# 0 Essentials

## Matrix/Vector

**Vectors:** Unit vector: $u^\top u = 1$ Orthogonal vectors: $u^\top v = 0$ **Range, Kernel, Nullity:** $range(\mathbf{A}) = \{\mathbf{z} | \exists \mathbf{x} : \mathbf{z} = \mathbf{A}\mathbf{x}\} = span(\text{columns of } A)$ $rank(\mathbf{A}) = dim(range(\mathbf{A}))$ $kernel(A) = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}$ (spans nullspace) $nullity(\mathbf{A}) = dim(kernel(\mathbf{A}))$
**Ranks:** $rank(XY) \leq rank(X) \forall X \in R^{mxn}, Y \in R^{nxk}$ eq. if $Y \in R^{nxn}, rank(Y) = n$ **Rank-nullity Theorem:** $dim(kernel(\mathbf{A})) + dim(range(\mathbf{A})) = n$
**Orthogonal mat.** $\mathbf{A}^{-1} = \mathbf{A}^\top$, $\mathbf{A}\mathbf{A}^\top = \mathbf{A}^\top\mathbf{A} = \mathbf{I}$, $\det(\mathbf{A}) \in \{+1, -1\}$, $\det(\mathbf{A}^\top\mathbf{A}) = 1$, preserves inner product, norm, distance, angle, rank, matrix orthogonality **Outer Product:** $\mathbf{u}\mathbf{v}^\top$, $(\mathbf{u}\mathbf{v}^\top)_{i,j} = \mathbf{u}_i\mathbf{v}_j$
**Inner Product:** $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top\mathbf{y} = \sum_{i=1}^N \mathbf{x}_i\mathbf{y}_i$. $\langle \mathbf{x} \pm \mathbf{y}, \mathbf{x} \pm \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle \pm 2\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle$
$(\mathbf{u}_i^T\mathbf{v}_j)\mathbf{v}_j = (\mathbf{v}_j\mathbf{v}_j^T)\mathbf{u}_i$ **Cross product:** $\vec{a} \times \vec{b} = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)^\top$
**Trace:** $trace(\mathbf{XYZ}) = trace(\mathbf{ZXY})$
**Transpose:** $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$, $(\mathbf{AB})^\top = \mathbf{B}^\top\mathbf{A}^\top$, $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$
**Cauchy-Schwarz inequality:** $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$
**Jensen inequality:** for convex function $f$, non negative $\lambda_i$ st. $\sum_{i=1}^n \lambda_i = 1$: $f(\sum_{i=1}^n \lambda_i x_i) \leq \sum_{i=1}^n \lambda_i f(x_i)$ Note: for concave, inequality sign switches **Convexity:** $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y), \forall \theta \in [0,1]$ **Least Squares equations:** $\arg\min_{\beta \in \mathbb{R}^{p+1}} \|\mathbf{y} - \mathbf{X}\beta\|^2$, $\hat{\beta} = (X^\top X)^{-1} X^\top y$
**Einstein matrix notation:** $(A \cdot B)_{ij} = \sum_{k=1}^n A_{ik} \cdot B_{kj}$
**Kullback-Leibler:** $KL(P||Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$

## Norms

- $\|\mathbf{x}\|_0 = |\{i | x_i \neq 0\}|$
- $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^N \mathbf{x}_i^2} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$
- $\|\mathbf{u} - \mathbf{v}\|_2 = \sqrt{(\mathbf{u} - \mathbf{v})^\top(\mathbf{u} - \mathbf{v})}$
- $\|\mathbf{x}\|_p = \left(\sum_{i=1}^N |x_i|^p\right)^{\frac{1}{p}}$ ; $\|\mathbf{x}\|_\infty = \max_{i=1,\ldots,n} |x_i|$
- $\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n \mathbf{m}_{i,j}^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2} = \|\sigma(\mathbf{A})\|_2 = \sqrt{trace(\mathbf{M}^T\mathbf{M})}$
- $\|\mathbf{M}\|_G = \sqrt{\sum_{ij} g_{ij} x_{ij}^2}$ (weighted Frobenius)
- $\|\mathbf{M}\|_1 = \sum_{i,j} |m_{i,j}|$
- $\|\mathbf{M}\|_2 = \sigma_{\max}(\mathbf{M}) = \|\sigma((M))\|_\infty$ (spectral)
- $\|\mathbf{M}\|_p = \max_{\mathbf{v} \neq 0} \frac{\|\mathbf{M}\mathbf{v}\|_p}{\|\mathbf{v}\|_p}$
- $\|\mathbf{M}\|_\star = \sum_{i=1}^{\min(m,n)} \sigma_i = \|\sigma(\mathbf{A})\|_1$ (nuclear)

