Ax=b: An Old Problem Through a New Lens

Karthik P. N.

A Familiar Setup

A Familiar Setup

An unknown vector $x \in \mathbb{R}^n$

A Familiar Setup

An unknown vector $x \in \mathbb{R}^n$

x is known to belong to a set L of the form

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

$$A = \begin{bmatrix} \cdots\cdots a_1^T \cdots\cdots \\ \cdots\cdots a_2^T \cdots\cdots \\ \vdots \\ \cdots\cdots a_k^T \cdots\cdots \end{bmatrix}_{k \times n} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}_{k \times 1}$$

A Familiar Setup

An unknown vector $x \in \mathbb{R}^n$

x is known to belong to a set L of the form

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

$$A = \begin{bmatrix} \cdots\cdots a_1^T \cdots\cdots \\ \cdots\cdots a_2^T \cdots\cdots \\ \vdots \\ \cdots\cdots a_k^T \cdots\cdots \end{bmatrix}_{k \times n} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}_{k \times 1}$$

Goal: Recover x based on the constraints in L

A Familiar Setup

An unknown vector $x \in \mathbb{R}^n$

x is known to belong to a set L of the form

Goal: Recover x based on the constraints in L

A Familiar Setup

An unknown vector $x \in \mathbb{R}^n$

x is known to belong to a set L of the form

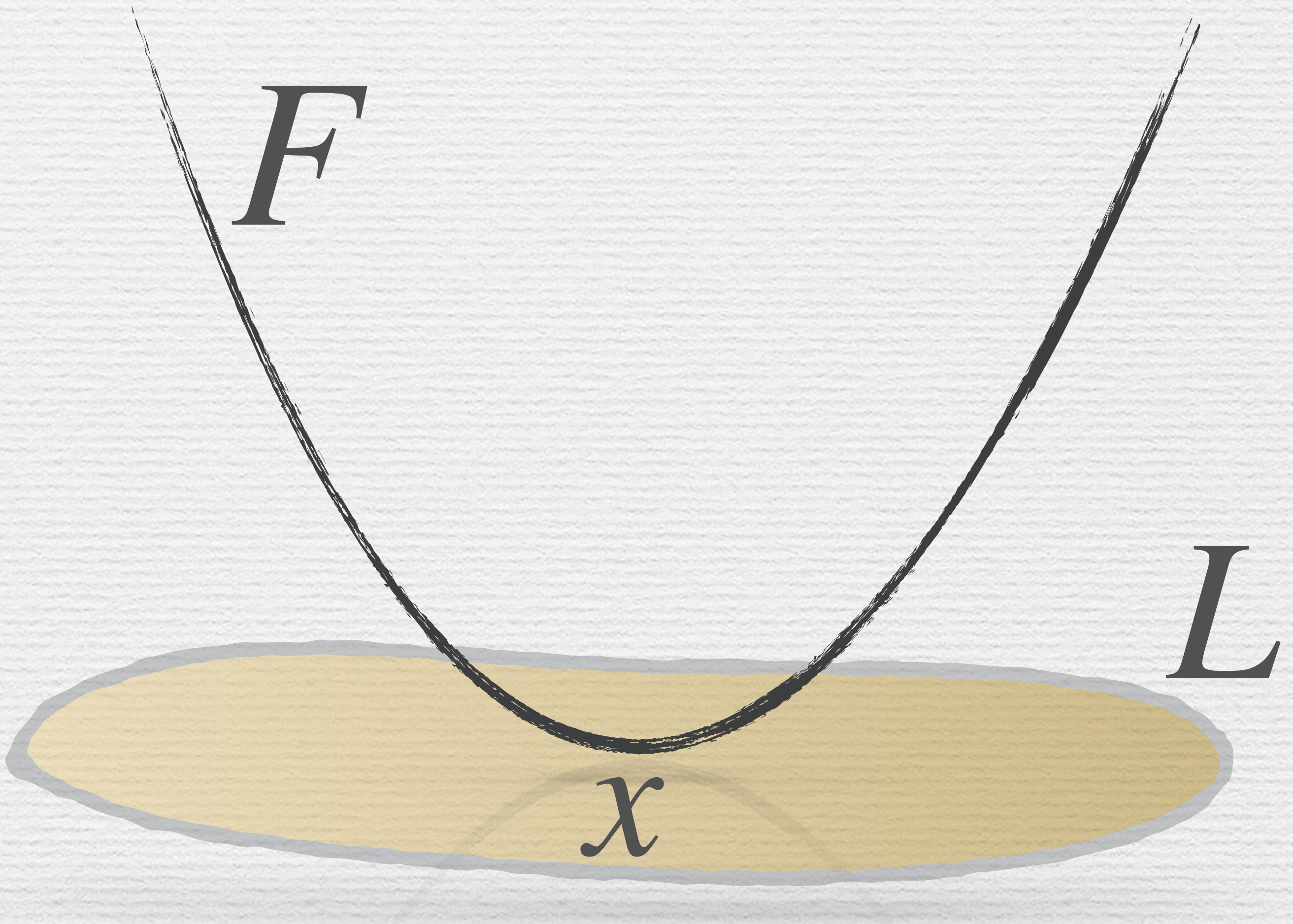
$$L = \{w \in \mathbb{R}^n : Aw = b\}$$

$$A = \begin{bmatrix} \cdots\cdots a_1^T \cdots\cdots \\ \cdots\cdots a_2^T \cdots\cdots \\ \vdots \\ \cdots\cdots a_k^T \cdots\cdots \end{bmatrix}_{k\times n} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}_{k\times 1}$$

