$Ax=b$: A Familiar Setup, Axioms and An Open Question

Karthik P. N.

Joint work with Prof. Rajesh Sundaresan

An example

An example

- Consider a problem of image reconstruction

An example

- Consider a problem of image reconstruction
- Image is represented as a function

An example

- Consider a problem of image reconstruction
- Image is represented as a function
- Domain is partitioned into a finite number of "pixels"

An example

- Consider a problem of image reconstruction
- Image is represented as a function
- Domain is partitioned into a finite number of "pixels"

$$f = \sum_{j=1}^{n} x_j f_j, \quad f_j = \text{ indicator of the } j\text{th pixel}$$

- Measurements

- Measurements $R_i f_j = a_{ij}, \quad R_i f = b_i, \quad i = 1,2,\ldots,k$

- Measurements $R_i f_j = a_{ij}, \quad R_i f = b_i, \quad i = 1,2,\ldots,k$

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

- Measurements $\quad R_i f_j = a_{ij}, \quad R_i f = b_i, \quad i = 1,2,\ldots,k$

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

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

- Measurements $R_i f_j = a_{ij}, \quad R_i f = b_i, \quad i = 1,2,\ldots,k$

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

$$✓A \qquad x \qquad b✓$$

- Measurements $\quad R_i f_j = a_{ij}, \quad R_i f = b_i, \quad i = 1,2,\ldots,k$

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

$A \qquad x \qquad b$

- Measurements $R_i f_j = a_{ij}, \quad R_i f = b_i, \quad i = 1,2,\ldots,k$

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

$$A \qquad x \qquad b$$

- Goal: Recover $x = (x_1, \ldots, x_n)^T$ from the measurements

Criterion for / goodness of recovery

Criterion for / goodness of recovery

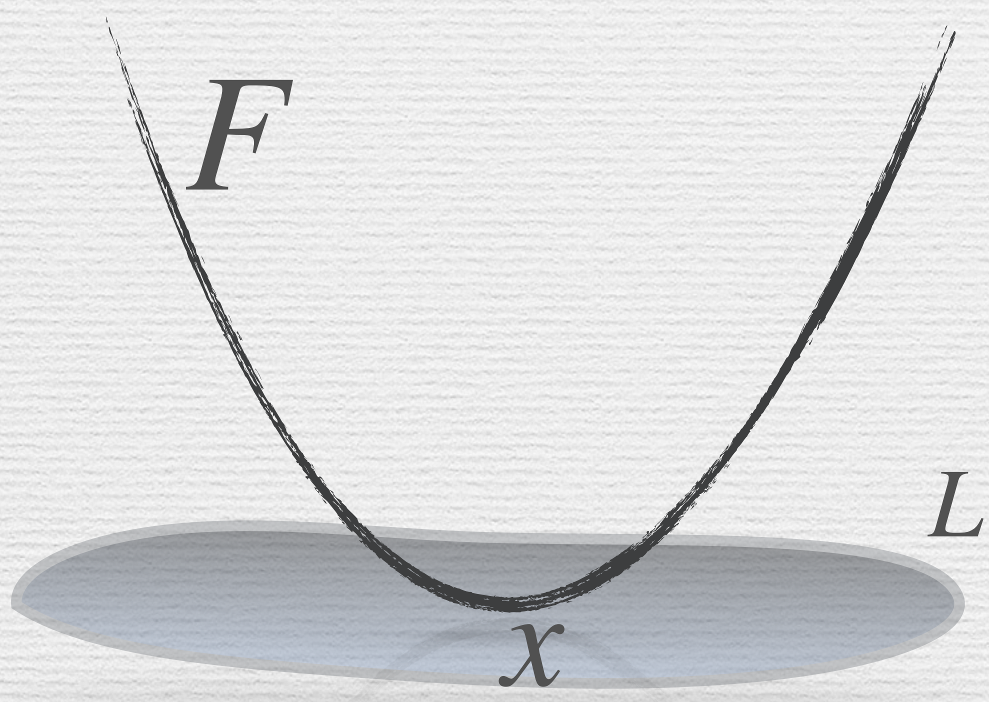
- Typically, one first defines a cost function

Criterion for / goodness of recovery

- Typically, one first defines a cost function
- Subsequently, x is recovered as one of the minimisers of the cost function

Criterion for / goodness of recovery

- Typically, one first defines a cost function
- Subsequently, x is recovered as one of the minimisers of the cost function



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

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

Function-minimisation approach: examples

Function-minimisation approach: examples

- Least squares:

Function-minimisation approach: examples

- Least squares:

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

Function-minimisation approach: examples

- Least squares:

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

- Maximum entropy:

Function-minimisation approach: examples

- Least squares:

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

- Maximum entropy:

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

Function-minimisation approach: examples

Function-minimisation approach: examples

- Mean squared error (MSE) minimisation with respect to a prior guess

Function-minimisation approach: examples

- Mean squared error (MSE) minimisation with respect to a prior guess

$$x = \arg\min_{w \in L} ||w - x^\star||^2 = \arg\min_{w \in L} \sum_{i=1}^{n} (w_i - x_i^\star)^2$$

Function-minimisation approach: examples

- Mean squared error (MSE) minimisation with respect to a prior guess

$$x = \arg\min_{w \in L} ||w - x^\star||^2 = \arg\min_{w \in L} \sum_{i=1}^{n} (w_i - x_i^\star)^2$$

- I-divergence minimisation with respect to a <u>nonnegative</u> prior guess

Function-minimisation approach: examples

- Mean squared error (MSE) minimisation with respect to a prior guess

$$x = \arg\min_{w \in L} ||w - x^\star||^2 = \arg\min_{w \in L} \sum_{i=1}^{n} (w_i - x_i^\star)^2$$

- I-divergence minimisation with respect to a nonnegative prior guess

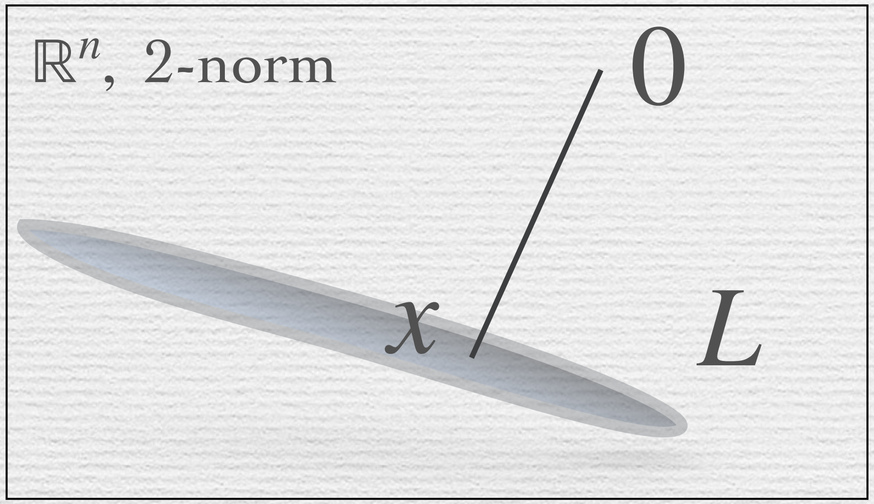
$$x = \arg \min_{w \in L \cap \mathbb{R}_+^n} \sum_{i=1}^{n} w_i \log \frac{w_i}{x_i^\star} - w_i + x_i^\star$$

Function-minimisation approach

- The aforementioned methods are examples of "projection rules" that involve updating some prior guess

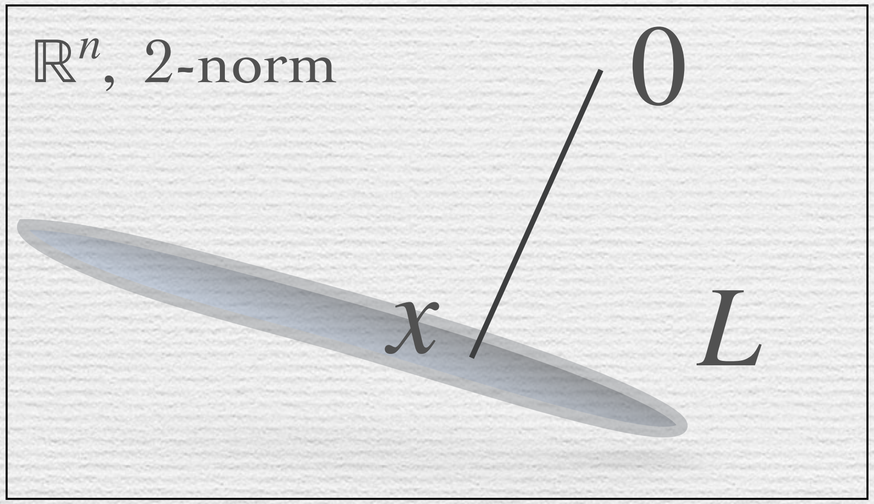
Function-minimisation approach

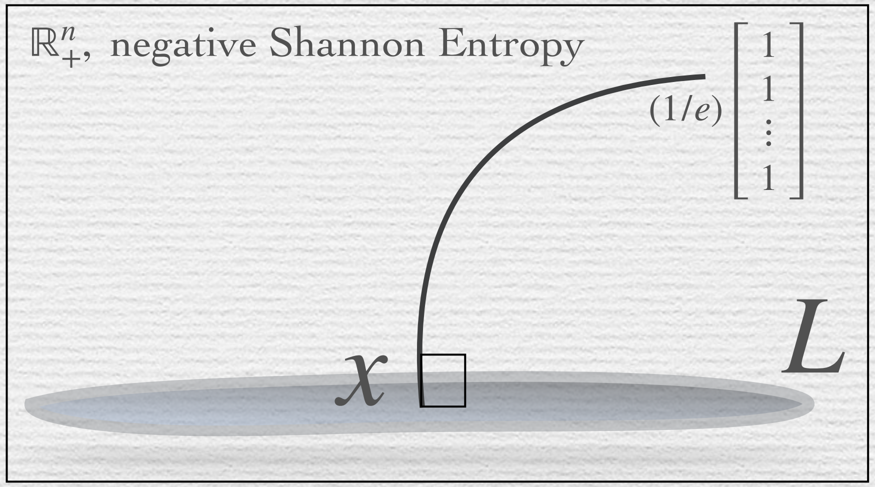
- The aforementioned methods are examples of "projection rules" that involve updating some prior guess



Function-minimisation approach

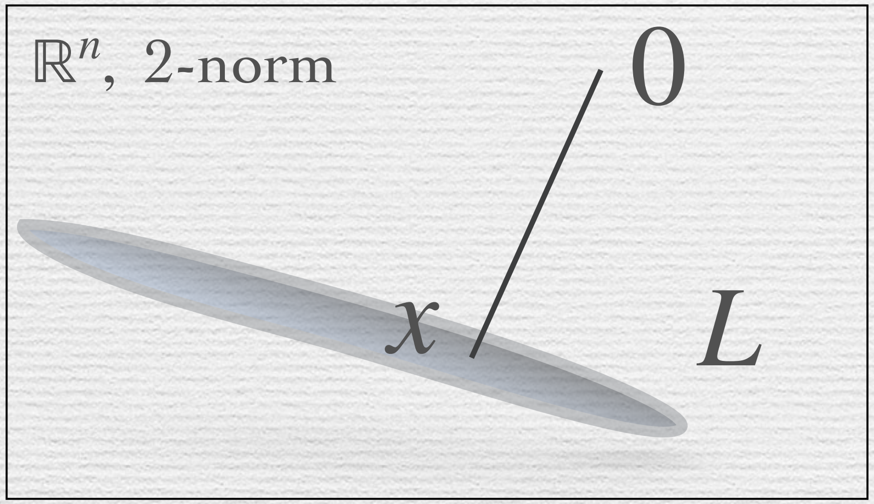
- The aforementioned methods are examples of "projection rules" that involve updating some prior guess

