# Optimal Top Two Method for Best Arm Identification and Fluid Analysis

NeurIPS 2024

Agniv Bandyopadhyay, TIFR Mumbai

Joint work with: Sandeep Juneja (Ashoka University) and
Shubhada Agrawal (Carnegie Mellon University)

# BEST-ARM IDENTIFICATION PROBLEM

$K$-unknown distributions (or *arms*) are given. **Objective** is to **identify the arm with highest mean**, a.k.a., the best arm, **consuming the minimum no. of samples**, with **probability of error atmost $\delta$.**

**Distributional Assumption:** We assume that instances are from a **single parameter exponential family.** Such families can be parameterised using their mean.

**$\delta$-correctness:** An algorithm is said to be $\delta$-correct, if for every instance and choice of confidence parameter $\delta$, the algorithm stops and identifies the best arm correctly with probability at least $1 - \delta$.

**Problem Statement:** To design a **$\delta$-correct algorithm** which **consumes the minimum no. of samples for every problem instance.**

# LOWER BOUND

**Theorem:** For every instance $\mu = (\mu_1, \mu_2, \ldots, \mu_K)$ having **the first arm as the best arm** and every choice of confidence $\delta$, **the sample complexity of any $\delta$-correct algorithm is lower bounded by**:

$$\mathcal{O}: \quad \min \sum_{a \in [K]} N_a$$

$$\text{s.t.} \quad \forall a \neq 1, \ \mathcal{I}_a = N_1 d(\mu_1, x_{1,a}) + N_a d(\mu_a, x_{1,a}) \geq \log(1/\delta),$$

where $x_{1,a} = \dfrac{N_1\mu_1 + N_a\mu_a}{N_1 + N_a}$. **Solution to the above problem is of the form $T^\star(\mu) \cdot \log(1/\delta)$**, **where $T^\star(\mu)$ is a constant depending only on the instance $\mu$.**

**Asymptotic Optimality:** An algorithm is asymptotically optimal if its sample complexity $\tau_\delta$ satisfies:

$$\limsup_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq T^\star(\mu)$$

for every instance $\mu$.

# EXISTING $\beta$-ASYMP. OPTIMAL ALGORITHMS

1. **empirically best arm** is pulled **with probability** $\beta$

2. **minimum empirical index arm** (index of arm $a$ is $\mathcal{I}_a = N_1 d(\hat{\mu}_1, \hat{x}_{1,a}) + N_a d(\hat{\mu}_a, \hat{x}_{1,a})$) **with probability** $1-\beta$**.** ([Jourdan et al., 22] )

3. **Stop when** the minimum empirical index $\min_{a \neq \hat{i}} \mathcal{I}_a \; = \; \log(1/\delta) +$ smaller order terms.

[Russo '16], [Jourdan et al. '22] showed **$\beta$-Top Two algorithms are asymptotically $\beta$-optimal** (optimal upto giving $\beta$-fraction of samples to the best arm). $\beta = 0.5$ gives sample complexity of atmost twice of the lower bound (see [Russo '16]).

# FIRST-ORDER CONDITIONS

**Theorem:** The **optimal allocations $N^\star$ solving $\mathcal{O}$** is **uniquely characterised by the conditions**:

$$g = \sum_{a \neq 1} \frac{d(\mu_1, x_{1,a})}{d(\mu_a, x_{1,a})} - 1 = 0$$

and index $\mathcal{I}_a$ of every alternative arm $a \neq 1$ are equal to $\log(1/\delta)$. Our algorithm tracks these first order conditions.

1. We call $g(\cdot)$ the **anchor function**

2. The optimal allocation is uniquely identified by the conditions:

   a. The anchor function $g(\cdot)$ must be zero

   b. Index $\mathcal{I}_a$ of all the sub-optimal arms are equal to each other and equal to $\log(1/\delta)$.

# ANCHORED TOP-TWO ALGORITHM (AT2)

At every iteration $N$ do:

- **Forced exploration:** Sample an arm if it has less than $N^\alpha$ samples ($\alpha \in (0,1)$ chosen in the beginning)
- **Choice of leader: If $g \geq 0$, sample the empirically best arm,** otherwise
- **Choice of challenger: If $g < 0$, sample the arm with minimum empirical index $\mathcal{I}_a$**
- **Stopping Condition (GLLR):** Terminate if the minimum index ( $\min_{a \neq \hat{i}_{best}} \mathcal{I}_a$) exceeds $\log(1/\delta)$+smaller order terms.