Goal: Recover x based on the constraints in L

The Typical Approach

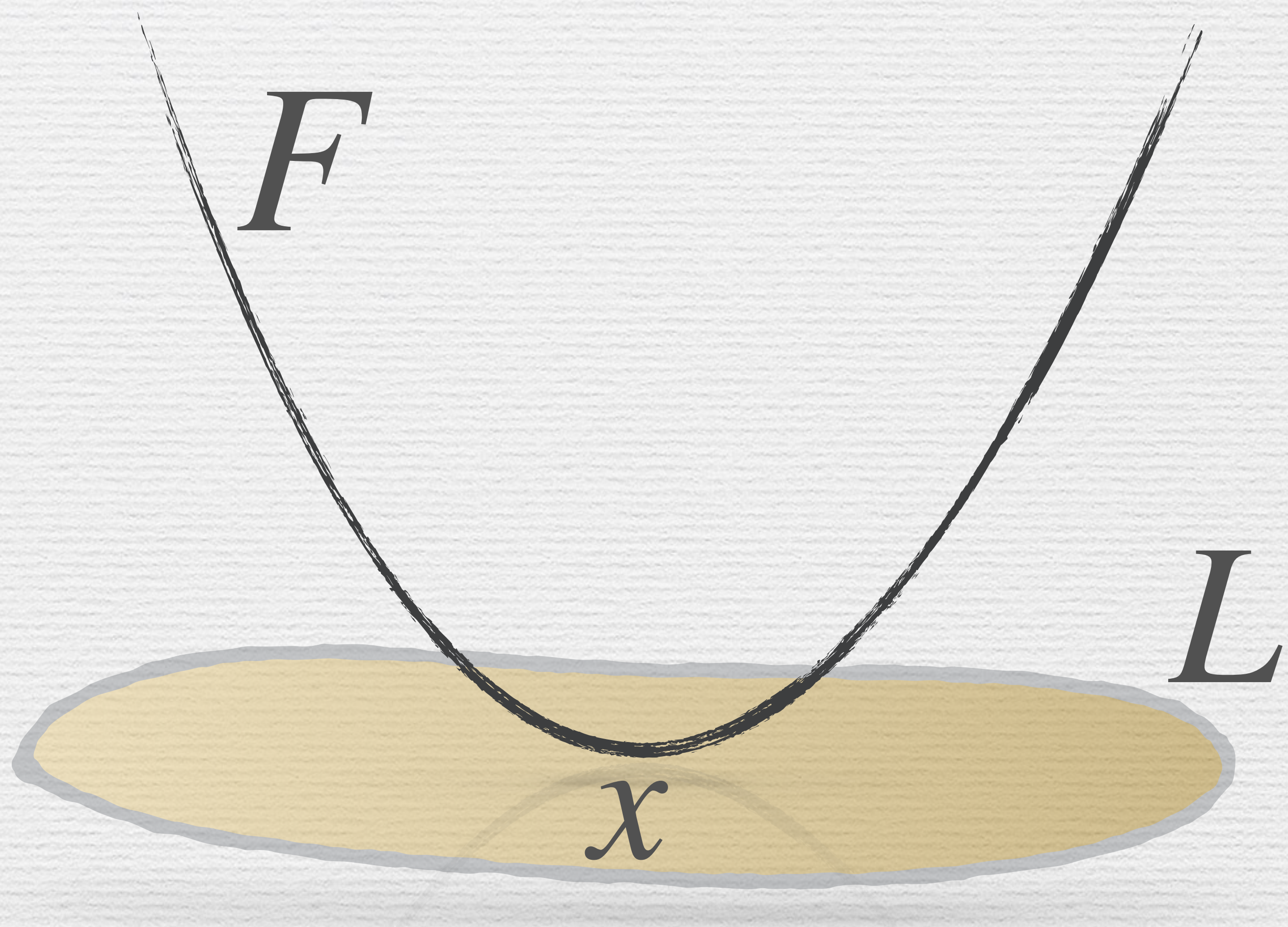
The Typical Approach



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