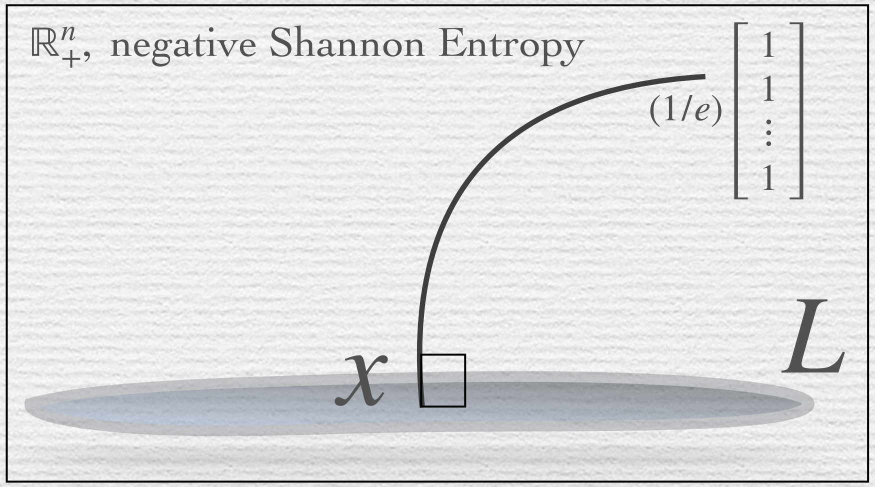


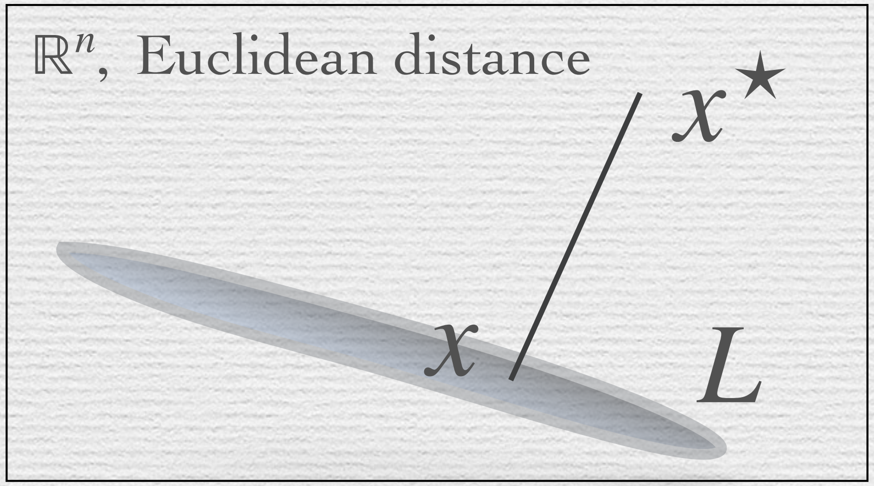


Function-minimisation approach

- The aforementioned methods are examples of "projection rules" that involve updating some prior guess

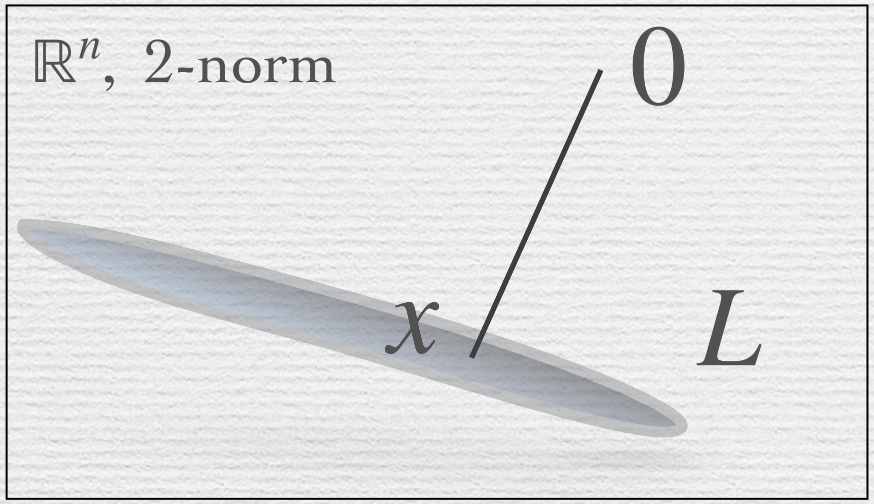


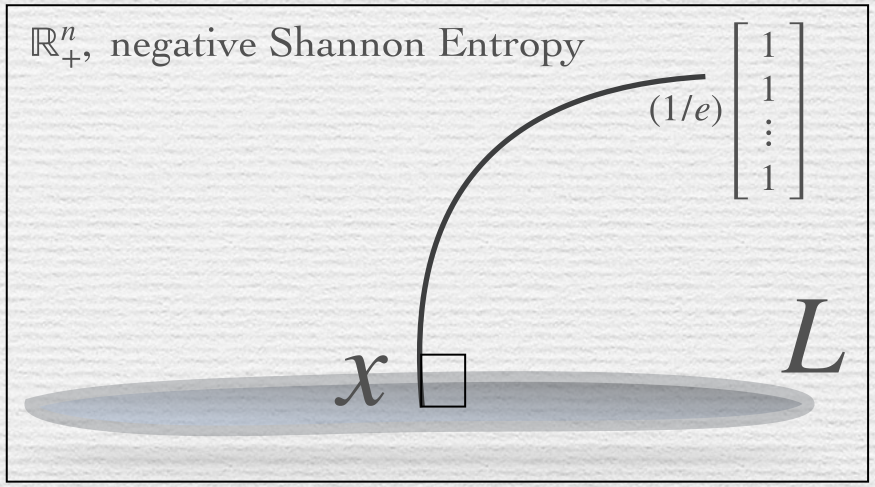


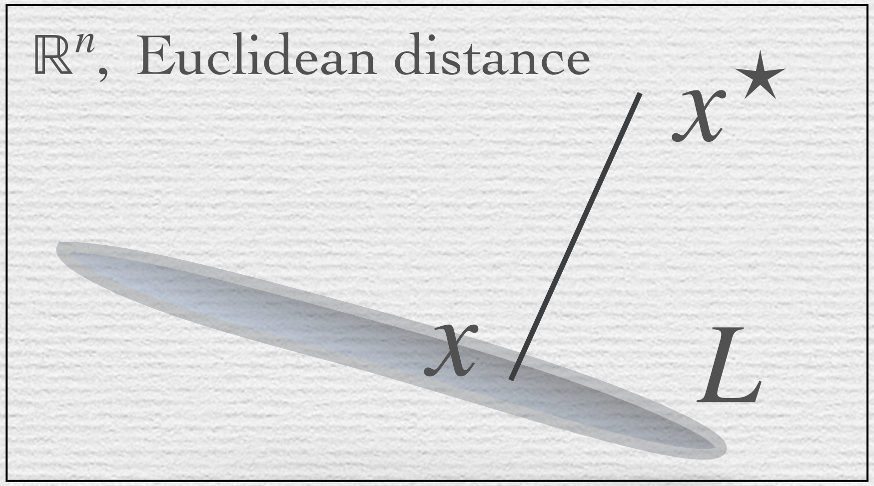


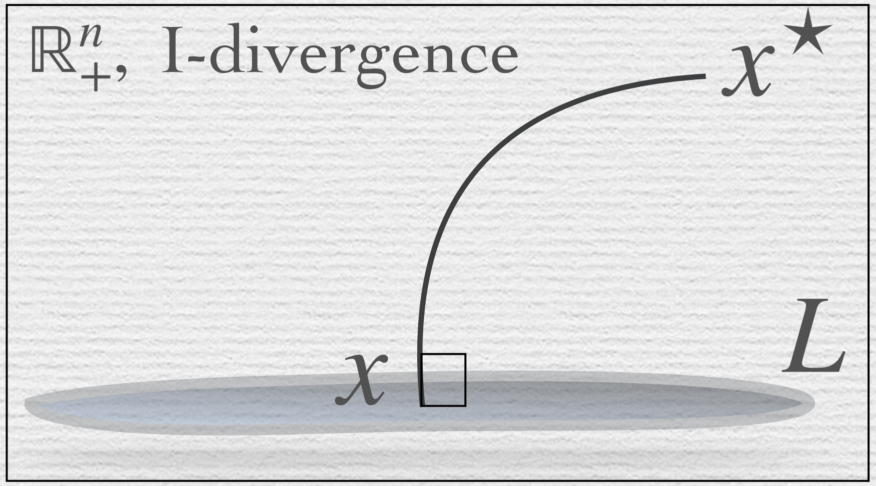
Function-minimisation approach

- The aforementioned methods are examples of "projection rules" that involve updating some prior guess



