Two-Timescale Learning Automata for Solving Stochastic Nonlinear Resource Allocation Problems

Anis Yazidi*, Hugo Lewi Hammer, and Tore Møller Jonassen

Department of Computer Science, Oslo and Akershus University College of Applied Sciences, Norway

Abstract. This papers deals with the the Stochastic Non-linear Fractional Equality Knapsack (NFEK) problem which is a fundamental resource allocation problem based on incomplete and noisy information [2, 3]. The NFEK problem arises in many applications such as in web polling under polling constraints, and in constrained estimation. The primary contribution of this paper is a *continuous* Learning Automata (LA)-based, *optimal*, efficient and yet simple solution to the NFEK problem. Our solution reckoned as the Two-Timescale based Learning Automata (T-TLA) solves the NFEK problem by performing updates on two different timescales. To the best of our knowledge, this is the first tentative in the literature to design an LA that operates with two-time scale updates. Furthermore, the T-TLA solution is distinct from the first-reported optimal solution to the problem due to Granmo and Oommen [2,3] which resorts to utilizing multiple two-action discretized LA, organized in a hierarchical manner, so as to be able to tackle the case of multi-materials. Hence, the T-TLA scheme mitigates the complexity of the state-of-the-art solution that involves partitioning the material set into two subsets of equal size at each level. We report some representative experimental results that illustrate the convergence of our scheme and its superiority to the state-of-the-art [2, 3].

Keywords: *Continuous Learning Automata, Two-timescale Learning, Stochastic Nonlinear Fractional Equality Knapsack, Resource Allocation* .

1 Introduction

This paper deals with the Stochastic Non-linear Fractional Equality Knapsack (NFEK) Problem which is the central underlying problem pertinent to allocating resources based on incomplete and noisy information. Such situations are not merely hypothetical – rather, they constitute the vast majority of allocation

* The author gratefully acknowledges the assistance of his PhD supervisor, Dr. John Oommen, a *Chancellor's Professor* from Ottawa, Canada. John helped a lot with the style and language of this paper. He also very graciously provided me with some of the text when it concerned the introductory sections and the background material. These portions have been included here, in some cases *verbatim*, with his kind permission. Thank you John!!

problems in the real-world. Resource allocation problems which involve such incomplete and noisy information are particularly intriguing. They cannot be solved by traditional optimization techniques, rendering them ineffective.

The NFEK problem, that was first solved optimally in [4], is not merely of academic interest. Indeed, it is found in many settings, for example, in the web polling problem and constrained estimation [2]. More specifically, in the case of web polling, the decision maker attempts to choose web pages in a manner that maximizes the number of changes detected, and the optimal allocation of the resources again involves "trial and error". Web pages may change with varying frequencies (that are unknown to the decision maker), and changes appear more or less randomly. Furthermore, as argued elsewhere [4], the probability that an individual web page poll uncovers a change on its own *decreases monotonically* with the polling frequency used for that web page. The NFEK also has applications in determining the optimal sample size required for estimation purposes. This paper briefly addresses these problems as application domain problems – they are discussed, in more detail, elsewhere [4].

The NFEK problem has two main peculiarities:

– First, the unit volume values of each material are treated as *stochastic* variables whose distributions are *unknown*.
– The expected value of a material may decrease after each addition to the knapsack.

The first optimal solution to the NFEK problem is due to Granmo and Oommen [2, 3], and resorts to the invoking a hierarchy of two-action discretized Learning Automata (LA). The solution was generalized using a hierarchical scheme in order to tackle the case of multi-materials. Although the solution proposed in [4] is elegant, its implementation is, unfortunately, complex because it involves updates at different levels of a balanced binary tree.

