Information-Based Complexity for Systems and Control: The Probabilistic Setting

Roberto Tempo

CNR-IEIIT
Consiglio Nazionale delle Ricerche
Politecnico di Torino
tempo@polito.it

CNR-IEIIT

Happy Birthday to Shinji!

CNR-IEIIT

Acknowledgments

Research on the probabilistic setting of Information-Based Complexity is joint work with

- Fabrizio Dabbene
- Mario Sznaier

CNR-IEIIT

This Lecture - 1

This talk deals with problems which are solvable only approximately because information is partial or contaminated

J.F. Traub, G.W. Wasilkowski, H. Wozniakowski, “Information-Based Complexity (IBC)”, 1988

This Lecture - 2

This talk deals with problems which are solvable only approximately because information is partial or contaminated

- Objectives:
  - derive optimal algorithms
  - compute approximation error
  - analyze computational complexity

This Lecture - 3

This talk deals with problems which are solvable only approximately because information is partial or contaminated

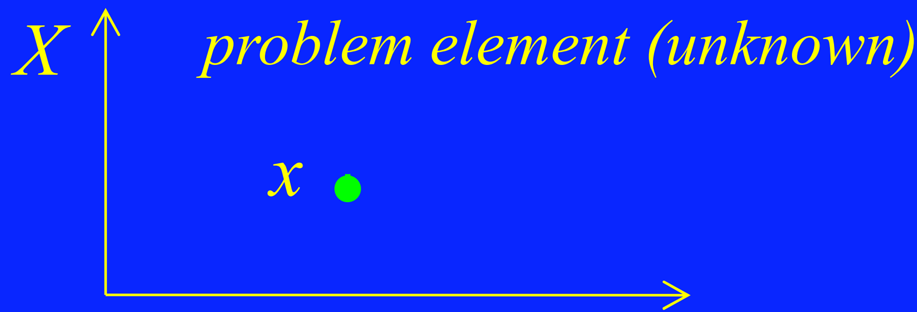
- Different settings:
  - worst-case (classical)
  - probabilistic (new)

CNR-IEIIT

This Lecture - 4

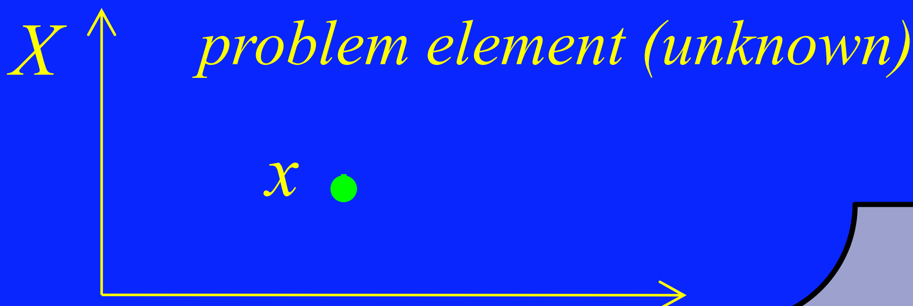
- Different areas may be covered, e.g. information theory, complexity, signal processing, numerical analysis, …
- IBC applications: integration problems, solutions of nonlinear equations, etc
- We are interested in systems and control and in the derivation of optimal algorithms for specific applications