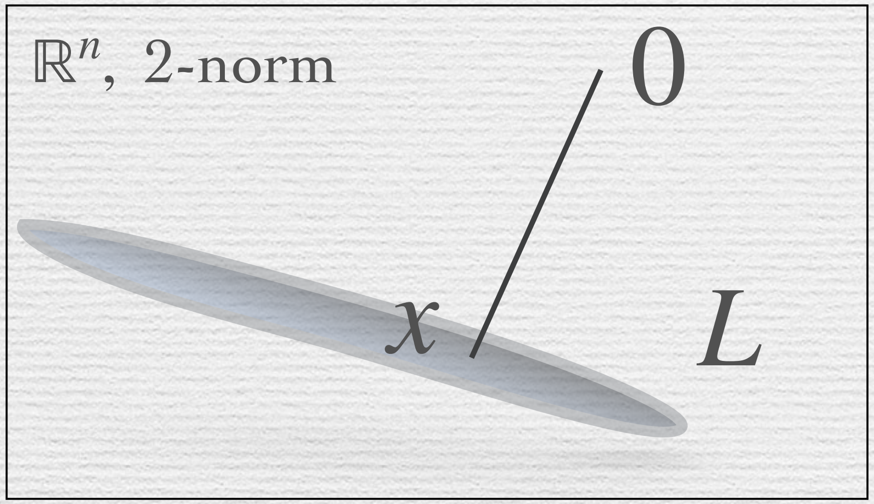


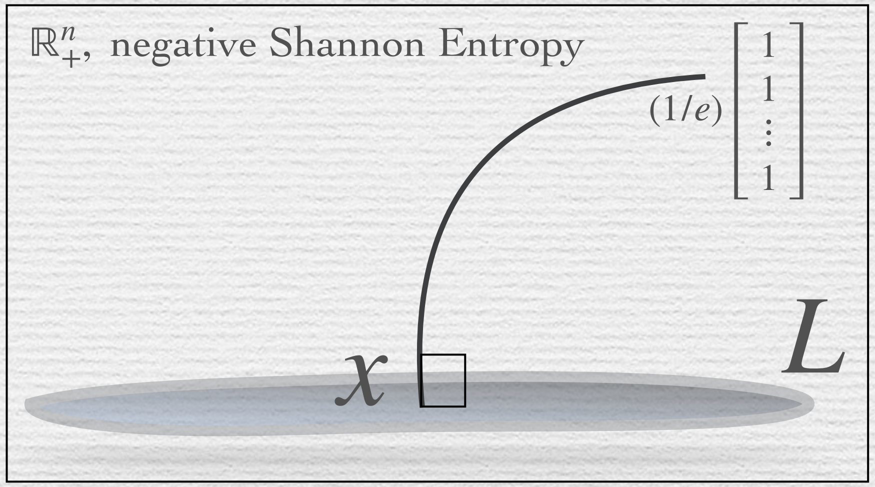




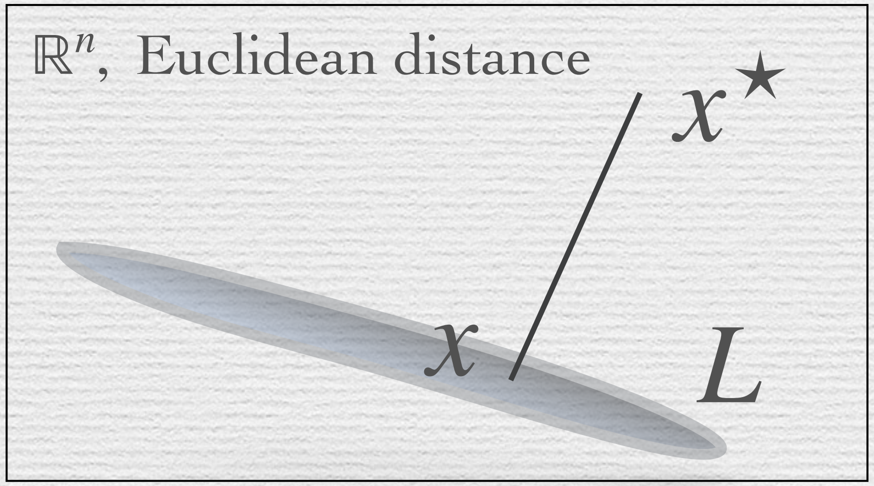
Function-minimisation approach

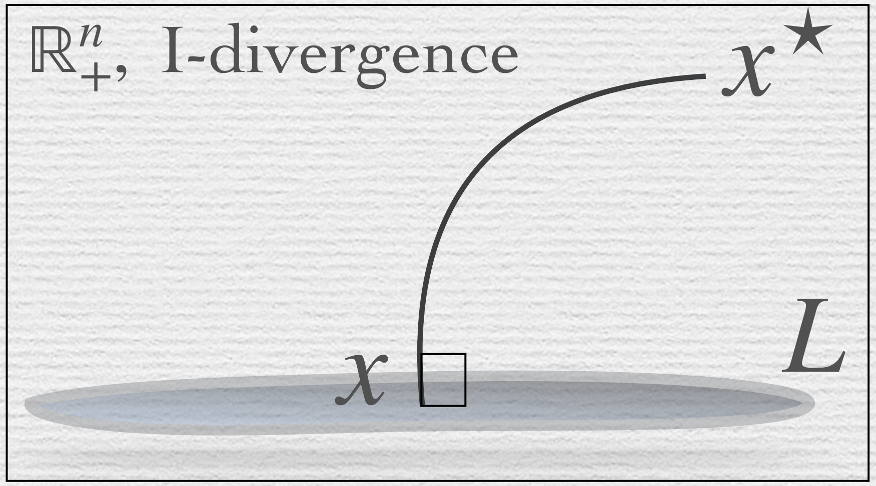
- The aforementioned methods are examples of "projection rules" that involve updating some prior guess





Thus, function-minimisation defines a projection rule





Projection rules: formal definition

- Let

$$\mathscr{L} = \left\{ L = \{w \in \mathbb{R}^n : Aw = b\} : A \text{ is a } k \times n \text{ matrix having rank } k,\ b \in \mathbb{R}^k \right\}$$

where $k = 0,1,2,\ldots,n$

Projection rules: formal definition

- Let

$$\mathscr{L} = \left\{ L = \{w \in \mathbb{R}^n : Aw = b\} : A \text{ is a } k \times n \text{ matrix having rank } k,\ b \in \mathbb{R}^k \right\}$$

where $k = 0,1,2,\ldots,n$

- A projection rule is a mapping

Projection rules: formal definition

- Let

$$\mathscr{L} = \left\{ L = \{w \in \mathbb{R}^n : Aw = b\} : A \text{ is a } k \times n \text{ matrix having rank } k,\ b \in \mathbb{R}^k \right\}$$

$$\text{where } k = 0,1,2,\ldots,n$$

- A projection rule is a mapping

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

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

Projection rules: formal definition

- Let

$$\mathscr{L} = \left\{ L = \{w \in \mathbb{R}^n : Aw = b\} : A \text{ is a } k \times n \text{ matrix having rank } k,\ b \in \mathbb{R}^k \right\}$$

$$\text{where } k = 0,1,2,\ldots,n$$

- A projection rule is a mapping

$$\Pi : \mathscr{L} \times \mathbb{R}^n \to \mathbb{R}^n$$
$$(L, x^\star) \mapsto \Pi(L \mid x^\star) \in L$$

$$\text{If } x^\star \in L, \text{ then } \Pi(L \mid x^\star) = x^\star$$

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

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

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

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

Definition: Π is generated by F, if

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

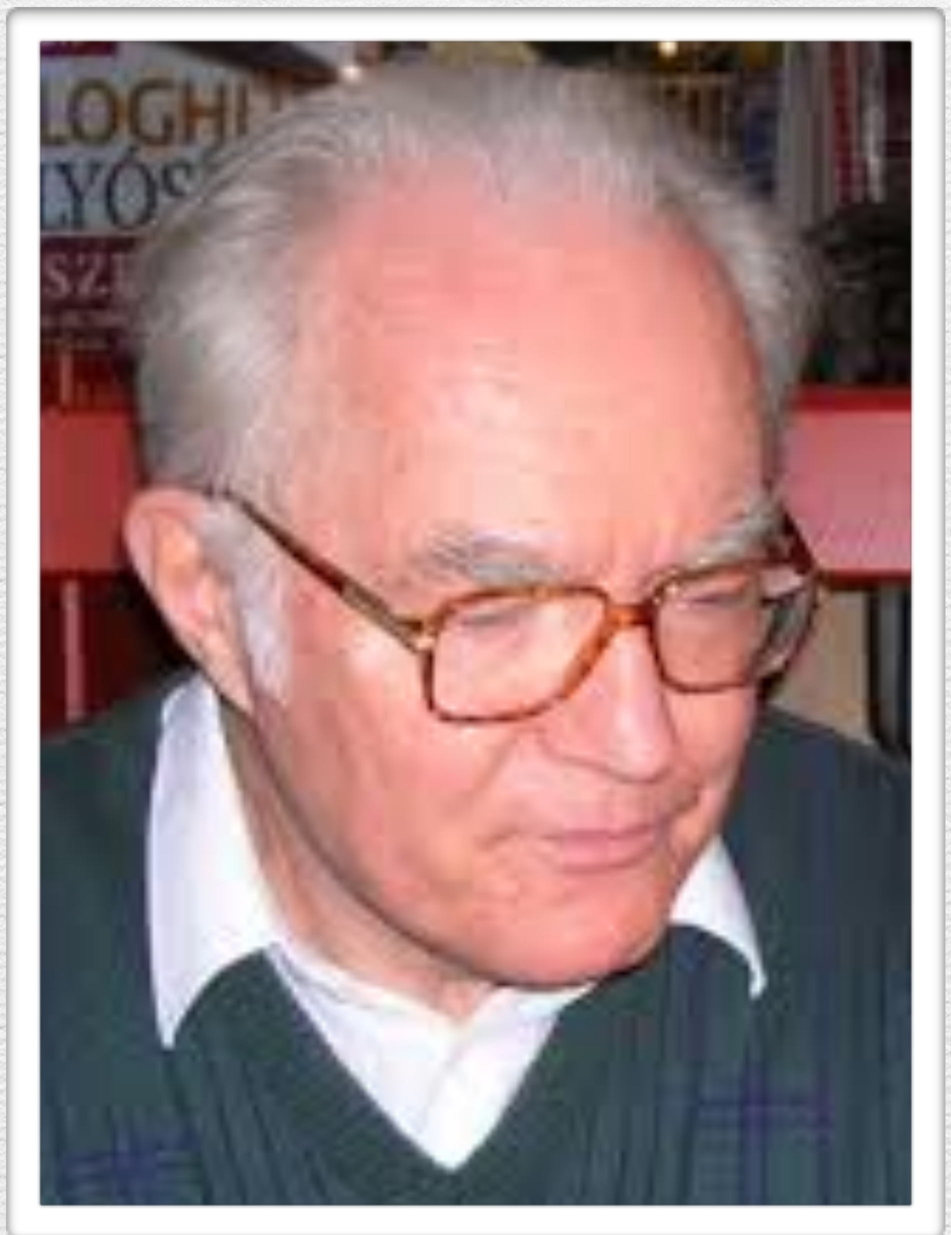
Is every projection rule generated by some function?

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.

Axiomatic approach

Axiomatic approach

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

Axiomatic approach

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

Axioms satisfied by a projection rule	Nature of the function generating the projection rule
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	$F(w \mid x^\star) =$ I-divergence
Regularity + Locality + Subspace Transitivity + Location Invariance + Scale Invariance	$F(w \mid x^\star) =$ Euclidean distance

The axiom of regularity

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

The axiom of regularity

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

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

The axiom of regularity

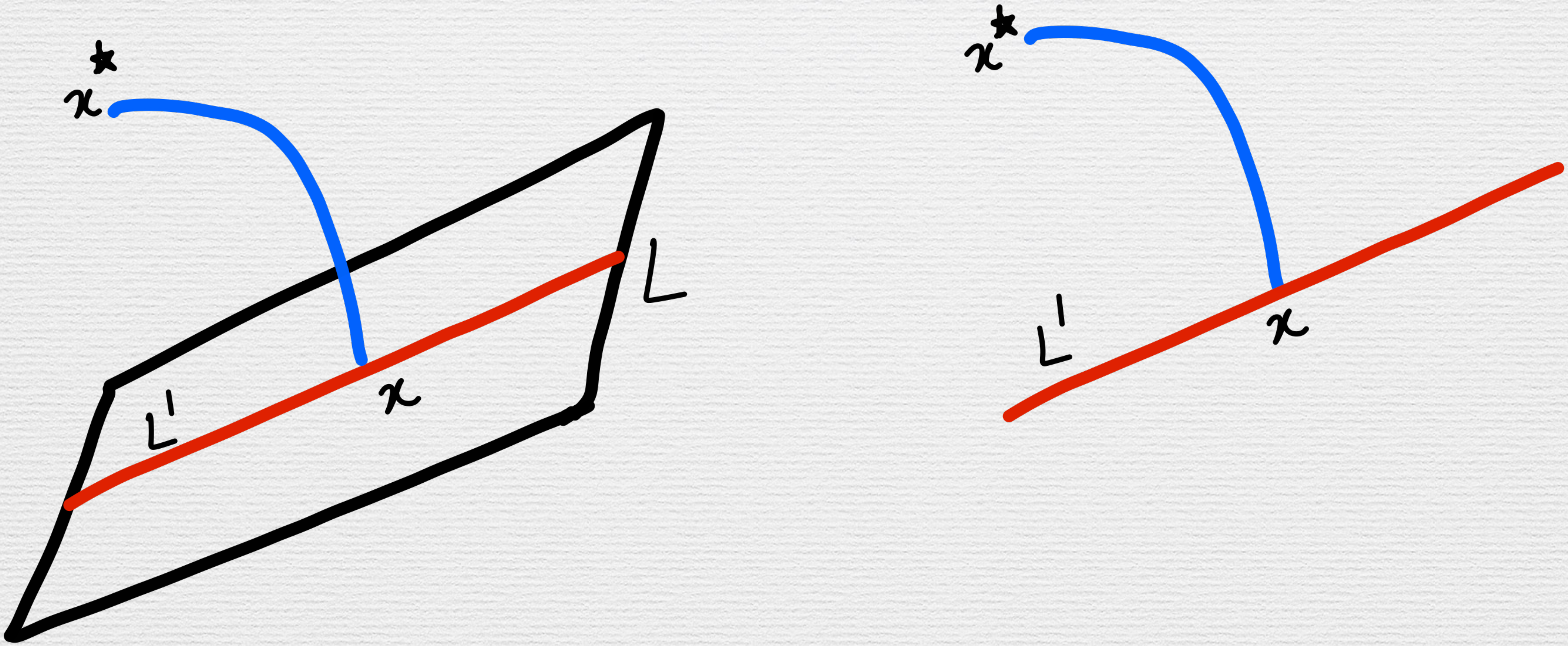
$\Pi : \mathscr{L} \times \mathbb{R}^n \to \mathbb{R}^n$ projection rule