1.1 Formal Problem Formulation

The Stochastic NFEK Problem: The generalization of the nonlinear equality knapsack problem is due to Granmo and Oommen [2, 3]. First of all, we let the material value per unit volume for any x_i be a *probability* function $p_i(x_i)$. Furthermore, we consider the distribution of $p_i(x_i)$ to be *unknown*. That is, each time an amount x_i of material i is placed in the knapsack, we are only allowed to observe an instantiation of $p_i(x_i)$ at x_i, and not $p_i(x_i)$ itself. Given this stochastic environment, we seek a solution to the Stochastic NFEK problem that is on-line and incremental, and that learns the mix of materials of maximal *expected* value, through a series of informed guesses. Thus, to clarify issues, we are provided with a knapsack of fixed volume c, which is to be filled with a mix of n different materials. However, unlike the NFEK, in the Stochastic NFEK Problem the unit volume value of a material i, $1 \leq i \leq n$, is a random quantity — it takes the value 1 with probability $p_i(x_i)$ and the value 0 with probability $1 - p_i(x_i)$, respectively. As an additional complication, $p_i(x_i)$ is nonlinear in the sense that it decreases monotonically with x_i, i.e., $x_{i_1} \leq x_{i_2} \Leftrightarrow p_i(x_{i_1}) \geq p_i(x_{i_2})$.

Since the unit volume values are random, we operate with expected unit volume values rather than the actual unit volume values themselves. With this understanding, and the above perspective in mind, the expected value of the amount x_i of material i, $1 \le i \le n$, becomes $f_i(x_i) = \int_0^{x_i} p_i(u)du$. Accordingly, the expected value per unit volume[1] of material i becomes $f_i'(x_i) = p_i(x_i)$. In this stochastic and non-linear version of the FK problem, the goal is to fill the knapsack so that the expected value $f(\boldsymbol{x}) = \sum_1^n f_i(x_i)$ of the material mix contained in the knapsack is maximized. Thus, we aim to:

$$\text{maximize} f(\boldsymbol{x}) = \sum_1^n f_i(x_i),$$

$$\text{where} f_i(x_i) = \int_0^{x_i} p_i(u)du, \text{ and } p_i(x_i) = f_i'(x_i),$$

$$\text{subject to} \sum_1^n x_i = c \text{ and } \forall i \in \{1, \ldots, n\}, x_i \ge 0.$$

A fascinating property of the above problem is that the amount of information available to the decision maker is limited — the decision maker is only allowed to observe the current unit value of each material (either 0 or 1). That is, each time a material mix is placed in the knapsack, the unit value of each material is provided to the decision maker. The actual outcome probabilities $p_i(x_i), 1 \le i \le n$, however, remain *unknown*. As a result of the latter, the expected value of the material mix must be maximized by means of trial-and-failure, i.e., by experimenting with different material mixes and by observing the resulting random unit value outcomes.

1.2 The Hierarchy of Twofold Resource Allocation Automaton (H-TRAA) Solution

The stochastic NFEK problem was first addressed in the literature in [4]. The first reported generic treatment of the stochastic NFEK problem itself can be found in [4]. The state-of-the-art scheme for hierarchically solving n-material problems [2, 3] involves a primitive module, namely the Twofold Resource Allocation Automaton (TRAA) for the *two-material* problem, that has been proven to be asymptotically optimal. The authors of [2, 3] demonstrated a mechanism by which the primitive TRAAs can be arranged in a hierarchy so as to solve *multi-material* Stochastic NFEK Problems.

The hierarchy of TRAAs, referred to as H-TRAA, assumes that $n = 2^\gamma, \gamma \in \mathbb{N}^+$. If the number of materials is less than this, one trivially assumes the existence of additional materials whose values are "zero", and which thus are not able to contribute to the final optimal solution. The hierarchy is organized as a

[1] We hereafter use $f_i'(x_i)$ to denote the derivative of the expected value function $f_i(x_i)$ with respect to x_i.

balanced binary tree with depth $D = \log_2(n)$. Each node in the hierarchy can be related to three entities: (1) a set of materials, (2) a partitioning of the material set into two subsets of equal size, and (3) a dedicated TRAA that allocates a given amount of resources among the two subsets. At depth D, then, each individual material can be separately assigned a fraction of the overall capacity by way of recursion, using a subtle mechanism described, in detail, in [3]. The principal theorem that guarantees the convergence of the H-TRAA [2,3] has cleverly shown that if all the individual TRAAs converge to their *local* optimum, then the global optimum is attained.

