EECS208 Discussion 1

Simon Zhai

Sep. 3

Reading: Chapters 1, 2, Appendix A, and B of *High-Dim Data Analysis with Low-Dim Models*.

1 Demos

See the demos in recordings: 1) ℓ^p-ball Illustration; 2) ℓ^0-norm Recovery; 3) ℓ^1-norm Minimization.

2 Linear Algebra

Reading: Appendix A.1.

2.1 Vector Space

Definition 2.1 (Vector Space (Definition A.1)) *A vector space $\mathbb{V}$ over a field of scalars $\mathbb{F}$ is a set $\mathbb{V}$ (with a distinguished zero element $\mathbf{0} \in \mathbb{V}$) endowed with two operations:*

- **vector addition** $+$, *which takes two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{V}$ and produce another vector $\boldsymbol{v} + \boldsymbol{w} \in \mathbb{V}$,*
- **scalar multiplication** $\cdot$ *(sometimes omitted), which takes a vector $\boldsymbol{v} \in \mathbb{V}$ and a scalar $\alpha \in \mathbb{F}$ and produce a vector $\alpha\boldsymbol{v} \in \mathbb{V}$,*

such that:

1. *the addition $+$ is associative $\boldsymbol{v} + (\boldsymbol{w} + \boldsymbol{x}) = (\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{x}$;*
2. *the addition $+$ is commutative: $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$;*
3. *zero is the additive identity: $\boldsymbol{v} + \mathbf{0} = \boldsymbol{v}$;*
4. *every element has an additive inverse: $\forall \boldsymbol{v} \in \mathbb{V}, \exists -\boldsymbol{v}$, such that $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}$;*
5. $\forall \alpha, \beta \in \mathbb{F}$, $\alpha(\beta\boldsymbol{v}) = (\alpha\beta)\boldsymbol{v}$
6. *multiplicative identity: $\exists 1 \in \mathbb{F}$, such that $1\boldsymbol{v} = \boldsymbol{v}$;*
7. $\alpha(\boldsymbol{v} + \boldsymbol{w}) = \alpha\boldsymbol{v} + \alpha\boldsymbol{w}$;
8. $(\alpha + \beta)\boldsymbol{v} = \alpha\boldsymbol{v} + \beta\boldsymbol{v}$.

Examples of Vector Space. 1) $\mathbb{V} = \mathbb{R}^n, \mathbb{F} = \mathbb{R}$; 2) $\mathbb{V} = \mathbb{C}^n, \mathbb{F} = \mathbb{C}$; and some other vector spaces.

2.2 Inner Product, Norm, and Orthogonal Complement

Definition 2.2 (Inner Product (Definition A.8)) *A function* $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \mapsto \mathbb{F}$ *is an inner prod if it satisfies:*

- **linearity**: $\langle \alpha \boldsymbol{v} + \beta \boldsymbol{w}, x \rangle = \alpha \langle v, x \rangle + \beta \langle w, x \rangle$;
- **conjugate symmetry**: $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \overline{\langle \boldsymbol{w}, \boldsymbol{v} \rangle}$;
- **positive definiteness**: $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$, *equality holds if and only if* $\boldsymbol{v} = \boldsymbol{0}$.

Definition 2.3 (Norm (Definition 2.1)) *A norm on a vector space* $\mathbb{V}$ *over* $\mathbb{R}$ *is a function* $\|\cdot\| : \mathbb{V} \mapsto \mathbb{R}$ *that is*

- *nonnegatively homogeneous:* $\|\alpha \boldsymbol{x}\| = \alpha \|\boldsymbol{x}\|$ *for all vectors* $\boldsymbol{x} \in \mathbb{V}$, *scalars* $\alpha \in \mathbb{R}$,
- *positive definite:*$\|\boldsymbol{x}\| \geq 0$, *and* $\|\boldsymbol{x}\| = 0$*if and only if* $\boldsymbol{x} = \boldsymbol{0}$,
- *subadditive:* $\|\cdot\|$ *satisfies the triangle inequality* $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{V}$.

Norm Induced by Inner Product.

