigma - U^*XV\|_F$ Denoting $N = U^*XV$, an $m \times n$ matrix of rank k, a direct calculation gives

$$\|\Sigma - N\|_F^2 = \sum_{i,j} |\Sigma_{i,j} - N_{i,j}|^2 = \sum_{i=1}^{r} |\sigma_i - N_{ii}|^2 + \sum_{i>r} |N_{ii}|^2 + \sum_{i\ne j} |N_{i,j}|^2$$

which is minimal when all the non diagonal terms of N equal to zero, and so are all N_{ii} for $i > r$. Finally, the minimum of $\sum_{i=1}^{r} |\sigma_i - N_{ii}|^2$ among all $N_{11}, \ldots, N_{rr}$ of which exactly k are nonzero is attained when $N_{ii} = \sigma_i$ for $i = 1, \ldots, , k$ and all other N_{ii} are zero. □

It is worth noting that the "complexity" of X_k can be further reduced as we can use lower dimensional matrices for its calculation, since $X_k =$

$U_k D_k V_k^*$ where is the $k \times k$ diagonal matrix with singular values $\sigma_1, \sigma_2, \ldots, \sigma_k$, and U_k and V_k are the first k columns of the matrices U, respectively V.

Indeed, splitting in blocks $U = [U_k, U']$ and $V = [V_k, V']$ we calculate

$$X_k = U \Sigma_k V^* = [U_k, U'] \begin{bmatrix} D_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_k^* \\ V'^* \end{bmatrix} = [U_k, U'] \begin{bmatrix} D_k V_k^* \\ 0 \end{bmatrix} = U_k D_k V_k^*$$