1.3 Contributions of this Paper

The contributions of this paper are the following:

1. We report an optimal solution to the stochastic NFEK problem based on the bridging the theory of LA with the theory two-timescale separation [1,5]. To the best of our knowledge, this paper provides the first attempt in the literature to bridge the latter two fields: LA on one hand and two-time scale scheme on the other hand.
2. In contrast to the H-TRAA solution [2,3], our T-TLA solution does not involve a hierarchy, and it is thus easier to implement. This is because, in fact, TRAAs must be arranged in a hierarchy in order for them to be able to solve a *multi-material* Stochastic NFEK Problems. Further, through empirical experiments, we confirm that the T-TLA provides desirable convergence properties that makes it competitive to the H-TRAA.

As a result of the above contributions, we believe that the T-TLA is a viable realistic strategy for solving demanding real-world knapsack-like problems such as the optimal allocation of sampling resources [2], and other problems related to the world wide web [4].

1.4 Paper Organization

The paper is organized as follows. In Section 2 we present the T-TLA for the *n-material* problem. We proceed in Section 3 to empirically verify that the T-TLA solution provides competitive convergence results to the H-TRAA while being, at the same time, simpler to implement. Finally, we offer suggestions for further work and conclude the paper in Section 4.

2 A T-TLA Solution to Resource Allocation

The *Stochastic Environment* for the n materials case can be characterized by:

1. The capacity c of the knapsack, which is normalized in this case;
2. $n - material$ unit volume value probability functions $[p_1(x_1), \ldots, p_n(x_n)]$.

In brief, if the amount x_i of material i is suggested to the Stochastic Environment, the Environment replies with a unit volume value $\delta_i = 1$ with probability $p_i(x_i)$ and a unit volume value $\delta_i = 0$ with probability $1 - p_i(x_i)$. To render the problem both interesting and non-trivial, we assume that $p_i(x_i)$ is unknown to the LA.

We shall first characterize the optimal solution to a Stochastic NFEK Problem provided in [2,3].

Lemma 1. *The material mix $\boldsymbol{x^*} = [x_1^*, \ldots, x_n^*]$ is a solution to a given Stochastic NFEK Problem if (1) the derivatives of the expected material amount values are all equal at $\boldsymbol{x^*}$, (2) the mix fills the knapsack, and (3) every material amount is positive, i.e.:*

$$f_1'(x_1^*) = \cdots = f_n'(x_n^*)$$
$$\sum_1^n x_i^* = c \text{ and } \forall i \in \{1, \ldots, n\}, x_i^* \geq 0.$$

The above lemma is based on the well-known principle of Lagrange Multipliers, and its proof is therefore omitted here for the sake of brevity.

Now, we shall present our solution to the stochastic NFEK [3].

The idea behind our T-TLA is to resort to a two-timescale based approach, where the polling probabilities x_i are updated on the "slower timescale" while $p_i(x_i)$ are estimated on a "faster timescale". In practice, the updating parameter (in this case λ) used for updating the probabilities x_i should be much smaller than the corresponding updating parameter θ for the task of estimation of the p_i. Thus, we can say that the fast-evolving dynamics of p_i sees x_i as almost constant, while the slowly evolving dynamics of x_i given sees p_i as almost equilibrated [1,5].

Another possible manner to to implement a two-time scale approach is to execute one update on the slower timescale loop for every few iterations on the faster timescale loop, i.e., the slower timescale loop is run less frequently.

We denote the decision variable for selecting an action at time instant t, $\alpha(t)$ that is, for $i \in [1..n]$. We say that the event $\{\alpha(t) = i\}$ has occurred if the action i is polled.

