Environmental Modelling & Software 69 (2015) 353–369

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Large-scale parallelization of the Borg multiobjective evolutionary algorithm to enhance the management of complex environmental systems ☆

David Hadka [a], Patrick Reed [b, *]

[a] Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA
[b] School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 5 March 2014
Received in revised form
25 September 2014
Accepted 6 October 2014
Available online 6 November 2014

Keywords:
Evolutionary algorithm
Borg MOEA
Multiobjective optimization
Large-scale parallelization

ABSTRACT

The Borg MOEA is a self-adaptive multiobjective evolutionary algorithm capable of solving complex, many-objective environmental systems problems efficiently and reliably. Water and environmental resources problems pose significant computational challenges due to their potential for large Pareto optimal sets, the presence of disjoint Pareto-optimal regions that arise from discrete choices, multimodal suboptimal regions, and expensive objective function calculations. This work develops two large-scale parallel implementations of the Borg MOEA, the master–slave and multi-master Borg MOEA, and applies them to a highly challenging risk-based water supply portfolio planning problem. The performance and scalability of both implementations are compared on up to 16384 processors. The multi-master Borg MOEA is shown to scale efficiently on tens of thousands of cores while dramatically improving the reliability of attaining high-quality solutions. Our results dramatically expand the scale and scope of complex environmental systems that can be addressed using many-objective evolutionary optimization.

1. Introduction

The role of evolutionary computation in water resources systems planning and management as described in Maier et al. (2014) as well as the recent review by Nicklow et al. (2010) is to advance our field's ability to address the mathematical complexities inherent to real decision contexts (i.e., multiple objectives, mixed real and discrete decisions, uncertainty, computational demanding simulations, etc.). Real water resources applications present substantial challenges given the evolving and highly uncertain impacts of climate change, rapid urbanization, and growing resource contention. These challenges present fundamental tradeoffs in water systems across time, space, and economic sectors (National Research Council, 2009, 2012). More formally from the optimization perspective, understanding the optimal balance between these tradeoffs requires evolutionary algorithms to approximate the Pareto set of solutions. These solutions encompass those water resources alternatives where improvement in one objective can only be improved by sacrificing performance in one or more other objectives (Cohon and Marks, 1975). The tradeoffs posed by managing water resources systems under change was directly discussed as a leading research challenge in the National Research Council's (NRC) recent vision for the future of global hydrology (National Research Council, 2012). The NRC highlights the need for translational science innovations that combine simulation, optimization, and high-performance computing to innovate real decision making. Complimentary to this vision, our core goals in this study are to advance: (1) a rigorous diagnostic framework for benchmarking massively parallel multiobjective evolutionary algorithms' (MOEAs) ability to discover the tradeoffs for highly challenging water resources problems and (2) fundamentally expand the capability of evolutionary multiobjective search to support decision making in computationally intensive applications. To address these goals, this study presents two parallel variants of the self-adaptive Borg MOEA (Hadka and Reed, 2013, 2012). Our parallelization study generalizes the Borg MOEAs auto-adaptive search to a massively parallel context to better discover effective search strategies that work cooperatively to tailor exploration for severely challenging water resources applications. Moreover, our parallel search diagnostics clarify the relative merits of the classical master–slave Borg MOEA versus a multi-master hierarchical parallelization

☆ Thematic Issue on Evolutionary Algorithms.

* Corresponding author.
E-mail addresses: dmh309@psu.edu (D. Hadka), patrick.reed@cornell.edu (P. Reed).

http://dx.doi.org/10.1016/j.envsoft.2014.10.014
1364-8152/

architecture (multi-master Borg MOEA) that allows scalable search on emerging high-performance computing platforms. These innovations have value for water resources systems planning as well as other fields currently applying parallel evolutionary search. Alba et al. (2013) provide a detailed review of parallel metaheuristics that highlights there is at present a need for studies addressing auto-adaptivity in parallel search, that provide careful theoretical assessment of the scalability of multiobjective algorithm architectures, provide tools for minimizing serial bottlenecks in the algorithms' architectures, and contribute rigorous statistical diagnostics for multiple parallelization strategies. The present study makes important contributions in each of these areas.

Both parallel variants of the Borg MOEA are applied to a highly challenging water resources management application: a risk-based water supply portfolio planning problem focused on the Lower Rio Grande Valley (LRGV, (Kasprzyk et al., 2009, 2012, 2013)). This problem is many-objective, non-linear, contains a mix of discrete and real decision variables, is severely-constrained, and it has stochastic objectives with expensive function evaluation times. In the largest systematic benchmarking of serial MOEAs in the water resources literature to date (Reed et al., 2013), the LRGV application's complex disjoint Pareto optimal regions caused ten state-of-the-art MOEAs (including the Borg MOEA) to fail to solve this problem reliably. As noted by Tang et al. (2007), problem difficulty is critical for distinguishing alternative MOEA parallelization strategies. In simple master–slave approaches where only the function evaluations are parallelized, the internal search of the original serial algorithm is not modified. Consequently, master–slave strategies simply provide the ability to compute more function evaluations in a fixed wallclock period. Tang et al. (2007) show that often perceived serial MOEA failures are simply the result of not having enough function evaluations given users' limits on search time and not the mathematical difficulty of the underlying water resources applications. They demonstrated this effect for hydrologic model calibration and combinatorial groundwater monitoring design applications.

