Revisiting Statistical Learning Theory for Control Systems Design

Roberto Tempo

IEIIT-CNR
Politecnico di Torino
tempo@polito.it

Statistical Learning Theory for Control Design of Uncertain Systems

- Statistical learning theory is a branch of the theory of empirical processes
- Significant results have been obtained with application in various areas, including neural networks, system identification, pattern recognition, …

Statistical Learning Theory for Control Design of Uncertain Systems

- The use of statistical learning theory for control design of uncertain systems was initiated by Vidyasagar[1]

- Motivation: Studying computationally *difficult* control problems…

[1] M. Vidyasagar (1998)

IEIIT-CNR

Computationally Difficult Problems

- Various robust control problems are computationally intractable
- This is generally meant for $\mathcal{NP}$-hard problems
- Examples:
  - static output feedback
  - stability of interval matrices
  - μ computation
  - system design with uncertainty

IEIIT-CNR

Example of a Difficult Control Problem

- Consider the *uncertain* system

$$\dot{x}(t) = A(\Delta)x(t) + Bu(t)$$

where $\Delta \in \mathcal{B}$

- The goal is to design a state feedback controller

$$u(t) = Kx(t)$$

such that the closed loop system

$$\dot{x}(t) = (A(\Delta) + BK)x(t)$$

is Hurwitz stable for all $\Delta \in \mathcal{B}$

Problems Addressed in this Talk

- Revisiting statistical learning theory for control design
- Two main problems remain unsolved:

  - conservatism of the existing sample size bounds is critical in practice

  - randomization of the controller parameters may be debatable

IEIIT-CNR

Notation

- Uncertainty Δ is bounded in a set $\mathcal{B}$
- Controller parameters θ varying in a set Θ
- $f_i : \Theta \times \mathcal{B} \rightarrow \mathbb{R}$ represents a system constraint
- We have n constraints f_i (but they are subject to uncertainty Δ)

IEIIT-CNR

Notation

- Uncertainty Δ is bounded in a set $\mathcal{B}$
- Controller parameters θ varying in a set Θ
- $f_i : \Theta \times \mathcal{B} \rightarrow R$ represents a system constraint
- We have n constraints f_i (but they are subject to uncertainty Δ)
- Examples of constraints
  - closed loop poles into the left half plane
  - $\mathcal{H}_\infty$ bound on the sensitivity function

Constrained Feedback Design with Uncertainty

- The problem is to determine controller parameters $\theta \in \Theta$ such that the constraints

$$f_i(\theta, \Delta) \leq 0,\ i = 1, \ldots, n, \text{ for all } \Delta \in \mathcal{B}$$

are satisfied (i.e. the problem is robustly feasible)

- If the problem is robustly feasible then the objective is to minimize a performance function $J(\theta)$ subject to the set of constraints

IEIIT-CNR

Binary Violation Function *g*

- We introduce the function *g*

  $g: \Theta \times \mathcal{B} \rightarrow \{0,1\}$

  which is a binary measurable function defined as follows

$$g(\theta, \Delta) = 0 \quad if \;\; f_i(\theta, \Delta) \leq 0,\; i = 1, \ldots, n$$
$$g(\theta, \Delta) = 1 \quad otherwise$$

IEIIT-CNR

Semi-Infinite Feasibility

- **Semi-Infinite Feasibility Problem**: Find θ, if it exists, in the feasible set

$$\{\theta \in \Theta: g(\theta, \Delta) = 0 \text{ for all } \Delta \in \mathcal{B}\} \qquad \text{(SIFP)}$$

where $g: \Theta \times \mathcal{B} \rightarrow \{0,1\}$ is a binary measurable function

Semi-Infinite Optimization

- **Semi-Infinite Optimization Problem**: If the feasible set is nonempty, find an optimal/suboptimal solution of the problem

$$\min_{\theta \in \Theta} \; J(\theta) \quad \text{subject to } g(\theta, \Delta) = 0 \text{ for all } \Delta \in \mathcal{B} \quad \text{(SIOP)}$$

where $J: \Theta \rightarrow (-\infty, \infty)$ is a measurable function

Probability of Violation

- Given $\theta \in \Theta$, the probability of violation for the function $g(\theta, \Delta)$ is defined as

$$E_g(\theta) = \mathrm{Prob}\{\Delta \in \mathcal{B}\colon g(\theta, \Delta) = 1\}$$

