NEURIPS 2026 / TMLR JOURNAL-TO-CONFERENCE

The Internal Growth Function:

A More General PAC Framework for Scenario Decision Making

Guillaume O. Berger and **Raphaël M. Jungers**
Department of Applied Mathematics, ICTEAM Institute, UCLouvain

UCLouvain
ICTEAM / FNRS

Sub-exponential internal growth guarantees PAC generalization, even with infinite VC dimension and no bounded compression scheme.

1 Decisions from scenarios

Draw $x_1, \dots, x_m \overset{\text{i.i.d.}}{\sim} \mu$ and choose a decision satisfying all sampled constraints. Will it satisfy a *new* constraint?
Let $A(\mathbf{x}) \subseteq X$ be its satisfied constraints:

$$x_i \in A(\mathbf{x}) \quad \text{and} \quad R_\mu(A(\mathbf{x})) = 1 - \mu(A(\mathbf{x})).$$

PAC goal: for every distribution μ, ensure

$$\Pr_{\mathbf{x}\sim\mu^m}[R_\mu(A(\mathbf{x})) > \epsilon] \leq \delta.$$

Here ϵ is the risk tolerance; $1-\delta$ is the confidence.

2 Count the internal traces

For a finite set $S \subseteq X$, train using **only elements of** S, then record which are satisfied:

$$\sigma_A(S) = \{A(\mathbf{x}) \cap S : \mathbf{x} \in S^*\}.$$

S^*: all finite tuples from S (repetitions allowed).

$$\tau_A(m) = \max_{\substack{S\subseteq X \\ |S|\leq m}} |\sigma_A(S)|$$

Classical growth allows training on all of X; internal growth restricts training to S.

Example: smallest enclosing interval
For $S = \{1,2,3\}$, exactly seven traces are possible.



Filled points are satisfied. The trace $\{1,3\}$ is impossible: an interval containing 1 and 3 also contains 2.

The complexity depends on the algorithm.

3 Slow growth implies PAC

Main theorem. For a consistent scenario decision algorithm, if $\epsilon m \geq 8$, then

$$\Pr[R_\mu(A(\mathbf{x})) > \epsilon] \leq 2e^{-\epsilon m/4}\tau_A(2m).$$

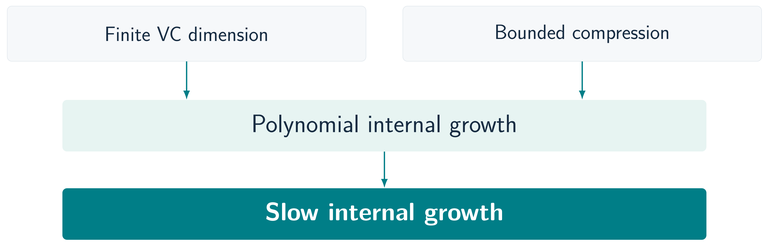
Consequently,

$$\underbrace{\log \tau_A(m) = o(m)}_{\text{slow internal growth}} \quad \Longrightarrow \quad A \text{ is PAC.}$$

How to use it: bound $\tau_A(2m)$, then choose m so the right-hand side is at most δ.

Why it works. A ghost sample and a union bound range over at most $\tau_A(2m)$ internal traces. Sub-exponential growth is dominated by the exponential probability decay.

A broader sufficient condition



Partial converse. If A is *stable and monotone*, then

$$A \text{ is PAC} \iff A \text{ has slow internal growth.}$$

Stable: adding a satisfied constraint changes nothing.
Monotone: adding constraints enlarges the satisfied set.
Open: is slow internal growth necessary for every PAC scenario decision algorithm?

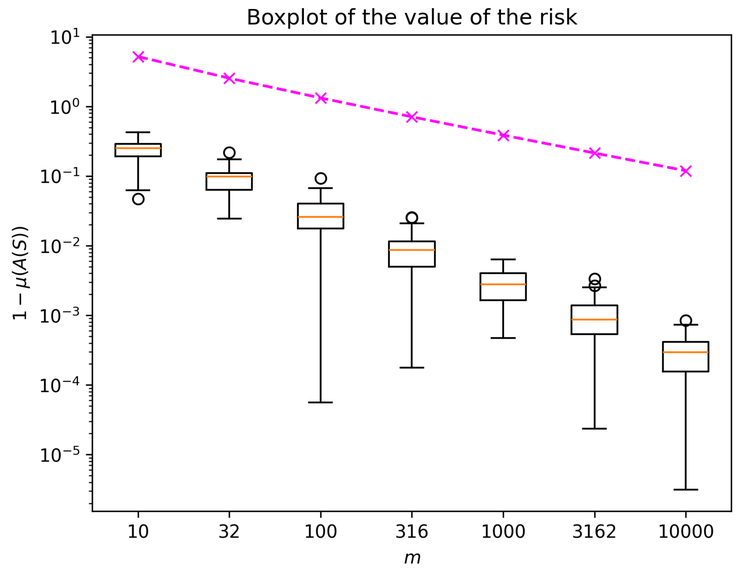
4 PAC beyond the old criteria

Two-stage nonconvex scenario optimization.
Choose $u \in \mathbb{N}^2$ and binary variables $(v_d)_{d\geq 0}$. Samples forbid values of u or require $v_d = 1$. Minimize

$$u_1 + u_2 + \sum_{d\geq 0} r^{d+1} v_d, \qquad 0 < r < \tfrac{1}{2},$$

subject to sampled constraints and $v_d = 1$ for $d \geq u_1 + u_2$.
Proved in the paper: $\tau_A(m) \in 2^{\Theta(\sqrt{m})}$. The algorithm is PAC, has infinite VC dimension, and admits no bounded compression scheme.

Explicit growth bound: $\tau_A(m) \leq 2^{2\sqrt{2m}}$.



Original experiment: 50 runs per sample size; geometric sampling with $p = 0.25$. Magenta: theoretical 95% risk bound (values above 1 are vacuous).

guillaume.berger@uclouvain.be | openreview.net/forum?id=HqPKJSAkrp
Supported by FNRS, ERC L2C (864017), Horizon Europe UniMaaS (101177842), and ARC SIDDARTA.

Read the paper
Definitions, proofs, and the k-VC characterization