The Typical Approach

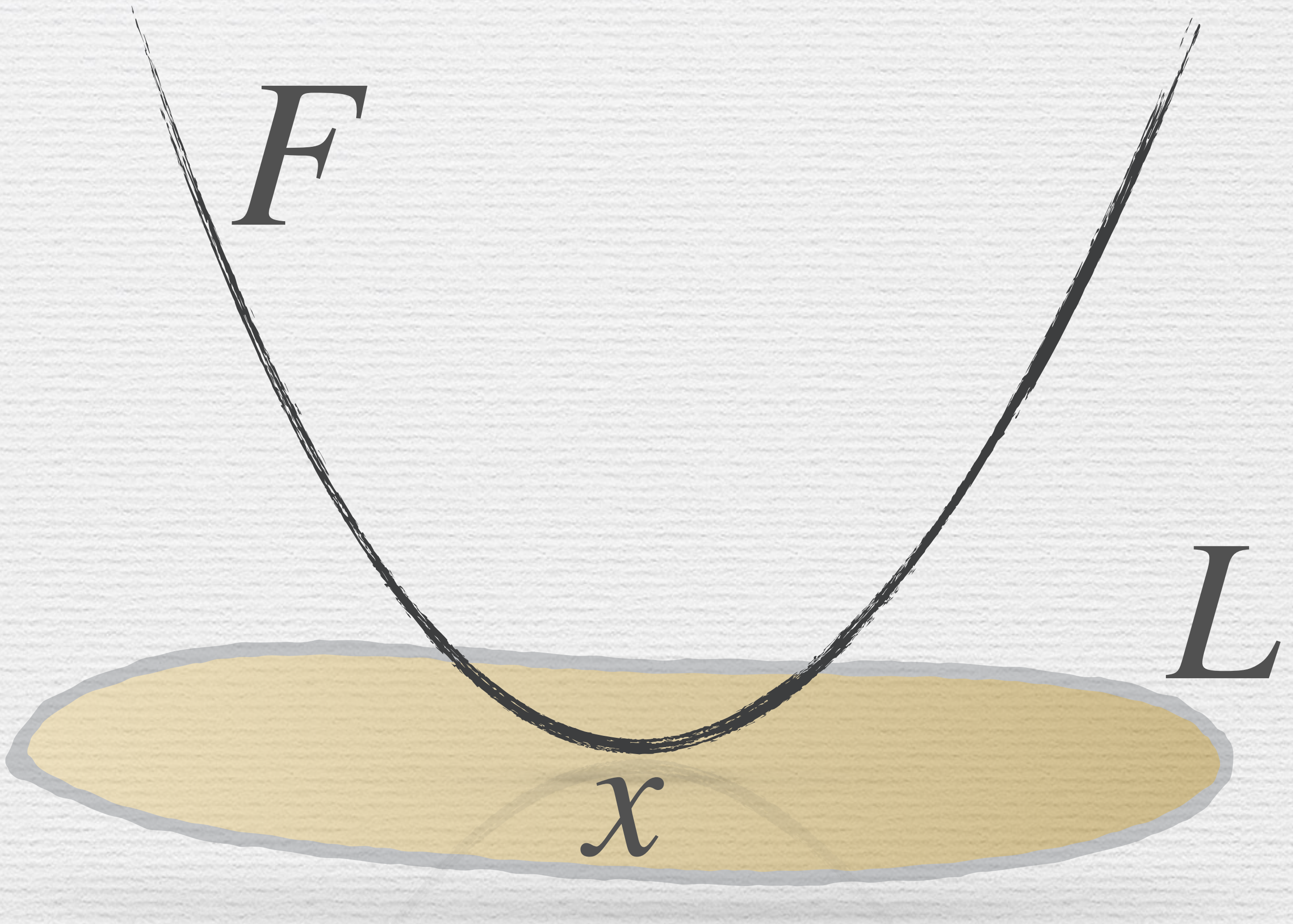
- Define a cost function
- Choose an element that minimises the value of the cost function