IEIIT-CNR

Empirical Mean of Violation

- Generate N i.i.d. samples (multisample)

$$\Delta^1, \Delta^2, \ldots, \Delta^N \in \mathcal{B}$$

according to a given probability measure

- Given $\theta \in \Theta$, the empirical mean of $g(\theta, \Delta)$ is defined as

$$\hat{E}_N(\theta) = 1/N \sum_{i=1}^{N} g(\theta, \Delta^i)$$

- Since g is a binary function $\hat{E}_N(\theta) \in [0,1]$

Randomized Strategy: Main Idea

- **Main Idea**: Develop a randomized PAC (probably approximately correct) strategy solving semi-infinite feasibility and optimization problems such that the empirical mean of violation is small or zero

IEIIT-CNR

Randomized Strategy - 1

- Generate N i.i.d. samples

$$\Delta^1, \Delta^2, \ldots, \Delta^N \in \mathcal{B}$$

according to a given probability measure

- The multisample $\{\Delta^1, \Delta^2, \ldots, \Delta^N\}$ is denoted as $\Delta^{1,\ldots,N}$

Randomized Strategy - 2

- Given a *level* $\rho \in [0,1)$, find (if possible) a feasible solution of the constraint

$$\hat{E}_N(\theta) \leq \rho \qquad \text{(RFP)}$$

- If a feasible solution exists, find an optimal/suboptimal solution of the randomized optimization problem

$$\min_{\theta \in \Theta} \; J(\theta) \quad \text{subject to} \quad \hat{E}_N(\theta) \leq \rho \qquad \text{(ROP)}$$

Comments - 1

- Solving the original semi-infinite feasibility and optimization problems is extremely difficult given the infinite number of constraints
- Using the concept of empirical mean, the feasibility and optimization problems of the randomized strategy have only one constraint with a finite sum (for fixed θ)

Comments - 2

- Considering a non-zero level ρ broadens the class of problems we study
- For example, in many manufacturing processes allowing a (small) probability of violation has the effect of reducing the production costs
- Given *accuracy* ε if $|E_g(\theta) - \hat{E}_N(\theta)| \leq \varepsilon$ with "high" *confidence* $\delta \in (0,1)$ for all $\theta \in \Theta$, $\hat{E}_N(\theta)$ is a "good" estimate of $E_g(\theta)$

Feasibility of Nonconvex Problems

Estimating the Probability of Violation

- For fixed $\theta \in \Theta$, given $\varepsilon \in (0,1)$, from the Hoeffding inequality we obtain

$$\mathrm{Prob}\{\Delta^{1,\ldots,N} : | E_g(\theta) - \hat{\mathrm{E}}_N(\theta) | \geq \varepsilon\} \leq 2\mathrm{e}^{(-2N\varepsilon^2)}$$

where e denotes the Euler number

- To guarantee confidence $\delta \in (0,1)$, we need to take N samples such that $2\mathrm{e}^{(-2N\varepsilon^2)} \leq \delta$ holds
- We obtain the (additive) Chernoff bound

$$N \geq 1/(2\varepsilon^2) \left[\log(2/\delta) \right]$$

Hoeffding Inequality versus Chernoff Bound

- The Hoeffding inequality provides a bound on the tail distribution

$$2e^{(-2N\varepsilon^2)}$$

- From the computational point of view, evaluating the minimum value of N that $2e^{(-2N\varepsilon^2)} \leq \delta$ is immediate (given ε and δ it is a one-parameter problem)
- The Chernoff bound provides a fundamental *explicit* relation (sample complexity) $N = N(\varepsilon, \delta)$ showing that $1/\varepsilon$ enters quadratically and $1/\delta$ logarithmically

IEIIT-CNR

Uniform Hoeffding Inequality - 1

- The Chernoff bound and the Hoeffding inequality hold only for fixed $\theta \in \Theta$
- In control problems, θ is a design variable ranging in the set Θ of admissible controller parameters
- Some results are available for finite families (i.e. a finite number of θ)

Uniform Hoeffding Inequality - 2

- We need to resort to more sophisticated techniques aiming at developing *uniform* law of large numbers and computing related bounds
- This objective is the essence of statistical learning theory