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

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

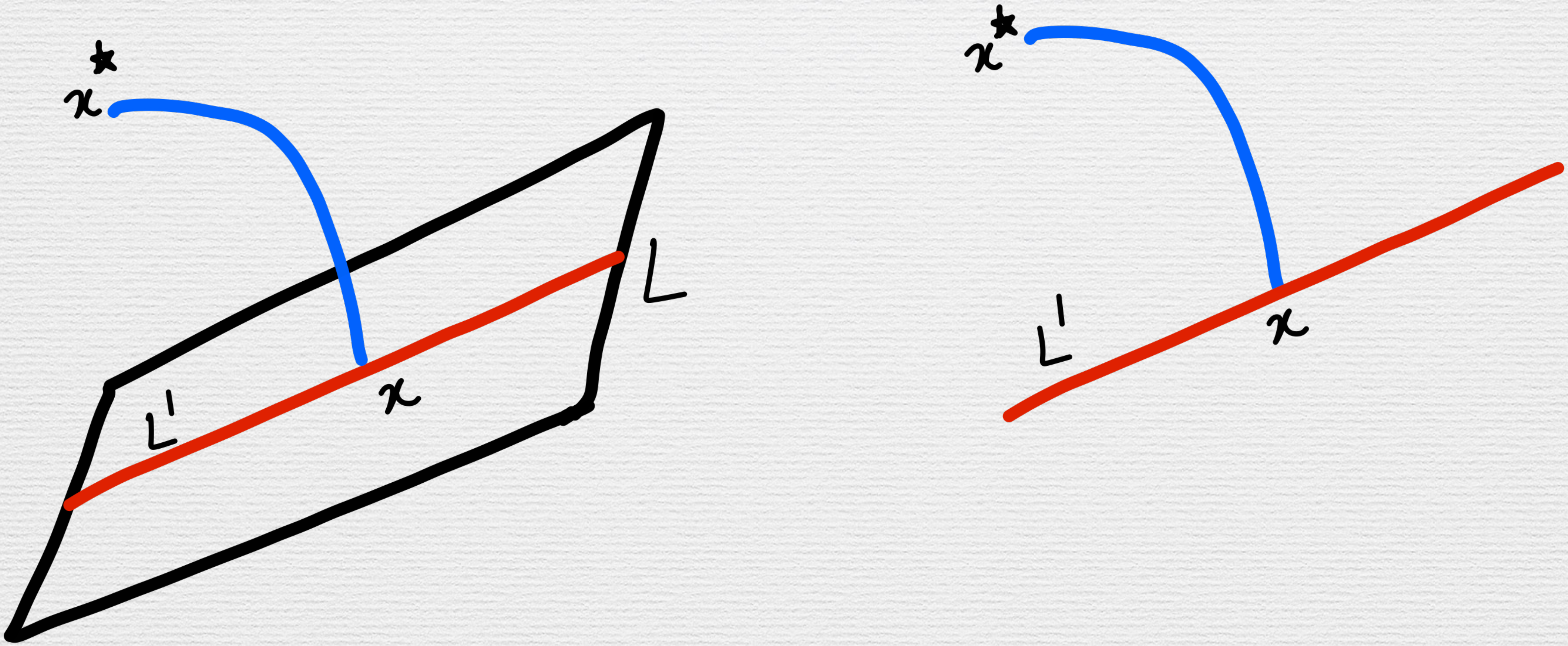
Regularity	(Consistency) $L' \subset L,\ \Pi(L \mid x^\star) \in L' \implies \Pi(L' \mid x^\star) = \Pi(L \mid x^\star)$
	(Distinctness) $L, \tilde{L} \in \mathscr{M},\ L \neq \tilde{L},\ x^\star \notin L \cap \tilde{L} \implies \Pi(L \mid x^\star) \neq \Pi(\tilde{L} \mid x^\star)$
	(Continuity) $\Pi(\,\cdot \mid x^\star)$ restricted to any fixed dimension is continuous

The axiom of consistency



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

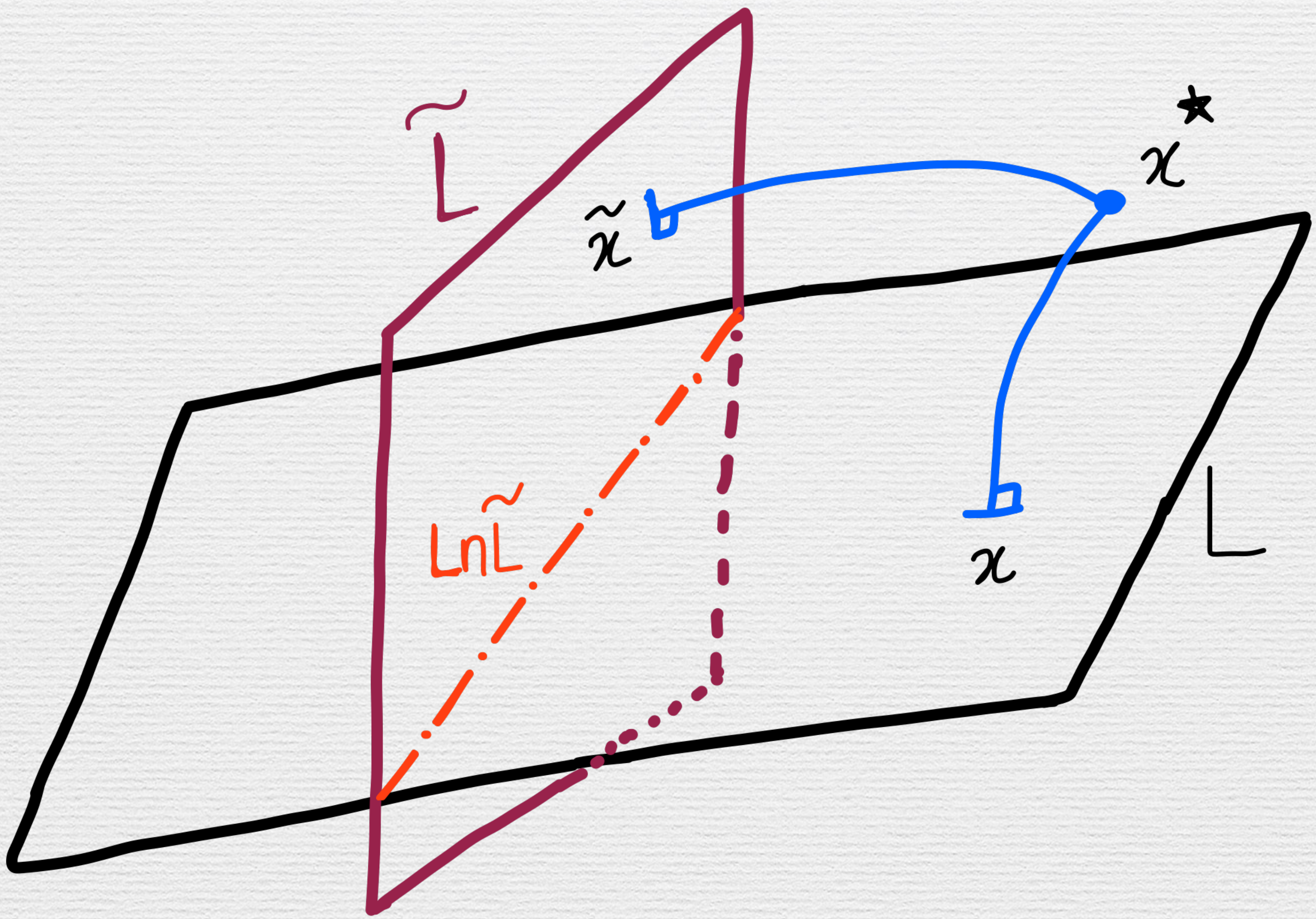
The axiom of consistency



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

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

The axiom of distinctness



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

Axiom of continuity

Axiom of continuity

- Let

Axiom of continuity

- Let

$$X = \left\{ (A, b) \in \mathbb{R}^{k \times n} \times \mathbb{R}^k : A \text{ has rank } k \right\}$$

Axiom of continuity

- Let
$$X = \left\{ (A, b) \in \mathbb{R}^{k \times n} \times \mathbb{R}^k : A \text{ has rank } k \right\}$$
- Define an equivalence relation on X as follows:

Axiom of continuity

- Let

$$X = \left\{ (A, b) \in \mathbb{R}^{k \times n} \times \mathbb{R}^k : A \text{ has rank } k \right\}$$

- Define an equivalence relation on X as follows:

$$(A, b) \sim (\bar{A}, \bar{b}) \Leftrightarrow \{w : Aw = b\} = \{w : \bar{A}w = \bar{b}\}$$

Axiom of continuity

- Let

$$X = \left\{ (A, b) \in \mathbb{R}^{k \times n} \times \mathbb{R}^k : A \text{ has rank } k \right\}$$

- Define an equivalence relation on X as follows:

$$(A, b) \sim (\bar{A}, \bar{b}) \Leftrightarrow \{w : Aw = b\} = \{w : \bar{A}w = \bar{b}\}$$

- Equivalence classes

Axiom of continuity

- Let

$$X = \left\{ (A, b) \in \mathbb{R}^{k \times n} \times \mathbb{R}^{k} : A \text{ has rank } k \right\}$$

- Define an equivalence relation on X as follows:

$$(A, b) \sim (\bar{A}, \bar{b}) \Leftrightarrow \{w : Aw = b\} = \{w : \bar{A}w = \bar{b}\}$$

- Equivalence classes

$$Y = \left\{ [(A, b)] \in \mathbb{R}^{k \times n} \times \mathbb{R}^{k} : A \text{ has rank } k \right\}$$

Axiom of continuity

- Let

$$X = \left\{ (A, b) \in \mathbb{R}^{k \times n} \times \mathbb{R}^k : A \text{ has rank } k \right\}$$

- Define an equivalence relation on X as follows:

$$(A, b) \sim (\bar{A}, \bar{b}) \Leftrightarrow \{w : Aw = b\} = \{w : \bar{A}w = \bar{b}\}$$

