Behind the Scenes of $Ax = b$: Axioms and an Open Question

P. N. Karthik

Research Fellow,
Institute of Data Science,
National University of Singapore

Based on joint work with Rajesh Sundaresan

23 March 2022

Linear Inverse Problem

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

Linear Inverse Problem

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}}_{b}$$

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

$$\checkmark A \qquad\qquad x \qquad\qquad b \checkmark$$

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

✓ A $\qquad$ x $\qquad$ b ✓

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

✓ A $\qquad$ x $\qquad$ b ✓

- Example 1: image reconstruction

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

✓ A x b ✓

- Example 1: image reconstruction
  - The domain of an image is its pixels (say n pixels)

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

✓ A $\quad$ x $\quad$ b ✓

- Example 1: image reconstruction
  - The domain of an image is its pixels (say n pixels)
  - f_j: indicator of the jth pixel; $f = \sum_{j=1}^{n} x_j \ f_j$

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

✓ A $\quad$ x $\quad$ b ✓

- Example 1: image reconstruction
  - The domain of an image is its pixels (say n pixels)
  - f_j: indicator of the jth pixel; $f = \sum_{j=1}^{n} x_j \; f_j$
  - Linear measurements $R_i f_j = a_{ij}, \quad R_i f = b_i, \quad i = 1, \ldots, k$

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

✓ A $\qquad$ x $\qquad$ b ✓

- Example 1: image reconstruction
  - The domain of an image is its pixels (say n pixels)
  - f_j: indicator of the jth pixel; $f = \sum_{j=1}^{n} x_j \, f_j$
  - Linear measurements $R_i f_j = a_{ij}$, $R_i f = b_i$, $i = 1, \ldots, k$
  - Goal: to recover $x_1, \ldots, x_n$ from the measurements

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

$$A \qquad\qquad x \qquad\qquad b$$

- Example 1: image reconstruction
  - The domain of an image is its pixels (say n pixels)
  - f_j: indicator of the jth pixel; $f = \sum_{j=1}^{n} x_j \, f_j$
  - Linear measurements $R_i f_j = a_{ij}, \;\; R_i f = b_i, \;\; i = 1,\ldots,k$
  - Goal: to recover $x_1, \ldots, x_n$ from the measurements
- Example 2: learning the frequency of alphabets from a corpus

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

✓ A $\qquad$ x $\qquad$ b ✓

- Example 1: image reconstruction
  - The domain of an image is its pixels (say n pixels)
  - f_j: indicator of the jth pixel; $f = \sum_{j=1}^{n} x_j \ f_j$
  - Linear measurements $R_i f_j = a_{ij}, \ \ R_i f = b_i, \ \ i = 1,\ldots,k$
  - Goal: to recover $x_1, \ldots, x_n$ from the measurements

- Example 2: learning the frequency of alphabets from a corpus
  - n alphabets $a_1, \ldots, a_n$ occurring with unknown probabilities $x_1, \ldots, x_n$

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

A x b

- Example 1: image reconstruction
    - The domain of an image is its pixels (say n pixels)
    - f_j: indicator of the jth pixel; $f = \sum_{j=1}^{n} x_j \, f_j$
    - Linear measurements $R_i f_j = a_{ij}, \quad R_i f = b_i, \quad i = 1,\ldots,k$
    - Goal: to recover $x_1, \ldots, x_n$ from the measurements

- Example 2: learning the frequency of alphabets from a corpus
    - n alphabets $a_1, \ldots, a_n$ occurring with unknown probabilities $x_1, \ldots, x_n$
    - Corpus: body of text containing strings of alphabets

Linear Inverse Problem

$$
\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}
$$

$$A \qquad\qquad x \qquad\qquad b$$

- Example 1: image reconstruction
  - The domain of an image is its pixels (say n pixels)
  - f_j: indicator of the jth pixel; $f = \sum_{j=1}^{n} x_j \, f_j$
  - Linear measurements $R_i f_j = a_{ij}, \;\; R_i f = b_i, \;\; i = 1, \ldots, k$
  - Goal: to recover $x_1, \ldots, x_n$ from the measurements

- Example 2: learning the frequency of alphabets from a corpus
  - n alphabets $a_1, \ldots, a_n$ occurring with unknown probabilities $x_1, \ldots, x_n$
  - Corpus: body of text containing strings of alphabets
  - Linear measurements on the corpus: $\mathbb{E}[T_i(Z)] = b_i, \;\; i = 1, \ldots, k$

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

A $\qquad$ x $\qquad$ b

- Example 1: image reconstruction
  - The domain of an image is its pixels (say n pixels)
  - f_j: indicator of the jth pixel; $f = \sum_{j=1}^{n} x_j \, f_j$
  - Linear measurements $R_i f_j = a_{ij}, \quad R_i f = b_i, \quad i = 1,\ldots,k$
  - Goal: to recover $x_1, \ldots, x_n$ from the measurements

- Example 2: learning the frequency of alphabets from a corpus
  - n alphabets $a_1, \ldots, a_n$ occurring with unknown probabilities $x_1, \ldots, x_n$
  - Corpus: body of text containing strings of alphabets
  - Linear measurements on the corpus: $\mathbb{E}[T_i(Z)] = b_i, \ i = 1,\ldots,k$
  - Goal: to recover $x_1, \ldots, x_n$ satisfying the moment constraints

Linear Inverse Problem

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

✓ A $\quad$ x $\quad$ b ✓

- Example 1: image reconstruction
  - The domain of an image is its pixels (say n pixels)
  - f_j: indicator of the jth pixel; $f = \sum_{j=1}^{n} x_j \, f_j$
  - Linear measurements $R_i f_j = a_{ij}, \;\; R_i f = b_i, \;\; i = 1,\ldots,k$
  - Goal: to recover $x_1, \ldots, x_n$ from the measurements

Is there a principled approach to solve the problem?

- Example 2: learning the frequency of alphabets from a corpus
  - n alphabets $a_1, \ldots, a_n$ occurring with unknown probabilities $x_1, \ldots, x_n$
  - Corpus: body of text containing strings of alphabets
  - Linear measurements on the corpus: $\mathbb{E}[T_i(Z)] = b_i, \;\; i = 1,\ldots,k$
  - Goal: to recover $x_1, \ldots, x_n$ satisfying the moment constraints

Approach 1: Function-Minimisation

Approach 1: Function-Minimisation

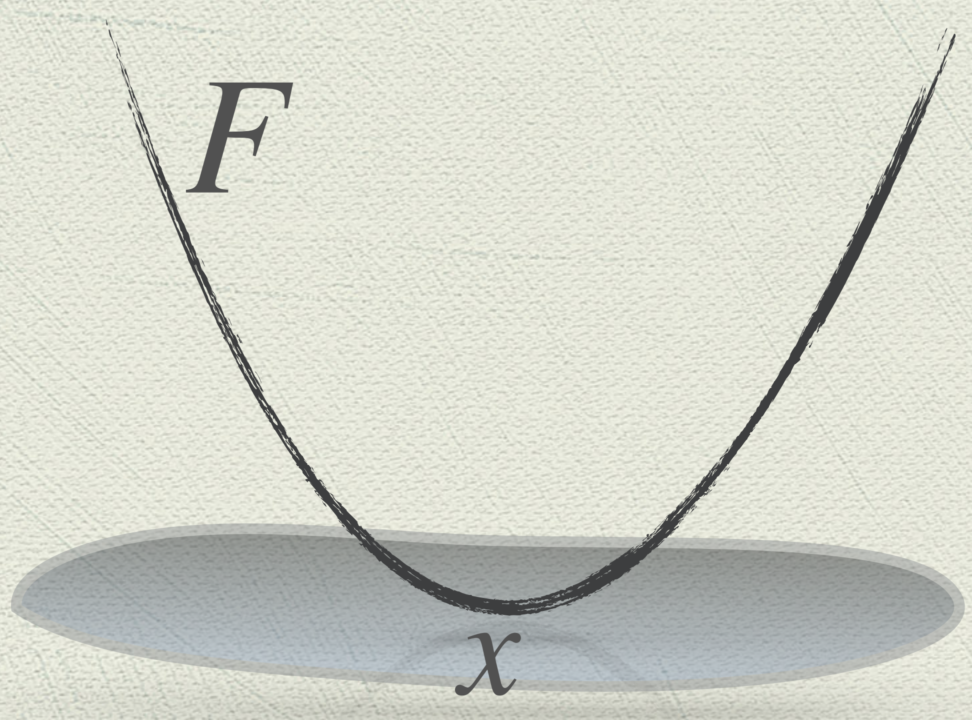
- Define a "cost" function F

Approach 1: Function-Minimisation

- Define a "cost" function F
- Recover x as a minimiser of F

Approach 1: Function-Minimisation

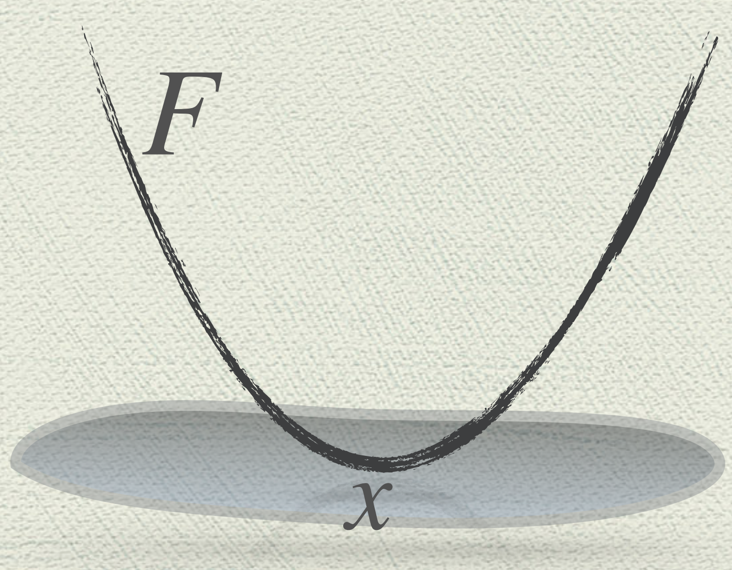
- Define a "cost" function F
- Recover x as a minimiser of F