Probability of Two-Sided Failure

- Given N, $\varepsilon \in (0,1)$ and g, the probability of two-sided failure, is defined as

$$q_g(N, \varepsilon) = \text{Prob}\{\Delta^{1,\ldots,N} : \sup_{\theta \in \Theta} | E_g(\theta) - \hat{E}_N(\theta) | > \varepsilon\}$$

- Remark: Taking "sup" provides a uniform Hoeffding inequality
- For technical reasons we take strict inequalities

UCEM Property and VC Dimension[1]

- The function g: $\Theta \times \mathcal{B} \to \{0,1\}$ enjoys the Uniform Convergence of Empirical Means (UCEM) property if

$$q_g(N, \varepsilon) \to 0 \text{ as } N \to \infty$$

  for each $\varepsilon \in (0,1)$

- The notion of VC dimension (VC_g) is crucial in this context because it establishes the “richness” of a family of functions g and it enters directly into the sample complexity

[1] V.N. Vapnik and A. Ya. Chervonenkis (1981) – V.N. Vapnik (1998)

IEIIT-CNR

Example: Bounding the VC Dimension

- Theorem[1]: Suppose that g: $\Theta \times \mathcal{B} \to \{0,1\}$ is a Boolean expression consisting of polynomials

$$\beta_1(\theta, \Delta), \ldots, \beta_k(\theta, \Delta)$$

in θ_i for $i = 1,\ldots, n$, and their degree is no larger than α. Then

$$VC_g \leq 2n \log_2(4e\alpha k) = d$$

- Example: For fixed Δ we have

$$g = \beta_1(\theta) = 3 + 2\,\theta_1^2 - 5\,\theta_2^4\,\theta_3 + \ldots + 4\,\theta_1^2\,\theta_2\,\theta_4^7 \qquad k=1,\ n=4,\ \alpha = 7$$

[1] M. Vidyasagar (2001)

IEIIT-CNR

Bounding the Probability of Two-Sided Failure

- Theorem[1]: Let $g: \Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $\mathrm{VC}_g \leq d < \infty$. Suppose that $\varepsilon, \delta \in (0,1)$ and $N \geq d$. Then

$$q_g(N, \varepsilon) \leq 4\,(2eN/d)^d\, e^{(-N\varepsilon^2/8)}$$

- Moreover, $q_g(N, \varepsilon) \leq \delta$ if

$$N \geq \max\left\{16/\varepsilon^2 \log(4/\delta),\ 32d/\varepsilon^2 \log(32e/\varepsilon^2)\right\}$$

[1] M. Vidyasagar (2001)

IEIIT-CNR

Sample Complexity Improvements

- **Theorem**[1]: Let g: $\Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $\mathrm{VC}_g \leq d < \infty$. Suppose that $\varepsilon, \delta \in (0,1)$. Then $q_g(N, \varepsilon) \leq \delta$ if

$$N \geq 1.2/\varepsilon^2 \left[\log(4e^{2\varepsilon}/\delta) + d \log(12/\varepsilon^2) \right]$$

- **Remark**: This improves the previous result, but the sample complexity still grows as $1/\varepsilon^2 \log(1/\varepsilon^2)$. To obtain more substantial improvements, we introduce a new notion of probability of violation

[1] T. Alamo, R. Tempo and E.F. Camacho (2009)

IEIIT-CNR

Probability of One-Sided Constrained Failure

- Question: Suppose that N and ρ are given. Suppose that after drawing N samples, we find θ_0 such that $\hat{E}_N(\theta_0) \leq \rho$. What is the probability that the difference between $E_g(\theta_0)$ and $\hat{E}_N(\theta_0)$ is larger than ε?
- Given N, $\varepsilon \in (0,1)$, $\rho \in [0,1)$ and g, the probability of one-sided constrained failure is defined as

$$p_g(N, \varepsilon, \rho) =$$

$$\text{Prob}\{\Delta^{1,\ldots,N}: \text{ there exists } \theta \text{ such that } \hat{E}_N(\theta) \leq \rho,\ E_g(\theta) - \hat{E}_N(\theta) > \varepsilon\}$$

IEIIT-CNR

Bounding the Probability of One-Sided Constrained Failure

- **Theorem**[1]: Let g: $\Theta \times \mathcal{B} \to \{0,1\}$ be such that $VC_g \leq d < \infty$. Suppose that ε, $\delta \in (0,1)$ and $\rho \in [0,1)$ are given. Then $p_g(N, \varepsilon, \rho) \leq \delta$ if

$$N \geq 5(\rho + \varepsilon)/\varepsilon^2 \left[\log(4/\delta) + d \log(40(\rho + \varepsilon)/\varepsilon^2) \right]$$