## Derivatives

$\frac{\partial}{\partial \mathbf{x}}(\mathbf{b}^\top\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top\mathbf{b}) = \mathbf{b}$ $\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top\mathbf{x}) = 2\mathbf{x}$
$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top\mathbf{A}\mathbf{x}) = (\mathbf{A}^\top + \mathbf{A})\mathbf{x}$ $\frac{\partial}{\partial \mathbf{x}}(\mathbf{b}^\top\mathbf{A}\mathbf{x}) = \mathbf{A}^\top\mathbf{b}$
$\frac{\partial}{\partial \mathbf{X}}(\mathbf{c}^\top\mathbf{X}\mathbf{b}) = \mathbf{c}\mathbf{b}^\top$ $\frac{\partial}{\partial \mathbf{X}}(\mathbf{c}^\top\mathbf{X}^\top\mathbf{b}) = \mathbf{b}\mathbf{c}^\top$
$\frac{\partial}{\partial \mathbf{x}}(\|\mathbf{x} - \mathbf{b}\|_2) = \frac{\mathbf{x} - \mathbf{b}}{\|\mathbf{x} - \mathbf{b}\|_2}$ $\frac{\partial}{\partial \mathbf{x}}(\|\mathbf{x}\|_2^2) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top\mathbf{x}) = 2\mathbf{x}$
$\frac{\partial}{\partial \mathbf{X}}(\|\mathbf{X}\|_F^2) = 2\mathbf{X}$ $\frac{\partial}{\partial \mathbf{x}} \log(x) = \frac{1}{x}$

## Eigendecomposition

$\mathbf{A} \in \mathbb{R}^{N \times N}$ then $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$ with $\mathbf{Q} \in \mathbb{R}^{N \times N}$.
if fullrank: $\mathbf{A}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^{-1}$ and $(\mathbf{\Lambda}^{-1})_{i,i} = \frac{1}{\lambda_i}$.
if $\mathbf{A}$ symmetric: $A = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$ ($\mathbf{Q}$ orthogonal). Eigenvalue $\lambda$: solve $det(A - \lambda I) = 0$ Eigenvector $v$: solve $(A - \lambda I) * v = \vec{0}$

## Probability / Statistics

- $P(x) := Pr[X = x] := \sum_{y \in Y} P(x, y)$ • $P(x|y) := Pr[X = x | Y = y] := \frac{P(x,y)}{P(y)}$, if $P(y) > 0$ • $\forall y \in Y : \sum_{x \in X} P(x|y) = 1$ (property for any fixed $y$) • $P(x, y) = P(x|y)P(y)$ • posterior$P(A|B) = \frac{\text{prior}P(A) \times \text{likelihood}P(B|A)}{\text{evidence}P(B)}$ (Bayes' rule) • $P(x|y) = P(x) \Leftrightarrow P(y|x) = P(y)$ (iff $X$, $Y$ independent)
- $P(x_1, \ldots, x_n) = \prod_{i=1}^n P(x_i)$ (iff IID) • Variance $Var[X] := E[(X - \mu_x)^2] := \sum_{x \in X}(x - \mu_x)^2 P(x) = E(X^2) - E(X)^2$ $\text{Var}(aX) = a^2\,\text{Var}(X)$ • expectation $\mu_x := E[X] := \sum_{x \in X} xP(x)$ • $E[X + Y] = E[X] + E[Y]$ • standard deviation $\sigma_x := \sqrt{Var[X]}$