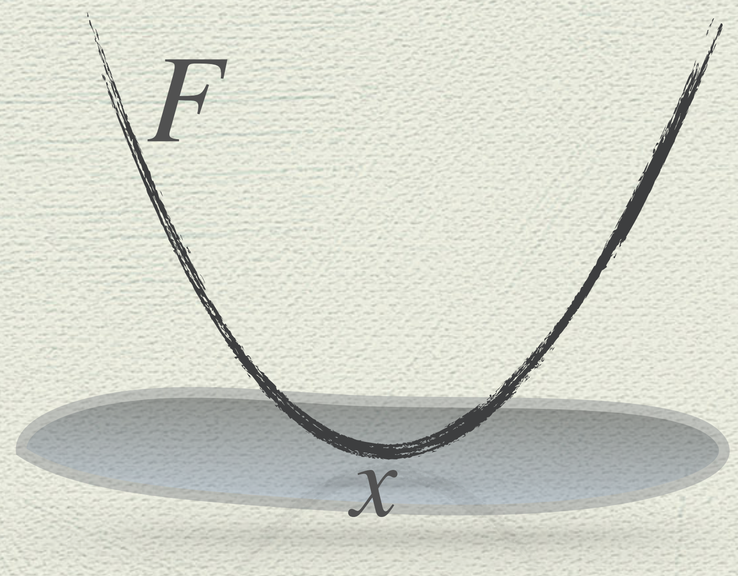
$$L = \{w \in \mathbb{R}^n : Aw = b\}$$

$$x = \arg\min_{w \in L} F(w)$$

Solution using Optimisation Techniques

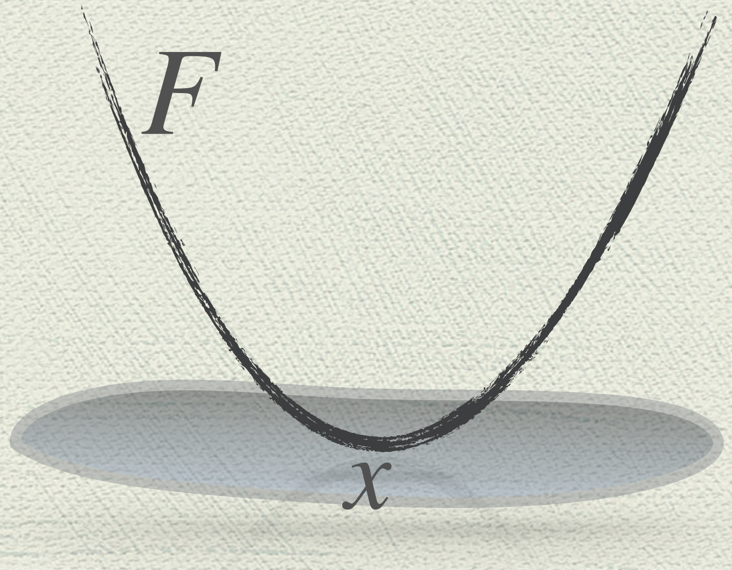


$$\min_{w \in L} F(w)$$



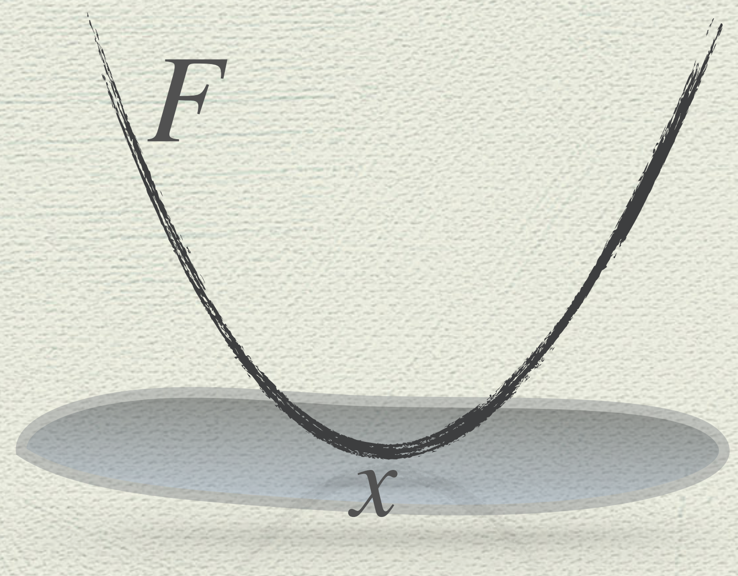
$$\min_{w \in L} F(w)$$

Solution using Optimisation Techniques



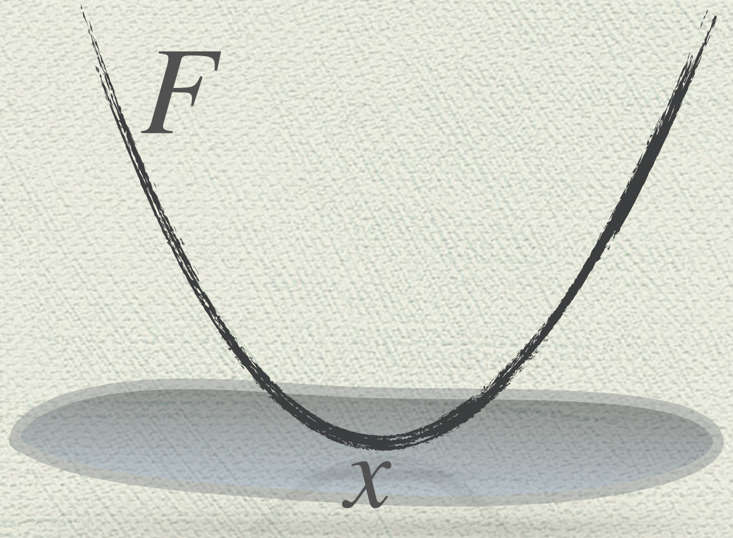
$$\min_{w \in L} F(w)$$

$$L = \{w \in \mathbb{R}^n : a^T w = b\}$$



$$\min_{w \in L} F(w)$$

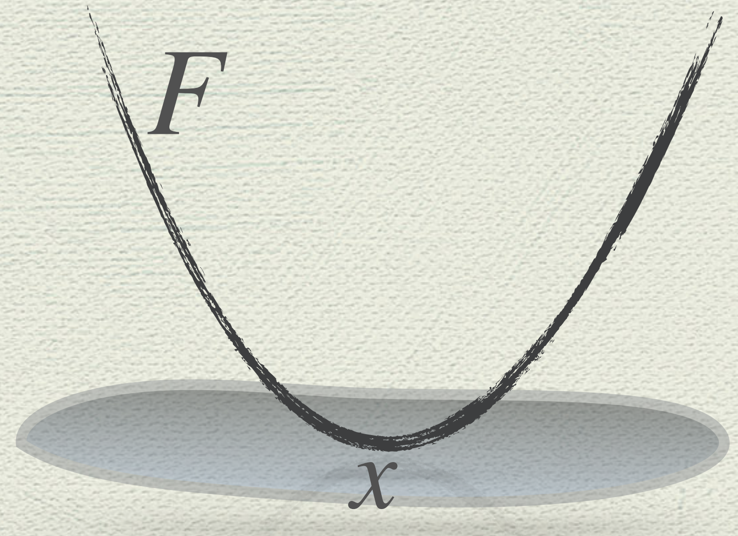
Solution using Optimisation Techniques



$$\min_{w \in L} F(w)$$

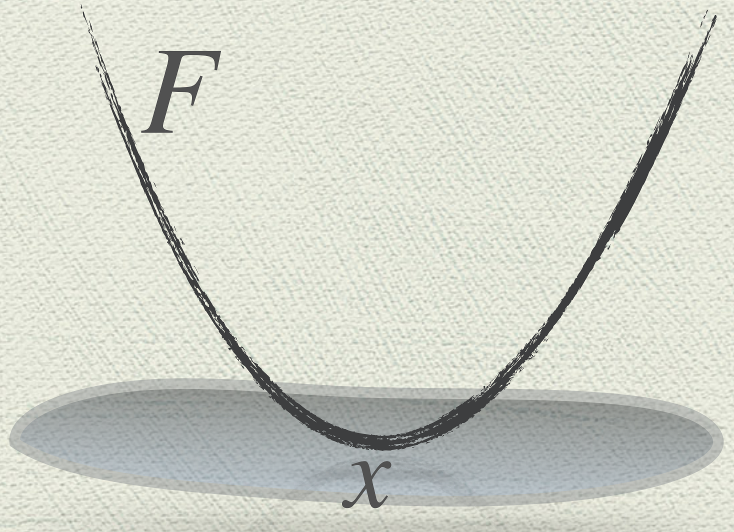
$$L = \{w \in \mathbb{R}^n : a^T w = b\}$$

$$G(w, \lambda) = F(w) - \lambda(a^T w - b)$$



$$\min_{w \in L} F(w)$$

Solution using Optimisation Techniques

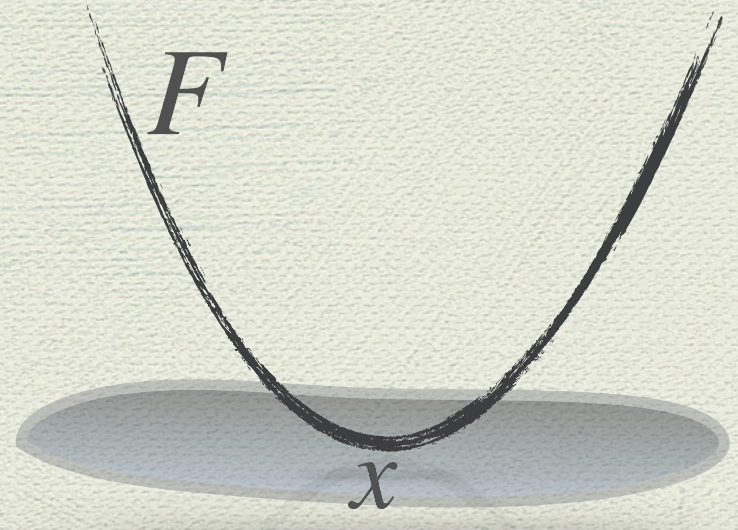


$$\min_{w \in L} F(w)$$

$$L = \{w \in \mathbb{R}^n : a^T w = b\}$$

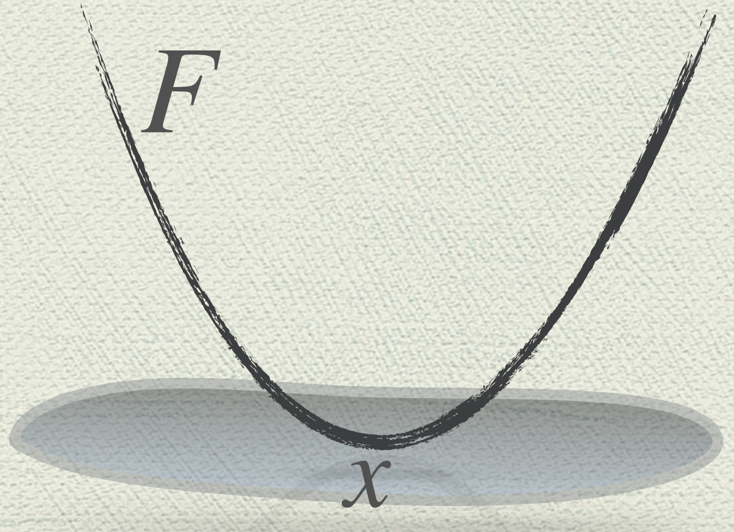
$$G(w, \lambda) = F(w) - \lambda(a^T w - b)$$

$$\nabla_w G(w, \lambda) = 0 \iff \nabla F(w) = \lambda a$$



$$\min_{w \in L} F(w)$$

Solution using Optimisation Techniques

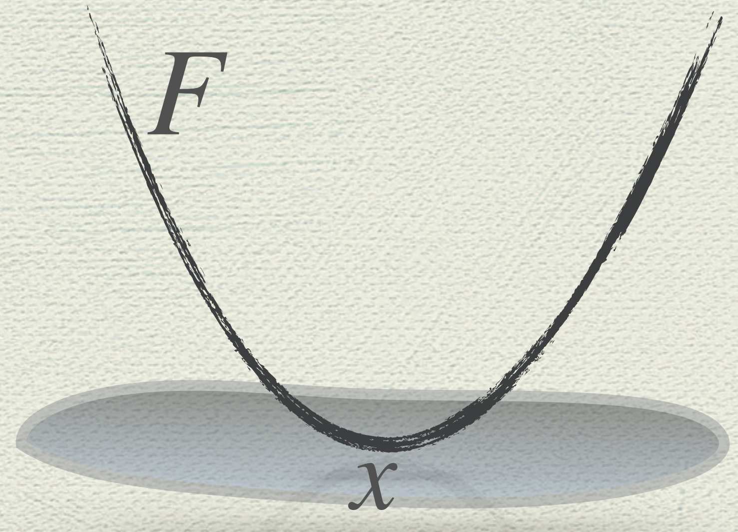


$$\min_{w \in L} F(w)$$

$$L = \{w \in \mathbb{R}^n : a^T w = b\}$$

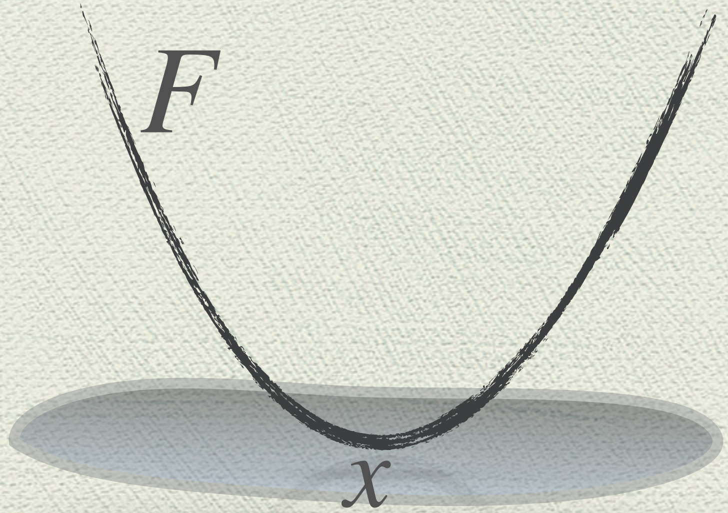
$$G(w, \lambda) = F(w) - \lambda(a^T w - b)$$

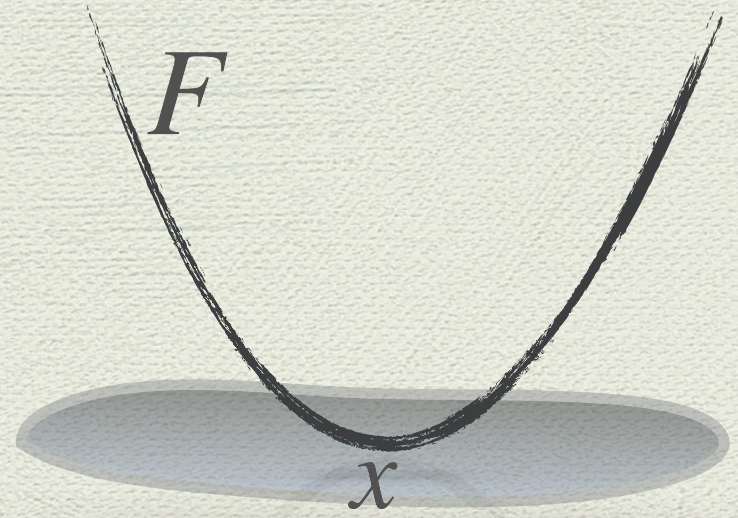
$$\nabla_w G(w, \lambda) = 0 \iff \boxed{\nabla F(w) = \lambda a}$$



$$\min_{w \in L} F(w)$$

Solution using Optimisation Techniques





$$\min_{w \in L} F(w)$$

$$L = \{w \in \mathbb{R}^n : a^T w = b\}$$

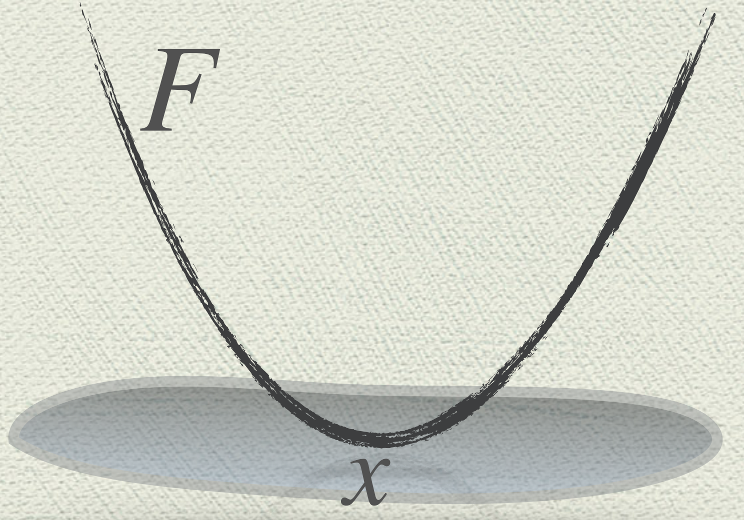
$$\min_{w \in L} F(w)$$

$$L = \{w \in \mathbb{R}^n : Aw = b\}$$

$$G(w, \lambda) = F(w) - \lambda(a^T w - b)$$

$$\nabla_w G(w, \lambda) = 0 \iff \boxed{\nabla F(w) = \lambda a}$$

Solution using Optimisation Techniques

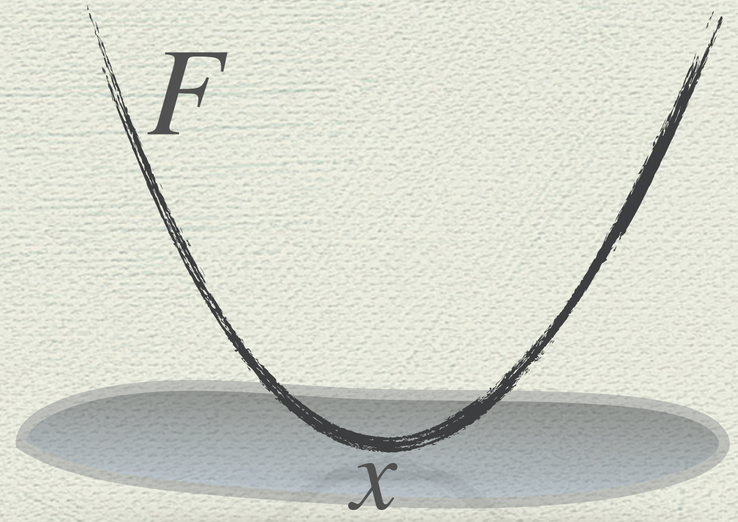


$$\min_{w \in L} F(w)$$

$$L = \{w \in \mathbb{R}^n : a^T w = b\}$$

$$G(w, \lambda) = F(w) - \lambda(a^T w - b)$$

$$\nabla_w G(w, \lambda) = 0 \iff \boxed{\nabla F(w) = \lambda a}$$



$$\min_{w \in L} F(w)$$

$$L = \{w \in \mathbb{R}^n : Aw = b\}$$

$$G(w, \lambda) = F(w) - \lambda^T(Aw - b)$$

$$\nabla_w G(w, \lambda) = 0 \iff \nabla F(w) = A^T \lambda$$

Commonly Used F

Commonly Used F

- Least squares: $F(w) = \|w\|^2$

Commonly Used F

- Least squares: $F(w) = \|w\|^2$

$$x = \arg\min_{w \in L} \|w\|^2 = \arg\min_{w \in L} \sum_{i=1}^{n} w_i^2$$

Commonly Used F

- Least squares: $F(w) = \|w\|^2$

$$x = \arg\min_{w \in L} \|w\|^2 = \arg\min_{w \in L} \sum_{i=1}^{n} w_i^2$$

- Maximum entropy: $F(w) = \sum_i w_i \log w_i$

Commonly Used F

- Least squares: $F(w) = \|w\|^2$

$$x = \arg\min_{w \in L} \|w\|^2 = \arg\min_{w \in L} \sum_{i=1}^{n} w_i^2$$

- Maximum entropy: $F(w) = \sum_i w_i \log w_i$

$$x = \arg\min_{w \in L \cap \mathbb{R}_+^n} \sum_{i=1}^{n} w_i \log w_i$$

Commonly Used F with Prior Guess

- Prior guess: $x^\star$

Commonly Used F with Prior Guess

- Prior guess: $x^\star$

- Least squares: $F(w \mid x^\star) = \|w - x^\star\|^2$

Commonly Used F with Prior Guess

- Prior guess: $x^\star$

- Least squares: $F(w \mid x^\star) = \|w - x^\star\|^2$

$$x = \arg\min_{w \in L} \|w - x^\star\|^2$$

Commonly Used F with Prior Guess

- Prior guess: $x^\star$

- Least squares: $F(w \mid x^\star) = \|w - x^\star\|^2$

$$x = \arg\min_{w \in L} \|w - x^\star\|^2$$

- Maximum entropy: $F(w \mid x^\star) = \sum_i w_i \log \frac{w_i}{x_i^\star}$

Commonly Used F with Prior Guess

- Prior guess: $x^\star$

- Least squares: $F(w \mid x^\star) = \|w - x^\star\|^2$

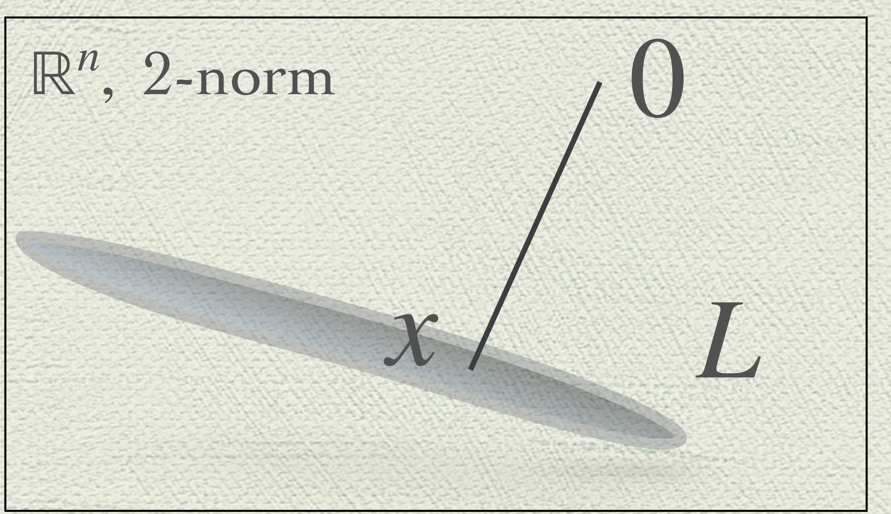
$$x = \arg\min_{w \in L} \|w - x^\star\|^2$$