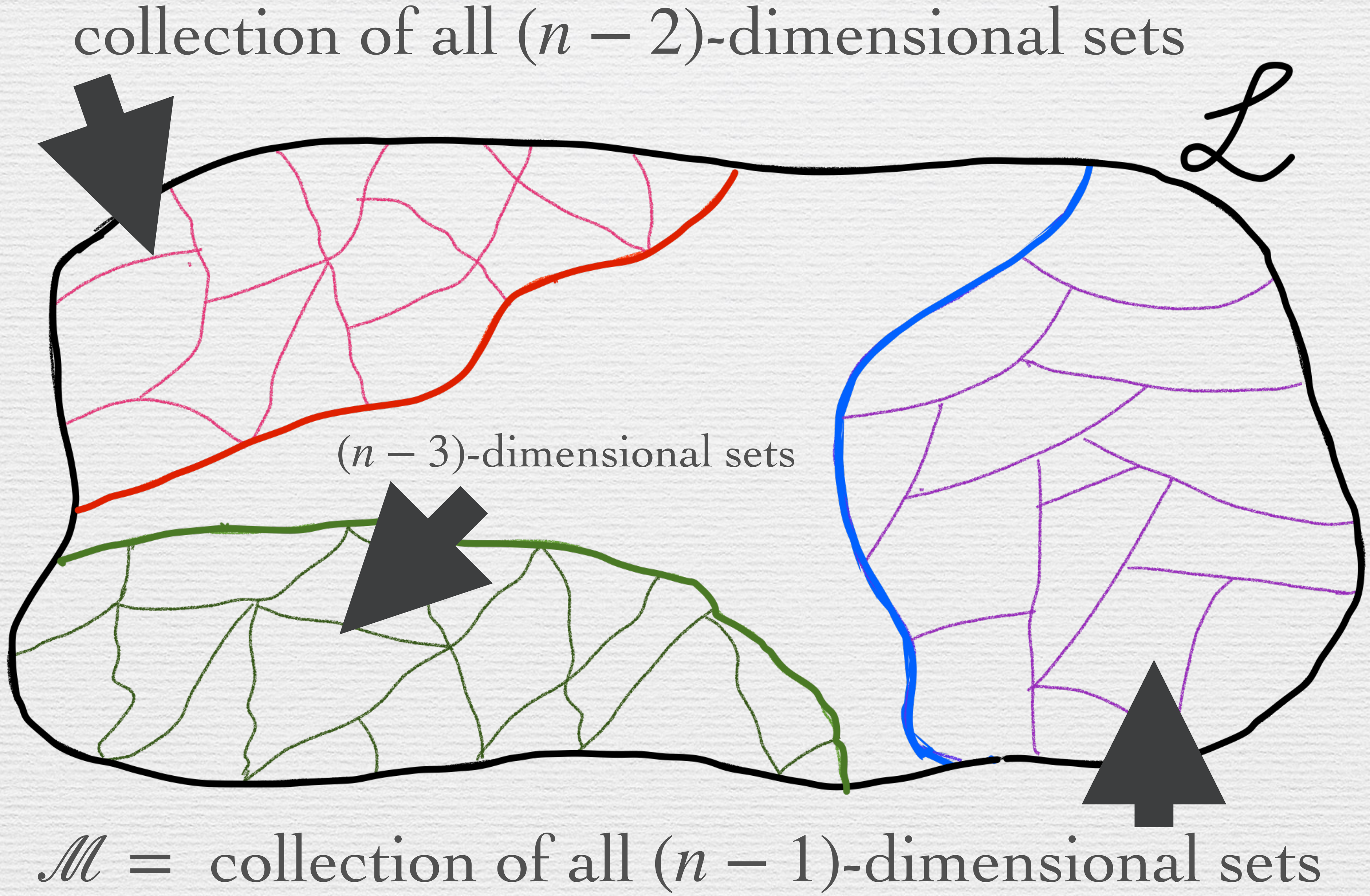
- Equivalence classes

$$Y = \left\{ [(A, b)] \in \mathbb{R}^{k \times n} \times \mathbb{R}^k : A \text{ has rank } k \right\}$$

- Topology on Y is the quotient topology derived out of the Euclidean topology on X

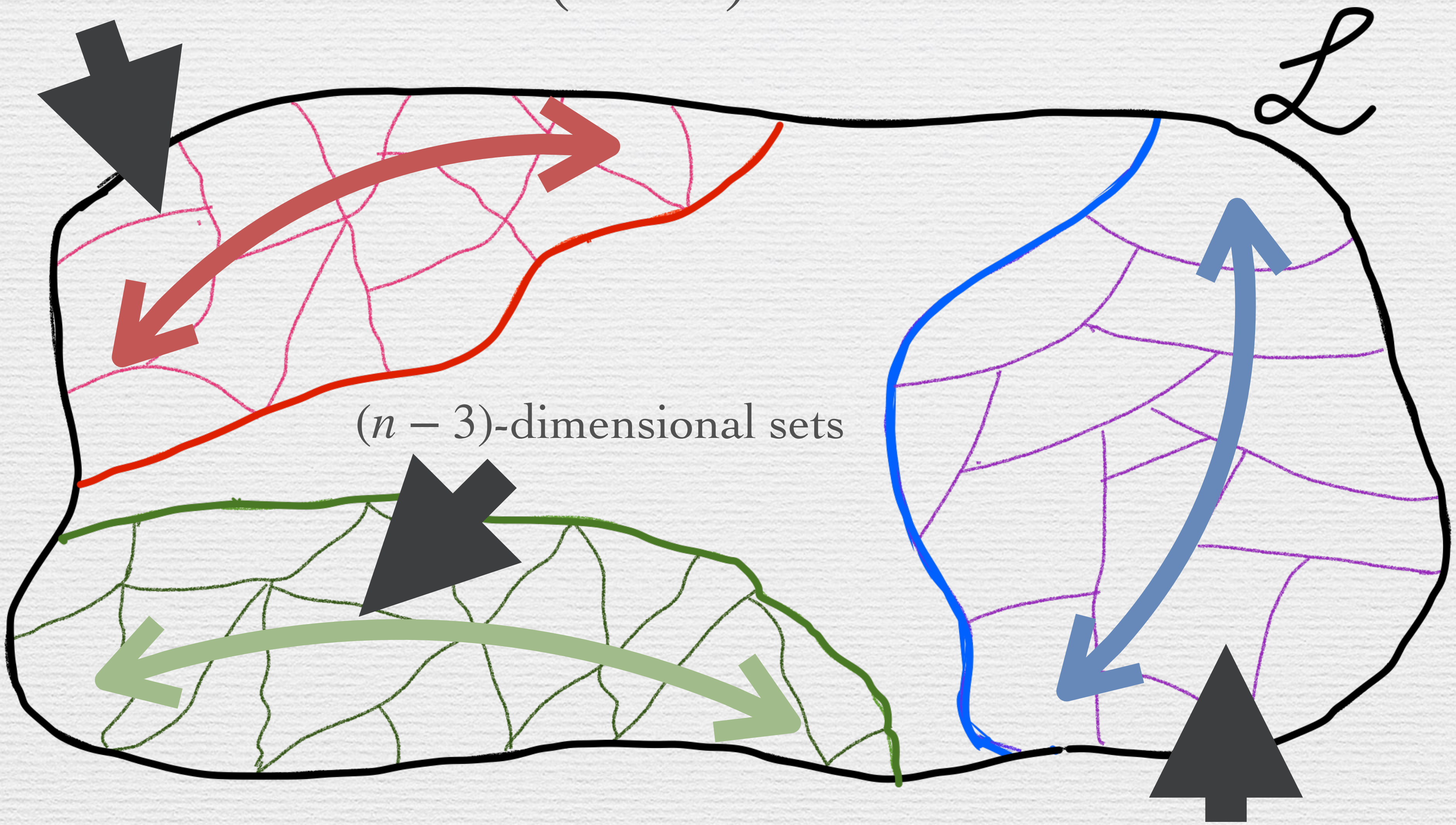
The axiom of continuity

The axiom of continuity



The axiom of continuity

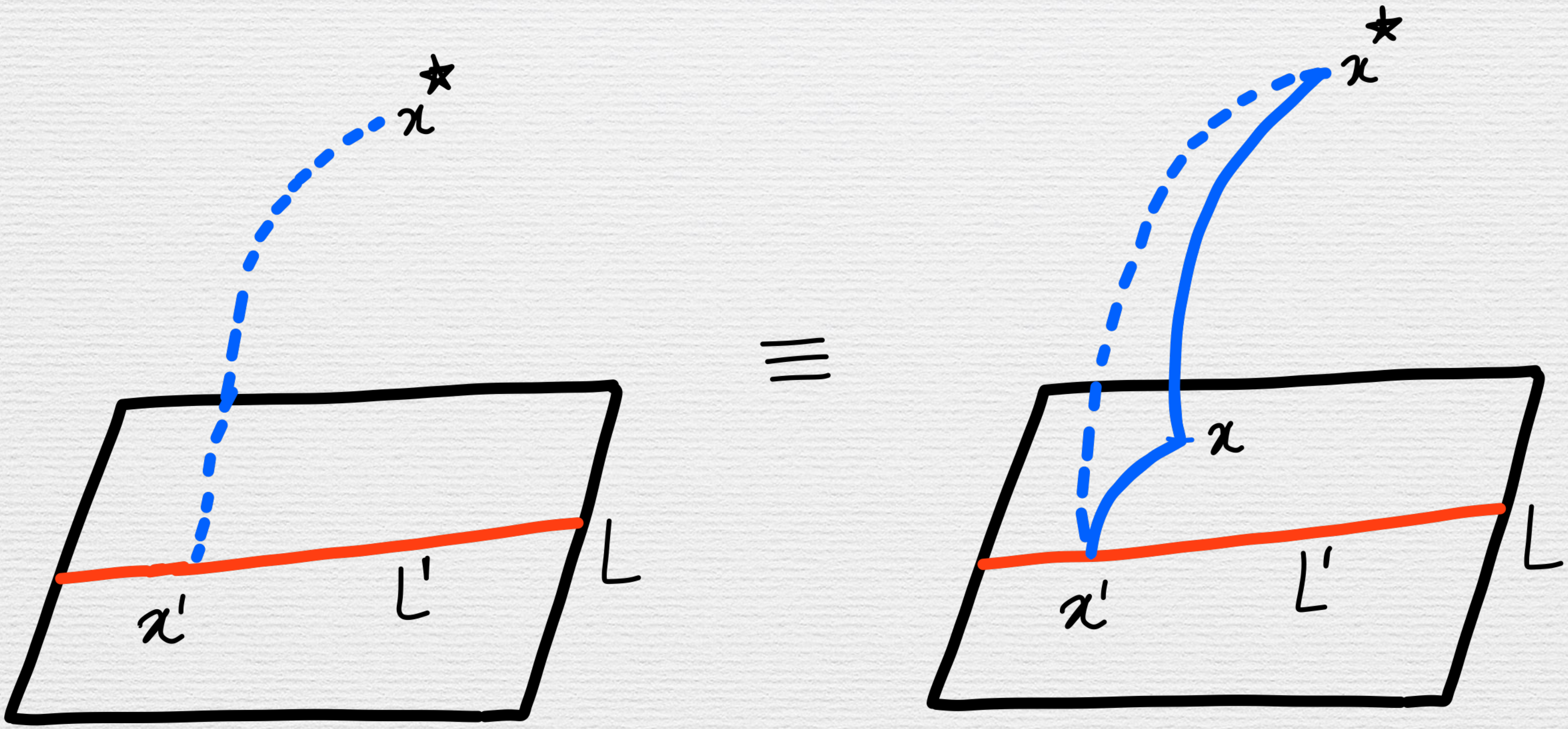
collection of all $(n-2)$-dimensional sets



$\mathscr{M} =$ collection of all $(n-1)$-dimensional sets

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

Axiom of subspace transitivity



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

Why are these axioms of interest to us?