**Theorem: AT2 algorithm is asymptotically optimal.**

# APPROXIMATING ALGO. THROUGH ITS FLUID ANALYSIS

We study the algorithm under an idealised setting where:

1. Mean of all the arms, i.e., $\mu_1, \mu_2, \ldots, \mu_K$ are known
2. Samples are treated as continuous object
3. **Once we reach $g = 0$, we stay there**
4. **Once index of two arms become equal, they stay equal and increase with the total sample allocation .**

# FLUID EQUATIONS

Let $N_a(N)$ be the allocation made to arm $a$ from the total allocation $N$, $B$ be the set of minimum index arms, and $I_B(N)$ be the minimum index.

In the fluid framework, **the allocations increase via the following system of ODEs until the minimum index hits the higher index**:

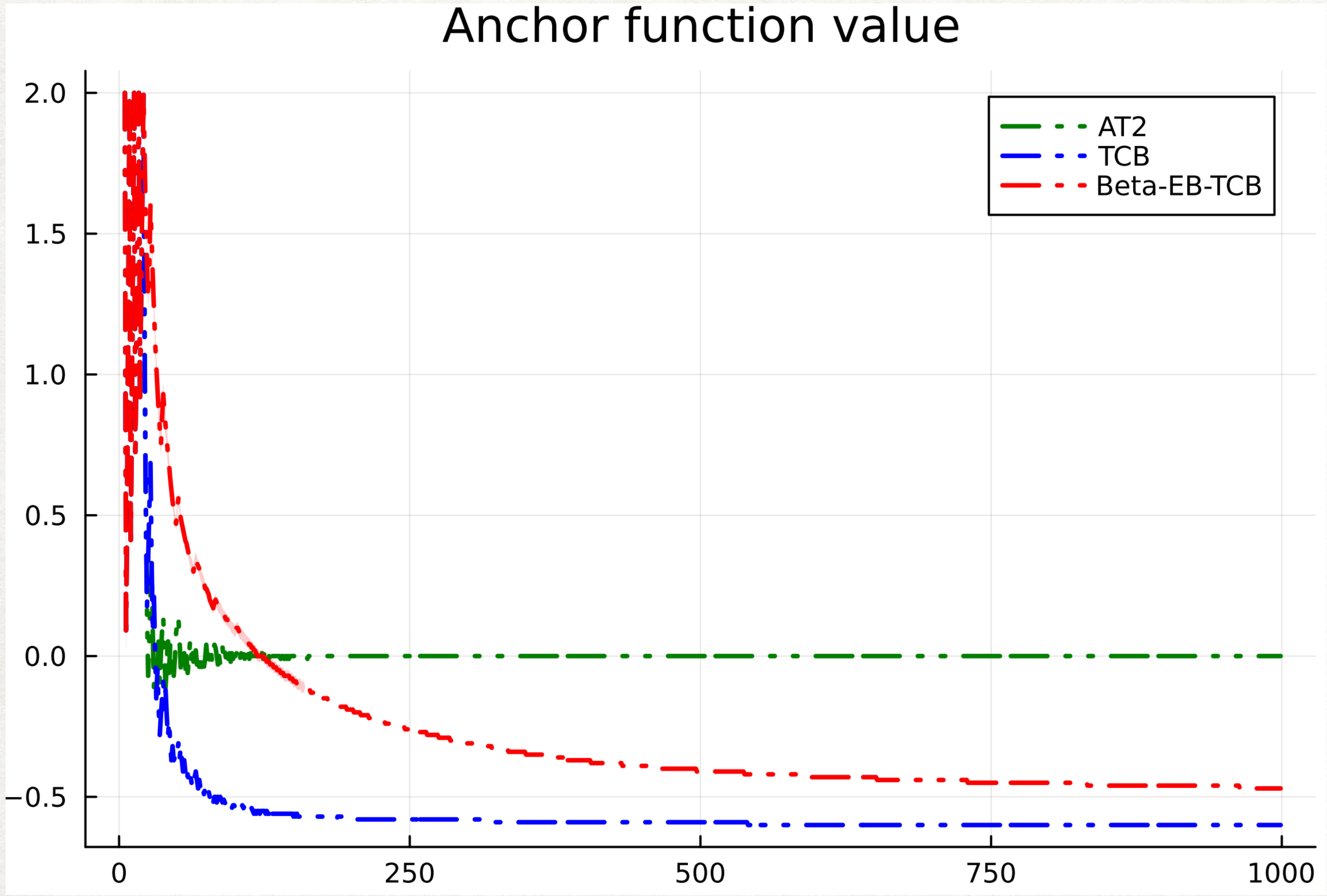
$$N_1' = \frac{N_1 h(B)}{(N_1 + \sum_{a\in B} N_a)h(B) + d_B^{-1}h(N)}, \qquad \forall b \in B, \quad N_b' = \frac{N_b h(B) + d_{b,b}^{-1}h(N)}{(N_1 + \sum_{a\in B} N_a)h(B) + d_B^{-1}h(N)},$$