Spaces and Operators



CNR-IEIIT

Spaces and Operators



example

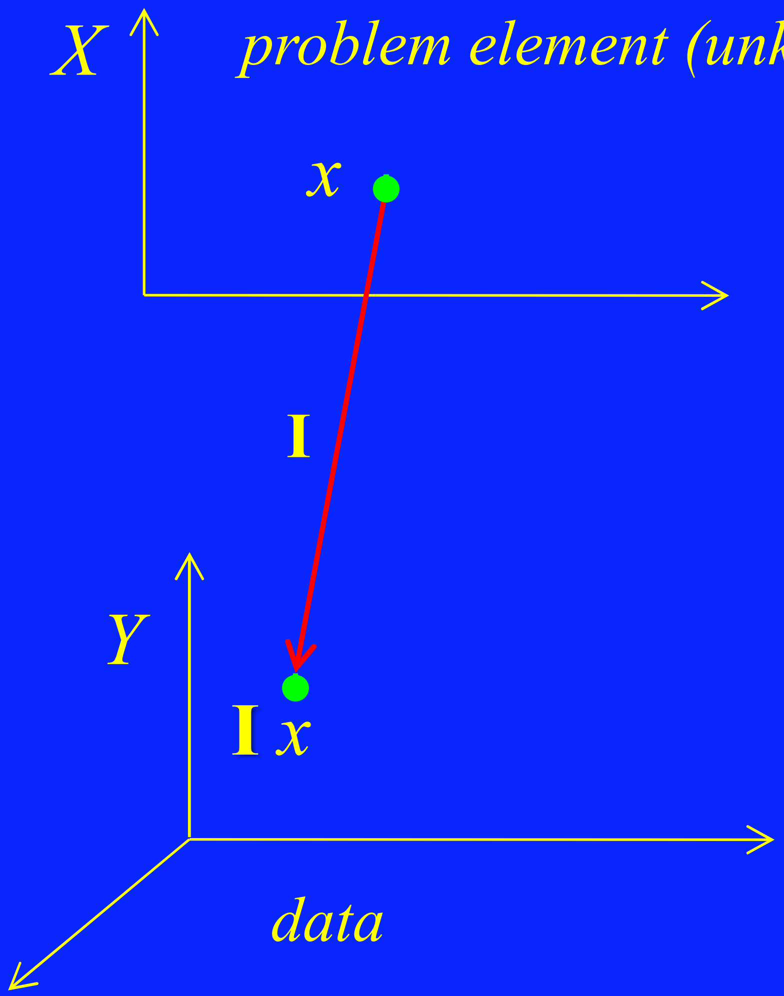
consider input-output pair

$\xi(x, t)$ of a dynamic system

$$\xi(x,t) = \sum_{i=1}^{n} x_i \varphi_i(t) = \Phi^{\mathrm{T}}(t)\, x$$

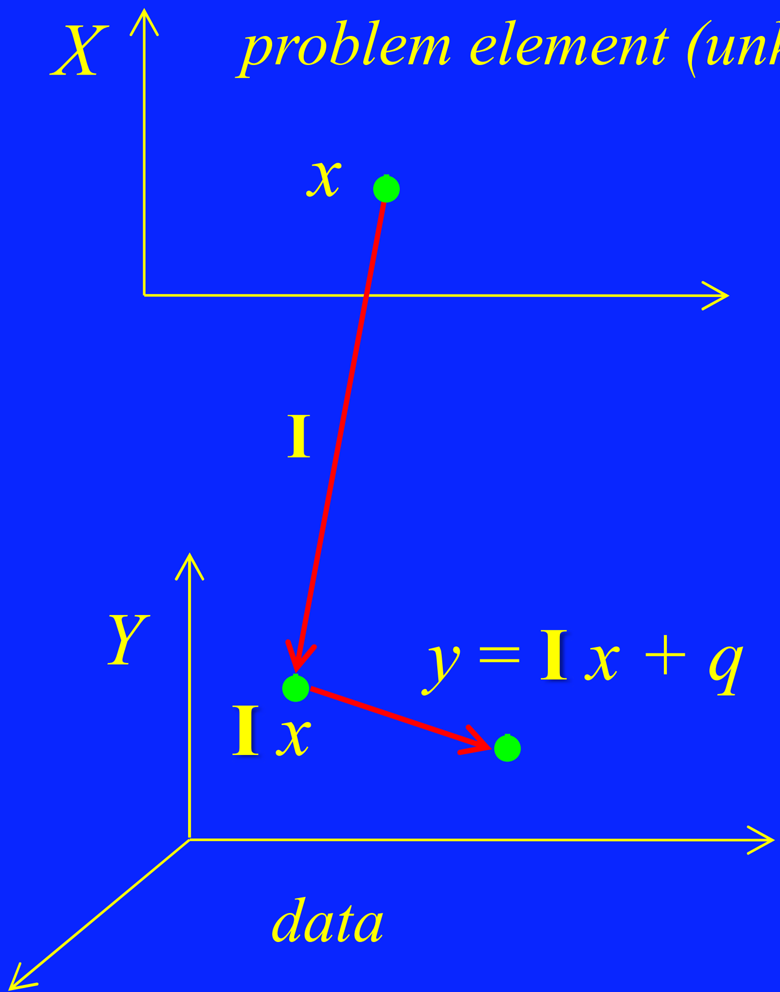
with given basis $\varphi_i(t)$

Spaces and Operators



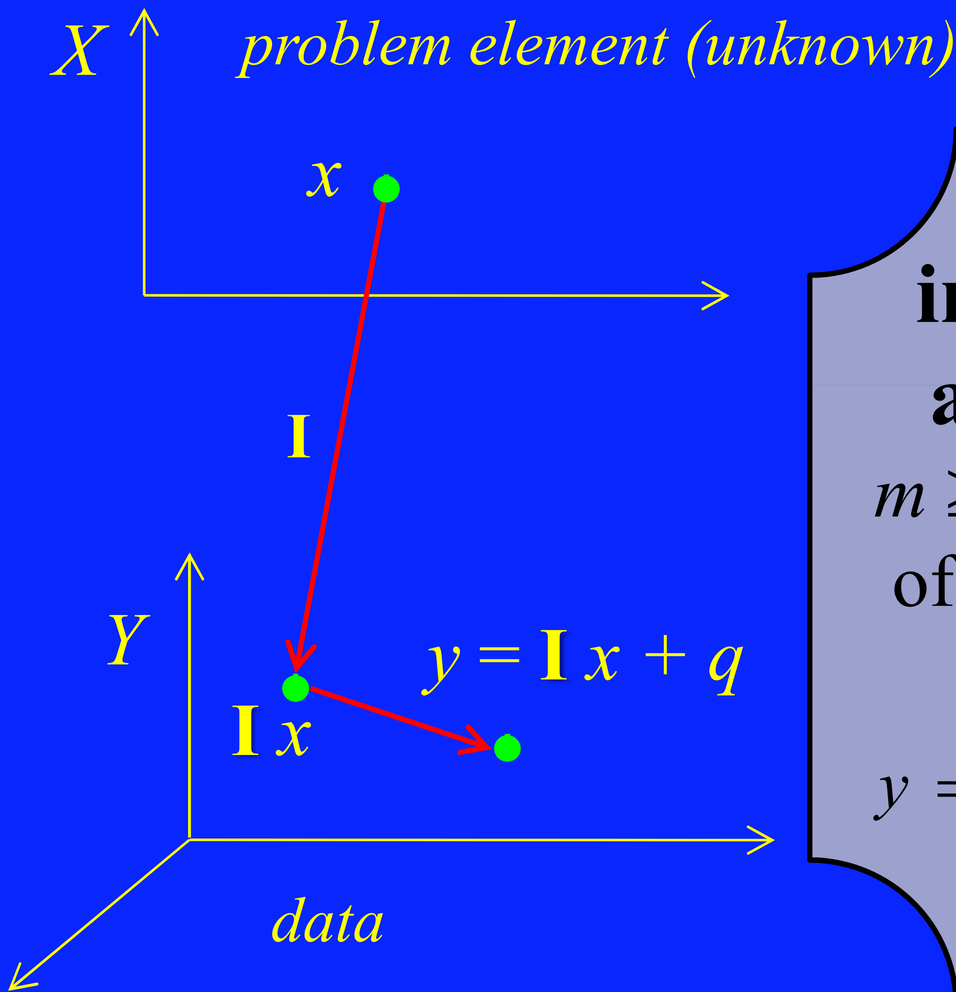
CNR-IEIIT

Spaces and Operators



CNR-IEIIT

Spaces and Operators



information operator and measurements

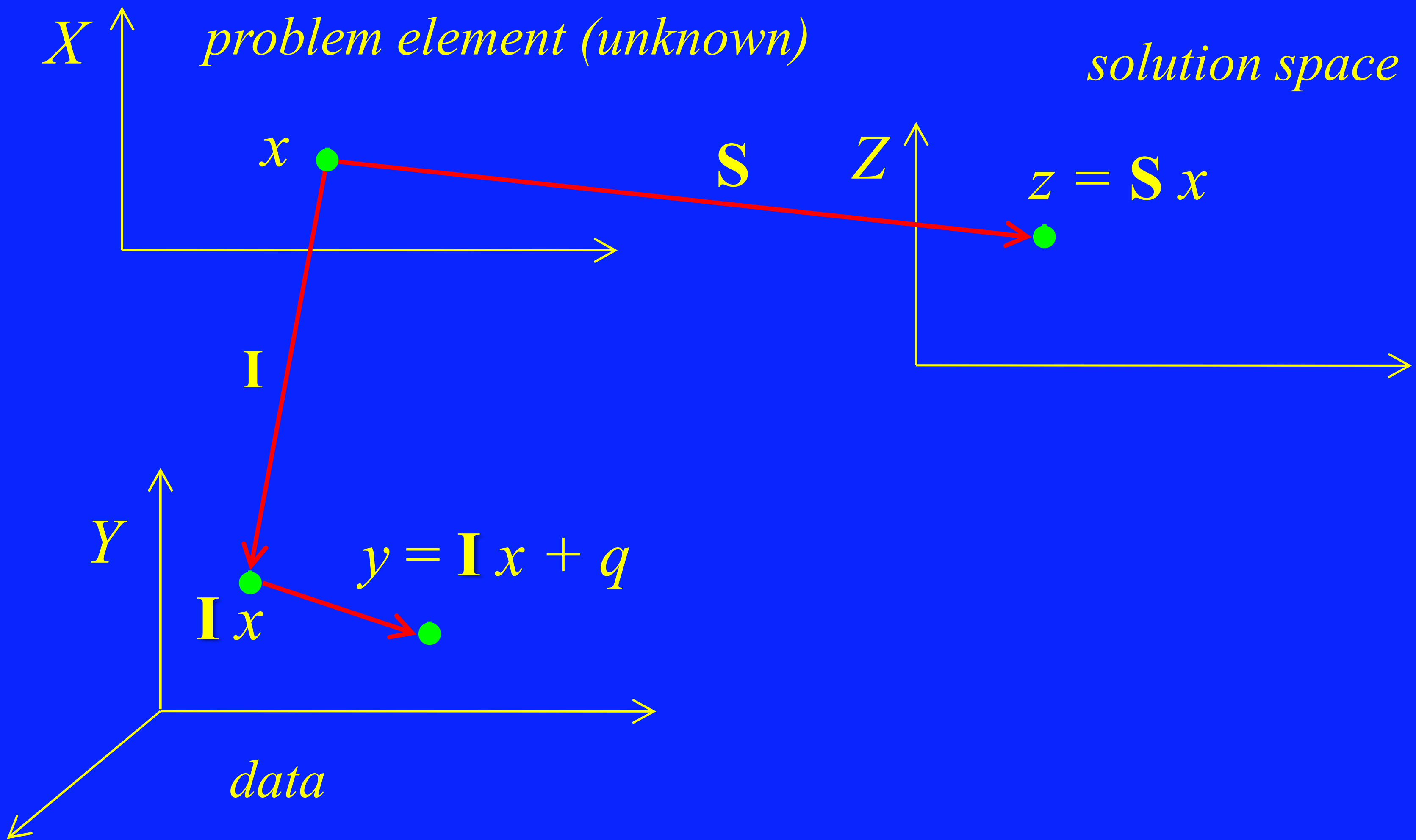
$m \geq n$ noisy measurements of $\xi(x, t)$ are available for