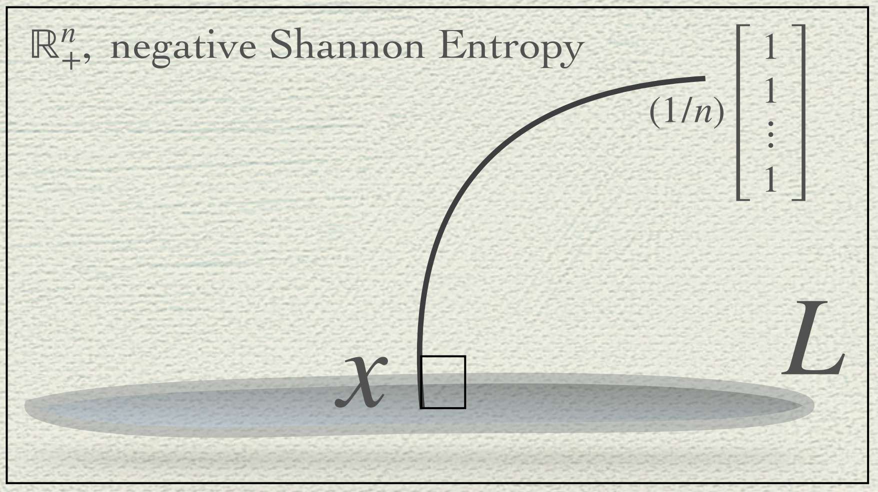
- Maximum entropy: $F(w \mid x^\star) = \sum_i w_i \log \frac{w_i}{x_i^\star}$

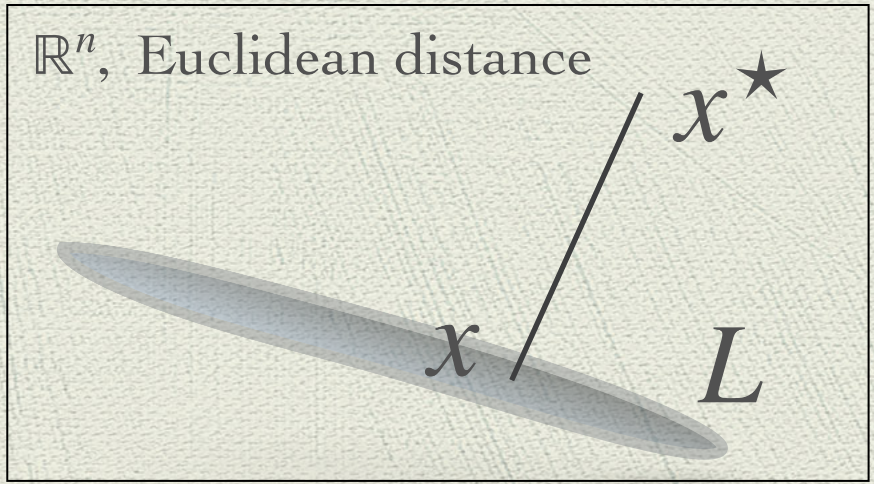
$$x = \arg\min_{w \in L \cap \mathbb{R}^n_+} \sum_i w_i \log \frac{w_i}{x_i^\star}$$