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

The Typical Approach

- Define a cost function
- Choose an element that minimises the value of the cost function



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

Examples: least squares, maximum entropy (max-ent)

Why Least Squares and Maximum Entropy? Why not xyz?

Why Least Squares and Maximum Entropy? Why not xyz?

- Why are least squares and max-ent rules widely accepted? What makes them so appealing?

Why Least Squares and Maximum Entropy? Why not xyz?

- Why are least squares and max-ent rules widely accepted? What makes them so appealing?

Least squares / maximum entropy rule

Rule xyz

Why Least Squares and Maximum Entropy? Why not xyz?

- Why are least squares and max-ent rules widely accepted? What makes them so appealing?

- Is there a way to classify rules as “naturally appealing” or “good”?

Least squares / maximum entropy rule

Rule xyz

An Axiomatic Approach

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.

An Axiomatic Approach

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.

An Axiomatic Approach

- Csiszár provides a method for classifying rules as good by defining a set of naturally appealing axioms

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.

An Axiomatic Approach

- Csiszár provides a method for classifying rules as good by defining a set of naturally appealing axioms
- Those rules that satisfy one or more axioms are classified as good

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.

An Axiomatic Approach

- Csiszár provides a method for classifying rules as good by defining a set of naturally appealing axioms
- Those rules that satisfy one or more axioms are classified as good
- Csiszár demonstrates that **<u>least squares and max-ent methods arise naturally</u>** as a result of satisfying some axioms

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]