Kumar, M. Ashok, and Rajesh Sundaresan. "Minimization Problems Based on Relative $\alpha $-Entropy I: Forward Projection." *IEEE Transactions on Information Theory* 61.9 (2015): 5063-5080.

Why are these axioms of interest to us?

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

Kumar, M. Ashok, and Rajesh Sundaresan. "Minimization Problems Based on Relative α-Entropy I: Forward Projection." *IEEE Transactions on Information Theory* 61.9 (2015): 5063-5080.

Why are these axioms of interest to us?

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

Regularity ✓

Kumar, M. Ashok, and Rajesh Sundaresan. "Minimization Problems Based on Relative α-Entropy I: Forward Projection." *IEEE Transactions on Information Theory* 61.9 (2015): 5063-5080.

Why are these axioms of interest to us?

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

Regularity	✓
Subspace transitivity	✓

Kumar, M. Ashok, and Rajesh Sundaresan. "Minimization Problems Based on Relative α-Entropy I: Forward Projection." *IEEE Transactions on Information Theory* 61.9 (2015): 5063-5080.

Why are these axioms of interest to us?

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

Regularity	✓
Subspace transitivity	✓
Locality	✗

Kumar, M. Ashok, and Rajesh Sundaresan. "Minimization Problems Based on Relative $\alpha $-Entropy I: Forward Projection." *IEEE Transactions on Information Theory* 61.9 (2015): 5063-5080.

Why are these axioms of interest to us?

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

Regularity	✓
Subspace transitivity	✓
Locality	✗

- These projection rules are generated by a divergence known as Sundaresan's divergence (relative α-entropy)

Kumar, M. Ashok, and Rajesh Sundaresan. "Minimization Problems Based on Relative $\alpha $-Entropy I: Forward Projection." *IEEE Transactions on Information Theory* 61.9 (2015): 5063-5080.

Sundaresan’s divergence

Sundaresan's divergence

- A parametric generalisation of KL divergence (or I-divergence)

Sundaresan's divergence

- A parametric generalisation of KL divergence (or I-divergence)

$$\mathscr{I}_\alpha(w, x) = \frac{\alpha}{1-\alpha} \log \left(\sum_{i=1}^{n} \frac{w_i}{||w||} \left(\frac{x_i}{||x||} \right)^{\alpha - 1} \right),$$

$$\text{where } ||x|| = \left(\sum_{i=1}^{n} x_i^\alpha \right)^{1/\alpha}, \quad \alpha > 0, \alpha \neq 1$$

Sundaresan's divergence

- A parametric generalisation of KL divergence (or I-divergence)

$$\mathcal{I}_\alpha(w, x) = \frac{\alpha}{1-\alpha} \log \left(\sum_{i=1}^{n} \frac{w_i}{||w||} \left(\frac{x_i}{||x||} \right)^{\alpha-1} \right),$$

$$\text{where } ||x|| = \left(\sum_{i=1}^{n} x_i^\alpha \right)^{1/\alpha}, \quad \alpha > 0, \alpha \neq 1$$

- This is **not a Bregman's divergence**. Hence, the corresponding projection rule is not local

Branch of a tree, or the trunk?

Branch of a tree, or the trunk?

- Are the projection rules generated by Sundaresan's divergence the only rules which are regular, subspace transitive and nonlocal?

Branch of a tree, or the trunk?

- Are the projection rules generated by Sundaresan's divergence the only rules which are regular, subspace transitive and nonlocal?
- Towards obtaining answers to the above question, we study the axiom of 'regularity' in more detail

Branch of a tree, or the trunk?

- Are the projection rules generated by Sundaresan's divergence the only rules which are regular, subspace transitive and nonlocal?
- Towards obtaining answers to the above question, we study the axiom of 'regularity' in more detail

Remainder of the talk is about regularity, and more….

What we want to know about regularity

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

Regularity + Locality ⟷ 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 ⟷ Projection rule generated by Bregman's divergence

What we want to know about regularity

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

Regularity + Locality ⟷ 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 ⟷ Projection rule generated by Bregman's divergence

Regularity ⟷ Projection rule generated by ???

Some implications of regularity

Some implications of regularity

- Let

Some implications of regularity

- Let

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

Some implications of regularity

- Let

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

- Recall

Some implications of regularity

- Let

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

- Recall

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

Some implications of regularity

- Let

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

- Recall

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

- Csiszár showed in his paper that

Some implications of regularity

- Let

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

- Recall

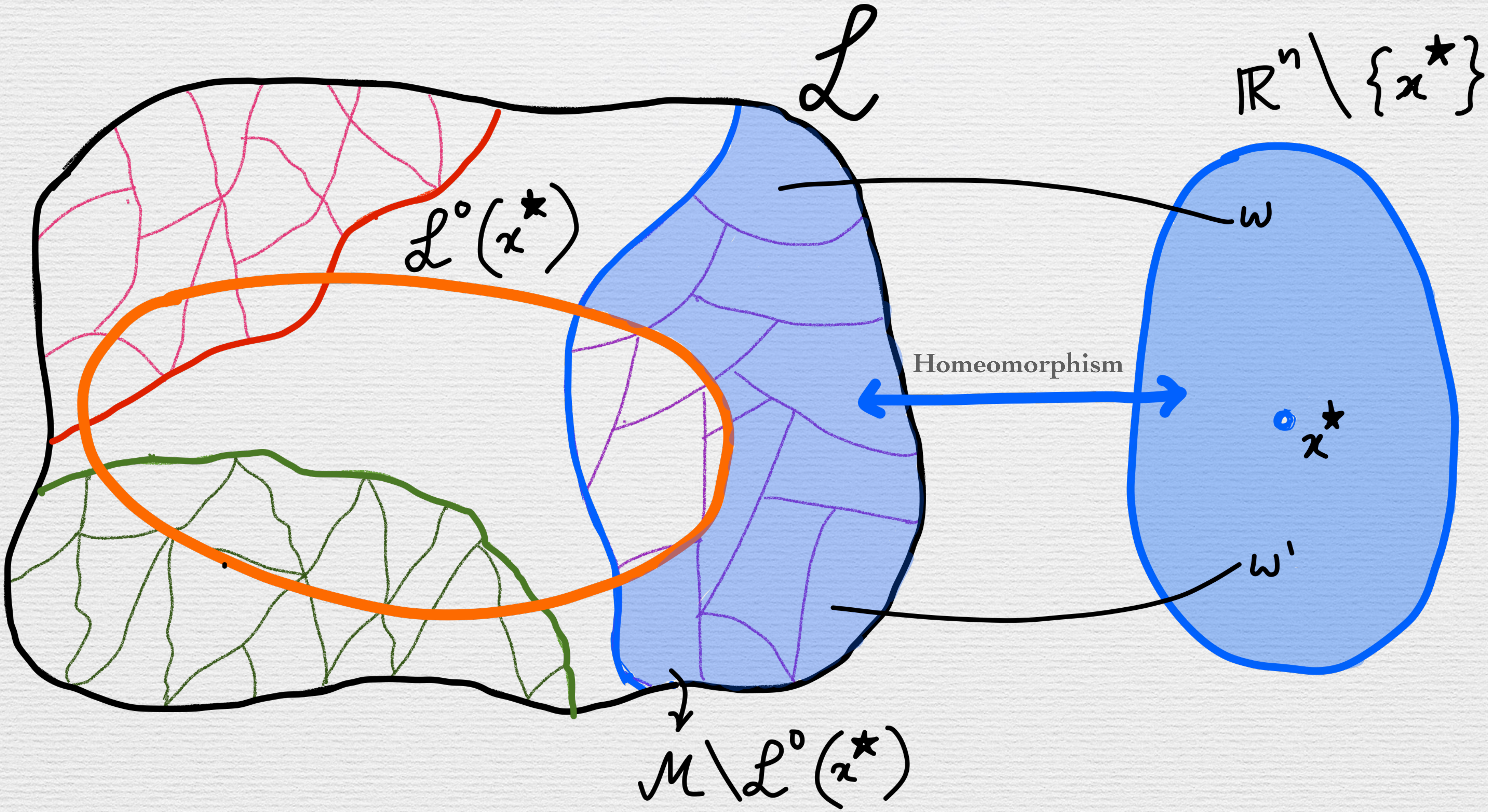
$$\mathscr{M} = \left\{ \{w \in \mathbb{R}^n : a^T w = b\}, a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R} \right\}$$

- Csiszár showed in his paper that

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

is a homeomorphism

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



Implications of regularity

This means that for every $w \neq x^\star$,

there exists a unique set L of dimension $(n-1)$

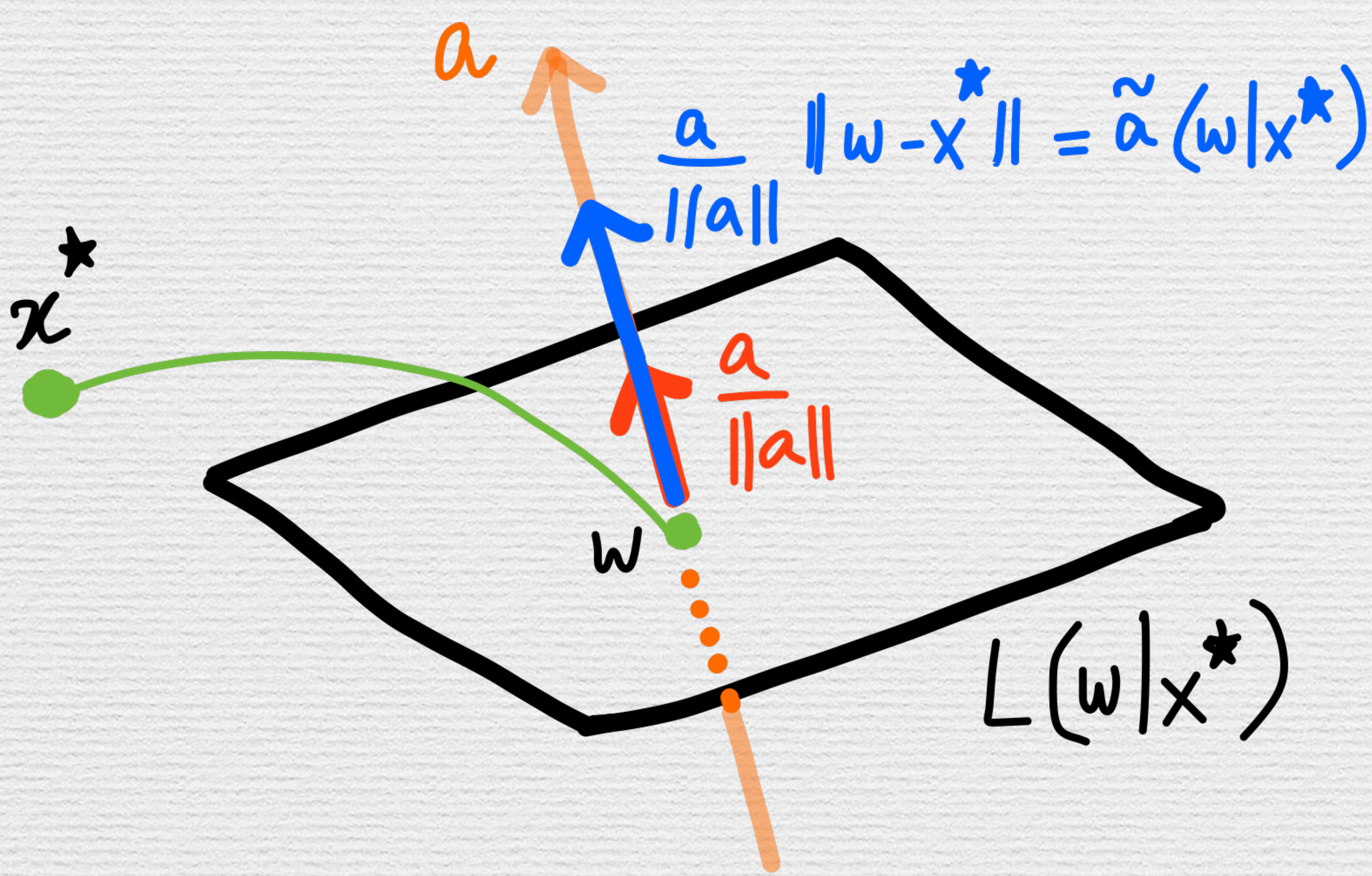
such that $\Pi(L \,|\, x^\star) = w$. Denote this L as $L(w \,|\, x^\star)$