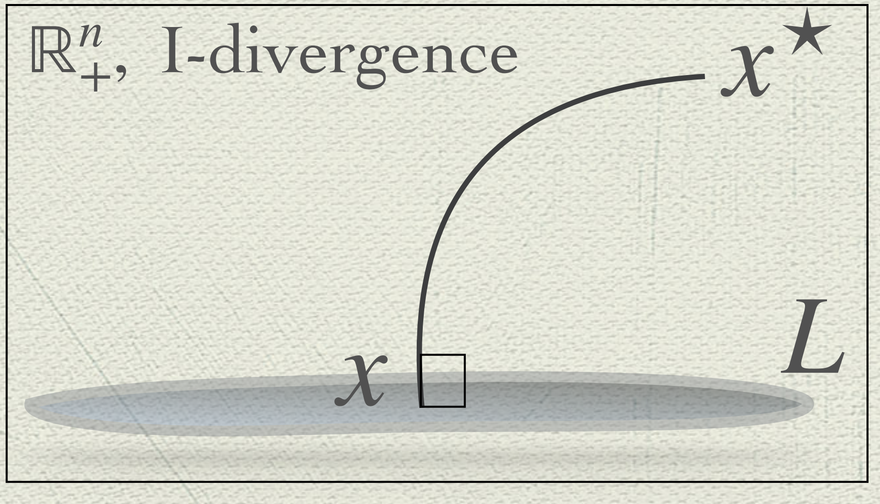
Projection Rules

Projection Rules

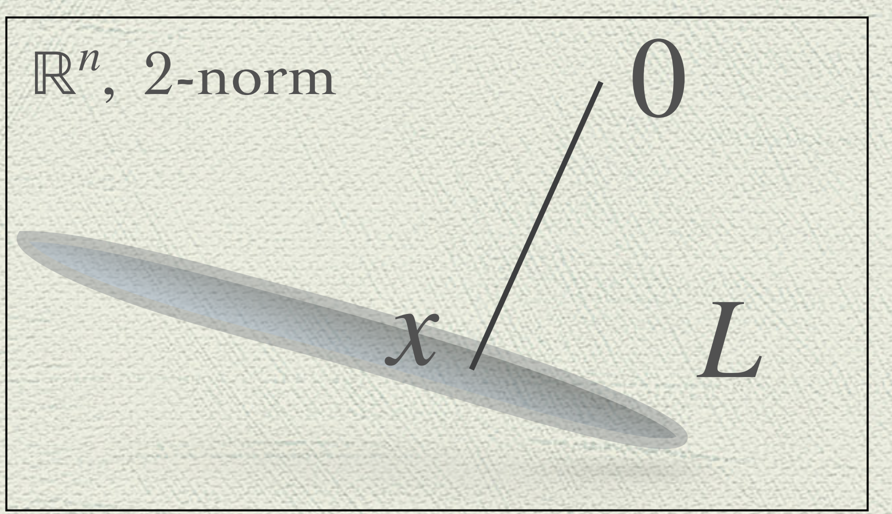


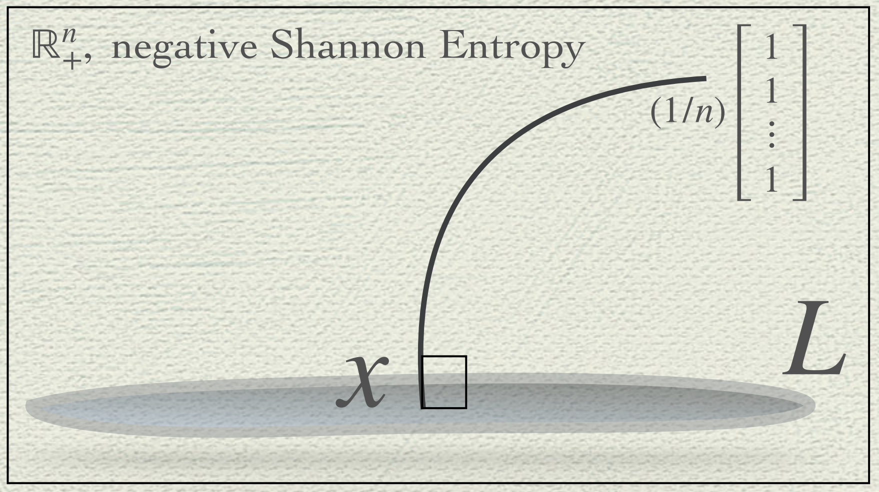




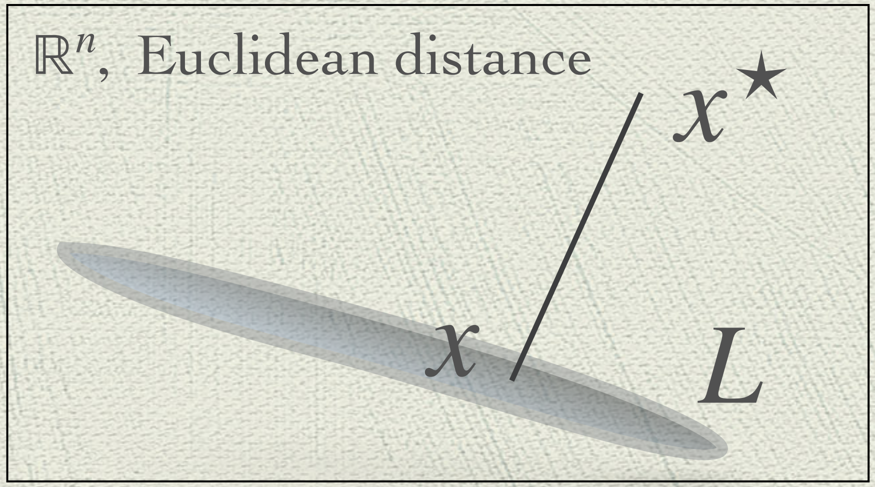


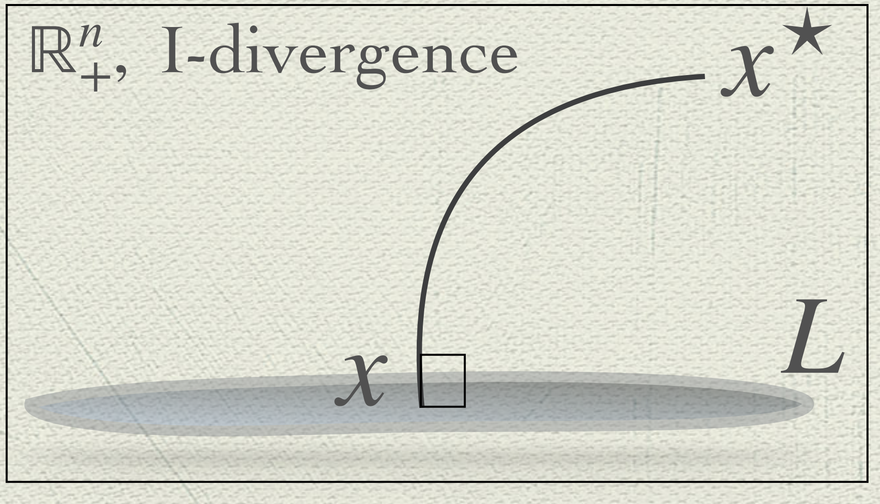
Projection Rules



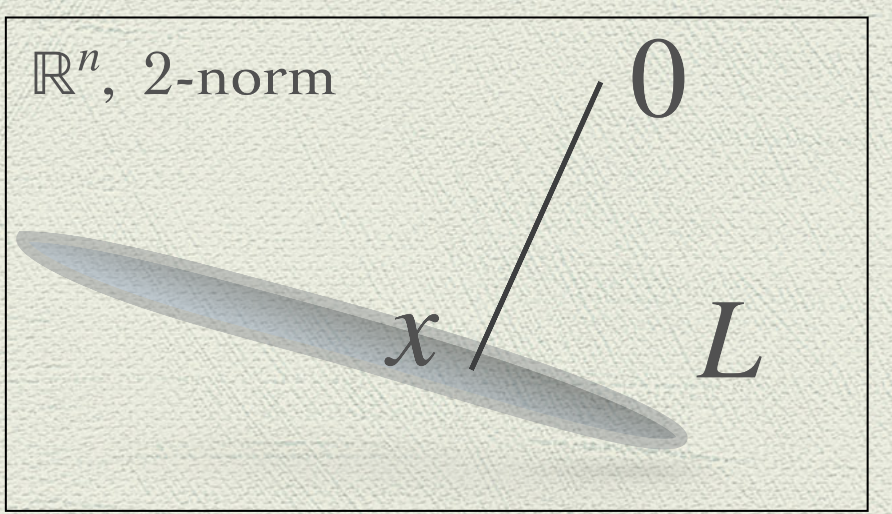


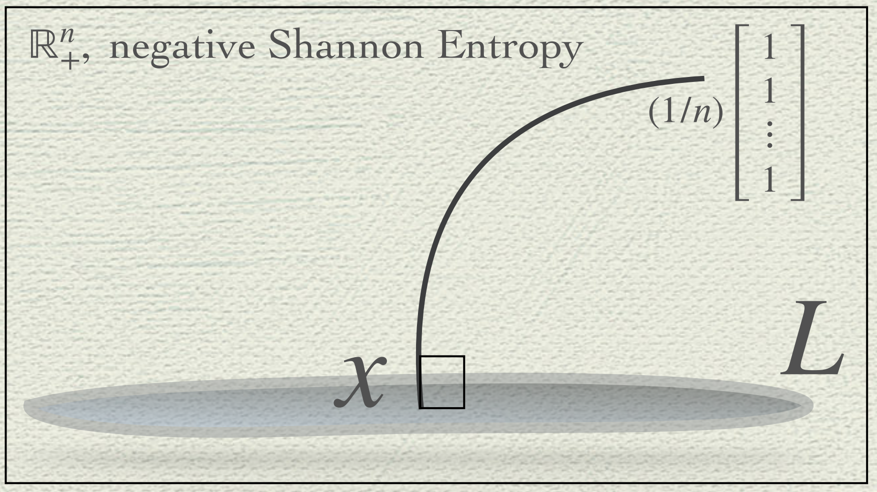
Function-minimisation
=
projection rule





Projection Rules

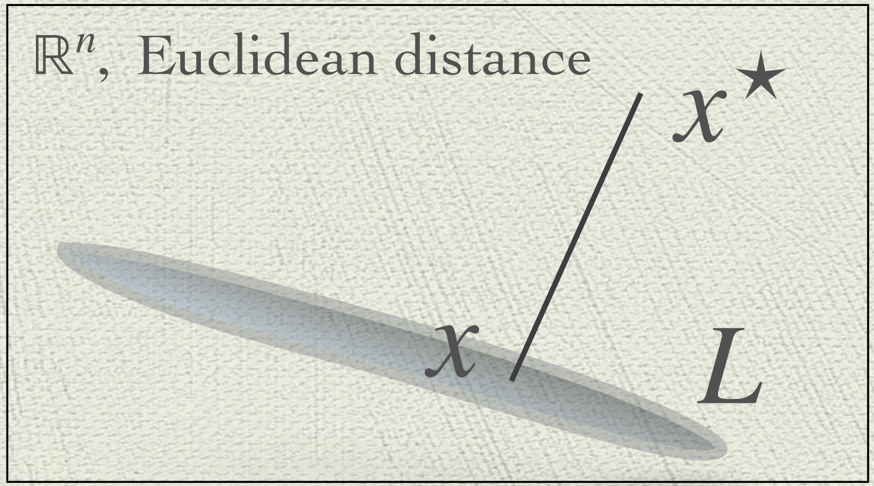




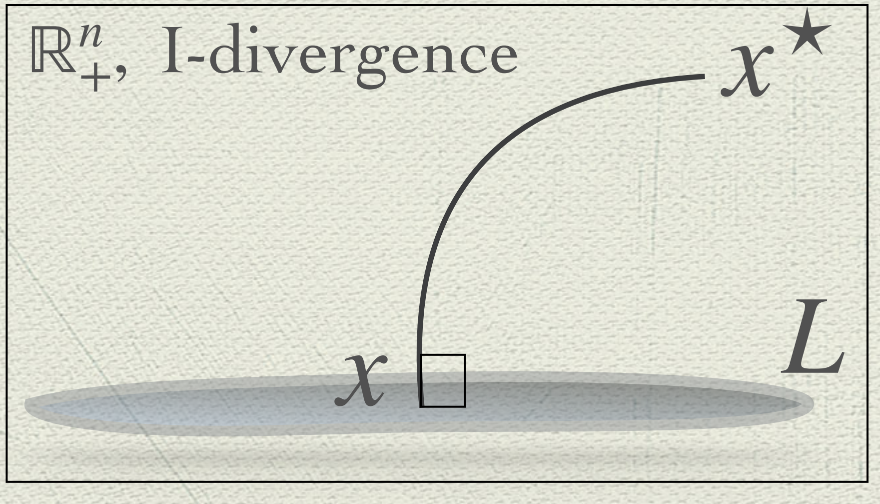
Function-minimisation

=

projection rule



$\Pi : (L, x^\star) \mapsto \Pi(L \,|\, x^\star) \in L$



Projection Rules and Function Minimisation

Projection Rules and Function Minimisation

$$\begin{aligned} F : \mathbb{R}^n \times \mathbb{R}^n &\to \mathbb{R} \\ (x^\star, w) &\mapsto F(w \,|\, x^\star) \end{aligned}$$

Projection Rules and Function Minimisation

$$F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$
$$(x^\star, w) \mapsto F(w \,|\, x^\star)$$

$\Pi : \mathscr{L} \times \mathbb{R}^n \to \mathbb{R}^n$ projection rule

$$\Pi : (L, x^\star) \mapsto \Pi(L \,|\, x^\star) \in L$$

Projection Rules and Function Minimisation

$$F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$
$$(x^\star, w) \mapsto F(w \,|\, x^\star)$$

$\Pi : \mathscr{L} \times \mathbb{R}^n \to \mathbb{R}^n$ projection rule

$$\Pi : (L, x^\star) \mapsto \Pi(L \,|\, x^\star) \in L$$

<u>Definition:</u> <u>Π is generated by F,</u> if

$$\Pi(L \,|\, x^\star) = \arg\min_{w \in L} F(w \,|\, x^\star)$$

Projection Rules and Function Minimisation

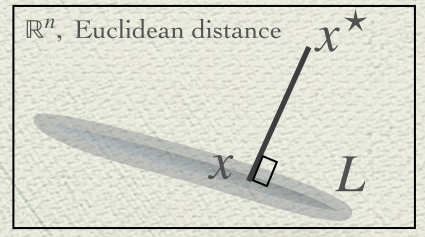
$$F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$
$$(x^\star, w) \mapsto F(w \,|\, x^\star)$$

$$\Pi : \mathscr{L} \times \mathbb{R}^n \to \mathbb{R}^n \text{ projection rule}$$

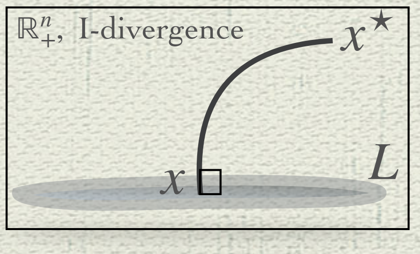
$$\Pi : (L, x^\star) \mapsto \Pi(L \,|\, x^\star) \in L$$

<u>Definition:</u> <u>Π is generated by F,</u> if

$$\Pi(L \,|\, x^\star) = \arg\min_{w \in L} F(w \,|\, x^\star)$$



$F(w \,|\, x^\star) = \|w - x^\star\|^2$



$F(w \,|\, x^\star) = \sum_i w_i \log \frac{w_i}{x_i^\star}$

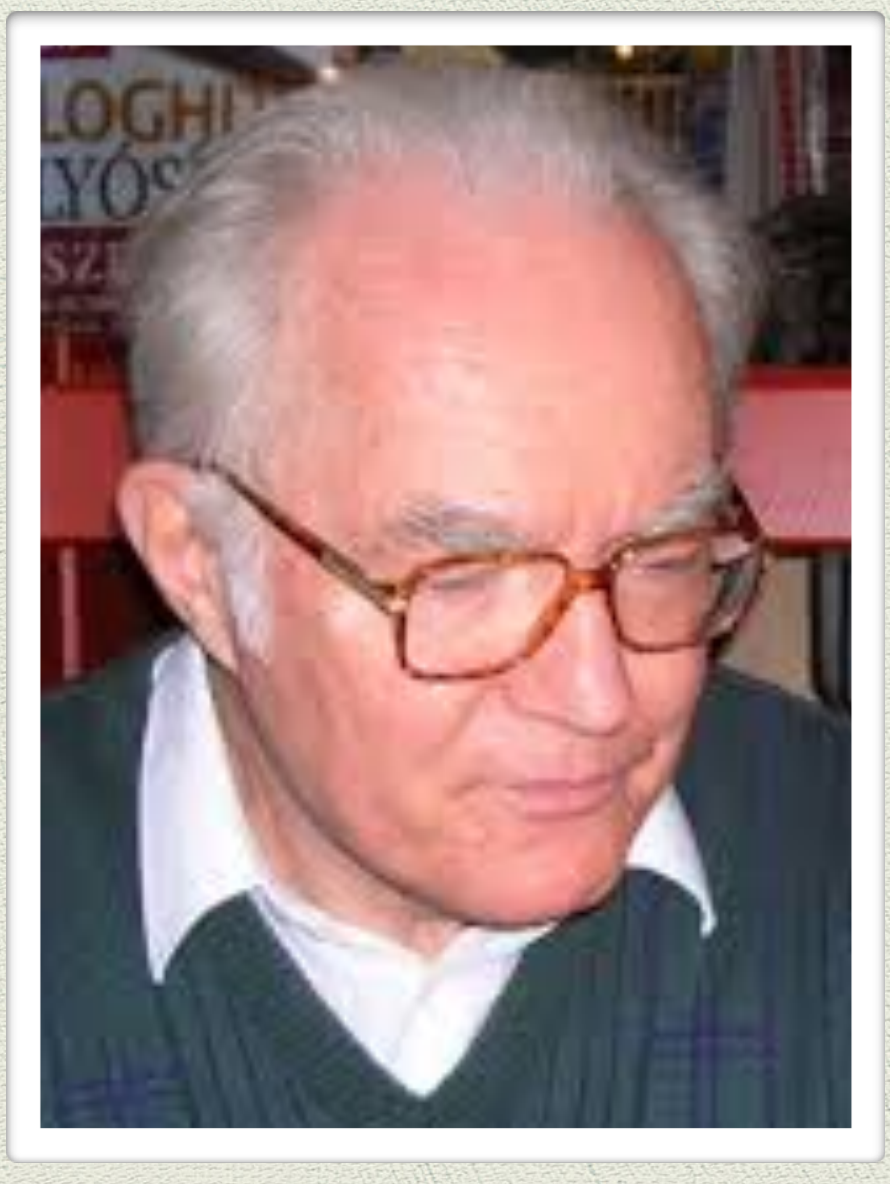
Is every projection rule generated by some function?

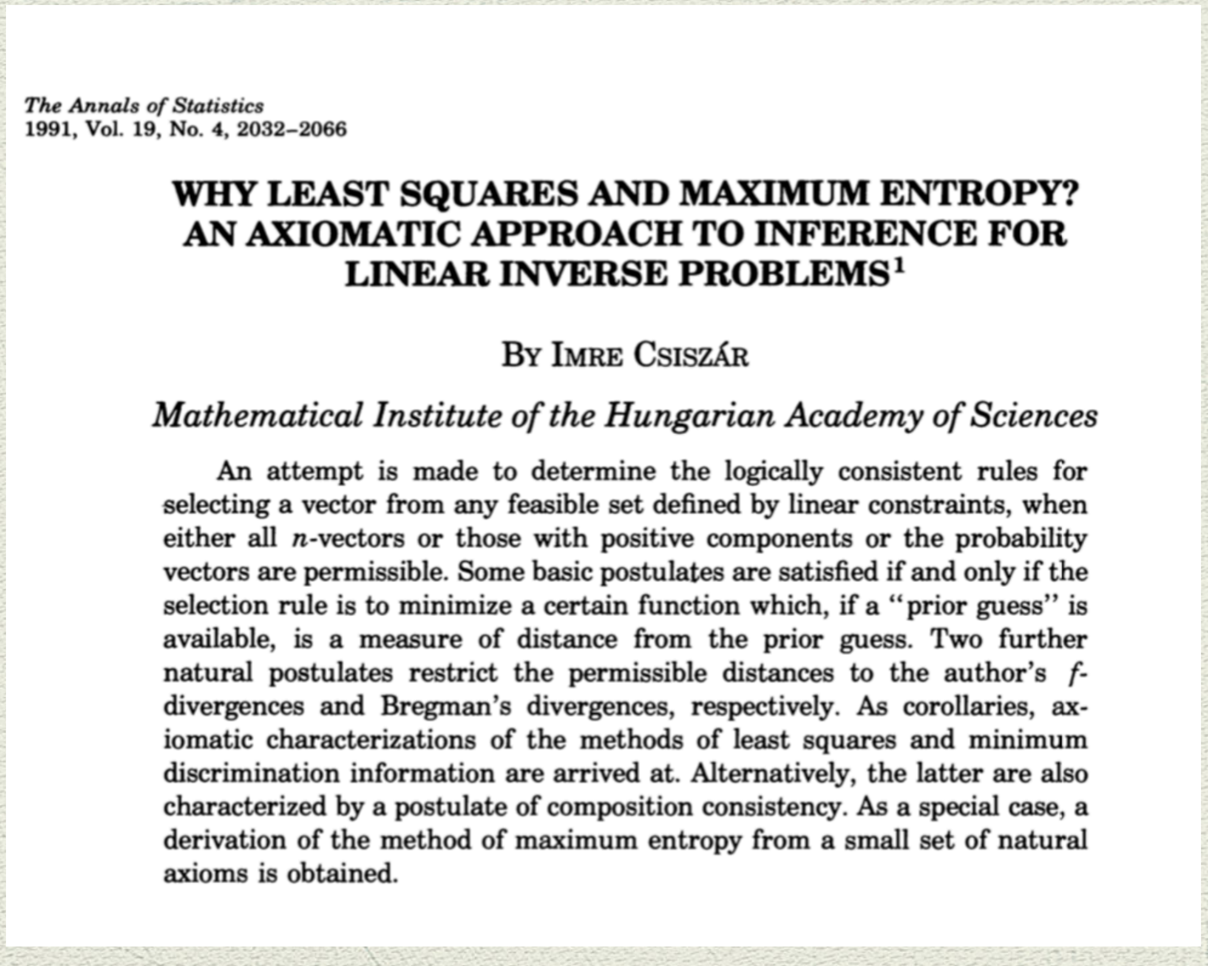
Is every projection rule generated by some function?

- Answers to this question were provided by Csiszár in 1991

Is every projection rule generated by some function?

- Answers to this question were provided by Csiszár in 1991





The Annals of Statistics
1991, Vol. 19, No. 4, 2032–2066

WHY LEAST SQUARES AND MAXIMUM ENTROPY? AN AXIOMATIC APPROACH TO INFERENCE FOR LINEAR INVERSE PROBLEMS[1]

BY IMRE CSISZÁR

Mathematical Institute of the Hungarian Academy of Sciences

An attempt is made to determine the logically consistent rules for selecting a vector from any feasible set defined by linear constraints, when either all n-vectors or those with positive components or the probability vectors are permissible. Some basic postulates are satisfied if and only if the selection rule is to minimize a certain function which, if a "prior guess" is available, is a measure of distance from the prior guess. Two further natural postulates restrict the permissible distances to the author's f-divergences and Bregman's divergences, respectively. As corollaries, axiomatic characterizations of the methods of least squares and minimum discrimination information are arrived at. Alternatively, the latter are also characterized by a postulate of composition consistency. As a special case, a derivation of the method of maximum entropy from a small set of natural axioms is obtained.

The seminal work of Csiszár

Approach 2: Axioms

Approach 2: Axioms

- Csiszár demonstrated that if projection rules satisfy some naturally appealing axioms, they must be generated by "nice" functions

Approach 2: Axioms

- Csiszár demonstrated that if projection rules satisfy some naturally appealing axioms, they must be generated by "nice" functions

Projection rules satisfies the axioms of	Projection rule is generated by
Regularity + Locality	$F(w\|x^\star) = \sum_{i=1}^{n} f_i(w_i\|x_i^\star)$, f_i continuously differentiable and strictly convex
Regularity+ Locality + Subspace Transitivity	$F(w\|x^\star) =$ Bregman's divergence
Regularity+ Locality + Subspace Transitivity + Statistical	$F(w\|x^\star) = \sum_{i=1}^{n} w_i \log \frac{w_i}{x_i^\star}$
Regularity + Locality + Subspace Transitivity + Location Invariance + Scale Invariance	$F(w\|x^\star) = \sum_{i=1}^{n} (w_i - x_i^\star)^2$

Why are these axioms of interest to us?

- Kumar and Sundaresan ('15) discovered a family of projection rules satisfying

Regularity ✓

Subspace transitivity ✓

Locality ✗

- These projection rules are generated by a parametric divergence known as relative α-entropy
(related to the well-known Rényi divergence)

M. A. Kumar and R. Sundaresan. (2015). Minimization problems based on relative α-entropy I: Forward projection. IEEE Transactions on Information Theory, 61(9), 5063-5080.

M. A. Kumar and R. Sundaresan. (2015). Minimization problems based on relative α-entropy II: Reverse projection. IEEE Transactions on Information Theory, 61(9), 5081-5095.

"Are the projection rules generated by relative α-entropy the only rules which are regular, subspace transitive and nonlocal?"

