Example – Sparse projection

– Find: $\quad \widehat{x} \in \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \quad \|x - y\|_2^2$ $\quad$ for given $y \in \mathbb{R}^p$

$\qquad\qquad$ s.t. $\quad \|x\|_0 \le k < p$

$Q \subset R^n$. means subset

Suppose that $f : X \to \mathbf{R}$ is a real-valued function whose domain is an arbitrary set X. The **set-theoretic support** of f, written **supp(f)**, is the set of points in X where f is non-zero

$$\operatorname{supp}(f) = \{x \in X \mid f(x) \neq 0\}.$$

– Step 1: Assume we know $\mathcal{S} := \operatorname{supp}(\widehat{x}) \subset \{1, \ldots, p\}$

$$\widehat{x}_{\mathcal{S}} = y_{\mathcal{S}}$$

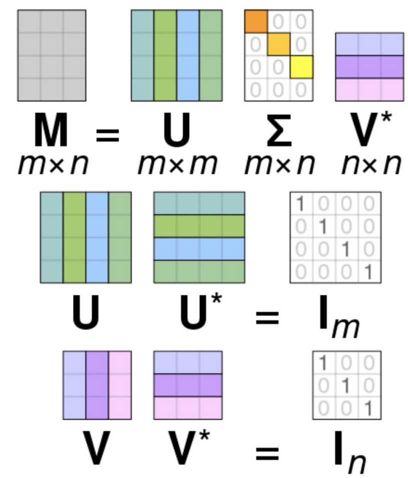
– Step 2: Transform original problem

$$\mathcal{S} \in \underset{\operatorname{card}(\mathcal{S}) \le k}{\operatorname{argmin}} \quad \|y_{\mathcal{S}} - y\|_2^2 \quad \equiv \quad \mathcal{S} \in \underset{\operatorname{card}(\mathcal{S}) \le k}{\operatorname{argmax}} \quad \|y_{\mathcal{S}}\|_2^2 := \sum_{i \in \mathcal{S}} y_i^2$$

Why?

– Step 3: Find efficient ways to solve the last part!

$\Sigma = \mathtt{diag}(\sigma_1, \ldots, \sigma_r)$ contains singular values where $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r$



Can we consider SVD as a special way to calculate eigen-values from those not square matrices?

orthogonal: $U^\top U = I$ and $V^\top V = I$

– Rank of a matrix: maximum # of independent columns or rows

Definition: If A is a square $n \times n$ matrix, then the **Trace of** A denoted $\operatorname{tr}(A)$ is the sum of all of the entries in the main diagonal, that is $tr(A) = \sum_{i=1}^{n} a_{ii}$. If A is not a square matrix, then the trace of A is undefined.

$$\langle A, B \rangle = \operatorname{Tr}(A^\top B) = \operatorname{Tr}(B^\top A), \forall A, B \in \mathbb{R}^{m \times n}$$

Positive semi-definite matrices: $\quad A \succeq 0$

1. $A \in \mathbb{R}^{n \times n}$
2. A is symmetric
3. $x^\top A x \ge 0, \; \forall x \in \mathbb{R}^n, \; x \neq 0$

$$\|A\|_F = \sqrt{\sum_{ij} A_{ij}^2} \qquad \|A\|_* = \sum_i^r \sigma_i \qquad \|A\|_2 = \max_i \sigma_i$$

(Frobenius norm) $\qquad$ (Nuclear norm) $\qquad$ (Spectral norm)

To denote an estimator, e.g., $\hat{\mu} = m$ for the sample mean.

$$f(x) \triangleq \|y - Ax\|_2^2$$

What does this triangle mean?

Example – Low-rank projection

– Find: $\quad \widehat{X} \in \underset{X}{\min} \quad \frac{1}{2}\|X - Y\|_F^2$ $\qquad$ for given $Y \in \mathbb{R}^{m \times n}$

$\qquad\qquad$ s.t. $\quad \operatorname{rank}(X) \le r$

optimal solution $x^\star$ has smallest value of f_0 among all vectors that satisfy the constraints

– Step 1: Observe $\|X - Y\|_F^2 = \|X - U\Sigma V^\top\|_F^2$

$$= \|U^\top X V - \Sigma\|_F^2$$

Why?

$$x = \theta_1 x_1 + \theta_2 x_2$$

with $\theta_1 \ge 0, \; \theta_2 \ge 0$

– Step 2: Transform original problem

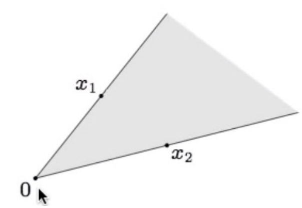
$$\underset{X}{\min} \quad \tfrac{1}{2}\|C - \Sigma\|_F^2 \qquad \text{where} \qquad C = U^\top X V$$

$\qquad$ s.t. $\quad \operatorname{rank}(C) \le r$

– Step 3: Optimal C is diagonal with singular values from Σ

How does this relate to sparse projection?