[1] T. Alamo, R. Tempo and E.F. Camacho (2009)

Bounding the Probability of One-Sided Constrained Failure

IEIIT-CNR

- **Theorem**: Let g: $\Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $VC_g \leq d < \infty$. Suppose that $\varepsilon, \delta \in (0,1)$ and $\rho \in [0,1)$ are given. Then $p_g(N, \varepsilon, \rho) \leq \delta$ if

$$N \geq 5(\rho + \varepsilon)/\varepsilon^2 \left[\log(4/\delta) + d \log(40(\rho + \varepsilon)/\varepsilon^2) \right]$$

Taking $\rho = 0$

Bounding the Probability of One-Sided Constrained Failure

- **Theorem**: Let g: $\Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $VC_g \leq d < \infty$. Suppose that ε, $\delta \in (0,1)$ and $\rho \in [0,1)$ are given. Then $p_g(N, \varepsilon, \rho) \leq \delta$ if

$$N \geq 5(\cancel{\rho} + \varepsilon)/\varepsilon^2 \left[\log(4/\delta) + d \log(40(\cancel{\rho} + \varepsilon)/\varepsilon^2) \right]$$

Taking $\rho = 0$

Bounding the Probability of One-Sided Constrained Failure

IEIIT-CNR

- **Theorem**: Let g: $\Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $VC_g \leq d < \infty$. Suppose that ε, $\delta \in (0,1)$ and $\rho \in [0,1)$ are given. Then $p_g(N, \varepsilon, \rho) \leq \delta$ if

$$N \geq 5(\rho + \varepsilon)/\varepsilon^2 \left[\log(4/\delta) + d \log(40(\rho + \varepsilon)/\varepsilon^2) \right]$$

Taking $\rho = 0$, we have a cancellation

IEIIT-CNR

Unconstrained Case $\rho = 0$

- **Corollary**: Let $g: \Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $VC_g \leq d < \infty$. Suppose that $\varepsilon, \delta \in (0,1)$ are given. Then $p_g(N, \varepsilon, 0) \leq \delta$ if

$$N \geq 5/\varepsilon \left[\log(4/\delta) + d \log(40/\varepsilon) \right]$$

Bounding the Probability of One-Sided Constrained Failure

- **Theorem**: Let g: $\Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $VC_g \leq d < \infty$. Suppose that $\varepsilon, \delta \in (0,1)$ and $\rho \in [0,1)$ are given. Then $p_g(N, \varepsilon, \rho) \leq \delta$ if

$$N \geq 5(\rho + \varepsilon)/\varepsilon^2 \left[\log(4/\delta) + d \log(40(\rho + \varepsilon)/\varepsilon^2) \right]$$

Taking $\rho = \varepsilon$

Bounding the Probability of One-Sided Constrained Failure

IEIIT-CNR

- **Theorem**: Let $g: \Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $VC_g \leq d < \infty$. Suppose that $\varepsilon, \delta \in (0,1)$ and $\rho \in [0,1)$ are given. Then $p_g(N, \varepsilon, \rho) \leq \delta$ if

$$N \geq 5(\varepsilon + \varepsilon)/\varepsilon^2 \left[\log(4/\delta) + d \log(40(\varepsilon + \varepsilon)/\varepsilon^2) \right]$$

Taking $\rho = \varepsilon$

Bounding the Probability of One-Sided Constrained Failure

- **Theorem**: Let g: $\Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $VC_g \leq d < \infty$. Suppose that $\varepsilon, \delta \in (0,1)$ and $\rho \in [0,1)$ are given. Then $p_g(N, \varepsilon, \rho) \leq \delta$ if

$$N \geq 5(\varepsilon + \varepsilon)/\varepsilon^2 \left[\log(4/\delta) + d \log(40(\varepsilon + \varepsilon)/\varepsilon^2) \right]$$