Regularity ✔

Subspace transitivity ✔

Locality ✘

Our Quest

- Csiszár's results provide a necessary and sufficient axiomatic characterisation of many projection rules

Regularity + Locality $\Longleftrightarrow$ Projection rule generated by $F(w\,|\,x^\star) = \sum_{i=1}^{n} f_i(w_i\,|\,x_i^\star)$

Regularity + Locality + Subspace Transitivity $\Longleftrightarrow$ Projection rule generated by Bregman's divergence

Our Quest

- Csiszár's results provide a necessary and sufficient axiomatic characterisation of many projection rules

Regularity + Locality $\Longleftrightarrow$ Projection rule generated by $F(w \mid x^\star) = \sum_{i=1}^{n} f_i(w_i \mid x_i^\star)$

Regularity + Locality + Subspace Transitivity $\Longleftrightarrow$ Projection rule generated by Bregman's divergence

$\boxed{\text{Regularity + Subspace Transitivity} \Longleftrightarrow \text{Projection rule generated by ??}}$

Our Quest

- Csiszár's results provide a necessary and sufficient axiomatic characterisation of many projection rules

Regularity + Locality $\Longleftrightarrow$ Projection rule generated by $F(w\,|\,x^\star) = \sum_{i=1}^{n} f_i(w_i\,|\,x_i^\star)$

Regularity + Locality + Subspace Transitivity $\Longleftrightarrow$ Projection rule generated by Bregman's divergence

Regularity + Subspace Transitivity $\Longleftrightarrow$ Projection rule generated by ??

Regularity $\Longleftrightarrow$ Projection rule generated by ??

The axiom of regularity

The axiom of regularity

$\Pi : \mathscr{L} \times \mathbb{R}^n \to \mathbb{R}^n$ projection rule

The axiom of regularity

$\Pi : \mathscr{L} \times \mathbb{R}^n \to \mathbb{R}^n$ projection rule

$$\mathscr{M} = \left\{ \{w \in \mathbb{R}^n : a^T w = b\},\ a \in \mathbb{R}^n,\ a \neq 0,\ b \in \mathbb{R} \right\}$$

The axiom of regularity

$\Pi : \mathscr{L} \times \mathbb{R}^n \to \mathbb{R}^n$ projection rule

$$\mathscr{M} = \left\{ \{w \in \mathbb{R}^n : a^T w = b\},\ a \in \mathbb{R}^n,\ a \neq 0,\ b \in \mathbb{R} \right\}$$

Π satisfies regularity if, for all $x^\star \in \mathbb{R}^n$,

Regularity

(Consistency)

$$L' \subset L,\ \Pi(L \,|\, x^\star) \in L' \implies \Pi(L' \,|\, x^\star) = \Pi(L \,|\, x^\star)$$

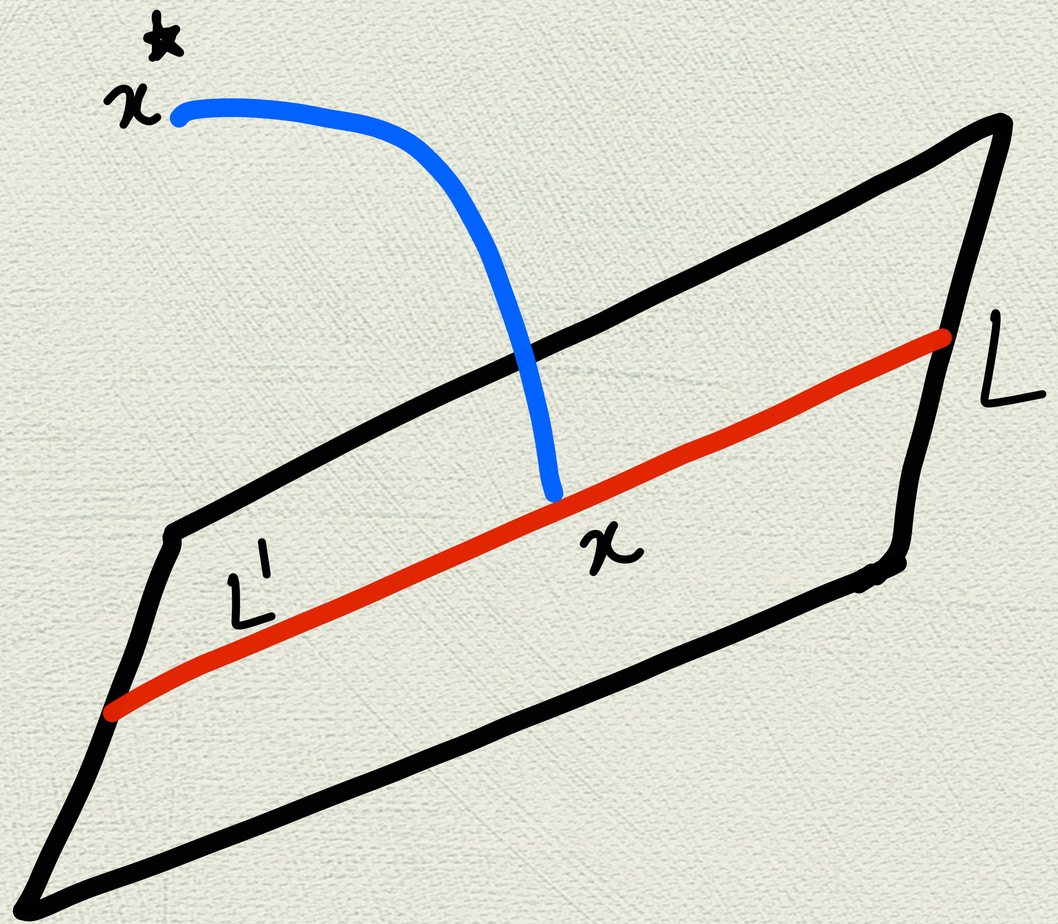
(Distinctness)

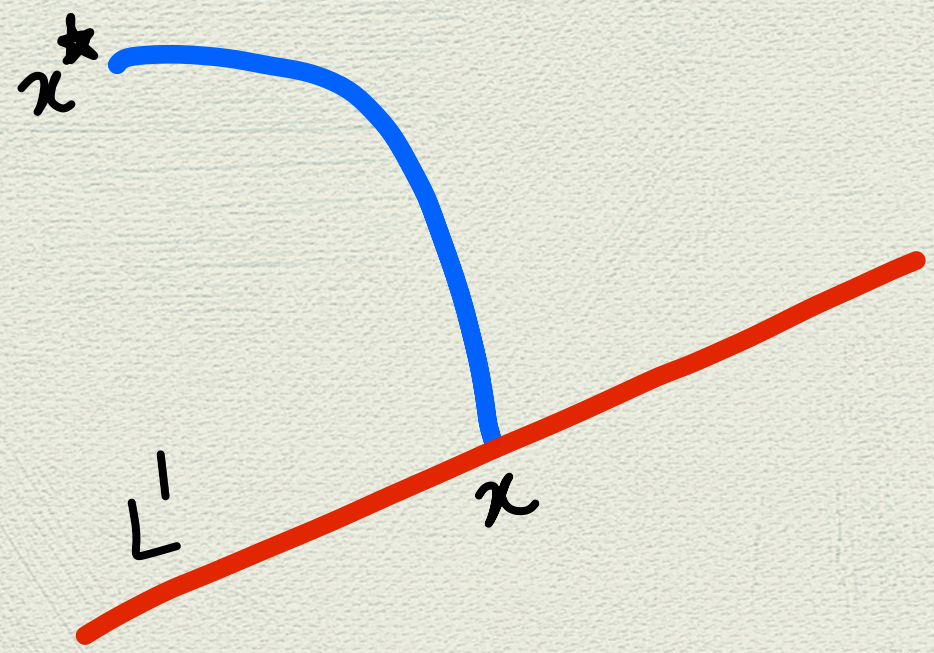
$$L, \tilde{L} \in \mathscr{M},\ L \neq \tilde{L},\ x^\star \notin L \cap \tilde{L} \implies \Pi(L \,|\, x^\star) \neq \Pi(\tilde{L} \,|\, x^\star)$$

(Continuity)

$\Pi(\,\cdot\,|\,x^\star)$ restricted to any fixed dimension is continuous

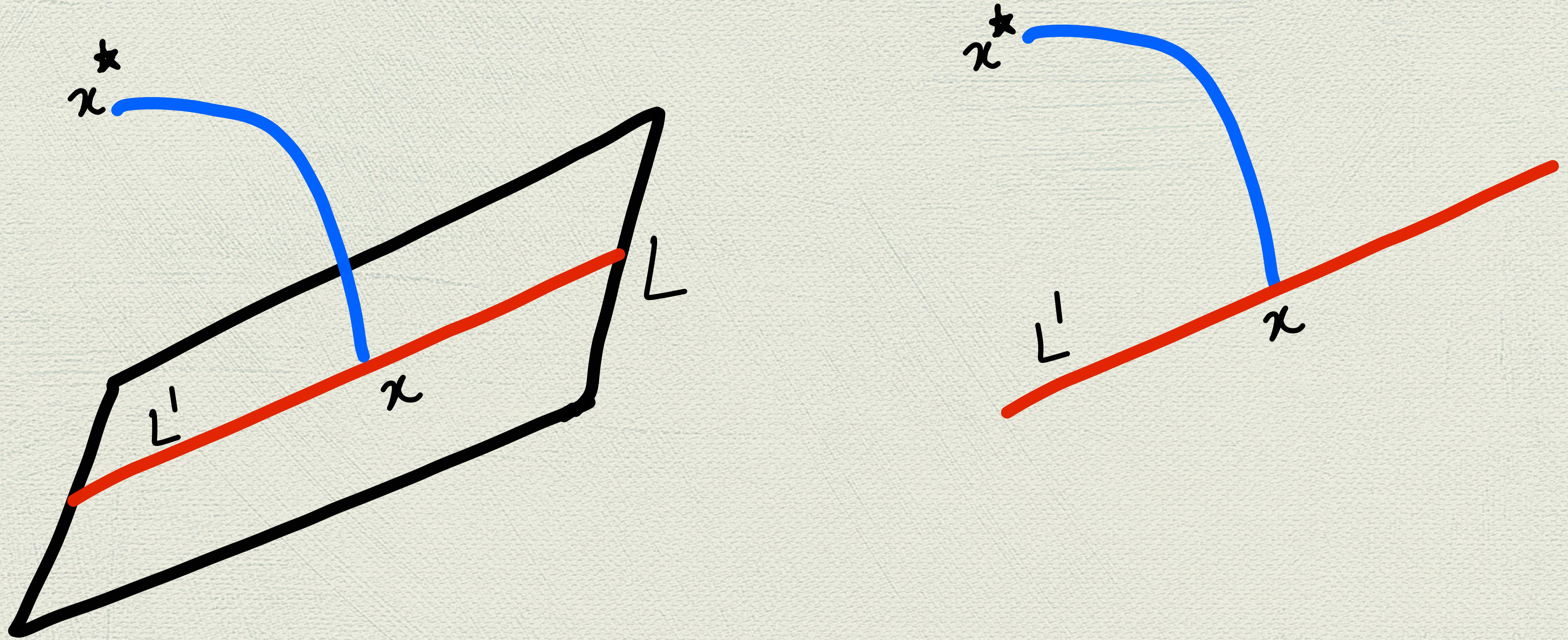
The axiom of consistency





$$L' \subset L,\ \Pi(L \,|\, x^\star) \in L' \implies \Pi(L' \,|\, x^\star) = \Pi(L \,|\, x^\star)$$

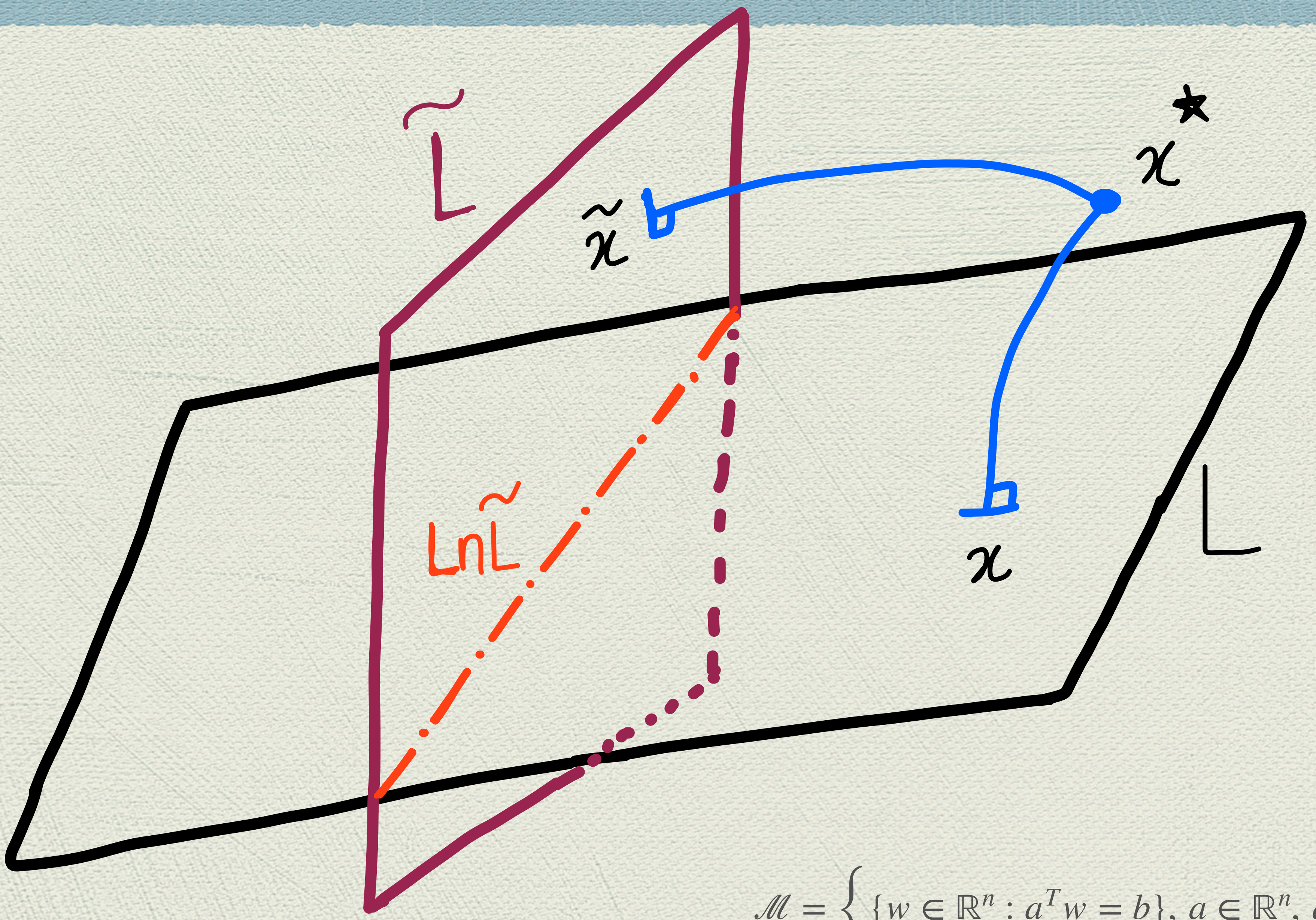
The axiom of consistency



$$L' \subset L,\ \Pi(L\,|\,x^\star) \in L' \implies \Pi(L'\,|\,x^\star) = \Pi(L\,|\,x^\star)$$