## Lagrangian Multipliers

Minimize $f(\mathbf{x})$ s.t. $g_i(\mathbf{x}) \leq 0, \ i = 1, .., m$ (**inequality constr.**) and $h_i(\mathbf{x}) = \mathbf{a}_i^\top\mathbf{x} - b_i = 0$ or $h_i(\mathbf{x}) = \sum_w x_{w,i} - b_i = 0, \ i = 1, .., p$ (**equality constraint**)
$L(\mathbf{x}, \alpha, \beta) := f(\mathbf{x}) + \sum_{i=1}^m \alpha_i g_i(\mathbf{x}) + \sum_{i=1}^p \beta_i h_i(\mathbf{x})$

# 1 Principal Component Analysis

$\mathbf{X} \in \mathbb{R}^{D \times N}$. $N$ observations, $K$ rank.
1. Empirical Mean: $\overline{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^N \mathbf{x}_n$.
2. Center Data: $\overline{\mathbf{X}} = \mathbf{X} - [\overline{\mathbf{x}}, \ldots, \overline{\mathbf{x}}] = \mathbf{X} - \mathbf{M}$.
3. Cov.: $\Sigma = \frac{1}{N}\sum_{n=1}^N (\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^\top = \frac{1}{N}\overline{\mathbf{X}}\overline{\mathbf{X}}^\top$.
4. Eigenvalue Decomposition: $\Sigma = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$.
5. Select $K < D$, only keep $\mathbf{U}_K, \lambda_K$.
6. Transform data onto new Basis: $\overline{\mathbf{Z}}_K = \mathbf{U}_K^\top\overline{\mathbf{X}}$.
7. Reconstruct to original Basis: $\tilde{\overline{\mathbf{X}}} = \mathbf{U}_k\overline{\mathbf{Z}}_K$.
8. Reverse centering: $\tilde{\mathbf{X}} = \tilde{\overline{\mathbf{X}}} + \mathbf{M}$.

For compression save $\mathbf{U}_k, \overline{\mathbf{Z}}_K, \overline{\mathbf{x}}$.
$\mathbf{U}_k \in \mathbb{R}^{D \times K}, \Sigma \in \mathbb{R}^{D \times D}, \overline{\mathbf{Z}}_K \in \mathbb{R}^{K \times N}, \overline{\mathbf{X}} \in \mathbb{R}^{D \times N}$
**Calculation of**: $var(X) = \frac{1}{N}\sum_{n=1}^N (X_i - \overline{X})^2$

## Iterative View

Residual $r_i$: $x_i - \tilde{x}_i = I - uu^T x_i$
Cov of $r$: $\frac{1}{n}\sum_{i=1}^n (I - uu^T)x_i x_i^T (I - uu^T)^T = (I - uu^T)\Sigma(I - uu^T)^T = \Sigma - 2\Sigma uu^T + uu^T\Sigma uu^T = \Sigma - \lambda uu^T$
1. Find principal eigenvector of $(\Sigma - \lambda uu^T)$
2. which is the second eigenvector of $\Sigma$
3. iterating to get $d$ principal eigenvector of $\Sigma$

## Power Method

Power iteration: $v_{t+1} = \frac{Av_t}{||Av_t||}$, $\lim_{t \to \infty} v_t = u_1$
Assuming $\langle u_1, v_0 \rangle \neq 0$ and $|\lambda_1| > |\lambda_j| (\forall j \geq 2)$

## Reconstruction Proof Sketch

Given: $\tilde{X} = U_K U_K^\top \overline{X}$ To prove: squared reconstruction error is the sum of the lowest $D - K$ eigenvalues of $\Sigma$. $err = 1/N\sum_{i=1}^N \|\tilde{x}_i - \overline{x_i}\|_2^2 = 1/N\|\tilde{X} - \overline{X}\|_F^2 = 1/N\|(U_K U_K^\top - I_d)\overline{X}\|_F^2 = 1/N * trace((U_K U_K^\top - I_d)\overline{X}\overline{X}^\top(U_K U_K^\top - I_d)^\top = 1/N * trace(([U_K; 0] - U)\Lambda([U_K; 0] - U)^\top)$