Alternatively, the LRGV test case selected in this study alternatively poses sufficient problem difficulty as will be shown in our results that the simple master–slave Borg MOEA implementation is fully statistically inferior to the proposed multi-master Borg MOEA parallelization that generalizes the algorithm's auto-adaptivity to modern leadership class supercomputing systems (i.e., thousands of processors). It is worth noting that virtually all of the recent parallel evolutionary computation efforts in the water resources systems literature employ simple master–slave variants of existing popular algorithms that have significant serial bottlenecks in their base algorithmic architectures (e.g., water distribution systems (Guidolin et al., 2012; Roshani and Filion, 2012; Zheng and Morad, 2012a, b); model calibration (Feyen et al., 2007; Tang et al., 2007; Vrugt et al., 2008; Zhang et al., 2013); and groundwater management (Kollat et al., 2011; Matott et al., 2006b, a; Reed and Kollat, 2013; Tang et al., 2007)). Moreover, a vast majority of these studies employed fewer than 500 processors and in the few instances of careful reporting of parallel scalability, their efficiencies decline severely with their maximum processor counts. Parallel efficiency refers to the ratio of actual speedup to theoretical speedup. For example, 100-percent efficiency when using 1000 processors requires a 1000-fold reduction in the wallclock time of a search application. The strong declines in parallel efficiency in the prior water resources literature are in fact theoretically expected given their use of master–slave architectures and existing MOEAs that have significant serial bottlenecks in their algorithmic architectures (e.g., generational selection, mating, mutation, and solution sorting). In master–slave architectures, an increasing number of workers increases communication costs and processing time at the master node, bounding the theoretical speedup attainable. Moreover, as described by Amdahl's law, the speedup of any program is limited by the serial portion of the program (Amdahl, 1967, 1988). Alternatively, this study contributes a self-adaptive multi-master Borg MOEA whose algorithmic architecture minimizes serial bottlenecks and exploits a parallelization strategy that reduces communication costs.

The master–slave and multi-master parallel variants of the Borg MOEA developed in this study make it possible to exploit leadership class computing systems with thousands or tens of thousands of processors to significantly improve convergence speed, solution quality, and reliability. As noted in the lead vision paper Maier et al. (2014), it is important to view the value of parallel metaheuristics in the broader continuum of advancing computing capabilities. Summary statistics for geographic and institutional access to increasingly powerful computing architectures shows the continued exponential growth in capability world-wide (Top 500 Supercomputer Sites, 2014). Moreover, these statistics clearly show that the supercomputers of today are the broadly available work stations of tomorrow. This study advances the water resources field's ability to exploit these trends in rapid and high-quality optimization.

Although this study focuses solely on parallelization, it should be noted that future studies can combine the master–slave and multi-master Borg MOEA advances with other efficiency enhancement strategies (e.g., response surface modeling (Razavi et al., 2012), emulation (Castelletti et al., 2012), global-local hybrid search (Sayeed and Mahinthakumar, 2005), pre-conditioning (Fu et al., 2013; Kollat and Reed, 2006), problem decomposition (Castelletti et al., 2012; Fu et al., 2013), etc.). Readers interested in these alternative efficiency enhancement strategies can reference the more detailed reviews by Nicklow et al. (2010); Maier et al. (2014). It is worth noting the relative concerns and consequences that users should consider when choosing parallelization and/or other commonly employed efficiency enhancement strategies for MOEAs. Beyond a simple focus on efficiency, an additional key question is whether or not an efficiency enhancement strategy precludes the exploration of important problem formulation hypotheses (Kasprzyk et al., 2013). The scope of problem formulations is fundamentally tied to the optimization strategies employed, and these interdependent choices are a significant potential source for negative decision biases if the system's complexities and uncertainties are overly simplified. Besides parallelization, response surface methods (see review for the water resources field in Razavi et al. (2012)) and their specific use in evolutionary algorithms termed fitness approximation (see general review for the metaheuristics field in Jin et al. (2002)) are the most popular strategies for reducing the computational demands. Jin et al. (2002) strongly demonstrate that these approximate evaluation techniques can actually severely degrade applications if not implemented carefully.

Response surfaces require offline or online training where the original computationally expensive model is used to evaluate the objectives for a statistical sampling of candidate decisions distributed throughout a problem's space of alternatives (Razavi et al., 2012). Response surface methods simply approximate the mapping from decisions to objectives. They do not provide insights into the spatial or temporal gradients of the states of the systems of interest (e.g., hydraulic heads or concentrations). This is problematic for data assimilation, uncertainty analysis, and state-based control optimization. Moreover in the multiobjective optimization context, every objective requires its own statistical design of experiments and a unique response surface, which poses a computational barrier onto itself. The water resources applications reviewed by Razavi et al. (2012) highlights response surface