Additional observations do not give a reason to change the original selection

The axiom of distinctness

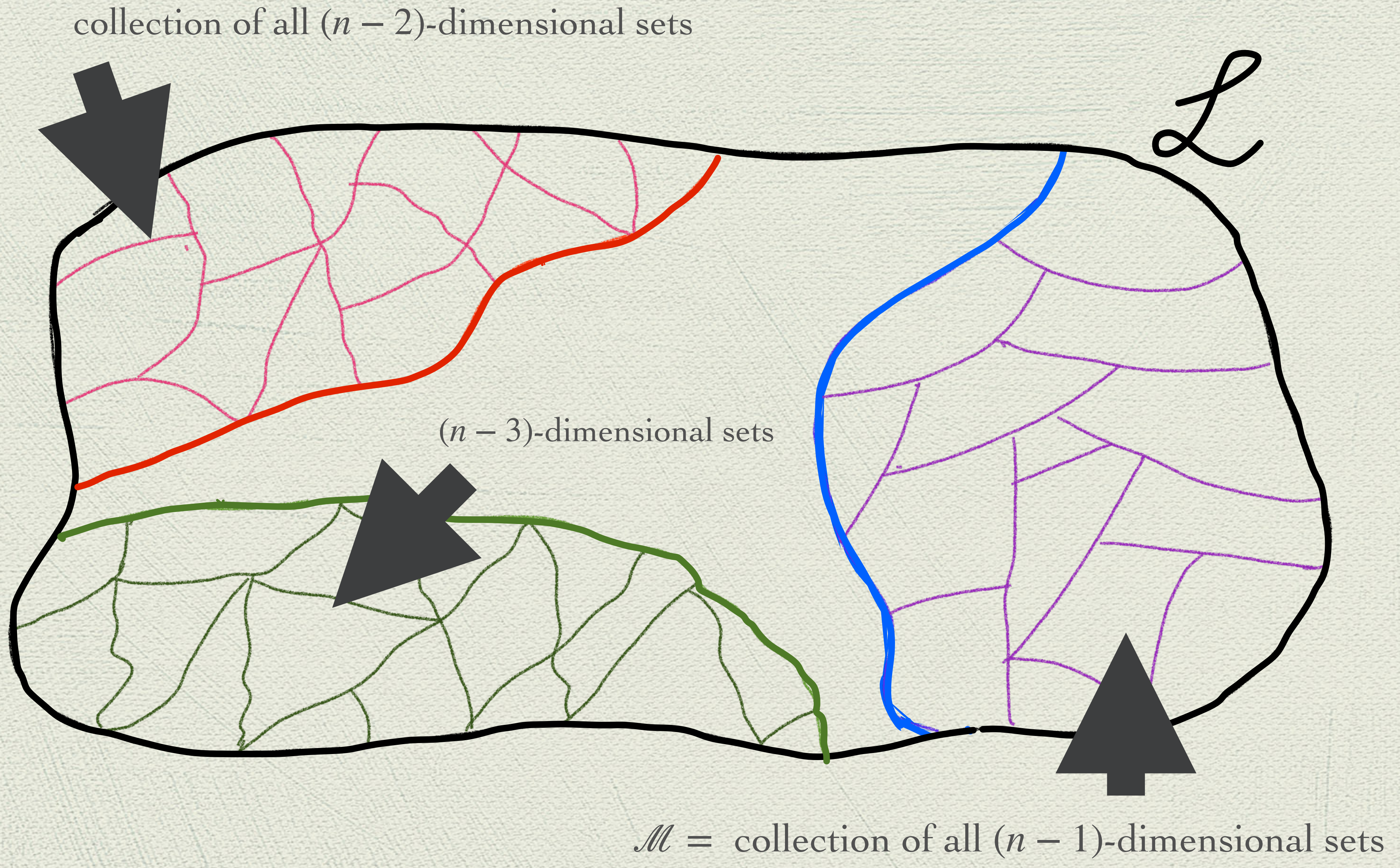


$$\mathscr{M} = \left\{ \{w \in \mathbb{R}^n : a^T w = b\},\ a \in \mathbb{R}^n,\ a \neq 0,\ b \in \mathbb{R} \right\}$$

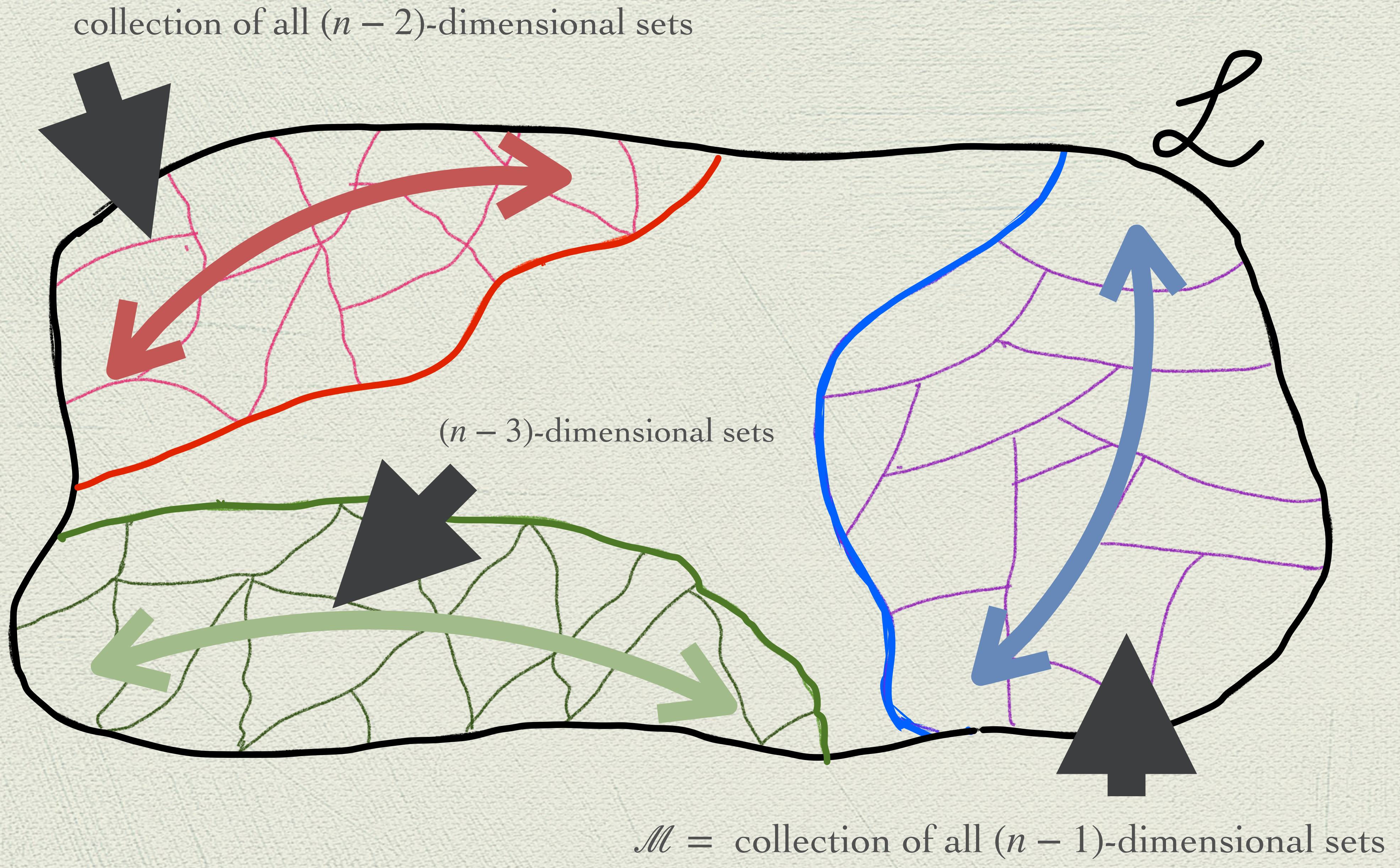
$$L, \tilde{L} \in \mathscr{M},\ L \neq \tilde{L},\ x^\star \notin L \cap \tilde{L} \implies \Pi(L \,|\, x^\star) \neq \Pi(\tilde{L} \,|\, x^\star)$$

The axiom of continuity

The axiom of continuity

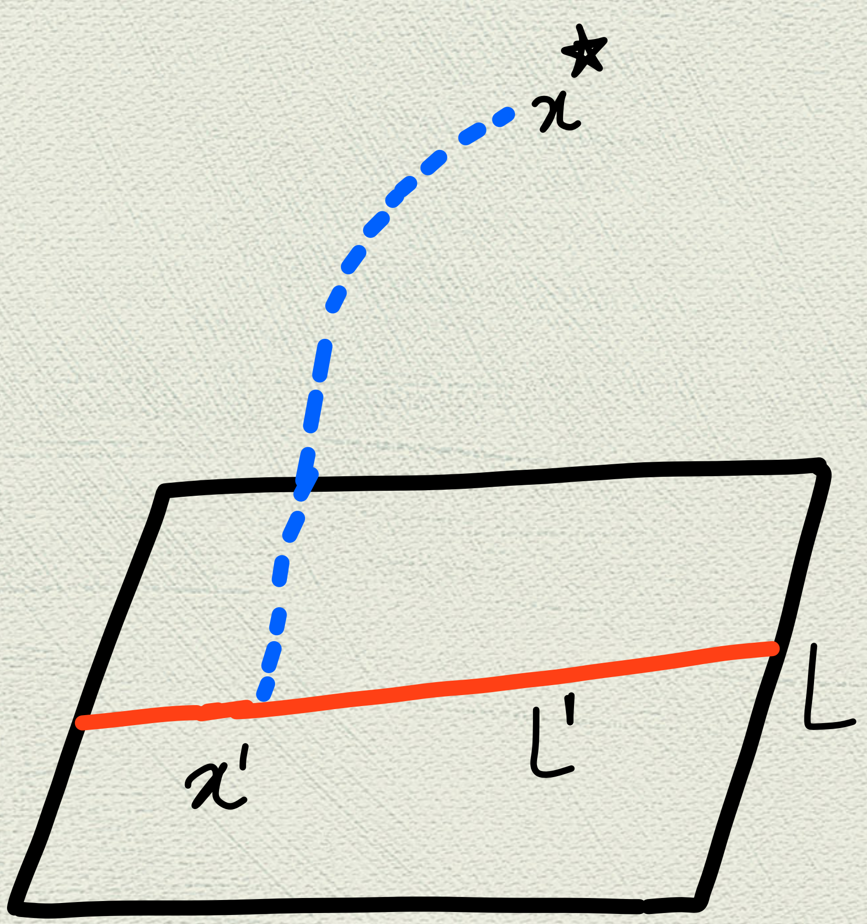


The axiom of continuity

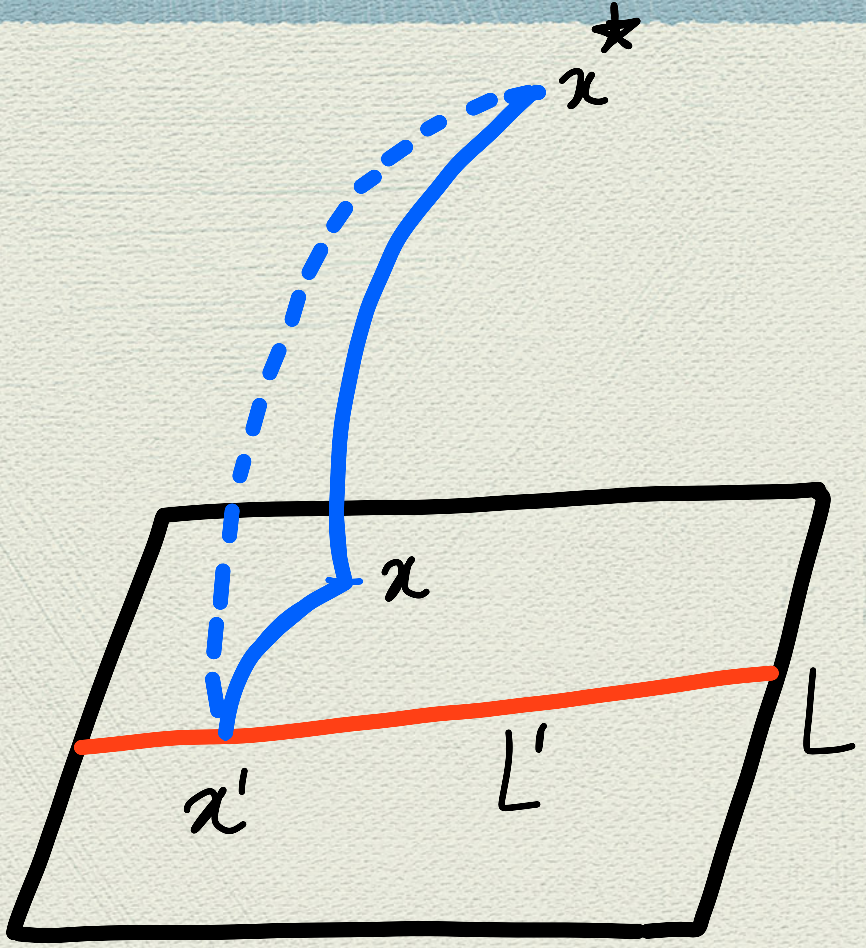


$\Pi(\,\cdot\,|\,x^{\star})$ restricted to any fixed dimension is continuous

Axiom of subspace transitivity



$\equiv$



$$L' \subset L \implies \Pi(L' \,|\, x^\star) = \Pi(\Pi(L \,|\, x^\star) \,|\, x^\star)$$

A Key Implication of Regularity

$$\mathscr{L}^0(x^\star) = \left\{ L \in \mathscr{L} : \Pi(L \,|\, x^\star) = x^\star \right\}, \quad x^\star \in \mathbb{R}^n$$

$$\mathscr{M} = \left\{ \{w \in \mathbb{R}^n : a^T w = b\},\, a \in \mathbb{R}^n,\, a \neq 0,\, b \in \mathbb{R} \right\}$$

A Key Implication of Regularity

$$\mathscr{L}^0(x^\star) = \left\{ L \in \mathscr{L} : \Pi(L \,|\, x^\star) = x^\star \right\}, \quad x^\star \in \mathbb{R}^n$$

$$\mathscr{M} = \left\{ \{w \in \mathbb{R}^n : a^T w = b\}, \, a \in \mathbb{R}^n, \, a \neq 0, \, b \in \mathbb{R} \right\}$$