$$t_1 < t_2 < \cdots < t_m$$

$$y = [\Phi(t_1) \cdots \Phi(t_m)]^T x + q$$

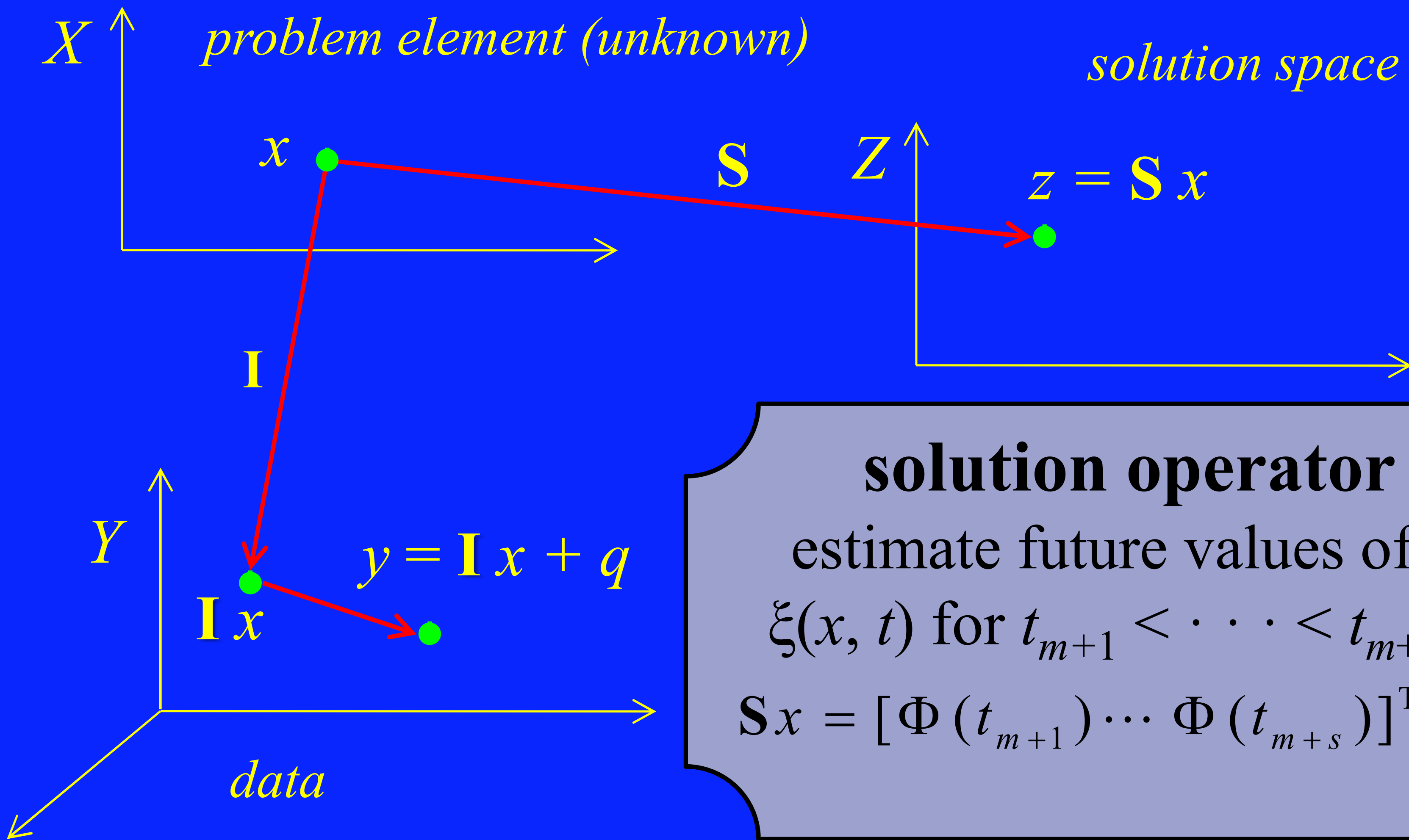
CNR-IEIIT

Spaces and Operators



CNR-IEIIT

Spaces and Operators



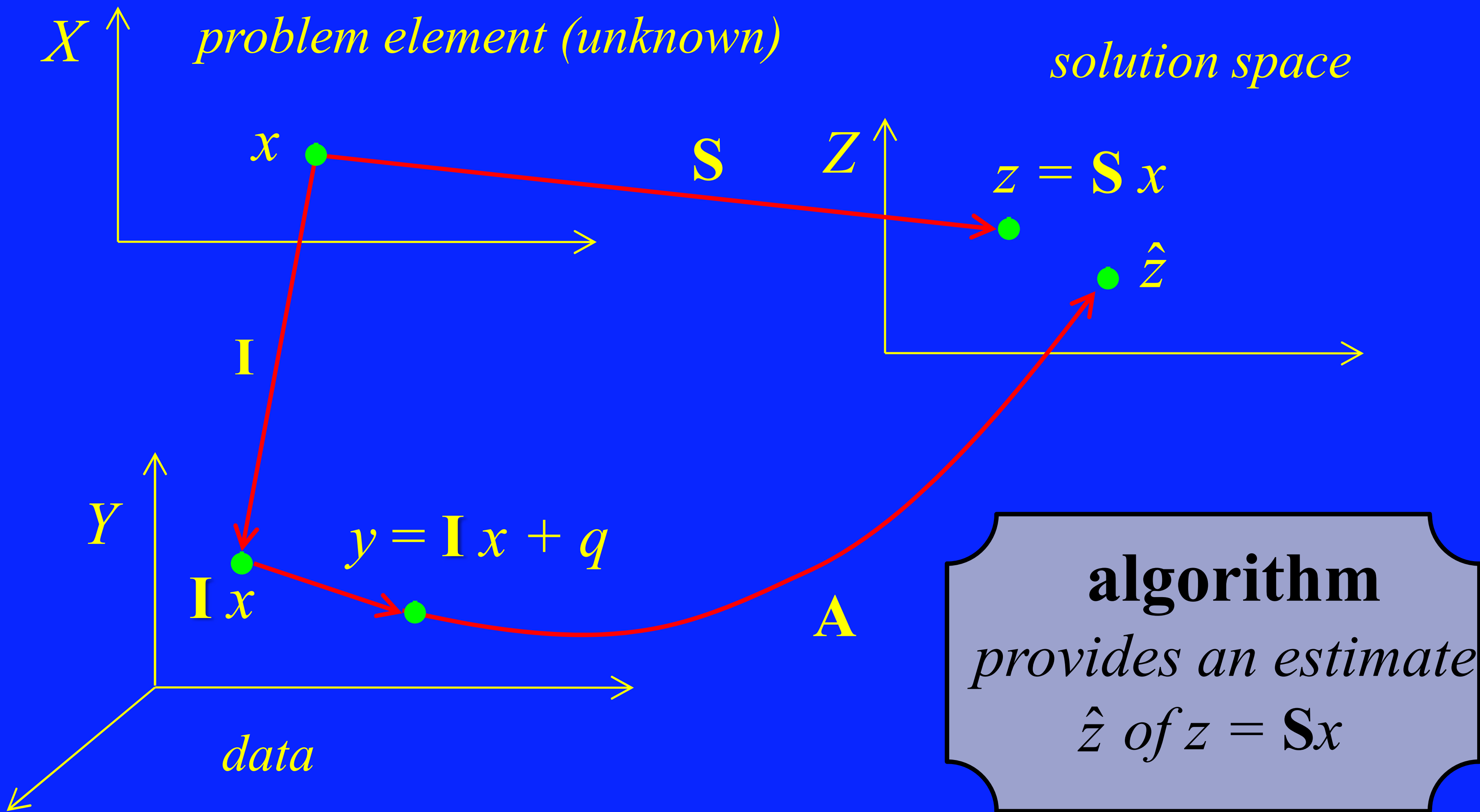
solution operator

estimate future values of

$\xi(x, t)$ for $t_{m+1} < \cdots < t_{m+s}$

$$\mathbf{S}x = [\Phi(t_{m+1}) \cdots \Phi(t_{m+s})]^{\mathrm{T}} x$$

Spaces and Operators



Ingredients

- Problem element $x \in X = \mathbf{R}^n$ with prior information K
- Information operator (linear) $\mathbf{I}$: $X \rightarrow Y = \mathbf{R}^m$ $(m \geq n)$
- Information $\mathbf{I}\,x$ corrupted by noise q
- Data $y = \mathbf{I}\,x + q$
- Bounding set Q for q
- Solution operator (linear) $\mathbf{S}$: $X \rightarrow Z = \mathbf{R}^s$ $(n \geq s)$
- Algorithm (nonlinear) $\mathbf{A}$: $Y \rightarrow Z$

CNR-IEIIT

Problem Element x

- Problem element (unknown) $x \in K \subseteq X$
- K represents prior information (if available)