Taking $\rho = \varepsilon$ we have another cancellation

Another Bound for $\rho = \varepsilon$

- **Corollary**: Let g: $\Theta \times \mathcal{B} \rightarrow \{0,1\}$ be such that $VC_g \leq d < \infty$. Suppose that ε, $\delta \in (0,1)$ and $\rho = \varepsilon$ are given. Then $p_g(N, \varepsilon, \rho{=}\varepsilon) \leq \delta$ if

$$N \geq 10/\varepsilon \left[\log(4/\delta) + d \log(80/\varepsilon) \right]$$

Comments

- In the unconstrained case ($\rho = 0$) and in the case ($\rho = \varepsilon$) the sample size is linear in d
- General parameterization $\rho = \varepsilon^l$ for $l > 0$
- N is a logarithmic function of $1/\delta$
- N depends on ε as $1/\varepsilon \log(1/\varepsilon)$
- N provides uniform sample complexity such that

$$\hat{E}_N(\theta) \leq \rho,\ E_g(\theta) > \hat{E}_N(\theta) + \varepsilon$$

and the probability of this event is at most δ

Optimization of Nonconvex Problems with $\rho = 0$

Semi-Infinite Optimization

- Semi-Infinite Optimization Problem: Find an optimal/suboptimal solution of the problem

$$\min_{\theta \in \Theta} J(\theta) \quad \text{subject to } g(\theta, \Delta) = 0 \text{ for all } \Delta \in \mathcal{B} \qquad \text{(SIOP)}$$

where $J\colon \Theta \to (-\infty, \infty)$ is a measurable function

- Find an optimal/suboptimal solution of the randomized optimization problem

$$\min_{\theta \in \Theta} J(\theta) \quad \text{subject to} \quad \hat{\mathrm{E}}_N(\theta) = 0 \qquad \text{(ROP)}$$

IEIIT-CNR

Binary Violation Function *g*

- We consider the unconstrained case $\rho = 0$
- The binary violation function g is a Boolean expression with k polynomials having largest degree equal to α

IEIIT-CNR

Bounding the Probability of Violation

- Theorem[1]: For any $\varepsilon \in (0,0.14)$ and $\delta \in (0,1)$, if

$$N \geq 4.1/\varepsilon \; [\, \log(21.64/\delta) + 4.39\, n \log_2(8e\alpha k/\varepsilon) \,]$$

then, with probability no smaller than $1-\delta$

- either (ROP) is unfeasible and then also (SIOP) is unfeasible

- or, (ROP) is feasible, then any optimal/suboptimal solution θ_0 satisfies

$$E_g(\theta_0) = \mathrm{Prob}\{\Delta \in \mathcal{B}: g(\theta_0, \Delta) = 1\} \leq \varepsilon$$

[1] T. Alamo, R. Tempo and E.F. Camacho (2009)

Comments - 1

- The function *g* consists of polynomials; constraints and objective function are nonconvex
- Sample size results are valid also for suboptimal (local) solutions (no need to compute a global solution)
- We can use various algorithms to obtain a local solution
- The approach consists of uncertainty randomization and deterministic optimization in controller space
- We avoid randomization of controller parameters

IEIIT-CNR

Comments - 2

- The restriction on $\varepsilon \in (0,0.14)$ allows us to obtain better bounds
- If ε is sufficiently "small" ($\varepsilon \leq (2e\alpha k)^{-1}$), the explicit bound

$$N \geq 4.1/\varepsilon \; [\, \log(21.64/\delta) + 8.78\, n \log_2(2/\varepsilon) \,]$$

  depends only on n, δ and ε
- N does not depend on the VC dimension or its upper bound d
- N is linear in n and logarithmic in $1/\delta$
- N depends on ε as $1/\varepsilon \log(1/\varepsilon)$

IEIIT-CNR

Convex Semi-Infinite Optimization

- **Convex Semi-Infinite Optimization Problem**: Find the optimal solution of the problem

$$\min_{\theta \in \Theta} \; c^T \theta \quad \text{subject to } g(\theta, \Delta) = 0 \text{ for all } \Delta \in \mathcal{B} \qquad \text{(SIOP)}$$

where $g(\theta, \Delta) = 0$ is *convex* in θ for all $\Delta \in \mathcal{B}$, the solution is unique and the level ρ is zero

Does Convexity Help?

- Convex SIOP

$$N \geq 2/\varepsilon \ \log(1/(2\delta)) + 2n + 2n/\varepsilon \ \log 4$$

Does Convexity Help?