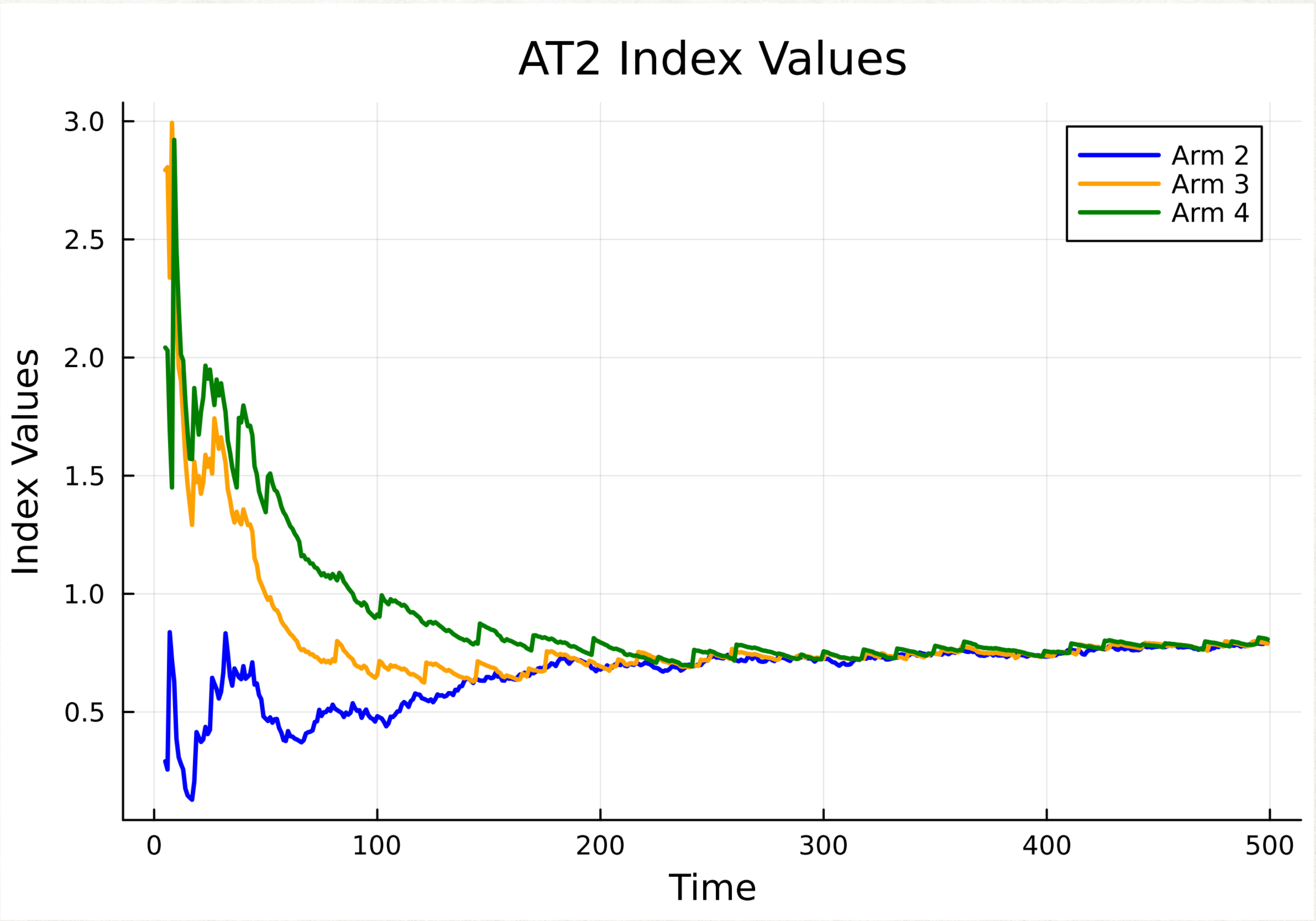
$$\text{and} \quad I_B' = \frac{I_B(N)h(B) + h(N)}{(N_1 + \sum_{a\in B} N_a)h(B) + d_B^{-1}h(N)},$$

$h(B), h(N), d_B$ are functions of the instance and the allocations $(N_a)$.