Once the action i is polled, the estimate $\hat{p_i}(t+1)$ of the reward probabilities is immediately updated using an exponential moving averaging based estimator:

$$\hat{p_i}(t+1) = \hat{p_i}(t) + \theta(\delta_i(t) - \hat{p_i}(t)) \tag{1}$$

where $\delta_i(t)$ is a random variable that takes a value 1 with $p_i(x_i(t))$ and 0 with $1 - p_i(x_i(t))$.

The reward estimates for the other actions are left unchanged, i.e.,

$$\hat{p_j}(t+1) = \hat{p_j}(t) \text{ for } j \neq i, j \in [1, n]$$

Thus, the evolution of the reward estimates can be described by the following set of stochastic iterative equations for $i \in [1..n]$:

$$\hat{p_i}(t+1) = \hat{p_i}(t) + \theta I_{\{\alpha(t)=i\}}(\delta_i(t) - \hat{p_i}(t)) \tag{2}$$

Now, we are ready to present the update equations for the polling probabilities x_i for $i \in [1..n]$.

The complete algorithm is described as follows:

1. Poll an action at time instant t denoted by $\alpha(t)$ according to the probability vector $[x_1, x_2, \ldots, x_n]$ and observe $\delta_i(t)$.
2. Update the reward probabilities estimates of the n actions according to the following equation, for $i \in [1..n]$:

$$\hat{p_i}(t+1) = \hat{p_i}(t) + \theta I_{\{\alpha(t)=i\}}(\delta_i(t) - \hat{p_i}(t)) \tag{3}$$

3. Update the polling probabilities for the next time instant $t+1$ according to:

$$\begin{aligned} x_1(t+1) &= x_1(t) - \lambda \left(\frac{1}{n} \sum_{i=1}^{n} \hat{p_i}(x_i(t)) - \hat{p_1}(x_1(t)) \right) \\ x_2(t+1) &= x_2(t) - \lambda \left(\frac{1}{n} \sum_{i=1}^{n} \hat{p_i}(x_i(t)) - \hat{p_2}(x_2(t)) \right) \\ &\vdots \\ x_n(t+1) &= x_n(t) - \lambda \left(\frac{1}{n} \sum_{i=1}^{n} \hat{p_i}(x_i(t)) - \hat{p_n}(x_n(t)) \right) \end{aligned} \tag{4}$$

Idea behind the proof The proof of the optimality of the above algorithm is quite involved and so we include only the overall behind the proof in the interest of space and brevity. The complete proof is included in the unabridged version of this paper [6]. According to Lemma 1, the optimal solution equalizes the reward probabilities $p_i(x_i^*)$.

Following the proof of two-timescale separation provided in [1], $\hat{p_i}(x_i(t))$ approximates $p_i(x_i(t))$ whenever λ is much smaller than θ reflecting the fact that the fast-evolving dynamics of p_i sees x_i as "almost constant". Moreover, the system of equations 4 can be proved to converged to the fixed point described by Lemma 1 using the theory of dynamical systems.

3 Experimental Results

We have conducted our experiments for one objective function (referred to as $E_i(x_i)$) being optimized. The function can be seen as representative for the class of concave objective functions that we address. We have also conducted experiments with a number of other objective functions, including those found in [2, 3] that are not reported here due to space limitations and can be found in [6]. However, it turns out that $E_i(x_i)$ is particularly useful in the sense that they permit us to appropriately model a large range of distinct material unit value functions.