Implications of regularity

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 $L(w \,|\, x^\star)$



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

Regularity and vector fields

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

Regularity and vector fields

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

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

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

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

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

Csiszár's results and vector fields

- The proof of Theorem 1 in Csiszár's paper reveals the following:

Regularity
+
Locality

Csiszár's results and vector fields

- The proof of Theorem 1 in Csiszár's paper reveals the following:

Regularity
+
Locality

➡

There exists a smooth function

$$F(w|x^\star) = \sum_{i=1}^{n} f_i(w_i|x_i^\star) \text{ such that}$$

$$\nabla F(w|x^\star) = \lambda \tilde{a}(w|x^\star) \text{ for all } w \neq x^\star$$

Csiszár's results and vector fields

- The proof of Theorem 1 in Csiszár's paper reveals the following:

Regularity
+
Locality

$\Rightarrow$

There exists a smooth function

$$F(w \mid x^\star) = \sum_{i=1}^{n} f_i(w_i \mid x_i^\star) \text{ such that}$$

$$\nabla F(w \mid x^\star) = \lambda \tilde{a}(w \mid x^\star) \text{ for all } w \neq x^\star$$

$\lambda \neq 0$, could depend on $w, x^\star$

Csiszár's results and vector fields

- The proof of Theorem 1 in Csiszár's paper reveals the following:

Regularity + Locality ⟹

There exists a smooth function

$$F(w\,|\,x^\star) = \sum_{i=1}^{n} f_i(w_i\,|\,x_i^\star) \text{ such that}$$

$$\nabla F(w\,|\,x^\star) = \lambda \tilde{a}(w\,|\,x^\star) \text{ for all } w \neq x^\star$$

$\lambda \neq 0$, could depend on $w, x^\star$

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

Csiszár's results and vector fields

- The proof of Theorem 1 in Csiszár's paper reveals the following:

Regularity
+
Locality

⟹

There exists a smooth function

$$F(w|x^\star) = \sum_{i=1}^{n} f_i(w_i|x_i^\star) \text{ such that}$$

$$\nabla F(w|x^\star) = \lambda \tilde{a}(w|x^\star) \text{ for all } w \neq x^\star$$

$\lambda \neq 0$, could depend on $w, x^\star$

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

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

Csiszár's results and vector fields

- The proof of Theorem 1 in Csiszár's paper reveals the following:

Regularity + Locality ⇒ There exists a smooth function

$$F(w|x^\star) = \sum_{i=1}^{n} f_i(w_i|x_i^\star) \text{ such that}$$

$$\nabla F(w|x^\star) = \lambda \tilde{a}(w|x^\star) \text{ for all } w \neq x^\star$$

$\lambda \neq 0$, could depend on $w, x^\star$

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

This vector field is conservative!!

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

Csiszár's results and vector fields

- The proof of Theorem 1 in Csiszár's paper reveals the following:

Regularity + Locality $\Rightarrow$

There exists a smooth function

$$F(w|x^\star) = \sum_{i=1}^{n} f_i(w_i|x_i^\star) \text{ such that}$$

$$\nabla F(w|x^\star) = \lambda \tilde{a}(w|x^\star) \text{ for all } w \neq x^\star$$

$\lambda \neq 0$, could depend on $w, x^\star$

This particular form is due to <u>locality</u> axiom

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

This vector field is conservative!!

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

Conservative vector fields and regularity axiom

- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to regularity**

Conservative vector fields and regularity axiom

- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to regularity**

$\Pi(\,\cdot\,|\,x^\star)$ regular

Conservative vector fields and regularity axiom

- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to regularity**

$\Pi(\,\cdot\,|\,x^\star)$ regular →

Conservative vector fields and regularity axiom

- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to regularity**

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

Conservative vector fields and regularity axiom

- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to regularity**

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

↓

Conservative vector fields and regularity axiom

- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to 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

Conservative vector fields and regularity axiom

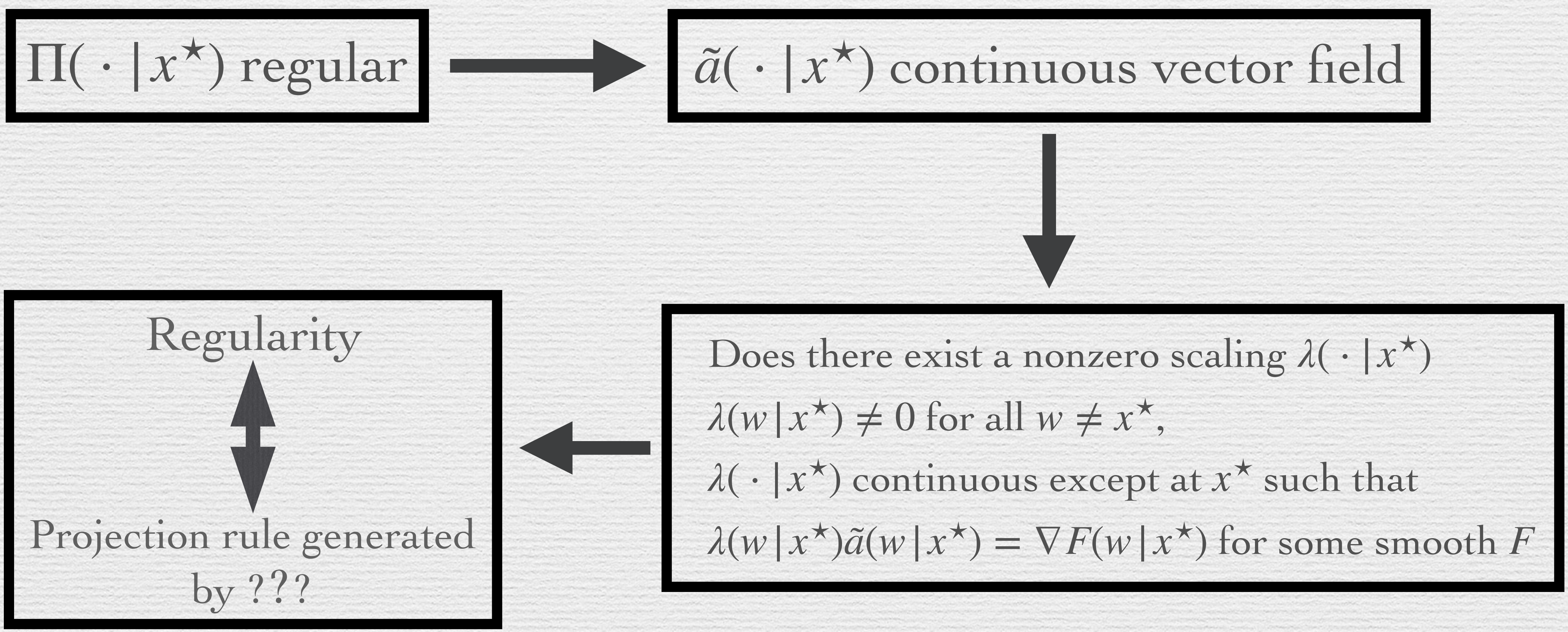
- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to regularity**

$\Pi(\,\cdot\,|x^\star)$ 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 ←

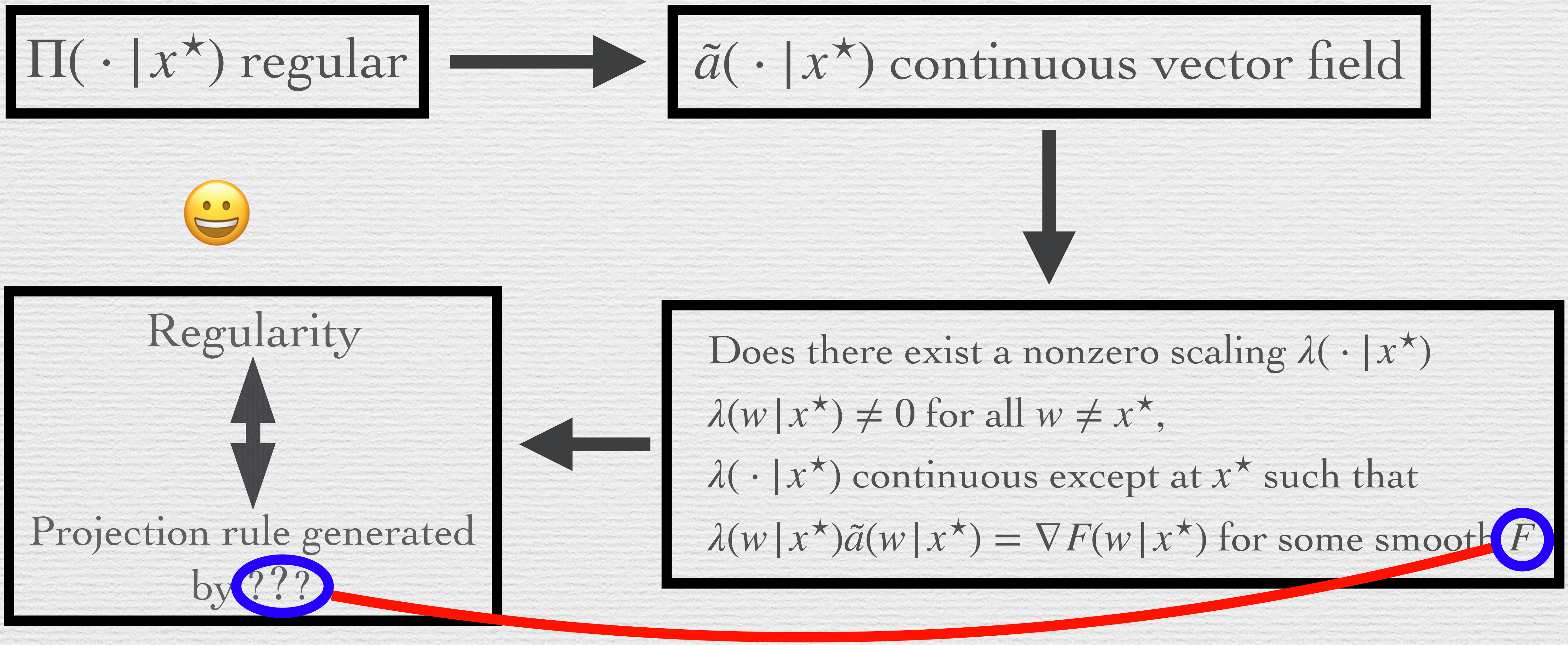
Conservative vector fields and regularity axiom

- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to regularity**



Conservative vector fields and regularity axiom

- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to regularity**



Conservative vector fields and regularity axiom

- While the specific form of the function may be due to locality, our feeling is that the **conservative property may be due to 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

Regularity

↕

Projection rule generated by ???