CNR-IEIIT

Problem Element x

- Problem element (unknown) $x \in K \subseteq X$
- K represents prior information (if available)

example

consider input-output pair $\xi(x, t)$ of a dynamic system

$$\xi(x,t) = \sum_{i=1}^{n} x_i \varphi_i(t) = \Phi^{\mathrm{T}}(t)x$$

$$K = \{x \in X : x_i > 0, i = 1, 2, \ldots, n\}$$

CNR-IEIIT

Noise q and Bounding Set Q

- We assume that

$$q \in Q = \{q \colon \|q\| \leq \rho\} \subset \mathbf{R}^m$$

where $\|.\|$ is the l_p norm

CNR-IEIIT

Approximation Error

- We study the approximation error

$$\| \mathbf{S}x - \mathbf{A}(y) \|$$

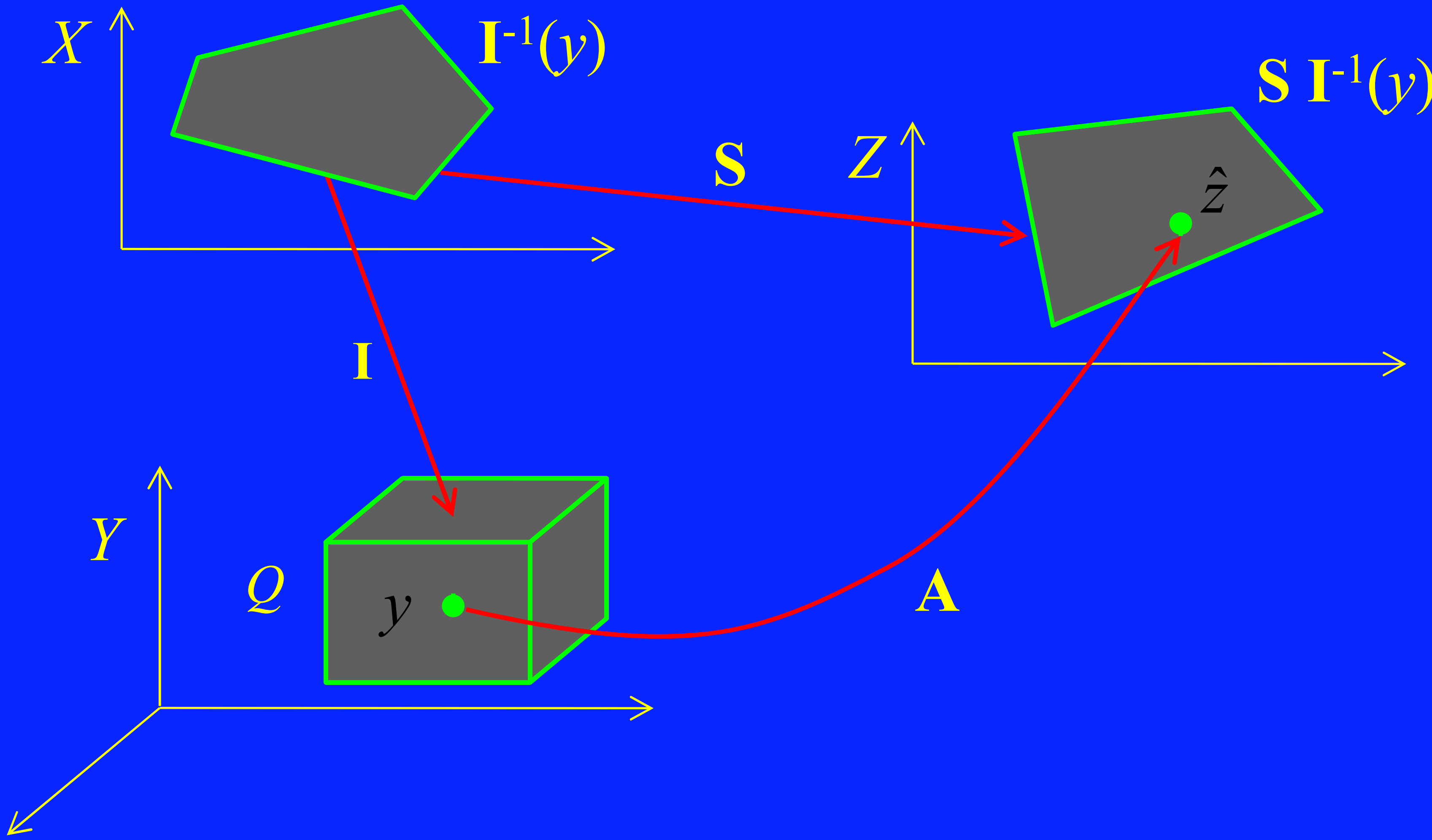
where ||.|| is the l_p norm

Consistency Set

- Consistency set is defined as

$$\mathbf{I}^{-1}(y) = \{ x \in K : \text{there exists } q \in Q \text{: } y = \mathbf{I}x + q \}$$

Consistency Set $I^{-1}(y)$



Worst-Case Setting

CNR-IEIIT

Objective for Worst-Case Setting

❖ **Objective:** construct an algorithm **A** (nonlinear)

$$\hat{z} = \mathbf{A}(y) \text{ of } z = \mathbf{S}\,x$$

and compute the worst-case radius $r^{wc}(\mathbf{A}, y)$

- prior information $x \in K \subseteq X$ (if available)
- data $y = \mathbf{I}\,x + q \in Y$
- bounding set $Q = \{q\colon \|q\| \leq \rho\} \subset \mathbf{R}^m$

Worst-Case Radius and Optimal Algorithm

CNR-IEIIT

- Given data y and the algorithm $\mathbf{A}$, the worst-case radius is defined as

$$r^{\mathrm{wc}}(\mathbf{A}, y) = \max_{x \in \mathbf{I}^{-1}(y)} \| \mathbf{S}x - \mathbf{A}(y) \|$$

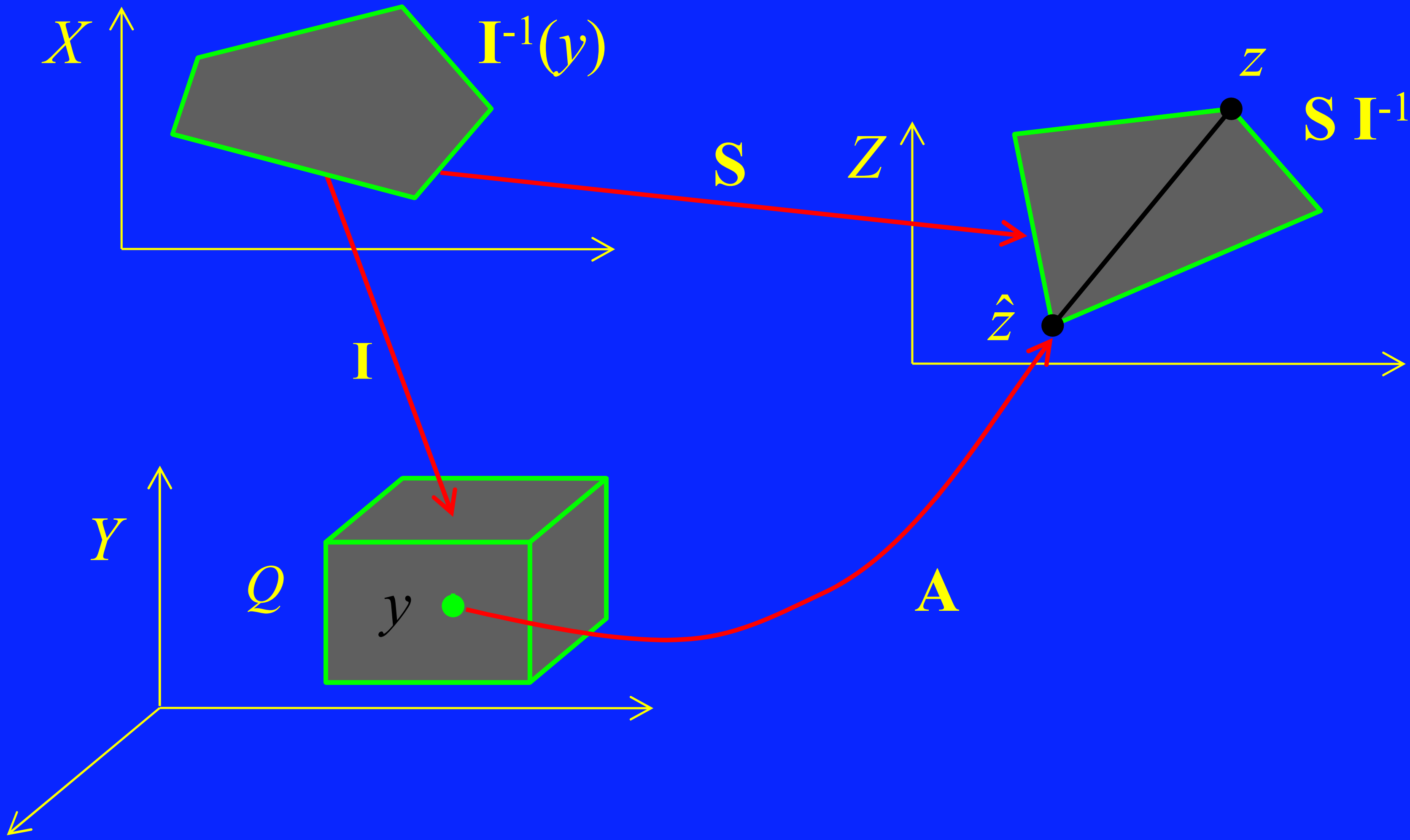
- Given y a worst-case optimal algorithm $\mathbf{A}_{\mathrm{o}}^{\mathrm{wc}}$ is defined as