More specifically, these objective functions have been given below for a material with index i as:

$$E_i = \frac{0.9}{i}(1 - \exp(-ix_i)) \tag{5}$$

In the above, the constants are based on the boundary conditions due the contributions of x_i at the boundary values. These constants, however, are not crucial in the optimization because the corresponding unit value functions are obtained as their respective derivatives. These are two probability functions given below for a material with index i as $E'_i(x_i)$, which fall exponentially as per equation (6) below:

$$E'_i(x_i) = 0.9 \cdot \exp(-ix_i) \tag{6}$$

To clarify how these functions work, consider the functions $E'_i(x_i)$. Then the relative profitability of material i decreases with x_i, its presence in the mixture, exponentially. Thus, if $x_2 = 0.3$ (i.e., material 2 fills 30% of the knapsack), the marginal profitability of increasing the amount of x_2 is $exp(-2 \cdot (0.3))$ = $exp(-0.6)$. Observe that with the notation, the profitability of materials that have a smaller index decreases *slower* than the profitability of materials that have a higher index.

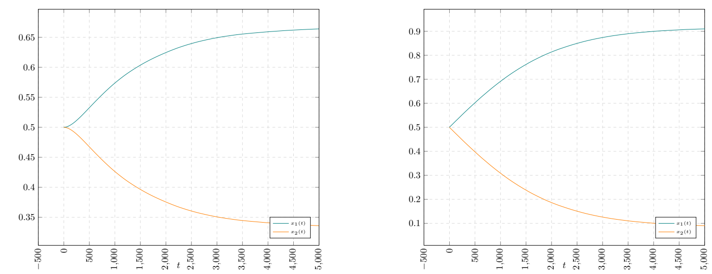
Given the above considerations, our aim is to find $\boldsymbol{x}^*$, the amounts of the materials that have to be included in the knapsack so as to maximize its value. Note that in general application domains, we may not able to observe $f'_i(x_i)$ directly — examining a potential solution may be the only way to reveal the success of the chosen allocation.

We will present some experimental results that compare our T-TLA to H-TRAA solution for binary and quaternary stochastic knapsacks. We performed ensembles of 1000 simulations each consisting of 5000 time steps.

Stochastic knapsack with 2 resources Figure 1a and Figure 1b depict the evolution of the polling probability for the case of two-resources $n = 2$ for our T-TLA and for the legacy H-TRAA solution respectively.

For the T-TLA solution we chose $\lambda = 0.001$ and $\theta = 0.01$. We chose the resolution of H-TRAA to be $N = 1000$, which corresponds here to $\frac{1}{\lambda}$ so that to allow fair comparison via an equal update parameters of both schemes. We observe from Figure 1a and Figure 1b that both approaches are able to converge to the optimal value $x^* = (2/3, 1/3)$ which is seen too from Lemma 1. This takes place after approximately more than 4000 time instants. Furthermore, Figure 2 reports the estimate of the reward probability which evolves at a faster timescale than the polling probabilities for our T-TLA solution. We observe from Figure 2 that the T-TLA solution successfully equalizes $\hat{p_1}(x_1(t))$ and $\hat{p_2}(x_2(t))$ after approximately more than 4000 time instants.

Stochastic Knapsack with 4 resources Similarly, Figure 3a and Figure 3b depict the evolution of the polling probability for the case of two-resources $n = 4$



(a) (b)

Fig. 1: Evolution of the polling probabilities for $n = 4$ for (a) the T-TLA solution and (b) the H-TRAA solution

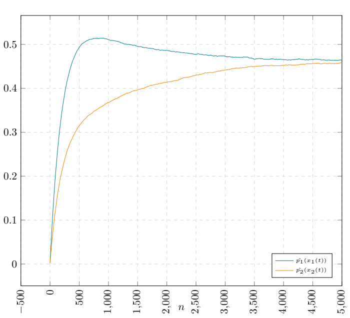
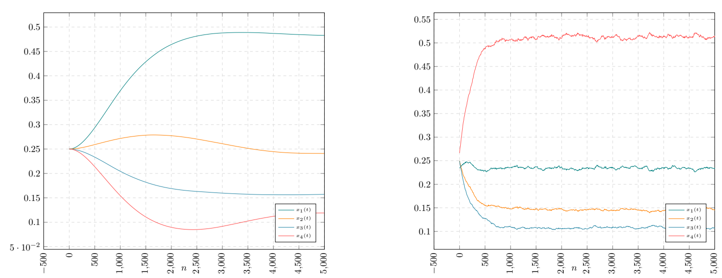


Fig. 2: Evolution of the reward probability estimates for $n = 2$.

for our T-TLA and for the legacy H-TRAA solution respectively. The T-TLA scheme was characterized by parameters $\lambda = 0.001$ and $\theta = 0.01$. We chose the resolution of H-TRAA to be $N = 31$. Please note that $\lambda =\approx \frac{1}{N^2}$ which reflects an equal "update steps" for both schemes when the number of levels in H-TRAA is 2.