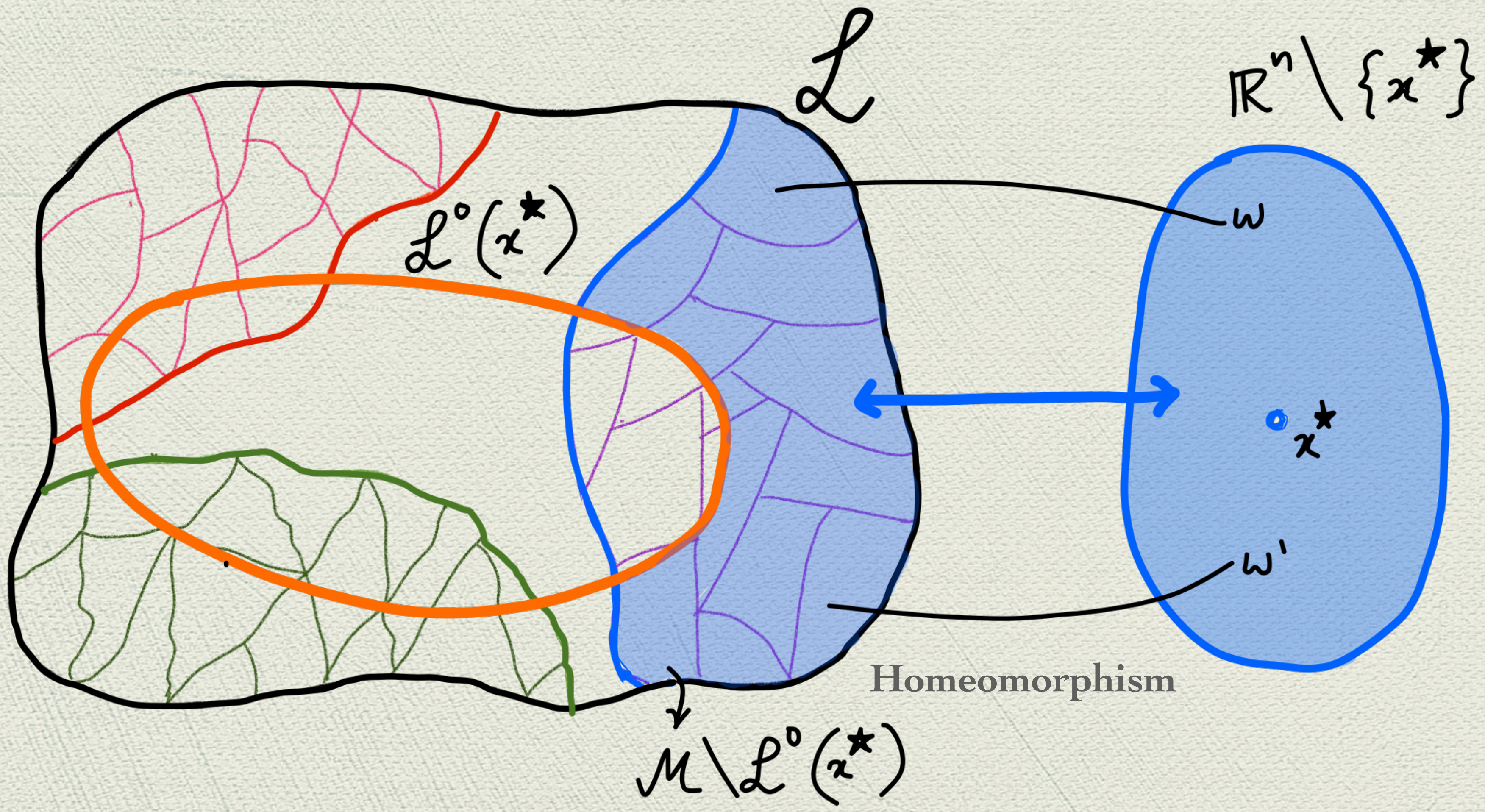
$\Pi(\,\cdot\,|\,x^\star)$: $\mathscr{M} \backslash \mathscr{L}^0(x^\star) \to \mathbb{R}^n \backslash \{x^\star\}$ is a homeomorphism

A Key Implication of Regularity

$\mathscr{L}^0(x^\star) = \left\{ L \in \mathscr{L} : \Pi(L \,|\, x^\star) = x^\star \right\}, \quad x^\star \in \mathbb{R}^n$

$\mathscr{M} = \left\{ \{w \in \mathbb{R}^n : a^T w = b\},\, a \in \mathbb{R}^n,\, a \neq 0,\, b \in \mathbb{R} \right\}$

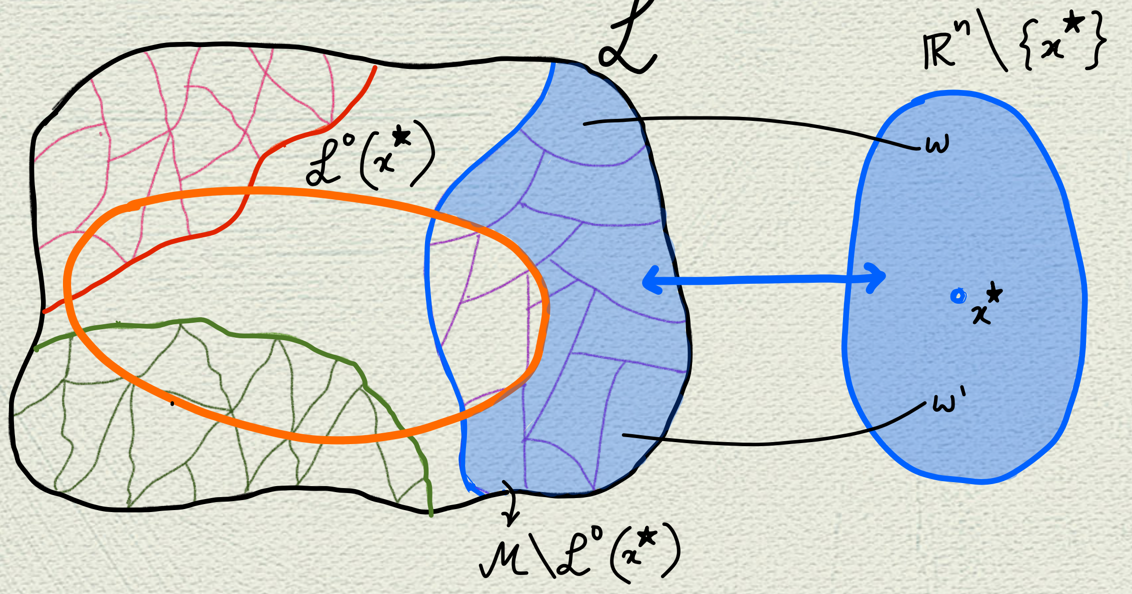
$\Pi(\,\cdot\,|\, x^\star) \colon \mathscr{M} \setminus \mathscr{L}^0(x^\star) \to \mathbb{R}^n \setminus \{x^\star\}$ is a homeomorphism



A Key Implication of Regularity

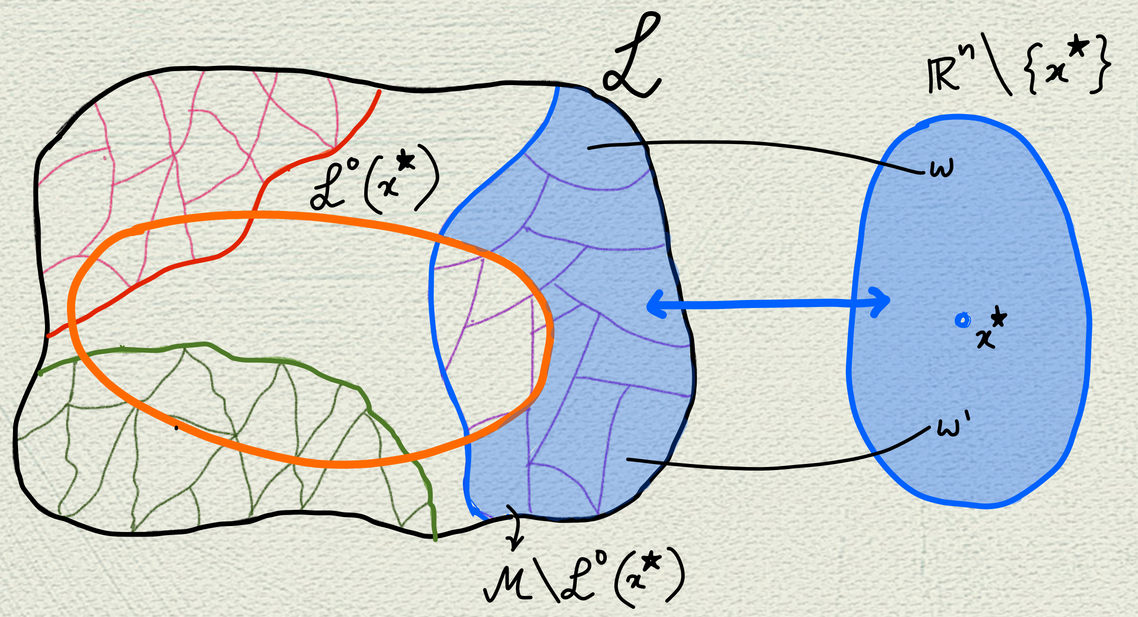
$\Pi(\cdot \mid x^\star)\colon \mathscr{M} \backslash \mathscr{L}^0(x^\star) \to \mathbb{R}^n \backslash \{x^\star\}$ is a homeomorphism

This means that for every $w \neq x^\star$,
there exists a unique set L of dimension $(n-1)$
such that $\Pi(L \mid x^\star) = w$. Denote this L as $L(w \mid x^\star)$

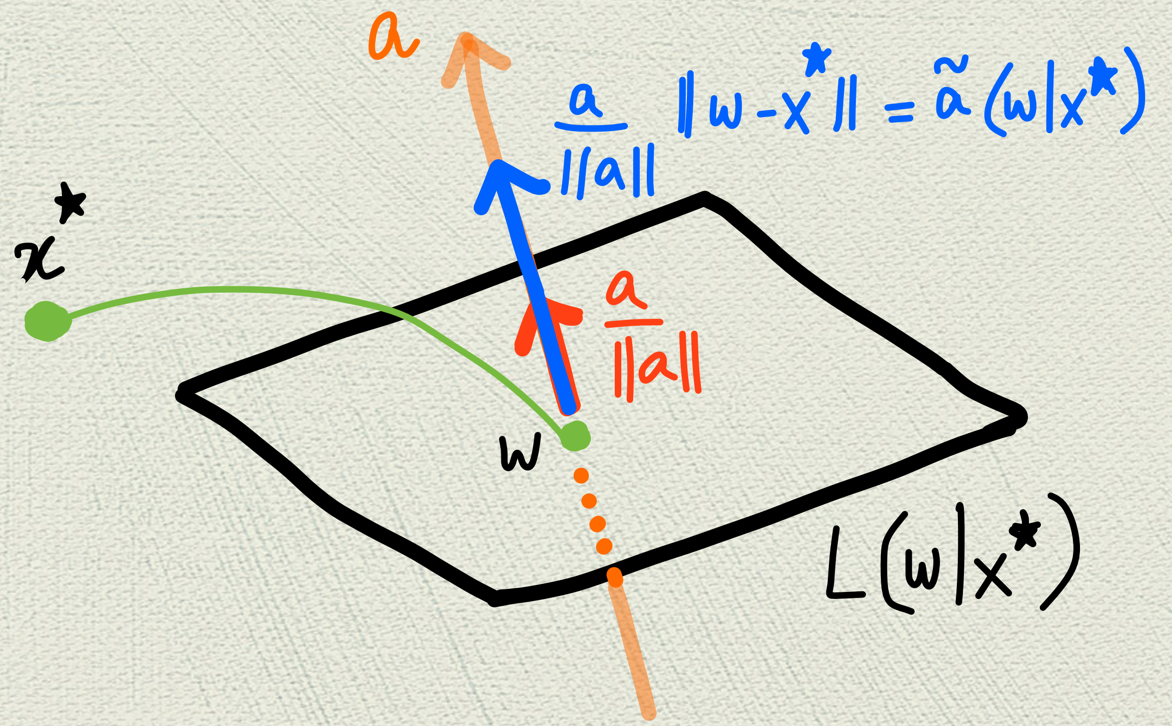


A Key Implication of Regularity

$\Pi(\,\cdot\,|\,x^\star)\colon \mathscr{M}\backslash\mathscr{L}^0(x^\star) \to \mathbb{R}^n\backslash\{x^\star\}$ is a homeomorphism



This means that for every $w \neq x^\star$,
there exists a unique set L of dimension $(n-1)$
such that $\Pi(L\,|\,x^\star) = w$. Denote this L as $L(w\,|\,x^\star)$

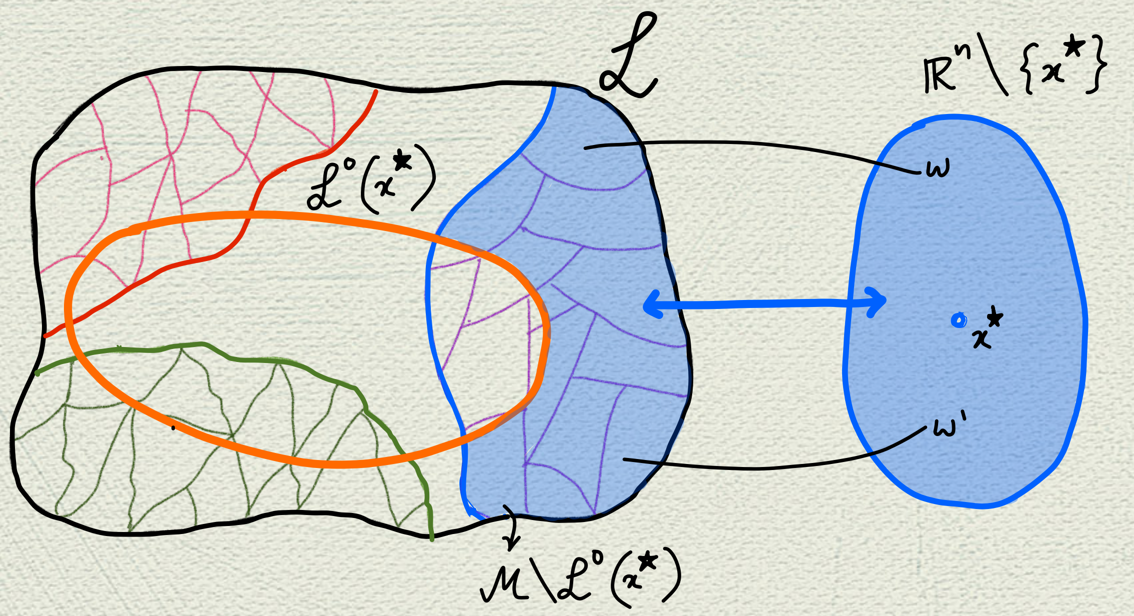


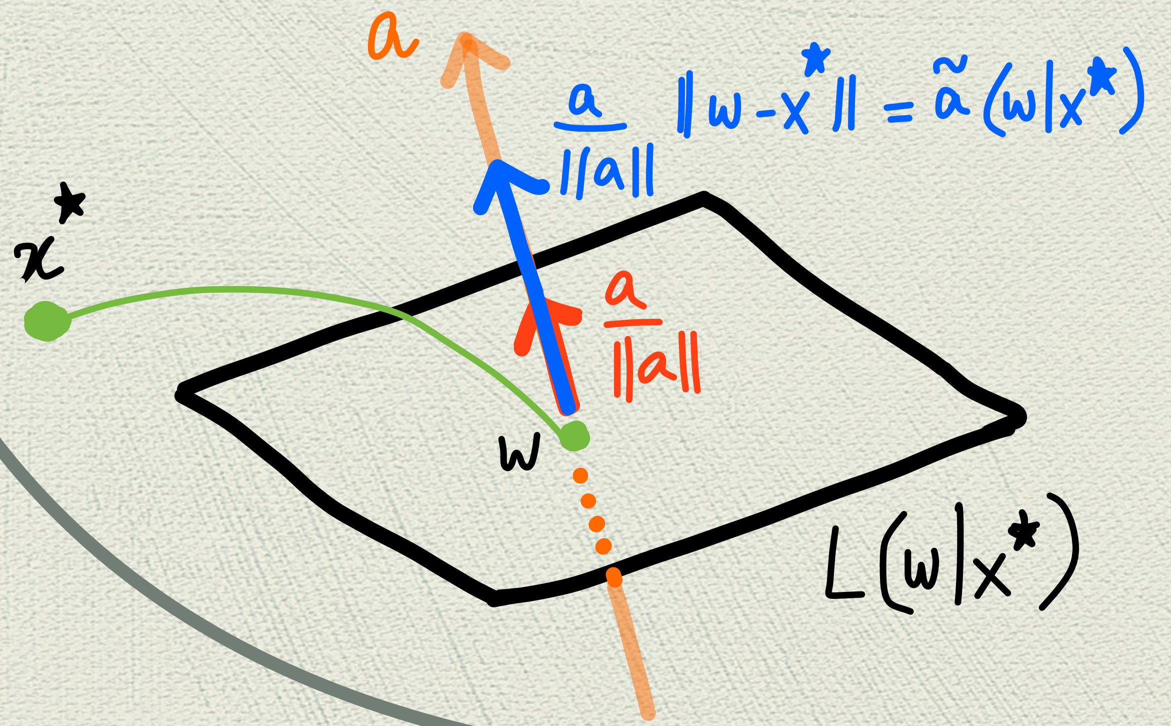
$$
\begin{aligned}
L(w|x^\star) &= \{y \in \mathbb{R}^n : a^T y = b\} \\
&= \{y \in \mathbb{R}^n : a^T y = a^T w\} \\
&= \{y \in \mathbb{R}^n : \tfrac{a^T}{\|a\|}(y-w) = 0\} \\
&= \{y \in \mathbb{R}^n : \tilde{a}(w|x^\star)^T (y-w) = 0\}
\end{aligned}
$$

