Adaptive, distributed control of constrained multi-agent systems

Stefan Bieniawski
253 Durand, Dept. of Aero/Astro
Stanford, CA 94305, stefanb@stanford.edu

David H. Wolpert
NASA Ames Research Center, Moffett
Field, CA 94035, dhw@ptolemy.arc.nasa.gov

Abstract

Product Distribution (PD) theory was recently developed as a framework for analyzing and optimizing distributed systems. In this paper we demonstrate its use for adaptive distributed control of Multi-Agent Systems (MAS’s), i.e., for distributed stochastic optimization using MAS’s. One common way to perform the optimization is to have each agent run a Reinforcement Learning (RL) algorithm. PD theory provides an alternative based upon using a variant of Newton’s method operating on the agent’s probability distributions. We compare this alternative to RL-based search in three sets of computer experiments. The PD-theory-based approach outperforms the RL-based scheme in all three domains.

1. Introduction

Product Distribution (PD) theory is an approach for analyzing, controlling, and optimizing distributed systems [4]. One application is distributed stochastic optimization using a MAS (which for our purposes is the same as adaptive, distributed control of a MAS). Often in stochastic optimization one uses probability distributions to help search for a point \( x \in X \) optimizing a function \( G(x) \). In contrast, in the PD approach one searches for a probability distribution \( P(x) \) that optimizes an associated Lagrangian, \( \mathcal{L}(P) \). Since \( P \) is a vector in a Euclidean space, the search can be done via techniques like gradient descent or Newton’s method — even if \( X \) is a categorical, finite space.

The game theoretic motivation for PD theory along with the specific Lagrangian can be found in [4, 3, 5]. Also found in [5] is the derivation of the probability updates based upon Newton’s method. The extension to constrained optimization, including the Lagrange multiplier update is described in [1]. The focus of this work is to evaluate PD theory in numerical experiments through a comparison with results using a Reinforcement Learning (RL) based algorithm. This type of algorithm, referred to here as Brouwer updating, has been previously studied for Multi-Agent Systems [6, 7].

2. Experiments

2.1. Queens Problem

The N-Queens problem, although not hard to solve, illustrates PD-theory’s application to distributed constraint satisfaction problems. The goal is to locate \( N \) queens on a N-by-N chessboard such that there are no conflicts between any of the queens, i.e., no shared rows, columns or diagonals. We compared the iterations to convergence for 50 random trials with \( N \) equal to 8. Figure 1 shows the cumulative probability distributions of iterations to convergence. Newton updating performs better than Brouwer updating.

2.2. Bar Problem

A modified version of Arthur’s El Farol Bar Problem has been used before to investigate the RL-based approach [6, 7]. In this scenario there are \( N \) agents, each selecting one of seven nights to attend a bar. The detailed problem formulation is the same as [7] except that the optimizations were performed at fixed temperature. Figure 2 compares the approaches over a wide temperature range. The performance after 1000 iterations is averaged over 20 cases. Error bars in the averages are all less than 0.25. Newton outperforms Brouwer at all temperatures.

2.3. Bin Packing Problem

We also compared the methods on a discrete constrained problem, the bin packing problem. This problem consists of assigning \( N \) items (the agents) of differing sizes into the smallest number of bins each with capacity \( c \). For the current study instances were obtained from the OR-Library [2]. Since in general the minimum number of bins is not known, the move space of the agents was set to the number of items.
3. Conclusion

Product Distribution (PD) theory provides a framework for analyzing and optimizing distributed systems. In this paper we compared PD-based to RL-based search in three sets of computer experiments. The PD-theory-based approach outperformed the RL-based scheme in all three domains.

References


The objective function used is

\[ G = \begin{cases} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{(c^2 - x_i)^2}{2} \right) & \text{if } x_i \leq c \\ \frac{1}{N} \sum_{i=1}^{N} \left( x_i - \frac{c^2}{2} \right)^2 & \text{if } x_i > c \end{cases} \]  

where \( x_i \) is the total size of the items in bin \( i \).

Two implementations of the Newton approach were considered; one with additional constraints enforcing the capacity for each bin and one without. For each problem variant 20 cases were used in determining the averages and error bars. Figure 3 compares all three schemes. The first plot shows the average number of bins over the optimum, the second the number of bins over capacity, and the third plot the average objective function. While Brouwer results in fewer bins it also results in more overfilled bins. In addition, the usage of the overfilled bins is much poorer than the Newton based schemes, as indicated by the average objective (lower is better). Both Newton schemes obtain similar average objectives, with the constrained version enforcing the capacity constraints by using extra bins.