$$r_{\mathrm{o}}^{\mathrm{wc}}(y) = r^{\mathrm{wc}}(\mathbf{A}_{\mathrm{o}}^{\mathrm{wc}}, y) = \inf_{\mathbf{A}} \max_{x \in \mathbf{I}^{-1}(y)} \| \mathbf{S}x - \mathbf{A}(y) \|$$

- $r_{\mathrm{o}}^{\mathrm{wc}}(y)$ is the worst-case radius of the optimal algorithm

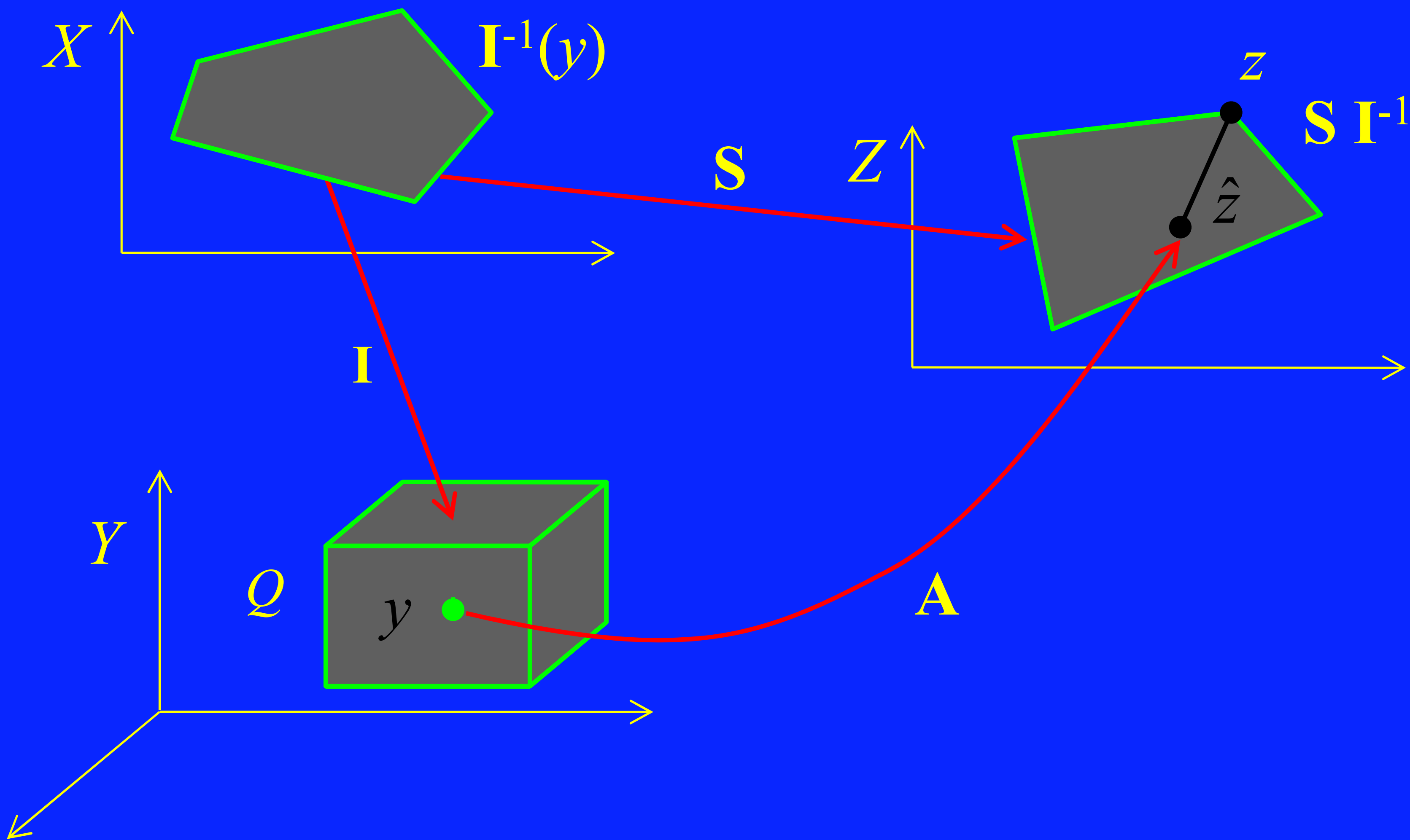
Error of the Algorithm

CNR-IEIIT



CNR-IEIIT

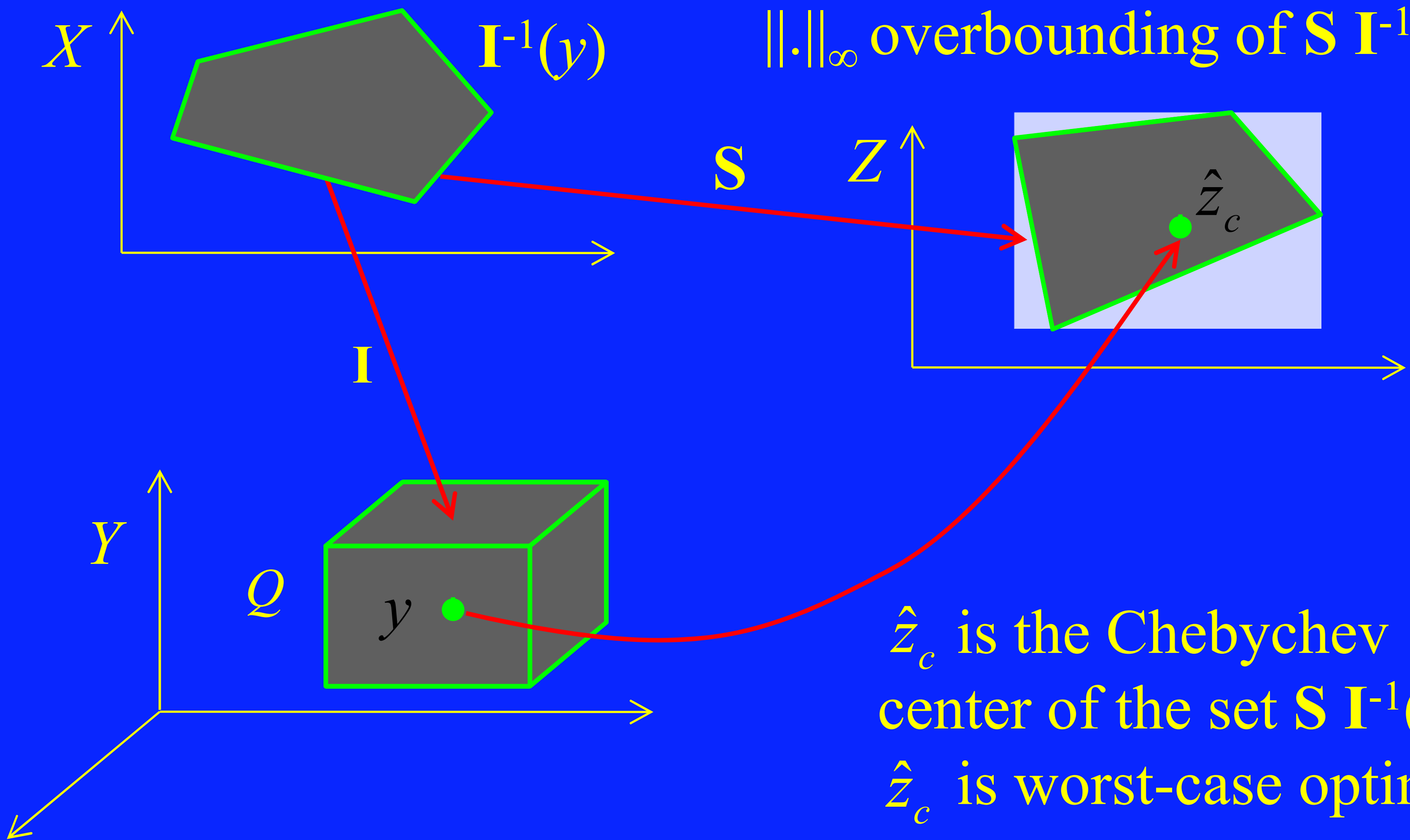
Worst-Case Optimal Algorithm



Optimal Algorithms and Chebychev Center

CNR-IEIIT

$\|.\|_\infty$ overbounding of $\mathbf{S}\,\mathbf{I}^{-1}(y)$



$\hat{z}_c$ is the Chebychev center of the set $\mathbf{S}\,\mathbf{I}^{-1}(y)$

$\hat{z}_c$ is worst-case optimal

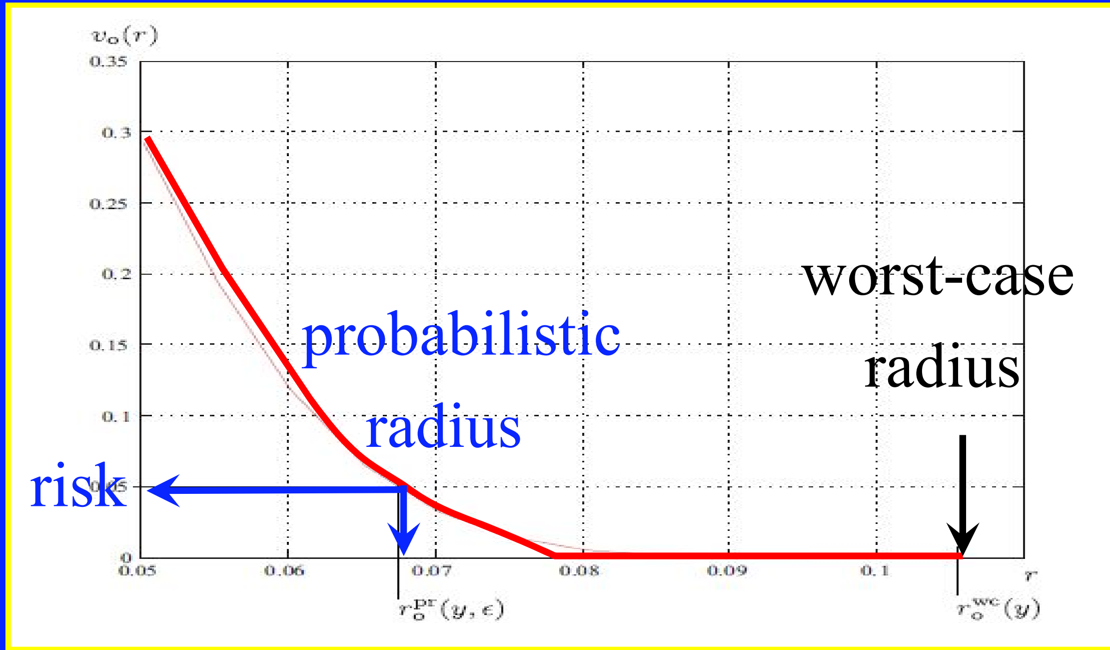
Probabilistic Setting

CNR-IEIIT

Probabilistic Setting: Conservatism Reduction

In the probabilistic setting we reduce the conservatism of the worst-case setting at the expense of a "small" risk

violation function $v_o(r)$ shows how the probabilistic risk $\varepsilon \in (0,1)$ changes as a function of the radius r

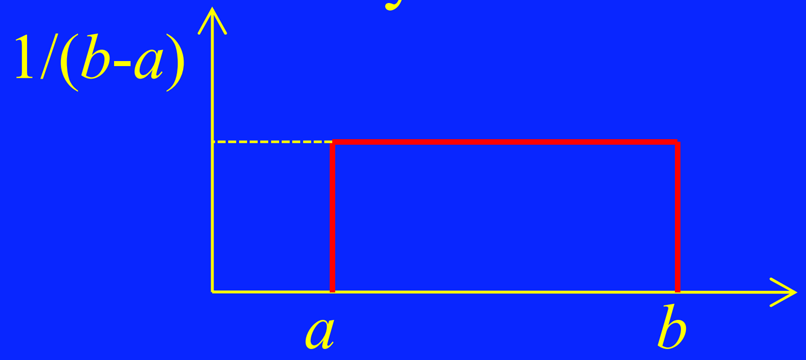


Random Noise