1. **Overall path is attained by concatenating the above system of ODEs.**
2. **After a finite amount of time, all the indexes becomes equal and $g$ becomes zero.**
3. **The AT2 algorithm closely mimics the fluid dynamics** after a random time of finite expectation and converges to the optimal allocation.

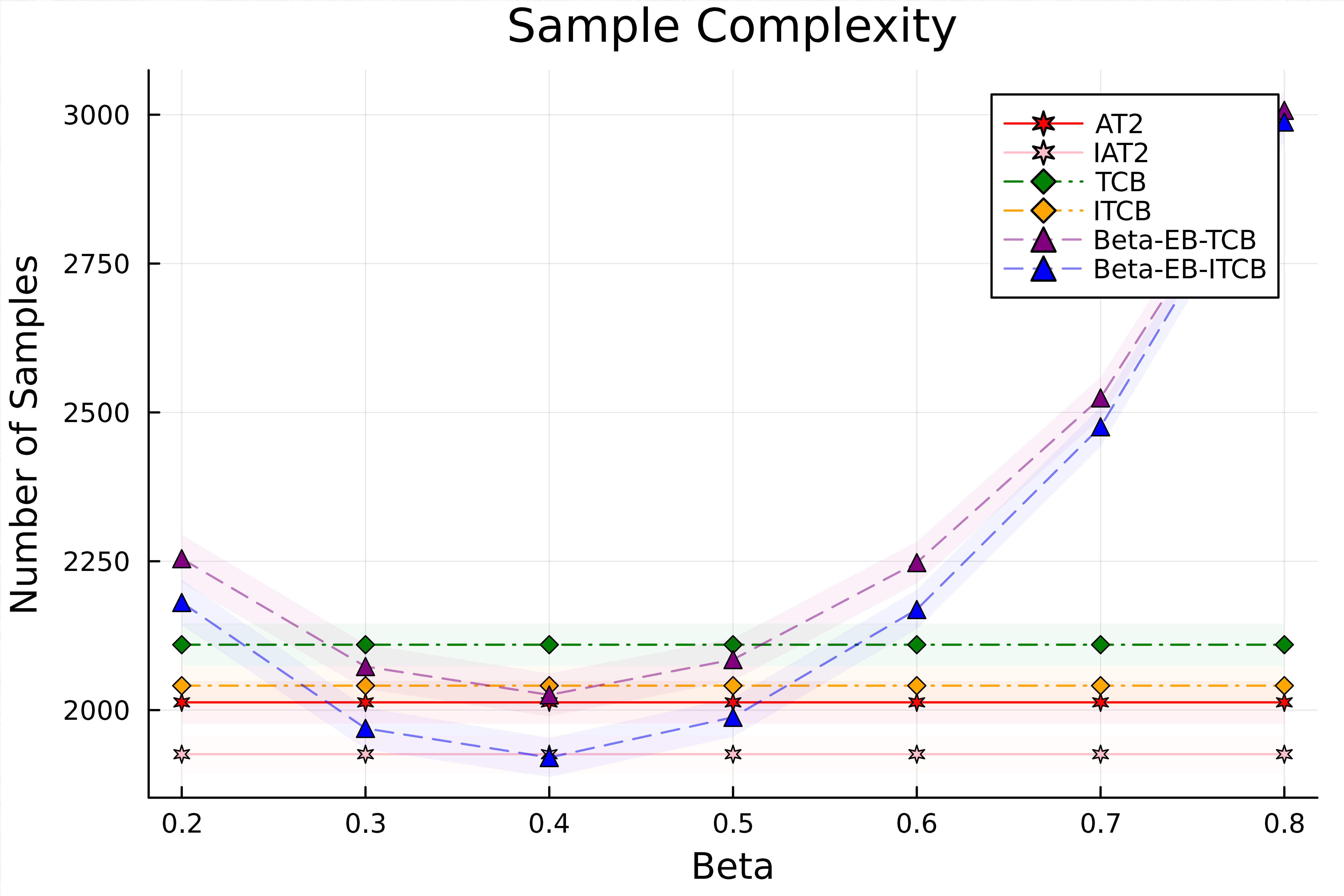
# AT2 ALGORITHM MIMICKING THE FLUID BEHAVIOUR




4 armed Gaussian instance with means [10, 8, 7, 6.5] and unit variance

# COMPARISON WITH EXISTING ALGORITHMS




Sample complexity comparison between ATT, TCB(I) (Tajer and Mukherjee), and $\beta$-Top-Two policies with different values of $\beta$.

**AT2 improves upon the existing algorithms.**

4 armed Gaussian instance with means [10, 8, 7, 6.5] and unit variance

Thank You !