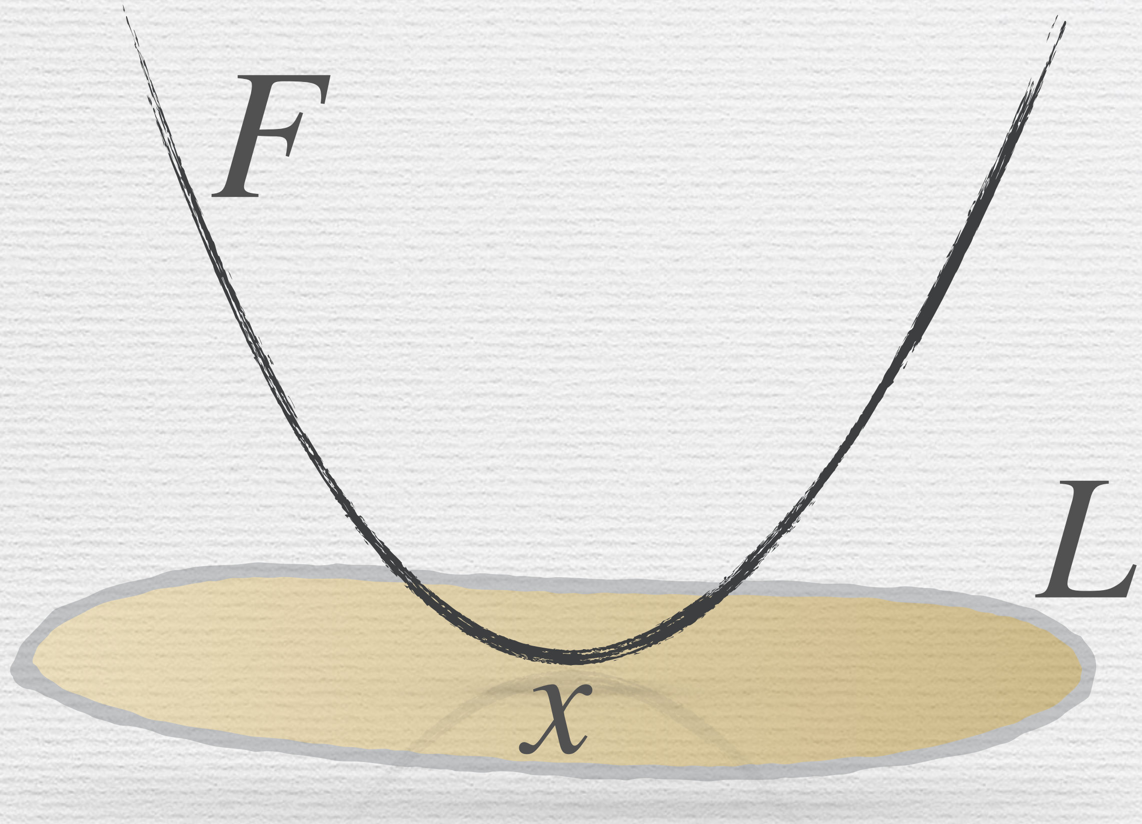
BY IMRE CSISZÁR

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.

One Problem, Two Approaches

One Problem, Two Approaches



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

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

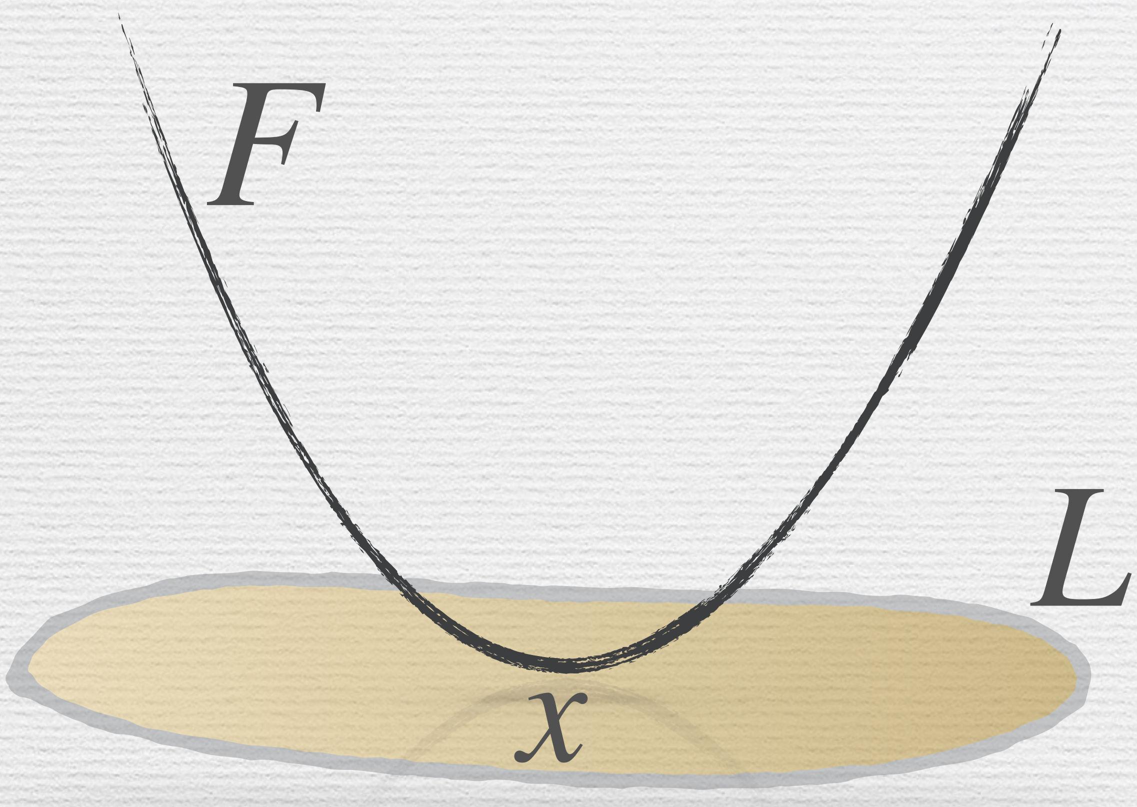
WHY LEAST SQUARES AND MAXIMUM ENTROPY? AN AXIOMATIC APPROACH TO INFERENCE FOR LINEAR INVERSE PROBLEMS[1]