- Assume that noise q is a random vector with uniform pdf $\mathcal{U}[Q]$ and support set Q

Uniform Density $\mathcal{U}[Q]$

- Univariate uniform density $\mathcal{U}[a,b]$



- Multivariate uniform density $\mathcal{U}[Q]$

$$\mathcal{U}[Q] = \begin{cases} \dfrac{1}{\mathrm{vol}(Q)} & \text{if } q \in Q \\ 0 & \text{otherwise} \end{cases}$$

CNR-IEIIT

Objective for Probabilistic Setting

- **Objective:** construct an algorithm **A** (nonlinear)

$$\hat{z} = \mathbf{A}(y) \;\; \text{of} \;\; z = \mathbf{S}\,x$$

and compute the probabilistic radius $r^{\mathrm{pr}}(\mathbf{A}, y, \varepsilon)$

- probabilistic risk $\varepsilon \in (0,1)$
- prior information $x \in K \subseteq X$ (if available)
- data $y = \mathbf{I}\,x + q \in Y$
- random vector q with uniform pdf
- bounding set $Q = \{q\colon \|q\| \leq \rho\} \subset \mathbf{R}^m$

CNR-IEIIT

Probabilistic Radius

- Given data y, accuracy $\varepsilon \in (0,1)$ and the algorithm **A**, the probabilistic radius is defined as

$$r^{\mathrm{pr}}(\mathbf{A}, y, \varepsilon) = \inf_{\Omega:\mu(\Omega)<\varepsilon} \max_{x\in\{\mathbf{I}^{-1}(y)\setminus\Omega\}} \|\mathbf{S}x - \mathbf{A}(y)\|$$