A Key Implication of Regularity

$\Pi(\,\cdot\,|x^\star): \mathscr{M}\backslash\mathscr{L}^0(x^\star) \to \mathbb{R}^n\backslash\{x^\star\}$ is a homeomorphism

This means that for every $w \neq x^\star$,
there exists a unique set L of dimension $(n-1)$
such that $\Pi(L\,|\,x^\star) = w$. Denote this L as $L(w\,|\,x^\star)$





$$
\begin{aligned}
L(w|x^\star) &= \{y \in \mathbb{R}^n : a^T y = b\} \\
&= \{y \in \mathbb{R}^n : a^T y = a^T w\} \\
&= \{y \in \mathbb{R}^n : \tfrac{a^T}{\|a\|}(y-w) = 0\} \\
&= \{y \in \mathbb{R}^n : \tilde{a}(w|x^\star)^T (y-w) = 0\}
\end{aligned}
$$

$w \mapsto \tilde{a}(w\,|\,x^\star)$ is continuous

Regularity and vector fields

- Mere regularity helps us extract a **vector field**

Regularity and vector fields

- Mere regularity helps us extract a **vector field**

$$\tilde{a}(\,\cdot\,|x^\star) : \mathbb{R}^n \to \mathbb{R}^n$$

$$\tilde{a}(\,\cdot\,|x^\star) \text{ continuous}$$

$$\tilde{a}(x^\star|x^\star) = 0,$$

$$\tilde{a}(w|x^\star) \neq 0 \text{ for all } w \neq x^\star$$

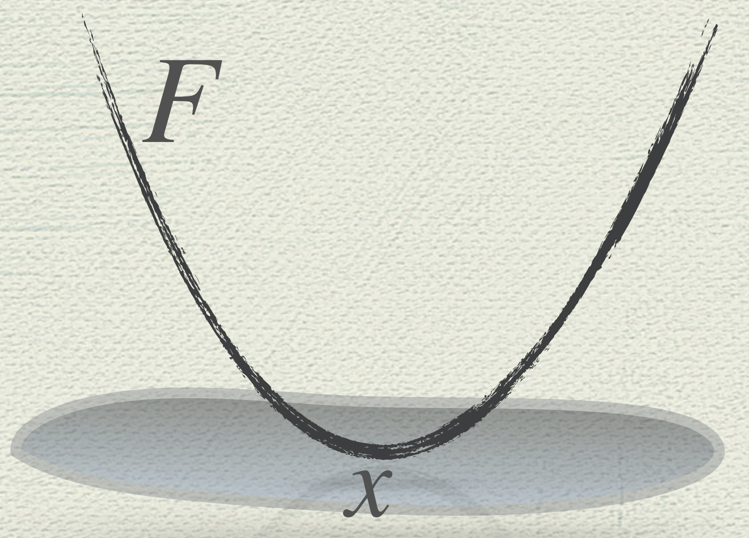
Existence of a Non-Zero Scaling Function

$\tilde{a}(\cdot\,|\,x^\star) : \mathbb{R}^n \to \mathbb{R}^n$

$\tilde{a}(\cdot\,|\,x^\star)$ continuous

$\tilde{a}(x^\star\,|\,x^\star) = 0,$

$\tilde{a}(w\,|\,x^\star) \neq 0$ for all $w \neq x^\star$



$$\min_{w \in L} F(w)$$

$$L = \{w \in \mathbb{R}^n : a^T w = b\}$$

$$G(w, \lambda) = F(w) - \lambda(a^T w - b)$$

$$\nabla_w G(w, \lambda) = 0 \iff \boxed{\nabla F(w) = \lambda a}$$

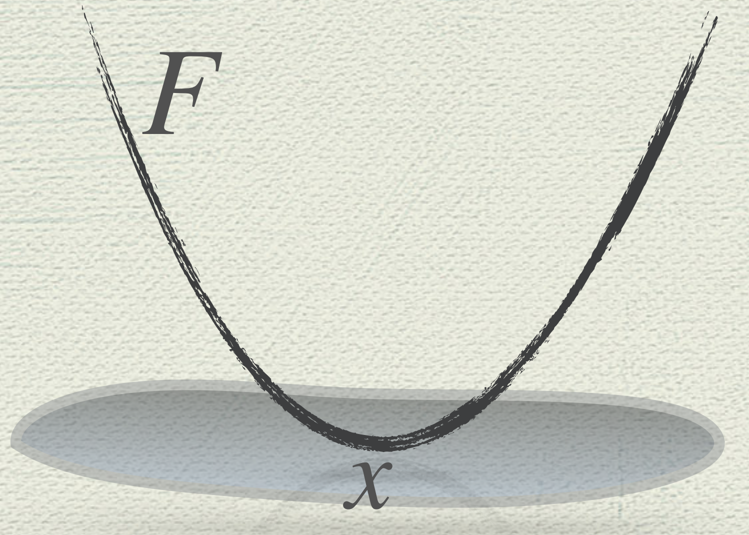
Existence of a Non-Zero Scaling Function

$\tilde{a}(\,\cdot\,|\,x^\star) : \mathbb{R}^n \to \mathbb{R}^n$

$\tilde{a}(\,\cdot\,|\,x^\star)$ continuous

$\tilde{a}(x^\star\,|\,x^\star) = 0,$

$\tilde{a}(w\,|\,x^\star) \neq 0$ for all $w \neq x^\star$



$$\min_{w \in L} F(w)$$

$$L = \{w \in \mathbb{R}^n : a^T w = b\}$$

Is there $\lambda : \mathbb{R}^n \to \mathbb{R}$ such that $\lambda(w) \cdot \tilde{a}(w\,|\,x^\star) = \nabla F(w\,|\,x^\star)$?

$$G(w, \lambda) = F(w) - \lambda(a^T w - b)$$

$$\nabla_w G(w, \lambda) = 0 \iff \boxed{\nabla F(w) = \lambda a}$$

Conservative Vector Fields and Regularity

Conservative Vector Fields and Regularity

$\boxed{\Pi(\,\cdot\,|\,x^{\star})\text{ regular}}$

Conservative Vector Fields and Regularity

$\boxed{\Pi(\,\cdot\,|\,x^\star)\text{ regular}} \longrightarrow$

Conservative Vector Fields and Regularity

$\boxed{\Pi(\,\cdot\,|\,x^\star) \text{ regular}} \longrightarrow \boxed{\tilde{a}(\,\cdot\,|\,x^\star) \text{ continuous vector field}}$

Conservative Vector Fields and Regularity

$\boxed{\Pi(\,\cdot\,|\,x^\star)\text{ regular}} \longrightarrow \boxed{\tilde{a}(\,\cdot\,|\,x^\star)\text{ continuous vector field}}$

$\downarrow$

Conservative Vector Fields and Regularity

$\boxed{\Pi(\,\cdot\,|\,x^\star)\text{ regular}} \longrightarrow \boxed{\tilde{a}(\,\cdot\,|\,x^\star)\text{ continuous vector field}}$

$\downarrow$

Does there exist a nonzero scaling $\lambda(\,\cdot\,|\,x^\star)$

$\lambda(w\,|\,x^\star) \neq 0$ for all $w \neq x^\star$,

$\lambda(\,\cdot\,|\,x^\star)$ continuous except at $x^\star$ such that

$\lambda(w\,|\,x^\star)\tilde{a}(w\,|\,x^\star) = \nabla F(w\,|\,x^\star)$ for some smooth F

Conservative Vector Fields and Regularity

$\boxed{\Pi(\,\cdot\,|\,x^\star) \text{ regular}}$ $\longrightarrow$ $\boxed{\tilde{a}(\,\cdot\,|\,x^\star) \text{ continuous vector field}}$

$\downarrow$

$\longleftarrow$

Does there exist a nonzero scaling $\lambda(\,\cdot\,|\,x^\star)$

$\lambda(w\,|\,x^\star) \neq 0$ for all $w \neq x^\star$,

$\lambda(\,\cdot\,|\,x^\star)$ continuous except at $x^\star$ such that

$\lambda(w\,|\,x^\star)\tilde{a}(w\,|\,x^\star) = \nabla F(w\,|\,x^\star)$ for some smooth F

Conservative Vector Fields and Regularity

$\Pi(\cdot \mid x^\star)$ regular $\longrightarrow$ $\tilde{a}(\cdot \mid x^\star)$ continuous vector field

$\downarrow$

Does there exist a nonzero scaling $\lambda(\cdot \mid x^\star)$
$\lambda(w \mid x^\star) \neq 0$ for all $w \neq x^\star$,
$\lambda(\cdot \mid x^\star)$ continuous except at $x^\star$ such that
$\lambda(w \mid x^\star)\tilde{a}(w \mid x^\star) = \nabla F(w \mid x^\star)$ for some smooth F

$\longleftarrow$

Projection rule generated by F

Conservative Vector Fields and Regularity

$\Pi(\,\cdot\,|\,x^\star)$ regular → $\tilde{a}(\,\cdot\,|\,x^\star)$ continuous vector field

↓

Does there exist a nonzero scaling $\lambda(\,\cdot\,|\,x^\star)$

$\lambda(w\,|\,x^\star) \neq 0$ for all $w \neq x^\star$,

$\lambda(\,\cdot\,|\,x^\star)$ continuous except at $x^\star$ such that

$\lambda(w\,|\,x^\star)\tilde{a}(w\,|\,x^\star) = \nabla F(w\,|\,x^\star)$ for some smooth F

←

Projection rule generated by F

😀

Conservative Vector Fields and Regularity

$\Pi(\cdot\,|\,x^\star)$ regular → $\tilde{a}(\cdot\,|\,x^\star)$ continuous vector field

Open question

Does there exist a nonzero scaling $\lambda(\cdot\,|\,x^\star)$

$\lambda(w\,|\,x^\star) \neq 0$ for all $w \neq x^\star$,

$\lambda(\cdot\,|\,x^\star)$ continuous except at $x^\star$ such that

$\lambda(w\,|\,x^\star)\tilde{a}(w\,|\,x^\star) = \nabla F(w\,|\,x^\star)$ for some smooth F

→ Projection rule generated by F

An Open Question

An Open Question

Given a continuous vector field, does there exist a scaling function such that the **scaled vector field is conservative**?

An Open Question

Given a continuous vector field, does there exist a scaling function such that the scaled vector field is conservative?

$a : \mathbb{R}^n \to \mathbb{R}^n$ continuous

$\lambda : \mathbb{R}^n \to \mathbb{R}$

$\lambda \cdot a = \nabla F$?

Instances Where the Answers are Known / Not Unknown

Given a continuous vector field, does there exist a scaling function such that the scaled vector field is conservative?

$a : \mathbb{R}^n \to \mathbb{R}^n$ continuous

$\lambda : \mathbb{R}^n \to \mathbb{R}$

$\lambda \cdot a = \nabla F$?

Instances Where the Answers are Known / Not Unknown

$$a(w) = \frac{w}{\|w\|} \qquad \lambda(w) = \|w\|$$

Given a continuous vector field, does there exist a scaling function such that the scaled vector field is conservative?

$a : \mathbb{R}^n \to \mathbb{R}^n$ continuous

$\lambda : \mathbb{R}^n \to \mathbb{R}$

$\lambda \cdot a = \nabla F?$

Instances Where the Answers are Known / Not Unknown

Given a continuous vector field, does there exist a scaling function such that the scaled vector field is conservative?

$$a : \mathbb{R}^n \to \mathbb{R}^n \text{ continuous}$$

$$\lambda : \mathbb{R}^n \to \mathbb{R}$$

$$\lambda \cdot a = \nabla F?$$

$$a(w) = \frac{w}{\|w\|} \qquad \lambda(w) = \|w\|$$

$$\lambda(w)\ a(w) = \nabla F(w) \qquad F(w) = \frac{1}{2}\|w\|^2$$

Instances Where the Answers are Known / Not Unknown

Given a continuous vector field, does there exist a scaling function such that the scaled vector field is conservative?

$$a : \mathbb{R}^n \to \mathbb{R}^n \text{ continuous}$$

$$\lambda : \mathbb{R}^n \to \mathbb{R}$$

$$\lambda \cdot a = \nabla F ?$$

$$a(w) = \frac{w}{\|w\|} \qquad \lambda(w) = \|w\|$$

$$\lambda(w)\ a(w) = \nabla F(w) \qquad F(w) = \frac{1}{2}\|w\|^2$$

$$a(x, y, z) = \begin{pmatrix} y \\ 0 \\ 1 \end{pmatrix} \implies \text{there is no } \lambda \text{ such that } \lambda \cdot a = \nabla F$$

Instances Where the Answers are Known / Not Unknown

Given a continuous vector field, does there exist a scaling function such that the scaled vector field is conservative?

$$a(w) = \frac{w}{\|w\|} \qquad \lambda(w) = \|w\|$$

$$\lambda(w)\ a(w) = \nabla F(w) \qquad F(w) = \frac{1}{2}\|w\|^2$$

$a : \mathbb{R}^n \to \mathbb{R}^n$ continuous

$\lambda : \mathbb{R}^n \to \mathbb{R}$

$\lambda \cdot a = \nabla F$?

$$\boxed{a(x,y,z) = \begin{pmatrix} y \\ 0 \\ 1 \end{pmatrix}} \Longrightarrow \text{ there is no } \lambda \text{ such that } \lambda \cdot a = \nabla F$$

Can this arise from a regular projection rule?

Summary

- Two approaches for solving systems of linear equations:
  - Function-minimisation approach
  - Axiomatic approach

Projection rules satisfies the axioms of	Projection rule is generated by
Regularity + Locality	$F(w \mid x^\star) = \sum_{i=1}^{n} f_i(w_i \mid x_i^\star)$, f_i continuously differentiable and strictly convex
Regularity+ Locality + Subspace Transitivity	$F(w \mid x^\star) =$ Bregman's divergence
Regularity+ Locality + Subspace Transitivity + Statistical	$F(w \mid x^\star) = \sum_{i=1}^{n} w_i \log \frac{w_i}{x_i^\star}$
Regularity + Locality + Subspace Transitivity + Location Invariance + Scale Invariance	$F(w \mid x^\star) = \sum_{i=1}^{n} (w_i - x_i^\star)^2$

- Regularity: a fundamental axiom
- Regularity has connections with conservative vector fields
- Given a continuous vector field that is not necessarily conservative, is there a continuous scaling that can make the product vector field conservative? Open!

Thank you!

karthik@nus.edu.sg