BY IMRE CSISZÁR

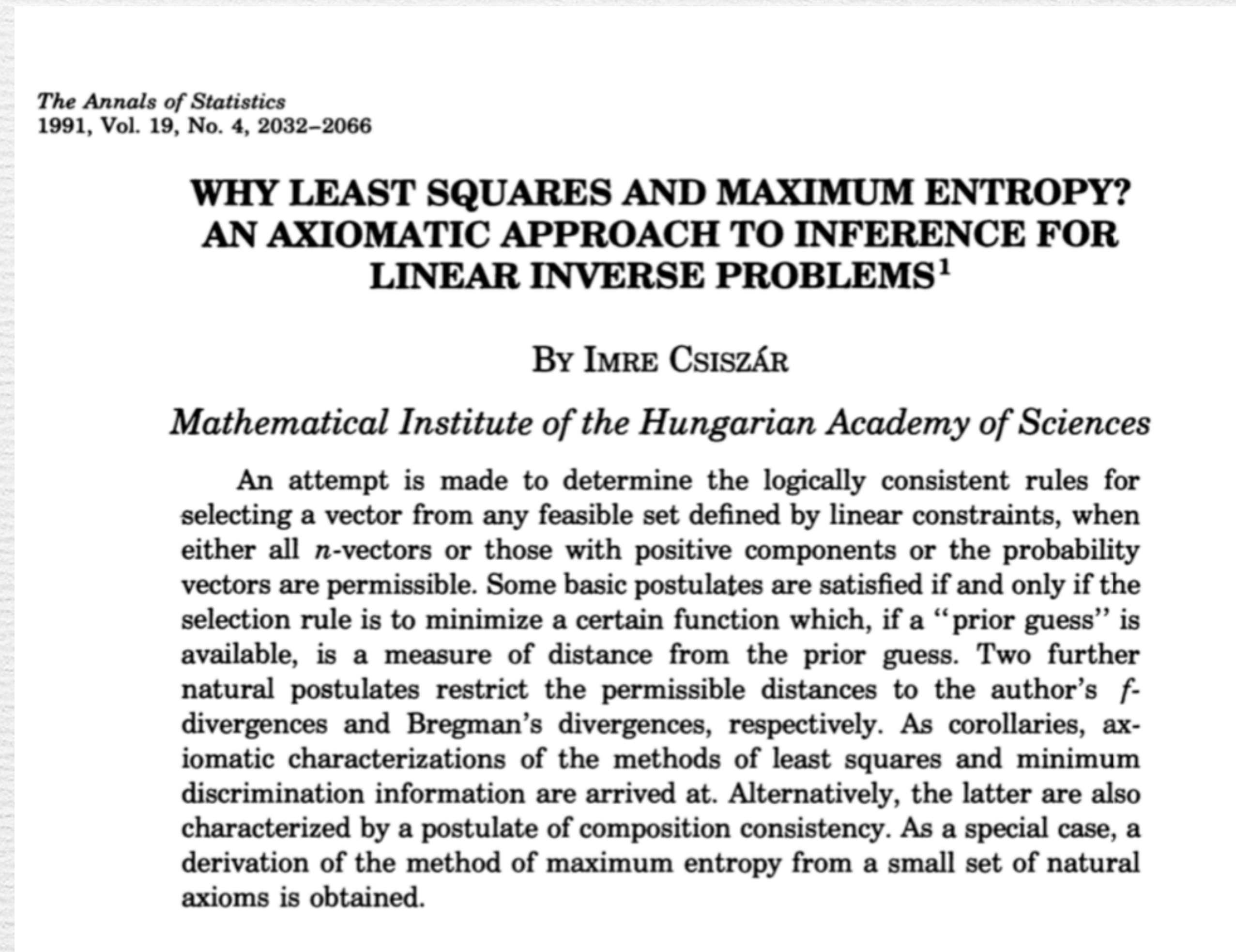
Mathematical Institute of the Hungarian Academy of Sciences

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.

One Problem, Two Approaches



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



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.

Is there a connection between the two approaches?

What the Main Talk Will be About

What the Main Talk Will be About

- Old patient v/s new doctor

What the Main Talk Will be About

- Old patient v/s new doctor
- Some axioms that are natural in many applications

What the Main Talk Will be About

- Old patient v/s new doctor
- Some axioms that are natural in many applications
- Conservative fields and an interesting open question

Thank You

Our Objective

To obtain an axiomatic characterisation for a class of rules larger than that considered by Csiszár