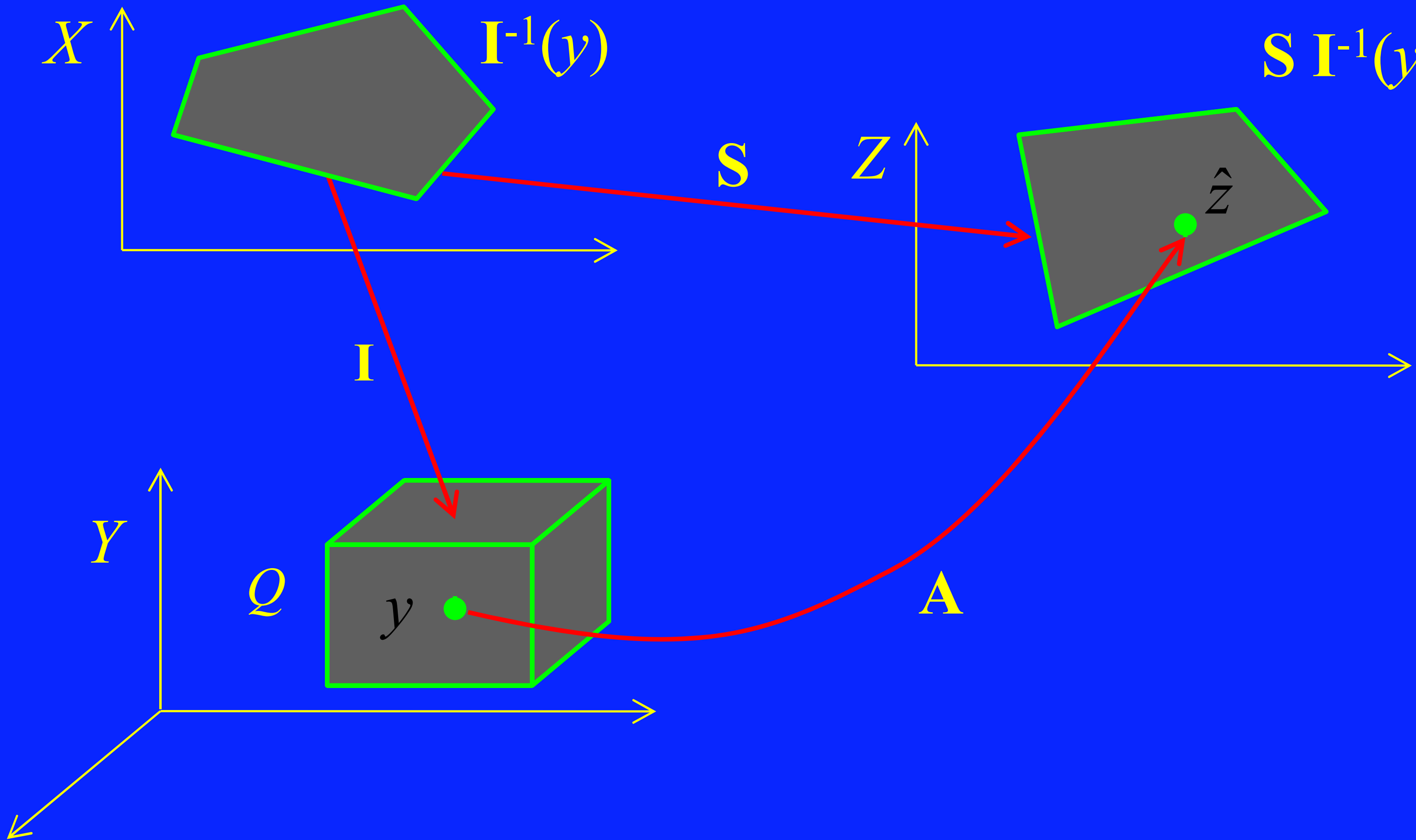
- **Remark:** we replace the consistency set $\mathbf{I}^{-1}(y)$ with

$$\{\mathbf{I}^{-1}(y)\setminus\Omega\}$$

- We discard sets Ω of measure $\mu(\Omega) < \varepsilon$
- Probabilistic radius is smaller than worst-case radius

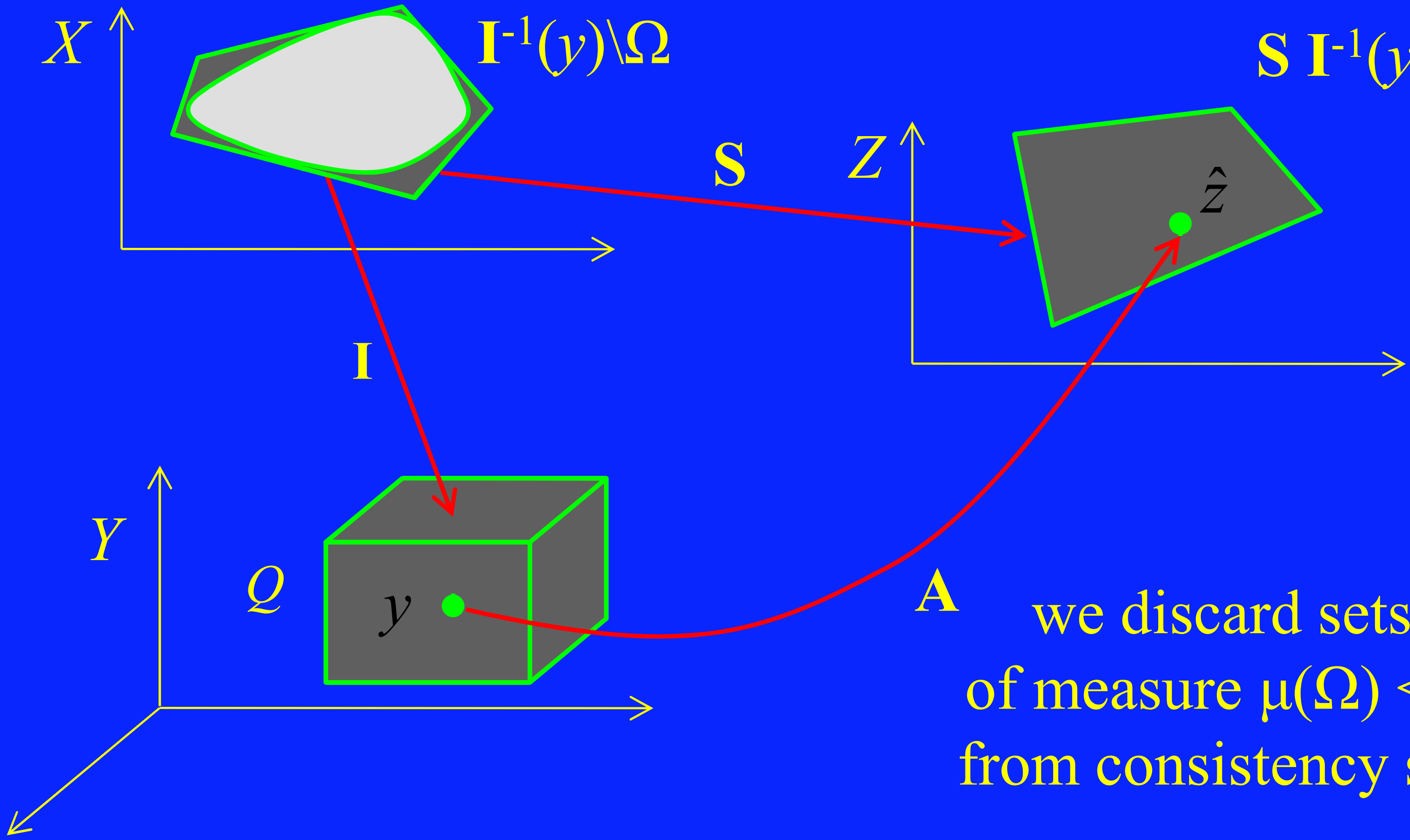
CNR-IEIIT

Consistency Set $\mathbf{I}^{-1}(y)$



CNR-IEIIT

Set $\mathbf{I}^{-1}(y) \setminus \Omega$



we discard sets Ω of measure $\mu(\Omega) < \varepsilon$ from consistency set

CNR-IEIIT

Probabilistic Optimal Algorithm

- Given data y and risk $\varepsilon \in (0,1)$ a probabilistic optimal algorithm $\mathbf{A}_o^{pr}$ is defined as

$$r_o^{pr}(y,\varepsilon) = r^{pr}(\mathbf{A}_o^{pr}, y, \varepsilon) = \inf_{\mathbf{A}} \inf_{\Omega:\mu(\Omega)<\varepsilon} \max_{x \in \{\mathbf{I}^{-1}(y)\setminus\Omega\}} \| \mathbf{S}x - \mathbf{A}(y) \|$$

- $r_o^{pr}(y,\varepsilon)$ is the probabilistic radius of optimal algorithm

CNR-IEIIT

Measure of Sets $\mathbf{I}^{-1}(y)$ and $\mathbf{S}\,\mathbf{I}^{-1}(y)$

Theorem: Let $q \sim \mathcal{U}[Q]$ and $K = \mathbf{R}^n$. Then, for any $y \in Y$

- $\mu\left\{x \in \mathbf{I}^{-1}(y)\right\}$ is uniform
- $\mu\left\{z = \mathbf{S}x \in \mathbf{S}\,\mathbf{I}^{-1}(y)\right\}$ is log-concave

❖ This result also holds when Q is a convex set

---

def log-concave: $\mu\left\{\lambda A + (1-\lambda)B\right\} \geq \mu\left\{A\right\}^{\lambda}\,\mu\left\{B\right\}^{1-\lambda}$