- Vector ℓ^2-norm: $\|\boldsymbol{x}\|_2^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \sum_{i=1}^n x_i^2, \forall x \in \mathbb{R}^n$.
- Matrix Frobenius norm: $\|\boldsymbol{X}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^m x_{i,j}^2 = \langle \boldsymbol{X}, \boldsymbol{X} \rangle_F = \mathsf{trace}(\boldsymbol{X}^\top \boldsymbol{X}), \forall \boldsymbol{X} \in \mathbb{R}^{m \times n}$.

Definition 2.4 (Orthogonal Complement) *For* $\mathsf{S} \subseteq \mathbb{V}$

$$\mathsf{S}^\perp = \{\boldsymbol{v} \in \mathbb{V} | \langle \boldsymbol{v}, \boldsymbol{s} \rangle, \forall \boldsymbol{s} \in \mathsf{S}\}. \tag{2.1}$$

Definition 2.5 (Range and Null Space of a Matrix (Equation A.5.3 - A.5.5))

$$\begin{aligned} \mathsf{null}(\boldsymbol{A}) &= \{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\} \\ \mathsf{range}(\boldsymbol{A}) &= \{\boldsymbol{A}\boldsymbol{x} | \boldsymbol{x} \in \mathbb{R}^n\} = \mathsf{col}(\boldsymbol{A}) \\ \mathsf{row}(\boldsymbol{A}) &= \{\boldsymbol{w}^\top \boldsymbol{A} | \boldsymbol{w} \in \mathbb{R}^m\}. \end{aligned} \tag{2.2}$$

2.3 SVD

Definition 2.6 (Compact SVD (Theorem A.34)) *Let* $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$ *be a matrix and* $r = \mathsf{rank}(\boldsymbol{X})$. *Then there exist* $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ *with ordering* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, *and matrices* $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}, \boldsymbol{V} \in \mathbb{R}^{n_2 \times r}$, *such that* $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}_r$ *and*

$$\boldsymbol{X} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^\top = \sum_{i=1}^r \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top. \tag{2.3}$$

Definition 2.7 (Full SVD (Theorem A.36)) *Let* $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$ *be a matrix. Then there exist orthogonal matrices* $\boldsymbol{U} \in \mathsf{O}(n_1)$ *and* $\boldsymbol{V} \in \mathsf{O}(n_2)$ *and scalars with ordering* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{n_1, n_2\}} > 0$, *such that if we let* $\boldsymbol{\Sigma} \in \mathbb{R}^{n_1 \times n_2}$ *with* $\Sigma_{ii} = \sigma_i$ *and* $\Sigma_{ij} = 0, \forall i \neq j$, *we have*

$$\boldsymbol{X} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^\top. \tag{2.4}$$

3 Statistics

Definition 3.1 (Mean, Variance) *Given a random variable* $x \in \mathcal{X}$ *with probability density function* $p(x)$, *the mean (expectation) is defined as*

$$\mathbb{E}x = \int_{x \in \mathcal{X}} x p(x) dx \tag{3.1}$$

and the variance is defined as

$$\mathsf{Var}(x) = \mathbb{E}(x - \mathbb{E}x)^2. \tag{3.2}$$

Definition 3.2 (Covariance and Covariance of a Vector) *Given two random variable $x \in \mathcal{X}, y \in \mathcal{Y}$, the covariance of (x, y) defined on the joint distribution $\mathcal{X} \times \mathcal{Y}$ is defined as*

$$\mathsf{Cov}(x, y) = \mathbb{E}[(x - \mathbb{E}x)(y - \mathbb{E}y)]. \tag{3.3}$$

For a vector $\boldsymbol{x} \in \mathbb{R}^n$, the covariance matrix of $\boldsymbol{x}$ is defined as

$$\begin{bmatrix} \mathsf{Cov}(x_1, x_1) & \mathsf{Cov}(x_1, x_2) & \dots & \mathsf{Cov}(x_1, x_n) \\ \mathsf{Cov}(x_2, x_1) & \mathsf{Cov}(x_2, x_2) & \dots & \mathsf{Cov}(x_2, x_n) \\ \\ \mathsf{Cov}(x_n, x_1) & \mathsf{Cov}(x_n, x_2) & \dots & \mathsf{Cov}(x_n, x_n) \end{bmatrix} \tag{3.4}$$

Example 3.3 (Mean and Covariance of Gaussian Vectors) *Suppose we have a Gaussian vector $\boldsymbol{x} \in \mathbb{R}$ satisfying $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, suppose $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, then the vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ satisfies $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top)$.*