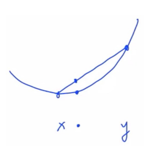
Convex cone



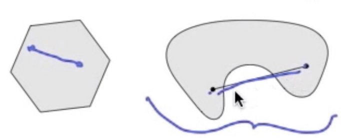
The definition of convex function:

A function $f : \mathcal{X} \to \mathbb{R}$ is said to be convex if it always lies below its chords, that is

$$\forall (x, y, \gamma) \in \mathcal{X} \times \mathcal{X} \times [0, 1], \; f((1-\gamma)x + \gamma y) \le (1-\gamma) f(x) + \gamma f(y).$$



examples (one convex, two nonconvex sets)



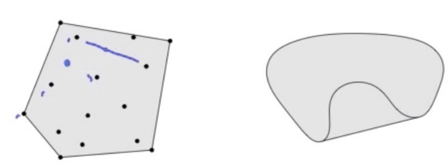
convex combination of $x_1, \ldots, x_k$: any point x of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

with $\theta_1 + \cdots + \theta_k = 1, \; \theta_i \ge 0$

Combination don't need to be unique

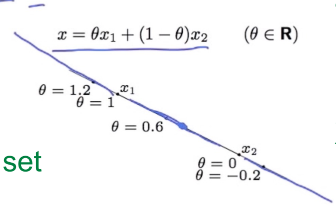
convex hull $\operatorname{conv} S$: set of all convex combinations of points in S



- hyperplanes are affine and convex; halfspaces are convex

line through x_1, x_2: all points

$$x = \theta x_1 + (1 - \theta) x_2 \qquad (\theta \in \mathbf{R})$$



Difference between Affine set and convex set, it theta in affine don't Need to be within 0 and 1, all the points On that line is okay, but in convex, theta Need to be between 0 and 1, i.e., between X1 and x2,

Affine set

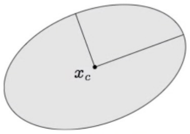
(Euclidean) ball with center x_c and radius r:

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \le r\} = \{x_c + ru \mid \|u\|_2 \le 1\}$$

ellipsoid: set of the form

$$\{x \mid (x - x_c)^T P^{-1}(x - x_c) \le 1\}$$

with $P \in \mathbf{S}^n_{++}$ (*i.e.*, P symmetric positive definite)



norm: a function $\|\cdot\|$ that satisfies

- $\|x\| \ge 0$; $\|x\| = 0$ if and only if $x = 0$
- $\|tx\| = |t|\,\|x\|$ for $t \in \mathbf{R}$
- $\|x + y\| \le \|x\| + \|y\|$

norm ball with center x_c and radius r: $\{x \mid \|x - x_c\| \le r\}$

$\preceq$ is componentwise inequality)

$Ax \preceq b$, All x meet Ax <= b.

the intersection of (any number of) convex sets is convex

suppose $f : \mathbf{R}^n \to \mathbf{R}^m$ is affine ($f(x) = Ax + b$ with $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$)

- the image of a convex set under f is convex

$$S \subseteq \mathbf{R}^n \text{ convex} \implies f(S) = \{f(x) \mid x \in S\} \text{ convex}$$

- the inverse image $f^{-1}(C)$ of a convex set under f is convex

$$C \subseteq \mathbf{R}^m \text{ convex} \implies f^{-1}(C) = \{x \in \mathbf{R}^n \mid f(x) \in C\} \text{ convex}$$

The perspective function $P : R^{n+1} \to R^n$ is defined mathematically as:

$P(x, t) = x/t, \quad domain\ of\ P = \{(x, t) | t > 0\}$

Intuitively it can be explained as Professor Stephen Boyd explains, "Divide first n elements of the vector by its last component".

images and inverse images of convex sets under perspective are convex

linear-fractional function $f : \mathbf{R}^n \to \mathbf{R}^m$:

$$f(x) = \frac{Ax + b}{c^T x + d}, \qquad \mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$$

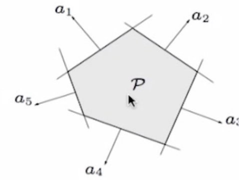
images and inverse images of convex sets under linear-fractional functions are convex

$\triangleq$ mean "is defined as" or equals by definition.

$$\frac{X - \mu}{\sigma}$$ Normalize.

more involved task. The main contribution of this paper is the derivation of gradient projections with ℓ_1 domain constraints that can be performed almost as fast as gradient projection with ℓ_2 constraints.

A **positive definite** matrix is a **symmetric** matrix with all **positive** eigenvalues.



polyhedron is intersection of finite number of halfspaces and hyperplanes

$\mathbf{S}^n$ is set of symmetric $n \times n$ matrices

$\mathbf{S}^n_+ = \{X \in \mathbf{S}^n \mid X \succeq 0\}$: positive semidefinite $n \times n$ matrices

$$X \in \mathbf{S}^n_+ \iff z^T X z \ge 0 \text{ for all } z$$

$\mathbf{S}^n_+$ is a convex cone

$\mathbf{S}^n_{++} = \{X \in \mathbf{S}^n \mid X \succ 0\}$: positive definite $n \times n$ matrices

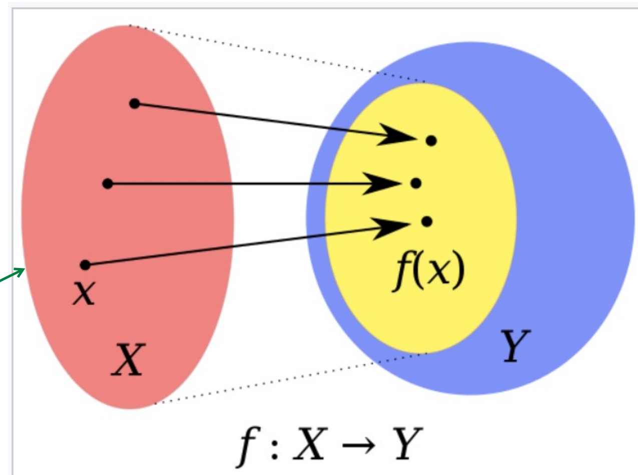


Illustration showing *f*, a function from the pink domain *X* to the blue codomain *Y*. The yellow oval inside *Y* is the image of *f*. Both the image and the codomain are sometimes called the range of *f*.

$$f(x) = \frac{1}{x_1 + x_2 + 1}x$$