Violation Function and its Computation

CNR-IEIIT

Optimal Violation Function

- Given $r > 0$ and $\mathbf{A}$, define a violation function

$$v(r,\mathbf{A}) = \mu\left\{x \in \mathbf{I}^{-1}(y) : \| \mathbf{S}x - \mathbf{A}(y) \| > r\right\}$$

- Given $r > 0$, the optimal violation function $v_{\mathrm{o}}(r)$ is

$$v_{\mathrm{o}}(r) = \inf_{\mathbf{A}} v(r,\mathbf{A}) = \inf_{\mathbf{A}} \mu\left\{x \in \mathbf{I}^{-1}(y) : \| \mathbf{S}x - \mathbf{A}(y) \| > r\right\}$$

CNR-IEIIT

Chance-Constrained Optimization

- Computation of the probabilistic radius of the optimal algorithm

$$r_{\mathrm{o}}^{\mathrm{pr}}(y,\varepsilon) = r^{\mathrm{pr}}(\mathbf{A}_{\mathrm{o}}^{\mathrm{pr}}, y, \varepsilon)$$

may be reformulated as a chance-constrained problem

$$r^{\mathrm{pr}}(\mathbf{A}_{\mathrm{o}}^{\mathrm{pr}}, y, \varepsilon) = \min\left\{ r : v_{\mathrm{o}}(r) \le \varepsilon \right\}$$

- For any $\varepsilon \in (0,1)$, we can compute $r^{\mathrm{pr}}(\mathbf{A}_{\mathrm{o}}^{\mathrm{pr}}, y, \varepsilon)$ solving a one-dimensional optimization problem in $r > 0$

CNR-IEIIT

Properties of Optimal Violation Function

- **Theorem:** Let $q \sim \mathcal{U}[Q]$, $K = \mathbf{R}^n$ and $\mathbf{S}$ = identity. For fixed $r > 0$, optimal violation is given by

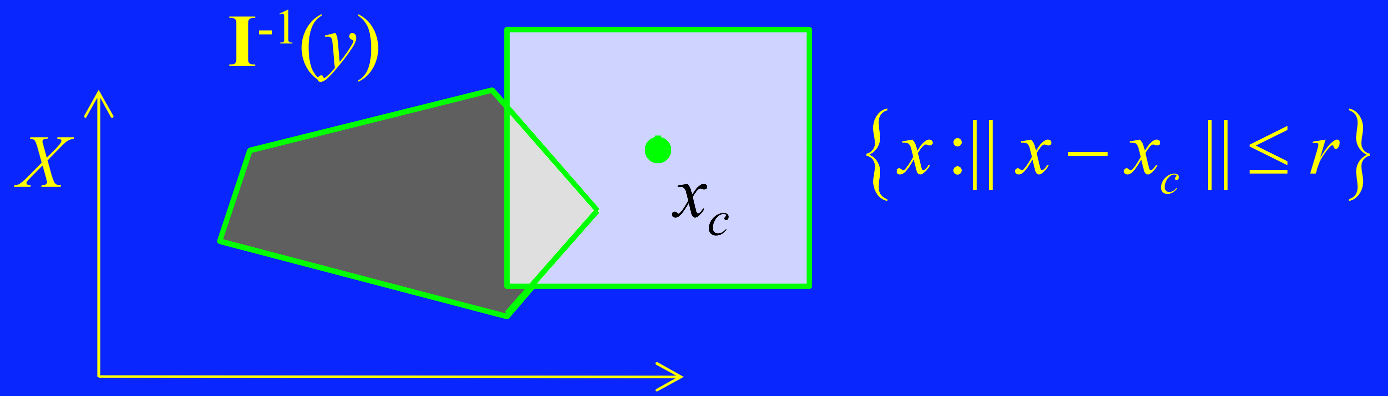
$$v_o(r) = \inf_{x_c} \frac{\operatorname{vol}\left\{\mathbf{I}^{-1}(y) \setminus \{x : \| x - x_c \| \le r\}\right\}}{\operatorname{vol}\left\{\mathbf{I}^{-1}(y)\right\}}$$

- This optimization problem is quasi-convex for all well-defined x_c (i.e. non-zero volume)
- Optimal violation $v_o(r)$ is right-continuous and non-increasing for $r > 0$

CNR-IEIIT

Optimal Violation

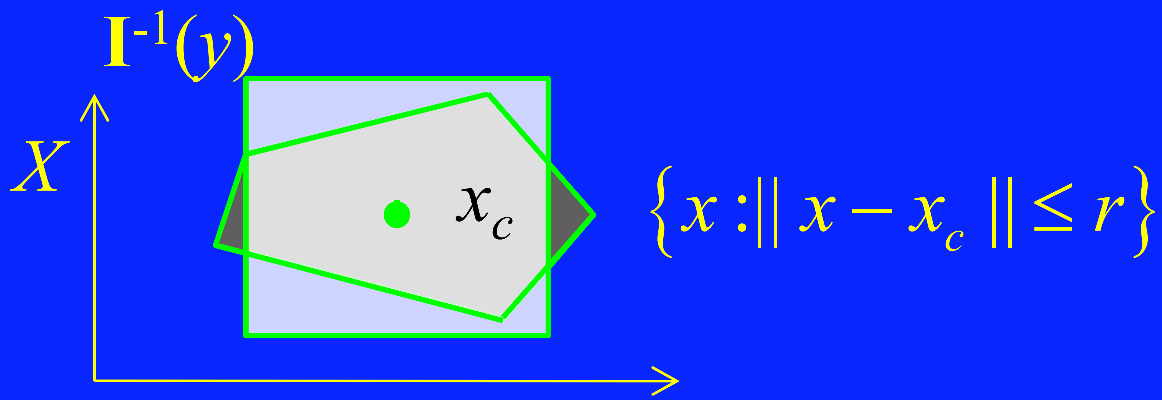
proof based on the Brunn-Minkovski inequality for intersection of convex sets



$$v_o(r) = \inf_{x_c} \frac{\mathrm{vol}\{\mathbf{I}^{-1}(y) \setminus \{x : \| x - x_c \| \le r\}\}}{\mathrm{vol}\{\mathbf{I}^{-1}(y)\}}$$

CNR-IEIIT

Optimal Violation



$$v_o(r) = \inf_{x_c} \frac{\mathrm{vol}\left\{\mathbf{I}^{-1}(y) \setminus \{x : \| x - x_c \| \leq r\}\right\}}{\mathrm{vol}\left\{\mathbf{I}^{-1}(y)\right\}}$$

CNR-IEIIT

Extensions and Algorithms

- Extensions for non-identity solution operator **S**
- Computation of optimal violation $v_{\mathrm{o}}(r)$ requires computation of vol$\left\{\mathbf{I}^{-1}(y) \setminus \{x : \| x - x_c \| \leq r\}\right\}$
- Volume computation of polytopes is NP-hard

- Two relaxation algorithms:
  - Hard - deterministic (SDP-based)
  - Soft - probabilistic (randomized-based)

CNR-IEIIT

Conclusions

- This probabilistic setting is completely new within systems and control
- Applications:
  - classical: identification, estimation, …
  - modern: PageRank computation in Google

CNR-IEIIT

Conclusions

- (modern): PageRank computation in Google

H. Ishii and R. Tempo "Distributed Randomized Algorithms for the PageRank Computation," IEEE TAC, 2010

example

$$x^* = \left[(1-m)(H+\Delta) + \frac{m}{n} U \right] x^*$$

x^* is PageRank, m=0.15, n is the number of pages, H is the hyperlink matrix, Δ represents link failures and U is a rank-one matrix

CNR-IEIIT

Main References

- F. Dabbene and R. Tempo, "Probabilistic and Randomized Tools for Control Design," The Control Handbook (W. S. Levine Ed.), *Taylor & Francis*, 2010
- G. Calafiore, F. Dabbene and R. Tempo "Research on Probabilistic Design Methods," Automatica, 2011
- R. Tempo, G. Calafiore and F. Dabbene, "Randomized Algorithms for Analysis and Control of Uncertain Systems," *Springer-Verlag*, London, 2005 (second edition in preparation)

http://staff.polito.it/roberto.tempo/