An open question on conservative vector fields

An open question on conservative vector fields

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

An open question on conservative vector fields

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

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

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

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

Under more assumption

- Suppose we are in 3-dim, and we know

$$a : \mathbb{R}^3 \to \mathbb{R}^3 \text{ is continuously differentiable}$$

Under more assumption

- Suppose we are in 3-dim, and we know

$$a : \mathbb{R}^3 \to \mathbb{R}^3 \text{ is continuously differentiable}$$

Is there a differentiable scaling function λ such that $\lambda \cdot a = \nabla F$?

Under more assumption

- Suppose we are in 3-dim, and we know

$$a : \mathbb{R}^3 \to \mathbb{R}^3 \text{ is continuously differentiable}$$

Is there a differentiable scaling function λ such that $\lambda \cdot a = \nabla F$?	$\Longleftrightarrow$	$\nabla \times (\lambda a) = 0$

Under more assumption

- Suppose we are in 3-dim, and we know

$$a : \mathbb{R}^3 \to \mathbb{R}^3 \text{ is continuously differentiable}$$

Is there a differentiable scaling function λ such that $\lambda \cdot a = \nabla F$?	⟷	$\nabla \times (\lambda a) = 0$

$$a_1(w)\frac{\partial \lambda(w)}{\partial w_2} - a_2(w)\frac{\partial \lambda(w)}{\partial w_1} = \lambda(w)\left(\frac{\partial a_2(w)}{\partial w_1} - \frac{\partial a_1(w)}{\partial w_2}\right),$$
$$a_2(w)\frac{\partial \lambda(w)}{\partial w_3} - a_3(w)\frac{\partial \lambda(w)}{\partial w_2} = \lambda(w)\left(\frac{\partial a_3(w)}{\partial w_2} - \frac{\partial a_2(w)}{\partial w_3}\right),$$
$$a_3(w)\frac{\partial \lambda(w)}{\partial w_1} - a_1(w)\frac{\partial \lambda(w)}{\partial w_3} = \lambda(w)\left(\frac{\partial a_1(w)}{\partial w_3} - \frac{\partial a_3(w)}{\partial w_1}\right)$$

Summary

- Axiomatic characterisation of projection rules
- 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

The axiom of locality

- Consider two sets L and $\tilde{L}$ of the form

$$L = \{w \in \mathbb{R}^n : w_J \in L_0, w_{J^c} \in L'\}$$

$$\tilde{L} = \{w \in \mathbb{R}^n : w_J \in L_0, w_{J^c} \in L''\}$$

where $J \subset \{1,2,\ldots,n\}$ is arbitrary

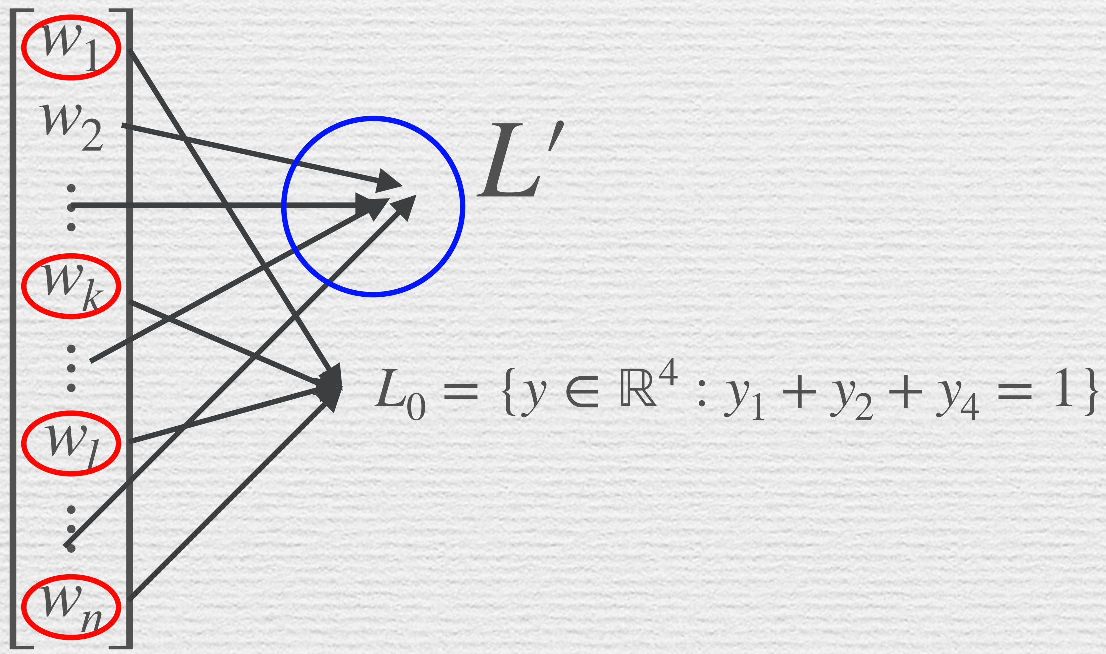
The axiom of locality

- Consider two sets L and $\tilde{L}$ of the form

$$L = \{w \in \mathbb{R}^n : w_J \in L_0, w_{J^c} \in L'\}$$

$$\tilde{L} = \{w \in \mathbb{R}^n : w_J \in L_0, w_{J^c} \in L''\}$$

where $J \subset \{1,2,\ldots,n\}$ is arbitrary



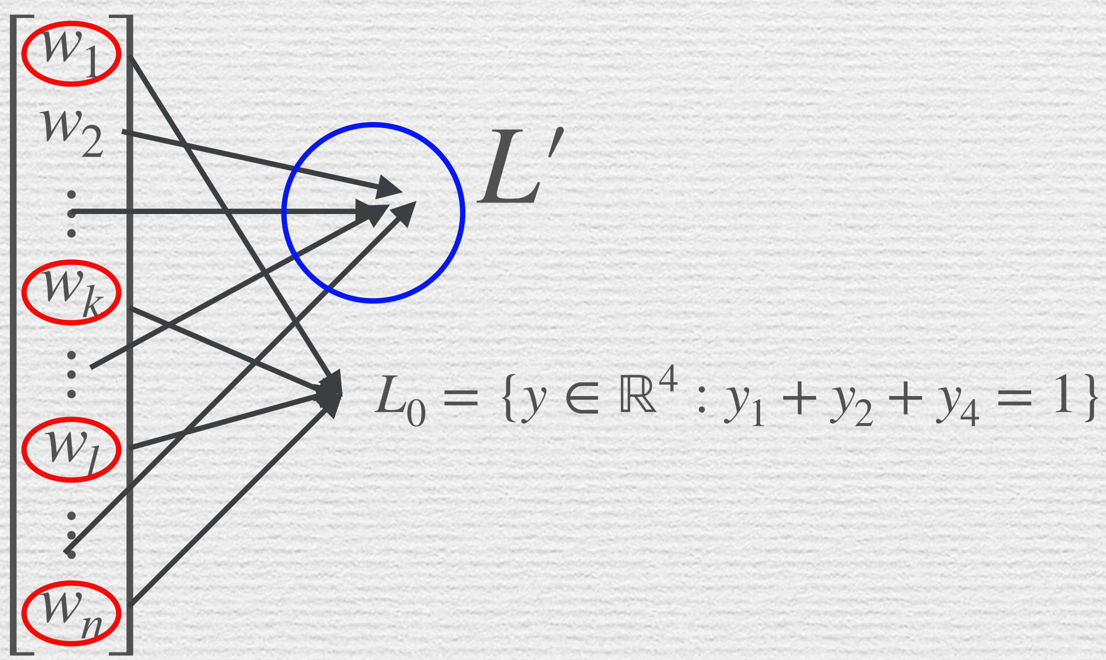
The axiom of locality

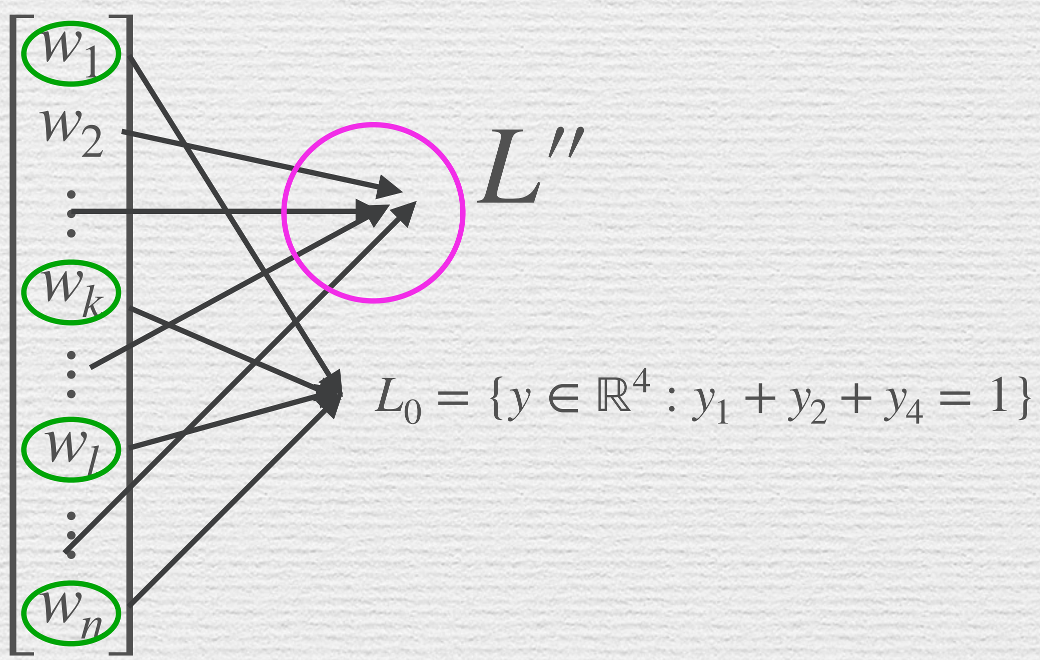
- Consider two sets L and $\tilde{L}$ of the form

$$L = \{w \in \mathbb{R}^n : w_J \in L_0, w_{J^c} \in L'\}$$

$$\tilde{L} = \{w \in \mathbb{R}^n : w_J \in L_0, w_{J^c} \in L''\}$$

where $J \subset \{1,2,\ldots,n\}$ is arbitrary





The axiom of locality

- Consider two sets L and $\tilde{L}$ of the form

$$L = \{w \in \mathbb{R}^n : w_J \in L_0, w_{J^c} \in L'\}$$

$$\tilde{L} = \{w \in \mathbb{R}^n : w_J \in L_0, w_{J^c} \in L''\}$$

where $J \subset \{1,2,\ldots,n\}$ is arbitrary

Then, $(\Pi(L\,|\,x^\star))_J = (\Pi(\tilde{L}\,|\,x^\star))_J$