We see from Figure 3a and Figure 3b that the polling probability vector converges to the optimal vector $x^* = (0.48, 0.24, 0.16, 0.12)$ which is also confirmed by Lemma 1. However, we observe too that the H-TRAA outperforms the T-TLA solution in terms of convergence speed. Nevertheless, as seen in Figure 3b, the H-TRAA is unable to get of rid of some fluctuations despite that we are averaging over an ensemble of 1000 experiments. These fluctuations of the H-TRAA merely reflect a larger variance than the T-TLA. Figure 4 reports the estimate of the reward probability for our T-TLA solution. As expected, the estimates get equalized over time and converge to the same optimal value.



(a) (b)

Fig. 3: Evolution of the polling probabilities for $n = 4$ for (a) the T-TLA solution and (b) the H-TRAA solution.

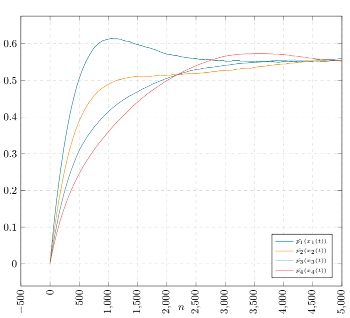


Fig. 4: Evolution of the reward probability estimates for $n = 4$

4 Conclusion

In this paper, we have presented an optimal and efficient solution to Stochastic Non-linear Fractional Equality Knapsack (NFEK) Problem, which is a fundamental resource allocation problem based on incomplete and noisy information [2, 3]. Unlike the existing solutions [2, 3], our primary contribution is a *two-timescale* Learning Automata (LA)-based, *optimal*, efficient and yet simple solution to the NFEK problem. Our solution is distinct from the one reported in solutions [2,3] that uses multiple two-action discretized LA, organized in a hierarchical manner, so as to be able to tackle the case of multi-materials. The T-TLA does not need a hierarchical partitioning, and does not require us to maintain dedicated two-action discretized LA that allocate a given amount of resources among the two subsets. Preliminary experimental results confirm the optimality of the solution. We hope that the current work will pave the way towards more development in bridging LA theory with two-time scale schemes.

References

1. Albert Benveniste, Michel Métivier, and Pierre Priouret. *Adaptive algorithms and stochastic approximations*, volume 22. Springer Science & Business Media, 2012.
2. O.-C. Granmo and B. J. Oommen. Optimal sampling for estimation with constrained resources using a learning automaton-based solution for the nonlinear fractional knapsack problem. *Applied Intelligence*, 33(1):3–20, 2010.
3. O.-C. Granmo and B. J. Oommen. Solving stochastic nonlinear resource allocation problems using a hierarchy of twofold resource allocation automata. *IEEE Transactions on Computers*, 59(4):545–560, 2010.
4. O.-C. Granmo, B. J. Oommen, S. A. Myrer, and M. G. Olsen. Learning automata-based solutions to the nonlinear fractional knapsack problem with applications to optimal resource allocation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(1):166–175, 2007.
5. PS Sastry, M Magesh, and KP Unnikrishnan. Two timescale analysis of the alopex algorithm for optimization. *Neural Computation*, 14(11):2729–2750, 2002.
6. A. Yazidi, L. H. Hammer, and T. M. Jonassen. On two-timescale approach for solving resource allocation problems. *Unabridged journal version of this paper, 2016. To be Submitted for publication.*, 2016.