- Convex SIOP

$$N \geq 2/\varepsilon \ \log(1/(2\delta)) + 2n + 2n/\varepsilon \ \log 4$$

- Non-convex SIOP

$$N \geq 4.1/\varepsilon \ [\ \log(21.64/\delta) + 8.78\ n \log_2(2/\varepsilon)\ ]$$

IEIIT-CNR

Does Convexity Help?

- Convex SIOP

$$N \geq 2/\varepsilon \ \log(1/(2\delta)) + 2n + 2n/\varepsilon \ \log 4$$

- Non-convex SIOP

$$N \geq 4.1/\varepsilon \ [\ \log(21.64/\delta) + 8.78\ n \log_2(2/\varepsilon)\ ]$$

- The bound for convex SIOP has been recently improved[1]

$$N \geq 1/\varepsilon\ [\ e/(e-1)\ ]\ [\ \log(1/\delta) + n - 1\ ]$$

[1] T. Alamo, R. Tempo and A. Luque (2009)

Conclusions

- Statistical Learning Theory provides a powerful tool for handling nonconvex semi-infinite feasibility and optimization problems
- The idea is to develop a randomized strategy which is based on uncertainty randomization and provides a probably approximate solution to these problems
- The methods are non-sequential and bounds on the sample size are derived
- These bounds improve significantly the existing ones

IEIIT-CNR

Future Work

- For convex problems various sequential randomized algorithms have been developed
- These algorithms are based on

1. generation of an uncertainty random sample
2. update of a probabilistic oracle

- Examples of oracles are stochastic gradient, ellipsoid method and cutting plane
- Convergence properties are shown
- This approach is satisfactory for feasibility problems, but not yet for optimization problems

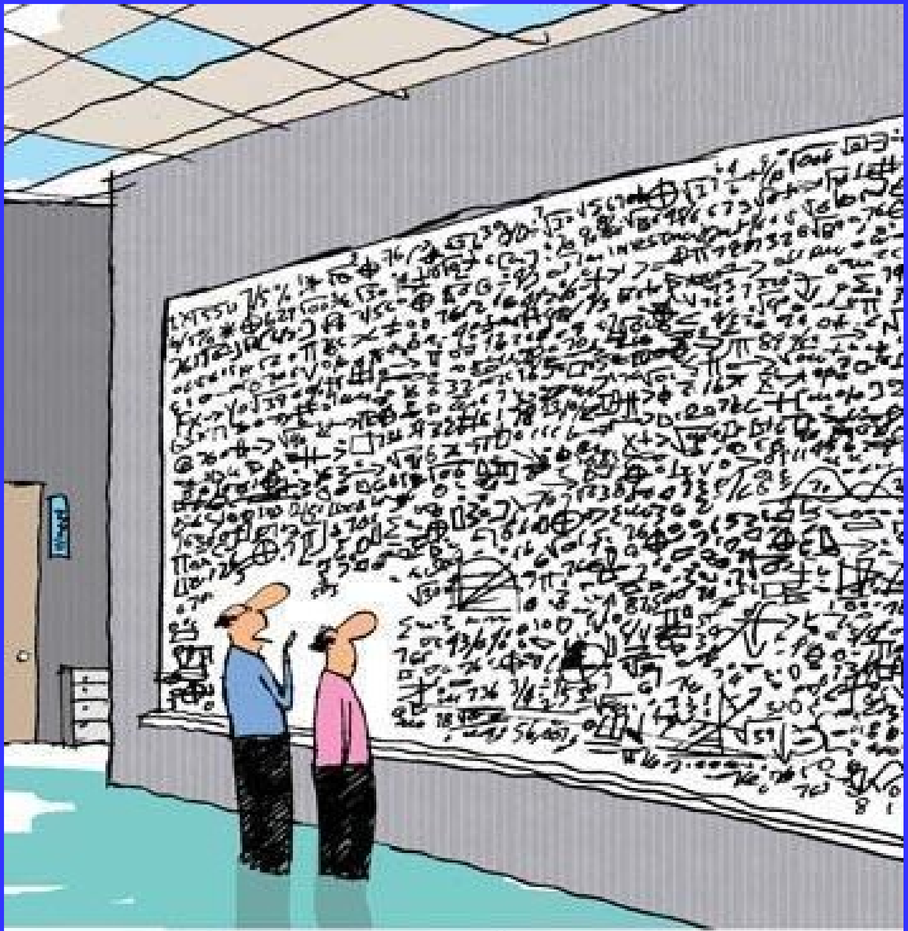
IEIIT-CNR

Acknowledgment

- The results presented here are based on joint work with Teodoro Alamo and Eduardo F. Camacho

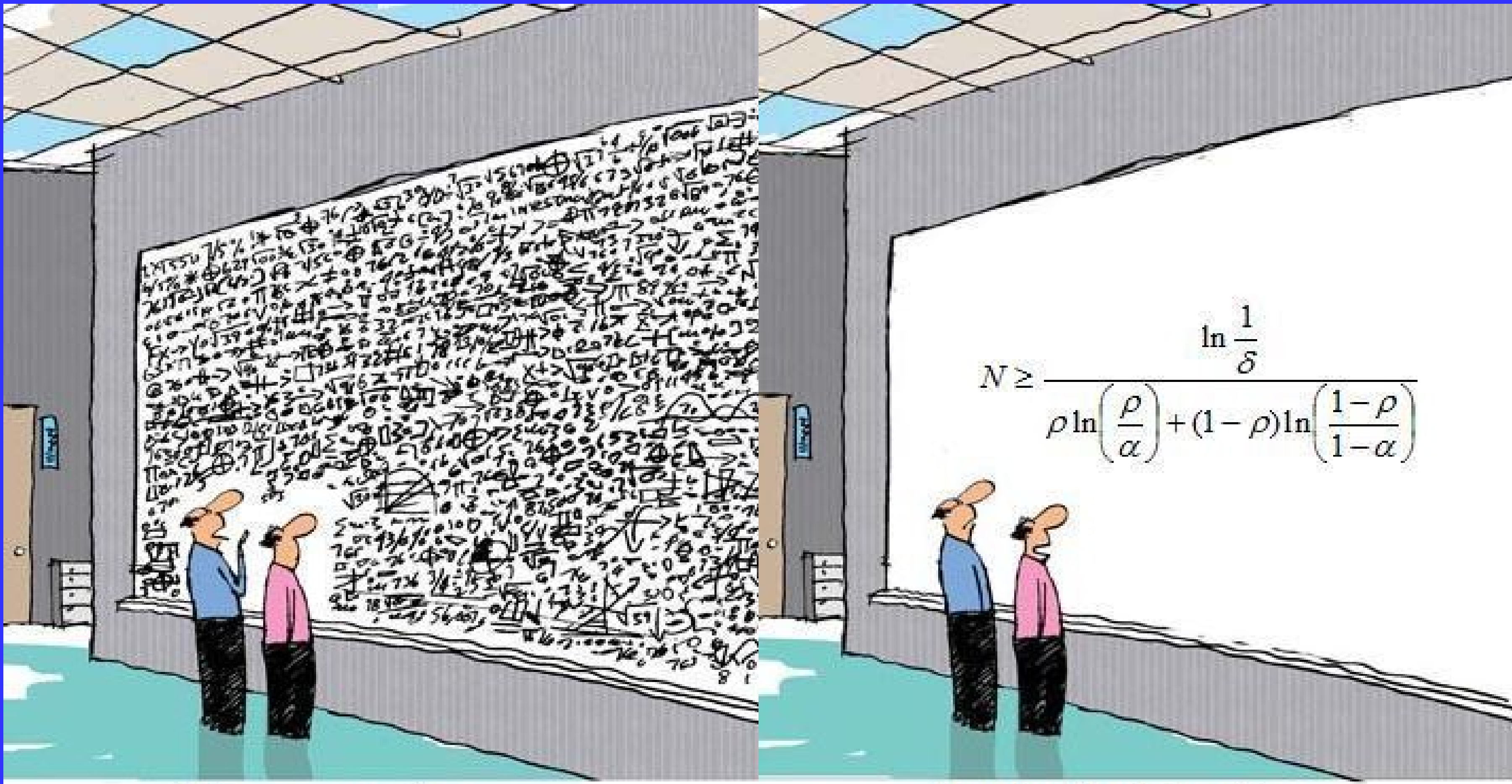
T. Alamo, R. Tempo and E. F. Camacho, "Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems," IEEE Transactions on Automatic Control, November 2009 (to appear)

Conclusion



Other approaches

Conclusion



Other approaches **Randomized algorithms**